Integration of system phenotypes in microbiome networks to identify candidate synthetic communities: a study of the grafted tomato rhizobiome

Ravin Poudel[1234#], Ari Jumpponen[5], Megan Kennelly[4], Cary Rivard[6], Lorena Gomez-Montano[1234], and Karen A. Garrett[1234#].

[1]*Department of Plant Pathology, University of Florida, Gainesville, FL 32611*

[2]*Institute for Sustainable Food Systems, University of Florida, Gainesville, FL 32611*

[3]*Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611*

[4]*Department of Plant Pathology, Kansas State University, Manhattan, KS 66506*

[5]*Division of Biology and Ecological Genomics Institute, Kansas State University, Manhattan, KS 66506*

[6]*Department of Horticulture and Natural Resources, Kansas State University, Olathe, KS 66061*

[#]*Correspondence: Karen A. Garrett and Ravin Poudel*

E-mail addresses: karengarrett@ufl.edu and rp3448@ufl.edu

Running title: Phenotype-based network for microbial consortia design

ABSTRACT Understanding factors influencing microbial interactions, and designing methods to identify key taxa, are complex challenges for achieving microbiome-based agriculture. Here we use a grafted tomato system to study how grafting and the choice of rootstock influence root-associated fungal communities. We then construct a phenotype-OTU network analysis (PhONA) using random forest and network models of the observed sequence-based fungal Operational Taxonomic Units (OTUs) and associated tomato yield data. PhONA provides a framework to select a testable and manageable number of OTUs to support microbiome-based agriculture. We studied three tomato rootstocks (BHN589, RST-04-106 and Maxifort) grafted to a BHN589 scion and profiled the associated endosphere and rhizosphere fungal communities by sequencing the fungal 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 higher fungal species richness than the other rootstocks or controls. Differentially abundant OTUs (DAOTUs) specific to each rootstock were identified for both endosphere and rhizosphere compartments. PhONA integrates the information about individual OTU obtained via differential abundance tests and role analyses. It identified OTUs that were directly important for predicting tomato yield, and other OTUs that were indirectly linked to yield through their links to these OTUs. Fungal OTUs that are directly or indirectly linked with tomato yield represent candidate components of synthetic communities for testing 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 random forest and network models. By incorporating yield data in the network, PhONA identified OTUs that were directly predictive of tomato yield, and other OTUs that were indirectly linked to yield through their links to these OTUs. Follow-up functional studies of taxa associated with effective rootstocks identified using 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 genomics data.

KEYWORDS fungi, phenotypes, microbiome networks, model integration, tomato

Interactions are key to defining system behaviors, structures, and outcomes. In microbial systems, interactions among microbes 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 (1-6). In agriculture, plant–microbe interactions improve plant productivity by providing access to nutrients (7-9), infection of plant pathogens (5, 10), triggering plant growth promoting factors (11, 12), and enhancing plant resistance (13, 14) and tolerance to abiotic stresses (15-17). Although the importance of microbes and host-microbe interactions to host health and ecological processes is known, interaction-based approaches to manage crop-production remain a scientific frontier. Past attempts to translate information about microbes to design biocontrol agents or biofertilizers have often had limited efficacy and durability (18, 19). Most microbes 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. 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 engineering synthetic communities.

Among tools to better understand microbial ecology, 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 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, such as diversity indices, network models can be used to identify hub taxa that may be key to maintaining microbial assemblages and diversity (20), or to

evaluate changes in community complexity and interactions in response to experimental treatments (21). The use of microbiome networks to describe general community structures and key properties is valuable and often the most practical option when additional information is missing or the goal is to compare across studies with different types of data (22, 23). The utility of network analysis for identifying candidate assemblages for biocontrol can be enhanced by incorporating nodes that represent more types of features (24, 25). For instance, a novel association of host metadata with the microbiome was revealed in an integrated microbiome-metadata network (26), 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, with the potential to identify microbial consortia that are predictive of host phenotypes. Because such models include host phenotypes, they facilitate finding candidate sets of OTUs that may directly or indirectly affect key OTUs or phenotypic traits. Visualization of networks is often valuable for this purpose; however, the real value of phenotype-based network models is their ability to identify 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 evaluated in lab or field studies.

While phenotype-based network models have the potential to predict 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 host genotype (27-32), 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 (33). Many of the microbes

selected may be taxa that are evolutionarily essential for the survival and function of plants (34, 35). 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 (36-38). Results that indicate microbial filtering by different crop hosts, plant compartments, geographic locations, and environmental factors (39, 40)are promising for minimizing the search space, or necessary sample numbers, to identify candidate taxa for synthetic communities. For instance, factors that explain greater 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 candidate taxa.

In our current study, building on previously described agricultural experiments in grafted tomato systems (41, 42), we characterized the root-associated fungal (RAF) communities and implemented a new approach to select potential candidate fungal taxa that are predictive of tomato yield and/or that are in significant association with such fungal taxa. The new phenotype-OTU network analysis (PhONA) is a network 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 the OTUs predictive of phenotype using a random forest model and indicates positive or negative association (Pearson's correlation coefficient) of the predicted OTUs with the host phenotype, and then combines the OTUs associated with the phenotype with other OTUs in a network model (Fig. 1). Random forest, a non-parametric ensemble learning method, has often been used in microbiome studies (43, 44) for classification of sample states and is suited for non-linear data types involving complex interactions (45).

Phenotype-based selection of microbial consortia is an effective approach to select simplified yet representative microbial taxa and could support the design of microbiome-based

products. Changes in abundance (46), successive selection over multiple generations (3), or analyses of binary host-microbe relationships (47) are some the recent phenotype-based applications to select candidate taxa for biological applications. In contrast to these approaches, first PhONA utilizes a statistical model to select OTUs predictive of host phenotype, then further integrates the predicted OTUs in the network framework. PhONA provides an interaction-based understanding of microbial communities to support microbiome studies. In addition, we contrasted the RAF community's diversity and interactions among the productive rootstocks and the controls, for endosphere and rhizosphere compartments. Based on our yield data, and previous studies of microbial interactions (21), we expected a greater number of fungal OTUs and a greater number of microbial associations for more productive rootstocks. Moreover, in our previous studies of bacterial communities in the tomato rhizobiome, we observed compartment-specific (endosphere vs rhizosphere) effects of grafting and rootstocks on bacterial community composition and diversity (42), and expected similar effects on RAF diversity and composition.

METHODS

Experimental Plots, Rootstocks, and Study Sites. We studied grafted tomato plants in high tunnels in an experimental design similar to the one described by Poudel et al. (42). Tomato plants were grafted following a tube-grafting protocol described in Meyer et al. (41). Our study included three rootstocks (BHN589; RST-04-106, and Maxifort) in four treatments: 1) nongrafted BHN589 plants; 2) selfgrafted BHN589 plants (plants grafted to their own rootstock); 3) BHN589 grafted to RST-04-106; and 4) BHN589 grafted to Maxifort. We chose BHN589 as scion primarily based on its popularity due to high yield and high-quality fruit with a long shelf

142 life. For rootstocks, we selected Maxifort because it is a productive and popular rootstock, and
143 RST-04-106 as a new rootstock variety based on tomato breeders' recommendations.
144 Our study included two sites: Olathe Horticulture Research and Extension Center
145 (OHREC) and Common Harvest Farm, a farm managed by a collaborating farmer. At each study
146 site, the four graft treatments were assigned to four plots per block in a randomized complete
147 block design. Each plot consisted of 5-8 plants, and one middle plant per plot was sampled
148 during the peak growth stage. There were six blocks at OHREC, and four blocks at Common
149 Harvest Farm, such that for each year, a given rootstock was replicated 10 times. The experiment
150 was repeated for two years (2014 and 2015) with a similar design; however, the blocks and
151 rootstocks were randomized in each year. More information about each study site is provided in
152 Table S1.
153 **Sample Preparation, DNA Extraction, and Amplicon Generation.** To compare the
154 fungal communities, we selected a center plant from each plot and carefully dug it out such that
155 the majority of the roots remained intact. Endosphere and rhizosphere samples were prepared as
156 previously described (42) and the total genomic DNA was extracted using a DNA extraction kit
157 (MoBio UltraClean Soil DNA Isolation Kit; MoBio, Carlsbad, CA, USA) as per manufacturer's
158 protocol, with slight modification during the homogenization step (42). To recover high genetic
159 diversity, we opted for the two-step PCR approach. The primary PCR amplicons were generated
160 in 50 µL reactions under the following conditions: 1 µM forward and reverse primers, 10 ng
161 template DNA, 200 µM of each dioxynucleotide, 1.5 mM $MgCl_2$, 10 µL 5x Phusion Green HF
162 buffer (Finnzymes,Vantaa, Finland), 24.5 µL molecular biology grade water, and 1 unit (0.5 µL)
163 Phusion Hot Start II DNA Polymerase (Finnzymes, Vantaa, Finland). PCR cycle parameters
164 consisted of a 98° C initial denaturing step for 30 seconds, followed by 30 cycles at 98° C for 10

seconds, 57° C annealing temperature for 30 seconds, and 72° C extension step for 30 seconds, followed by a final extension step at 72° C for 10 minutes. 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 ITS region of fungal rRNA genes using primers ITS1F-CTTGGTCATTTAGAGGAAGTAA and ITS4-TCCTCCGCTTATTGATATGC, (e.g. 48). The average amplicon length of the ITS region in fungi is about 600 bp and could not be fully covered with the Illumina MiSeq platform (v.3-chemistry) in a single read. Thus, in the following nested PCR, only ITS2 out of the whole ITS region was amplified using fITS7-ITS4 primers (49) 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: amplification cycles were reduced to ten, total reaction volume was reduced to 25 ul, and 5 ul of cleaned PCR product from the first PCR amplification was used as the DNA template. Similarly, the nested PCR was 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 PCR amplicons were combined per experimental unit, and the final pool was cleaned again using an Agencourt AmPure cleanup kit as above. Illumina MiSeq adaptors were ligated to the library and paired-end sequenced on a MiSeq Personal Sequencing System (Illumina, San Diego, CA, USA) using 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:……).

Bioinformatics and OTU Designation. The sequence library of fastq files was curated using the MOTHUR pipeline (Version 1.33.3; (50)) following steps modified from the MiSeq Standard Operating Protocol (SOP; www.mothur.org/wiki/MiSeq_SOP). Briefly, the forward and the reverse reads were assembled into contigs using the default alignment algorithm. Any sequences shorter than 250 base pairs or containing an ambiguous base call or more than eight homopolymers or missing MIDs were removed from the library. Barcoded sequences were assigned to experimental units, and fasta and group files for the 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 data 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 (51) referencing the UNITE plus INSD non-redundant ITS database (52). To minimize the bias resulting from unequal sequence counts per sample, samples were rarified at the sequence frequency of the sample yielding the lowest count (6,777). The final curated OTU database included 1,084,281 total sequences representing 16,151 fungal OTUs, including singletons (5376).

Statistical analyses. We evaluated the network of associations among fungal OTUs 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 strong evidence ($p < 0.01$) that their frequencies are correlated across samples. Reducing false associations due to compositional bias in network modeling of microbiome data is important for clearer interpretation (53). Thus, we used a Sparse Correlations for Compositional data (SparCC)

method to evaluate the pairwise associations (53), 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 100 bootstrapped datasets. Once the matrix defining all the pairwise associations was derived, we selected only those OTUs for which the absolute value of the association was greater than 0.5 (and $p < 0.01$) in the network analyses for each of the rootstock genotypes.

To identify the OTUs associated with tomato yield in each rootstock, a regression-based random forest model was fitted to the observed data. Marketable tomato yield data reported by Meyer et al. (41) was the response variable and fungal OTUs were potential predictors. In order to minimize the influence of rare and transient OTUs, only those OTUs that occurred in more than 20% of the samples were filtered then transformed using Z-scores. The transformed data were used for training the model. OTUs in the resulting model were ranked by their importance based on increased mean square error (%IncMSE) of the model. Random forest uses out-of-bag (OOB) samples (set of samples not used for constructing the tree) to estimate the MSE error for each decision tree. The average MSE across the decision trees is then evaluated with the error due to the permutation of a variable (predictor) in the model, where an increase in error (%IncMSE) represents feature importance. From the list of OTUs ranked by their importance to the model, we selected all the OTUs with %IncMSE > 0 for constructing PhONA. PhONA integrates the results from the random forest model for yield with the OTU-OTU association network. We then plotted the resulting network using the igraph package (54) 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 (24, 55). Categorization of individual OTUs based on differential

abundance tests or role analyses were indicated visually as a component of the PhONA, to provide an analysis integrating information from multiple analyses.

To evaluate the effects of rootstocks on fungal diversity, Shannon entropy and species richness were evaluated using the vegan package (56) implemented as part of the phyloseq package (57) in R (58). Differences in diversity across the rootstocks were compared using a mixed model ANOVA in the lme4 package in R (59). Study site and sampling year were treated as random factors, blocked by study sites, whereas the rootstock and compartments were treated as fixed factors. Changes in fungal community composition across the samples were estimated based on a Bray-Curtis dissimilarity matrix and visualized in non-metric 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 1000 permutations) using the adonis function in the vegan package (56). 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 selfgrafts 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 (60). A differential abundance test was performed within the controls (selfgraft vs. nongraft) to identify the OTUs responsive to the grafting procedure itself.

RESULTS

General Root-Associated Fungal Community in the Grafted Tomato System. Once rare taxa (<10 sequence counts, which accounted for more than 90% of the observed OTUs) were removed, the community consisted of 1586 OTUs and 1,063,017 sequences. Of these sequences, 4.8% remained unclassified at the phylum level (Fig. S1). The classified sequences represented Ascomycota (52.5%), Basidiomycota (25.6%), Zygomycota (11.5%), Chytridiomycota (3.6%), Glomeromycota (1.7%), and Rozellomycota (0.07%) (Fig. 1). 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, Cantharellales, Mortierellales, and Hypocreales. Analyses at the family level revealed that Pyronemataceae, Mortierellaceae, Ceratobasidiaceae, and Pleosporaceae were the most common taxa in the overall community. At the genus level, *Mortierella* spp., *Thanatephorus* spp., and *Alternaria* spp. were the most abundant genera.

Effects of Grafting and Rootstock on α-Diversity. There was strong evidence for an effect of rootstocks on OTU richness ($F_{df,3} = 8.6$, $p < 0.001$) and Shannon entropy ($F_{df,3} = 3.2$, $p = 0.02$) for tomato RAF communities. Mean species richness was higher in both the endosphere ($p = 0.01$) and rhizosphere ($p = 0.001$) of one of the hybrid rootstocks, Maxifort, compared to the nongrafted control. Shannon entropy followed trends similar to richness with a higher estimate for Maxifort; however, there was only evidence for higher Shannon entropy in Maxifort for the rhizosphere ($p = 0.004$), but not for the endosphere ($p = 0.6$) (Fig. 2). 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 (42), we expected a significant effect of rootstocks on the community compositions. Rootstock explained 2% of the variation in the RAF community composition (PERMANOVA; $F_{df,3} = 1.39$, $R^2 = 0.02$, $p = 0.02$), while compartment, study site, and year explained a greater proportion of the variation than rootstock (Fig. S3 and Table S2). Endosphere-rhizosphere compartments accounted for 8.9% of the variation. Study site and interannual variation explained 8.4% and 5.4% of the total variation, respectively (Table S2).

Comparison of 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 = 135) as in the rhizosphere (n = 72) (Figs. 2 and S4). Comparison across rootstocks indicated a greater number of DAOTUs in Maxifort (n = 92) than in RST-04-106 (n = 64) and the selfgraft control (n = 51). Compared to the hybrid rootstocks, the number of depleted taxa was greater in the selfgraft control (n = 27). Among the enriched OTUs in Maxifort, 31 OTUs belonged to Basidiomycota, 23 to Ascomycota, and 12 to Glomeromycota, whereas four basidiomycete, three ascomycete, and two zygomycete OTUs were depleted in Maxifort. In the case of RST-04-106, enriched taxa included 21 OTUs in Basidiomycota, 25 OTUs in Ascomycota and four OTUs in Glomeromycota, whereas the depleted OTUs included five in Zygomycota. Between the self- and nongraft controls, seven OTUs in Ascomycota, eight OTUs in Basidiomycota, and five OTUs in Zygomycota were enriched, whereas 10 OTUs in Basidiomycota, six OTUs in Ascomycota, four OTUs in Zygomycota, and four OTUs in Glomeromycota were depleted.

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 for one of the hybrid rootstocks, Maxifort, compared to both controls and RST-04-106 (Figs. 3, S5, S6, and Table S3). Comparing compartments, complexity was higher in the rhizosphere than in the endosphere (Figs. 3, S5, S6, and Table S3). In addition to the total number of links, the link type (either positive or negative) differed among the rootstocks in both compartments (Table S3), 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 as compared to 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 number of modules defined using a simulated annealing (SA) algorithm, were similar in both the endosphere and rhizosphere compartments and across the rootstocks (Table S3). 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, and key nodes were more common in the hybrid rootstocks than in the non- and selfgrafted controls (Figs. 4 and 5).

Random Forest and PhONA. Using random forest models, we identified the compartment-specific OTUs useful for predicting tomato yield in each rootstock. The number of OTUs with positive %IncMSE (importance) differed between the compartments and among the rootstocks. The number of OTUs with positive %IncMSE was higher in the rhizosphere (54.3 ± 17. 1; mean ± sd) than in the endosphere (28 ± 6.5), with the highest number for Maxifort. This might indicate that rhizosphere OTUs are more important for determining host-specific traits.

For each treatment, the observed PhONA provided a graphical representation that can be used to select candidate taxa based on the network structure, phenotype, and node attributes evaluated in differential abundance tests and role analyses. The rational selection of taxa could even be improved with *a priori* knowledge about potential biological and ecological functions of the observed OTUs.

DISCUSSION

This study demonstrated the effect of rootstocks on RAF community composition and structure. General diversity-based analyses indicated the effect of rootstocks on RAF community composition. The more productive hybrid rootstock, Maxifort, supported higher fungal richness and Shannon entropy, as well as a greater number of DAOTUs than the controls, consistent with the expectation that higher diversity and a higher number of responsive taxa (DAOTUs) would be associated with a more productive genotype. Also consistent with our expectations, we observed higher microbial diversity and fewer responsive taxa (DAOTUs) in the rhizosphere than in the endosphere. The integrated host phenotype and OTU network in the PhONA indicated potential candidate taxa for each rootstock, and community structures in the endosphere and rhizosphere compartments. The general network analysis found more interactions and complex structures among fungal taxa in Maxifort, consistent with our expectation that increased community complexity would be associated with the more productive rootstock. Community complexity, when defined in terms of mean node degree, also differed between the compartments: the endosphere community was less complex than the rhizosphere community in all the rootstocks. Overall, our study i) showed that rootstocks and grafting are significant drivers of RAF community composition, diversity and structure, and ii) provided

PhONA as an analytical framework to select potential candidate taxa for microbiome-based agriculture, where the taxa were predictive of yield and in significant association with other OTUs in the community. From a practical standpoint, our results indicate the potential for using plant genotypes and agricultural practices to modulate plant-associated microbial communities, and the potential for the PhONA illustrated here to improve identification of candidate taxa to support microbiome-based crop production.

In contrast to network models that portray only microbe-microbe interactions, PhONA integrates the results of random forest models of microbe association with phenotypic traits, to support inference 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 have an indirect association with the host response variable through their associations with other taxa in the community that have direct associations with host traits. For instance, a node that has a positive association with the system phenotype node (in our case, yield) might have negative or positive association with other OTUs. Such OTUs with indirect positive associations with the desired phenotype might also be included in designing a consortium for biofertilizers. Using a PhONA, a rational consortium can be selected based on the phenotype of interest. Applying PhONA for the case of disease or pathogen-based phenotypes could be useful for designing a rational consortium for biocontrol. We also observed some OTUs with direct negative associations with the yield node. For instance, *g_Thanatephorus* is one of the nodes identified as depleted in a differential abundance test (Fig. 2), and had a negative association with yield for Maxifort (Fig. 3). Some members of *g_Thanatephorus (OTU00001)* are known plant pathogens. Similarly, the *g_Thanatephorus (OTU00012)* node was observed to be a connector and had a negative association with Maxifort yield (Fig. 3). Effort to control this

368 particular taxon, as well as the taxa that are have positive associations with *g_Thanatephorus*,
369 might contribute to maximizing yield. Although we did not observe any disease symptoms in our
370 experiments, OTUs negatively associated with the yield node might represent a case of
371 asymptomatic yield loss. Efforts to explore negatively associated OTUs might provide
372 opportunity to minimize asymptomatic yield loss in crops. The main goal of PhONA is to
373 provide a systems framework to generate hypotheses about the role of microbiome components
374 in host function and performance, and to support the potential for mechanistic/predictive models
375 to understand host-microbiome interactions. *In planta* experiments with fungal cultures are
376 essential to test the hypotheses generated by these models, to help to differentiate between
377 associations that are based on consistent biological interactions and not simply based on shared
378 (or opposing) niche preferences. It is important to be cautious in attributing biological
379 interactions to the key structures in network attributes, because the links in the PhONA may or
380 may not depict biological interactions. That is, many links may represent correlative
381 relationships rather than causal relationships(24). For instance, the hub node in the network is
382 often regarded as a key node, but the high number of links with the hub node in the association
383 network could be due to shared niches, biological interactions, or a mixture thereof. If the
384 associations are mostly due to shared niches, removing such a hub node will have a more limited
385 effect, whereas removal of a hub node involved in many biological interactions could lead to
386 fragmentation of the microbial community.

387 RAF community composition, diversity, and interactions differed between the
388 endosphere and rhizosphere compartments. Although the endosphere and the rhizosphere are
389 continuous, they are distinct in community composition and diversity. Compartment specificity
390 in community composition and diversity has been reported for other plant species, in both natural

and agricultural settings (27, 61, 62). Usually, bulk soil is considered a source of microbial communities, a subset of which is selected for in the rhizosphere (27, 63), mainly as a function of root exudates and rhizodeposits (38, 64-66). Selection of the rhizosphere microbiome could be specific (*e.g.*, antagonistic to plant pathogens) (67) or more general with less influence of host genotype. In comparison, the endosphere of host plants often supports lower microbial diversity compared to the rhizosphere (62, 63). Host tissues and immune systems act as a biotic filter (2). As a result, the microbiome is more specialized in the endosphere than in the rhizosphere. The 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. Plant genotypes can structure root-associated fungal communities (27, 62, 68). The commercial rootstocks in our study have been bred for defined purposes, specifically to provide resistance against specific soil-borne pathogens and pests. Small variations in the host genotype could alter the physiological and immunological responses in the root systems, thereby selecting genotype-specific root-associated fungal communities. For example, root exudates and metabolites could be specific to plant host (69-71), and thus could provide specific control of microbial communities (72, 73). In some cases, the host genotype effect can be directly attributed to root anatomy (74). Effective and efficient root types and architectures are desired agronomic traits to cope with biotic and abiotic stresses (75), and root systems vary among plant species and genotypes (74). Moreover, the control of plant genotypes on microbial communities in the root system is linked to nutrient positive feedbacks between the scion and rootstock. Plant genotypes supporting higher yield and biomass may support higher microbial diversity by excreting a

414 greater volume of photosynthates as root exudates and metabolites. Although we did not evaluate
415 root exudates, our study is consistent with a role for higher yield and biomass (as for the
416 Maxifort rootstock) associated with higher fungal diversity. Additionally, we observed an effect
417 of rootstocks on the RAF community composition. Collectively, the results support our
418 expectations of rootstock-specific control of the RAF community.

419 While the effects of plant genotype and compartment are evident, our current study of
420 fungi and previous study of bacteria (42) in the same system indicate that the magnitude of the
421 effect may vary between fungi and bacteria. The estimated effect of rootstocks on bacterial
422 communities (3%) was slightly greater than for fungal communities (2%), although the effects of
423 study sites and compartments were far greater in both the bacterial and fungal communities. As
424 in the RAF communities, compartment (endosphere vs. rhizosphere) and study sites played
425 major roles in structuring bacterial community composition (42); however, bacterial
426 communities in the endosphere and the rhizosphere were much more distinct than the fungal
427 endosphere and rhizosphere - possibly a result of their distinct growth forms (predominantly
428 mycelial or unicellular). Compartment and study site effects on the bacterial communities (42)
429 were greater than on the fungal communities in the same experiment: the estimated compartment
430 effect was three-fold higher (about 25-28%) for bacteria than fungi, and the estimated study site
431 effect was two-fold higher (about 17-18%). In another example plant system, *Agave* species, the
432 bacterial community was most strongly influenced by the plant compartment whereas the fungal
433 community was more strongly influenced by the study site (61). Lack of a strong study-site
434 effect on the fungal community in our study might indicate a lesser influence of edaphic factors
435 on fungal compared to bacterial communities, or the presence of a wider environmental range for
436 optimal fungal growth (76). Despite the differences in the magnitude of the effect for bacteria

versus fungi, in both studies there was strong evidence for the effect of plant genotypes, compartments, and study sites on the microbial community.

Our definition of complexity is based on interactions in networks, using a definition similar to that used in other microbiome network analyses (21, 30, 77). However, an increased number of interactions and complex network structures/motifs would tend to be observed whenever there are more nodes in these association networks. The higher number of OTUs associated with Maxifort would tend to result in higher complexity by this measure. Another potential measure of complexity is network density, the proportion of links observed in a network relative to the total number of possible links. In the case of the endosphere community, Maxifort did not have higher density estimates than the other rootstocks, so by that measure of complexity its network was not more complex. Statistical methods comparable to rarefaction, designed to equalize the number of nodes across networks or methods to balance OTU richness for sampling efforts(78), 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 level of association to maximize the ecosystem functions represent gaps and opportunities for 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. Studies directly applied to identify potential assemblages for agricultural applications could benefit from exploring results for a range of thresholds.

The value of information depends on its ability to reduce uncertainty and maximize the utility of predictions. Given the unprecedented complexity of microbial communities, it's difficult to make a rational selection of candidate taxa for biological applications. A

system/community level analytical scheme to incorporate information gained via multiple sources or statistical analyses is imperative. Combining results obtained from different analyses in the PhONA framework supports selection of a rational microbial consortium, enhancing the quality of decision-making. However, careful interpretation is needed, because often two or more OTUs with the same name/taxonomy may have totally different roles in the network. This can be due in part to the low resolving power of the current reference databases. Improvement of reference databases and strain level information will enhance PhONA in the future.

Our study indicates the rootstock-genotype specific effect on RAF diversity, composition, and interactions, and also shows how to integrate 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 is commonly 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 (79, 80) has the potential to support causal inference, including understanding of directionality in microbiome networks. PhONA utilizes a random forest model to link OTUs with a system phenotype, although any other models such as generalized linear models or regression could also be used. Given the nature of microbiome data with a high number of features (p) and relatively small number of samples (n), other models to address the n x p problem can readily be applied in PhONA when they improve on random forest models. Additionally, with a larger sample size, more rigorous approaches such as model validation by splitting data into training and test sets could also be used to enhance PhONA. Models to integrate multiomics data to predict the

structure of microbial assemblages and provide mechanistic understanding are another domain where an integrated network model like PhONA may offer a new perspective. As lab-based technologies and computational resources become less expensive, and are combined with improved predictive analytics such as PhONA, microbial community analyses can go beyond simple analyses of diversity studies to help make microbiome-based agriculture a reality.

Acknowledgements

We appreciate support from the Ceres Trust, a USDA NCR SARE Research and Education Grant (LNC13-355), the Kansas Agricultural Experiment Station, the National Institute for Mathematical and Biological Synthesis (NIMBioS), and the University of Florida. We thank members of Garrett lab for their comments on the early version of the manuscript. K.A.G., A.J., M.M.K., C.L.R., and R.P. 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.L.R., and L.G.M. revised the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. Berendsen RL, Pieterse CM, Bakker PA. 2012. The rhizosphere microbiome and plant health. Trends Plant Sci 17:478-86.
2. 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.
3. 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 9:980-9.
4. 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 JI. 2009. A core gut microbiome in obese and lean twins. Nature 457:480-4.
5. 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.
6. Gould AL, Zhang VV, Lamberti L, Jones EW, Obadia B, Korasidis N, Gavryushkin A, Carlson JM, Beerenwinkel N, Ludington WB. 2018. Microbiome interactions shape host fitness. Proc Natl Acad Sci USA 115:E11951-E11960.
7. 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.
8. Koide RT. 1991. Nutrient supply, nutrient demand and plant-response to mycorrhizal infection. New Phytol 117:365-386.
9. Mishra PK, Bisht SC, Mishra S, Selvakumar G, Bisht JK, Gupta HS. 2012. Coinoculation 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.

10. 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.

11. 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 100:1738-50.

12. Glick BR. 1995. The enhancement of plant-growth by free-living bacteria. Can J Microbiol 41:109-117.

13. Conrath U. 2009. Priming of induced plant defense responses. Plant Innate Immunity 51:361-395.

14. 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.

15. Marasco R, Rolli E, Ettoumi B, Vigani 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(10): e48479.

16. Yang J, Kloepper JW, Ryu CM. 2009. Rhizosphere bacteria help plants tolerate abiotic stress. Trends Plant Sci 14:1-4.

17. 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.

18. Mazzola M, Freilich S. 2017. Prospects for biological soilborne disease control: Application of indigenous versus synthetic microbiomes. Phytopathology 107:256-263.

19. 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.

20. 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(1): e1002352.

21. Shi S, Nuccio EE, Shi ZJ, He Z, Zhou J, Firestone MK. 2016. The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. Ecol Lett 19:926-36.

22. Ascunce MS, Shin K, Huguet-Tapia JC, Poudel R, Garrett KA, van Bruggen AHC, Goss EM. 2018. Penicillin trunk injection affects bacterial community structure in citrus trees. Microb Ecol 78:457–469.

23. Garrett KA, Alcala-Briseno 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.

24. 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.

25. Rottjers L, Faust K. 2018. From hairballs to hypotheses-biological insights from microbial networks. Fems Microbiology Reviews 42:761-780.

26. Wang Z, Bafadhel M, Haldar 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. European Respiratory Journal 47:1082-1092.

27. Cregger 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.

28. 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.

29. 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 USA 115:E1157-E1165.

30. 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.

31. 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.

32. 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.

33. 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 64:807-38.

34. 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.

35. Ortiz-Castro R, Contreras-Cornejo HA, Macias-Rodriguez L, Lopez-Bucio J. 2009. The role of microbial signals in plant growth and development. Plant Signal Behav 4:701-12.

36. Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, Bhatnagar S, Eisen JA, Sundaresan V. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. Proc Natl Acad Sci USA 112:E911-E920.

37. 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.

38. 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 CO2 and O3. Sci Rep 7:15019.

39. Herrera J, Poudel R, Nebel KA, Collins SL. 2011. Precipitation increases the abundance of some groups of root-associated fungal endophytes in a semiarid grassland. Ecosphere 2(4):art50.

40. Bokati D, Herrera J, Poudel R. 2016. Soil influences colonization of root-associated fungal endophyte communities of maize, wheat, and their progenitors. J Mycol 2016:1-9.

41. Meyer L. 2016. Grafting to increase high tunnel tomato productivity in the central united states. Master Thesis. http://hdl.handle.net/2097/32736.

42. Poudel R, Jumpponen A, Kennelly MM, Rivard CL, Gomez-Montano L, Garrett KA. 2019. Rootstocks shape the rhizobiome: Rhizosphere and endosphere bacterial communities in the grafted tomato system. Appl Environ Microbiol 85:01765–01718.

43. Roguet A, Eren AM, Newton RJ, McLellan SL. 2018. Fecal source identification using random forest. Microbiome 6:185.

44. 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(2):104.

45. Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT. 2007. Random forests for classification in ecology. Ecology 88:2783-2792.

46. Niu B, Paulson JN, Zheng XQ, Kolter R. 2017. Simplified and representative bacterial community of maize roots. Proc Natl Acad Sci USA 114:E2450-E2459.

47. Paredes SH, Gao TX, Law TF, Finkel OM, Mucyn T, Teixeira PJPL, Gonzalez IS, Feltcher ME, Powers MJ, Shank EA, Jones CD, Jojic V, Dangl JL, Castrillo G. 2018. Design of synthetic bacterial communities for predictable plant phenotypes. Plos Biology 16.

48. Gardes M, Bruns TD. 1993. ITS primers with enhanced specificity for basidiomycetes--application to the identification of mycorrhizae and rusts. Mol Ecol 2:113-8.

49. 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.

50. 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 75:7537-41.

51. 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.

52. Koljalg 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, Duenas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Luecking R, Martin MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Poldmaa K, Saag L, Saar I, Schuessler A, Scott JA, Senes C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson KH. 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol 22:5271-5277.

53. Friedman J, Alm EJ. 2012. Inferring correlation networks from genomic survey data. PLoS Comput Biol 8:e1002687.

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

55. Guimera R, Amaral LAN. 2005. Functional cartography of complex metabolic networks. Nature 433:895-900.

56. Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn 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.:39pp.

57. McMurdie PJ, Holmes S. 2013. phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8(4): e61217.

58. R Core Team. 2018. R: A language and environment for statistical computing. R foundation for statistical computing, vienna, austria. URL https://www.R-project.Org.

59. Bates D, Machler M, Bolker BM, Walker SC. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw 67:1-48.

60. Ferreira JA. 2007. The Benjamini-Hochberg method in the case of discrete test statistics. Int J Biostat 3:Article 11.

61. 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.

62. 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.

63. Gottel NR, Castro HF, Kerley M, Yang ZM, Pelletier DA, Podar M, Karpinets 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.

64. Badri DV, Vivanco JM. 2009. Regulation and function of root exudates. Plant Cell Environ 32:666-681.

65. 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.

66. van der Heijden MG, Schlaeppi K. 2015. Root surface as a frontier for plant microbiome research. Proc Natl Acad Sci USA 112:2299-300.

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. FEMS Microbiol Ecol 56:250-261.

68. Reazin C, Baird R, Clark S, Jumpponen A. 2019. Chestnuts bred for blight resistance depart nursery with distinct fungal rhizobiomes. Mycorrhiza 29:313-324.

69. 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.

70. 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.

71. U'Ren N. 2007. Types, amounts, and possible functions of compounds released into the rhizosphere by soil-grown plants, p 1-21, The rhizosphere doi:10.1201/9781420005585.ch1. CRC Press.

72. Beattie GA. 2018. Metabolic coupling on roots. Nat Microbiol 3:396-397.

73. 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.

74. 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.

75. Malamy JE. 2005. Intrinsic and environmental response pathways that regulate root system architecture. Plant Cell Environ 28:67-77.

76. Rousk J, Baath E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J 4:1340-1351.

77. 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.

78. 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.

79. 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.

80. Penalver Bernabe B, Cralle L, Gilbert JA. 2018. Systems biology of the human microbiome. Curr Opin Biotechnol 51:146-153.

FIG 1 A systematic diagram representing the Phenotype-OTU network analysis (PhONA), where an OTU-OTU association network (A) was combined with the nodes selected based on a random forest model (B) to create PhONA (C).

FIG 2 Enriched and depleted OTUs across tomato rootstock genotypes (nongraft BHN589, selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (RST-04-106 and Maxifort)) evaluated for the rhizosphere (A) and the endosphere (B), using OTU counts from self-grafts and non-grafts 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, and the position along the x axis represents the abundance fold-change contrast with controls (except for the selfgraft vs. 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 for BHN589 grafted on Maxifort. Node color indicates the phylum, except that the tomato-color node represents yield associated with the rootstock. Nodes connected to the rootstock yield node with black links are taxa that were predictive of rootstock yield, where dotted and solid lines indicate negative and positive associations (Pearson's correlation coefficient) with the yield node, respectively. Red and blue links represent negative and positive associations, respectively, between OTUs. Nodes are labeled with the lowest available taxonomic information. Square-shaped and star-shaped nodes represent DAOTU and key nodes detected in role analyses, respectively.

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 color indicates rootstock treatment.

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 color indicates rootstock treatment.

FIG S1 Relative abundance of endosphere and rhizosphere fungi at the phylum level recovered from four tomato rootstock treatments: nongraft BHN589, selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (RST-04-106 and Maxifort). Each individual bar represents a rootstock treatment, and the colored area within the bar represents the relative abundance of the corresponding phylum.

FIG S2 Comparison of overall fungal diversity (A) and richness (B) associated with tomato rootstock genotypes and controls, evaluated in the endosphere and rhizosphere. The plot is divided by the four tomato rootstock treatments: nongraft BHN589, selfgraft BHN589, and

BHN589 grafted on two hybrid rootstocks (RST-04-106 and Maxifort). Shannon entropy and species richness, measures of community diversity, were both higher for Maxifort ($p < 0.005$) compared to the self-graft and RST-04-106 in the rhizosphere, while there was not evidence for a difference in Shannon entropy in the endosphere ($p = 0.634$). Treatment means were separated using the "difflsmeans" function as specified in the lmerTest package in R. Tests for boxplots sharing a letter or letter case type had $p > 0.05$.

FIG S3 Non-metric multidimensional scaling (NMDS) ordination plot of samples labeled by tomato rootstock (nongraft BHN589, selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (RST-04-106 and Maxifort)), compartment (endosphere or rhizosphere), and study site, based on the Bray-Curtis dissimilarity distance matrix of fungal OTUs. Color indicates rootstock treatment, shape indicates study site, and size indicates compartment. Ellipses surrounding the samples indicate the 95% CI of the endosphere and rhizosphere sample centroids.

FIG S4 Number of DAOTUs in a contrast analysis, evaluated for the endosphere and rhizosphere compartments for four tomato rootstock treatments: nongraft and selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (Maxifort and RST-04-106). The green color in each bar represents the number of enriched taxa, and the red color represents the number of depleted taxa. The number of differentially changed taxa was greater for the endosphere than for the rhizosphere. Among the contrast pairs, hybrid rootstocks had a greater number of enriched taxa compared to depleted taxa. However, the number of depleted taxa was higher compared to

enriched taxa in the controls. Among the treatments, Maxifort had the highest number of DAOTUs in both compartments.

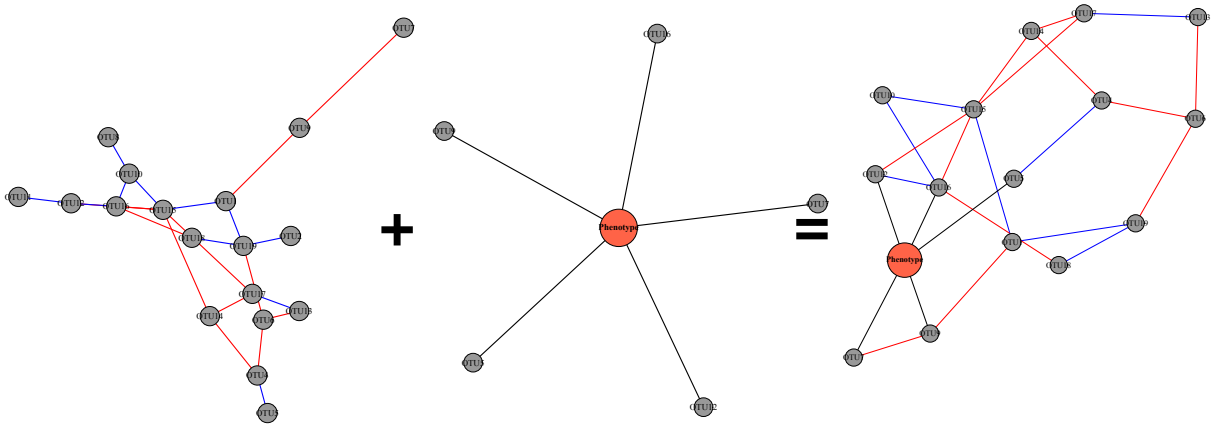
FIG S5 Phenotype-OTU network analysis (PhONA) of **endosphere** fungal taxa for tomato rootstock treatments: (A) nongraft and (B) selfgraft BHN589, and BHN589 grafted on (C) RST-04-106. Node color indicates the phylum, except that the tomato-color node represents yield associated with the rootstock. Nodes connected to the rootstock yield node with black links are taxa that were predictive of rootstock yield, where dotted and solid lines indicate negative and positive associations with the yield node, respectively. Red and blue links represent negative and positive associations, respectively, between OTUs. Nodes are labeled with the lowest available taxonomic information. Square-shaped and star-shaped nodes represent DAOTUs and key nodes detected in role analyses, respectively.

FIG S6 Phenotype-OTU network analysis (PhONA) of **rhizosphere** fungal taxa for tomato rootstock treatments: (A) nongraft and (B) selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks ((C) Maxifort and (D) RST-04-106). Node color indicates the phylum, except that the tomato-color node represents yield associated with the rootstock. Nodes connected to the rootstock yield node with black links are taxa that were predictive of rootstock yield, where dotted and solid lines indicate negative and positive associations with the yield node, respectively. Red and blue links represent negative and positive associations, respectively, between OTUs. Nodes are labeled with the lowest available taxonomic information. Square-shaped and star-shaped nodes represent DAOTUs and key nodes detected in role analyses, respectively.

TABLE S1 Sites included in the study, their soil type, and geographic coordinates.

TABLE S2 Results of the multivariate permutational analysis of variance (PERMANOVA) for fungal taxon abundance data. Permutation was based on the Bray-Curtis distance matrix generated for root associated fungal communities at the OTU level from four tomato rootstock treatments: nongraft and selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (Maxifort and RST-04-106) (1000 permutations). P values < 0.05 are in bold.

TABLE S3 Network attributes and links observed in the fungal association networks for four tomato rootstock treatments: nongraft and selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (Maxifort and RST-04-106) in each of rhizosphere and endosphere compartments.



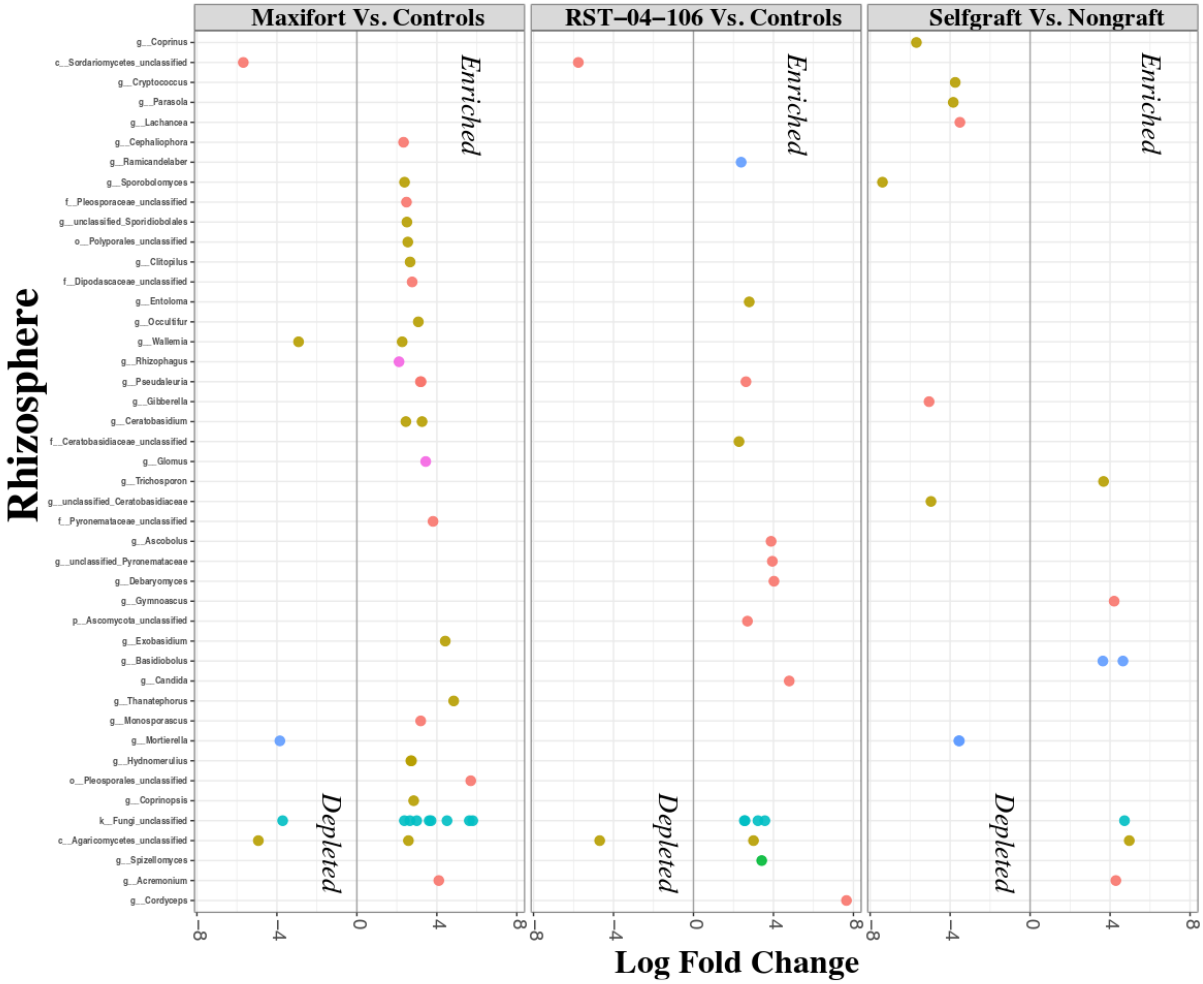
(A) OTU-OTU association network based on SparCC

(B) OTUs predictive of host phenotype, based on random forest model

(C) PhONA predicting candidate taxa based on association network **(A)** and phenotype-predictive OTUs **(B)**

FIG 1 A systematic diagram representing the Phenotype-OTU network analysis (PhONA), where an OTU-OTU association network (A) was combined with the nodes selected based on random forest model (B) to create PhONA (C).

(A)


Selfgraft Vs. Nongraft
RST−04−106 Vs. Controls
Maxifort Vs. Controls
Enriched
Depleted
Rhizosphere
g__Coprinus
c__Sordariomycetes_unclassified
g__Cryptococcus
g__Parasola
g__Lachancea
g__Cephaliophora
g__Ramicandelaber
g__Sporobolomyces
f__Pleosporaceae_unclassified
g__unclassified_Sporidiobolales
o__Polyporales_unclassified
g__Clitopilus
f__Dipodascaceae_unclassified
g__Entoloma
g__Occultifur
g__Wallemia
g__Rhizophagus
g__Pseudaleuria
g__Gibberella
g__Ceratobasidium
f__Ceratobasidiaceae_unclassified
g__Glomus
g__Trichosporon
g__unclassified_Ceratobasidiaceae
f__Pyronemataceae_unclassified
g__Ascobolus
g__unclassified_Pyronemataceae
g__Debaryomyces
g__Gymnoascus
p__Ascomycota_unclassified
g__Exobasidium
g__Basidiobolus
g__Candida
g__Thanatephorus
g__Monosporascus
g__Mortierella
g__Hydnomerulius
o__Pleosporales_unclassified
g__Coprinopsis
k__Fungi_unclassified
c__Agaricomycetes_unclassified
g__Spizellomyces
g__Acremonium
g__Cordyceps
−8
−4
0
4
8
Log Fold Change

(B)

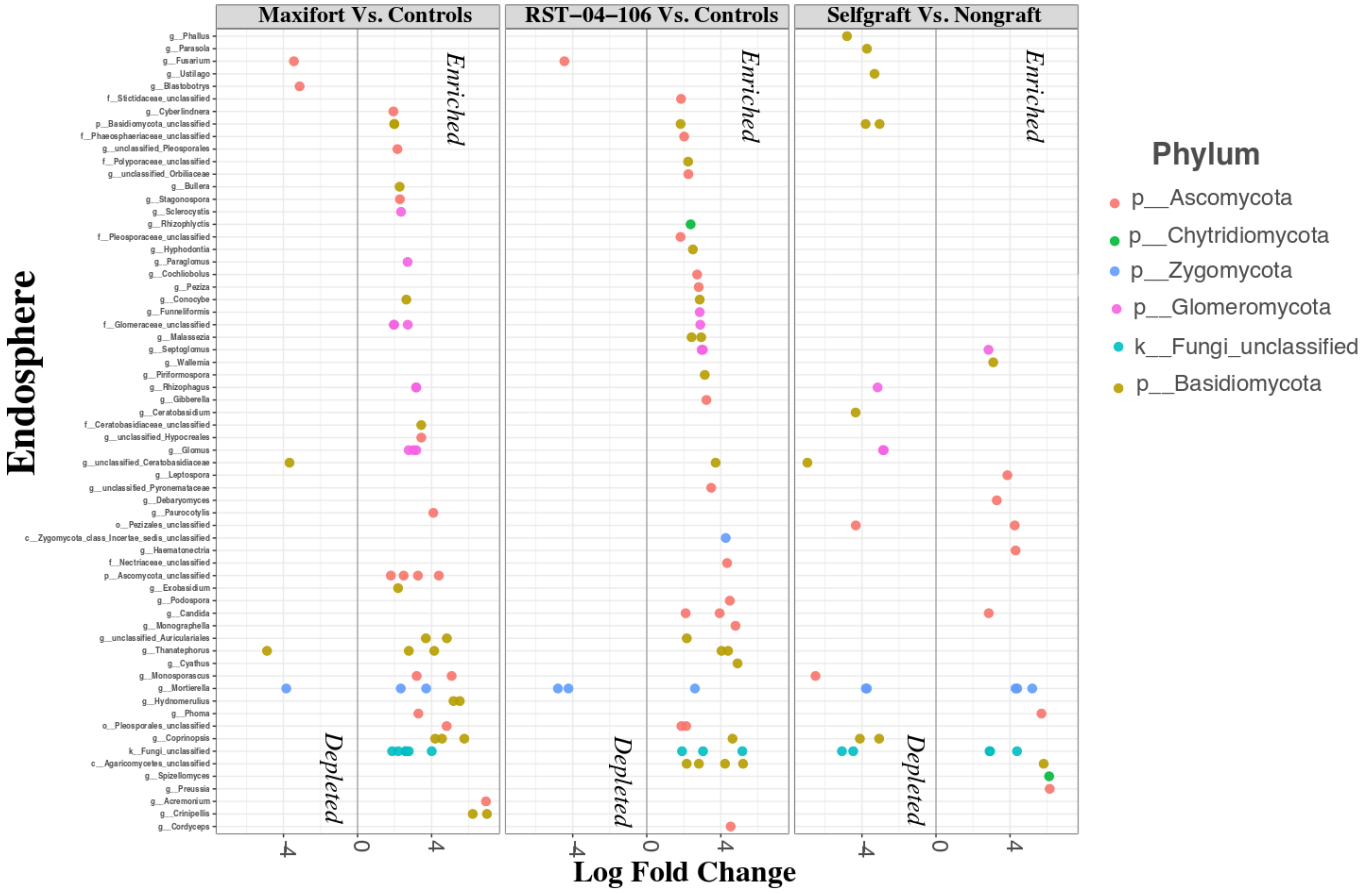


FIG 2 Enriched and depleted OTUs across tomato rootstock genotypes (nongraft BHN589, selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (RST-04-106 and Maxifort)) evaluated for the rhizosphere (A) and the endosphere (B), using OTU counts from self-grafts and non-grafts as controls. All the 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, and the position along the x axis represents the abundance fold change contrast with controls (except for the selfgraft vs. nongraft comparison, where the nongraft treatment was used as a control for the contrast).

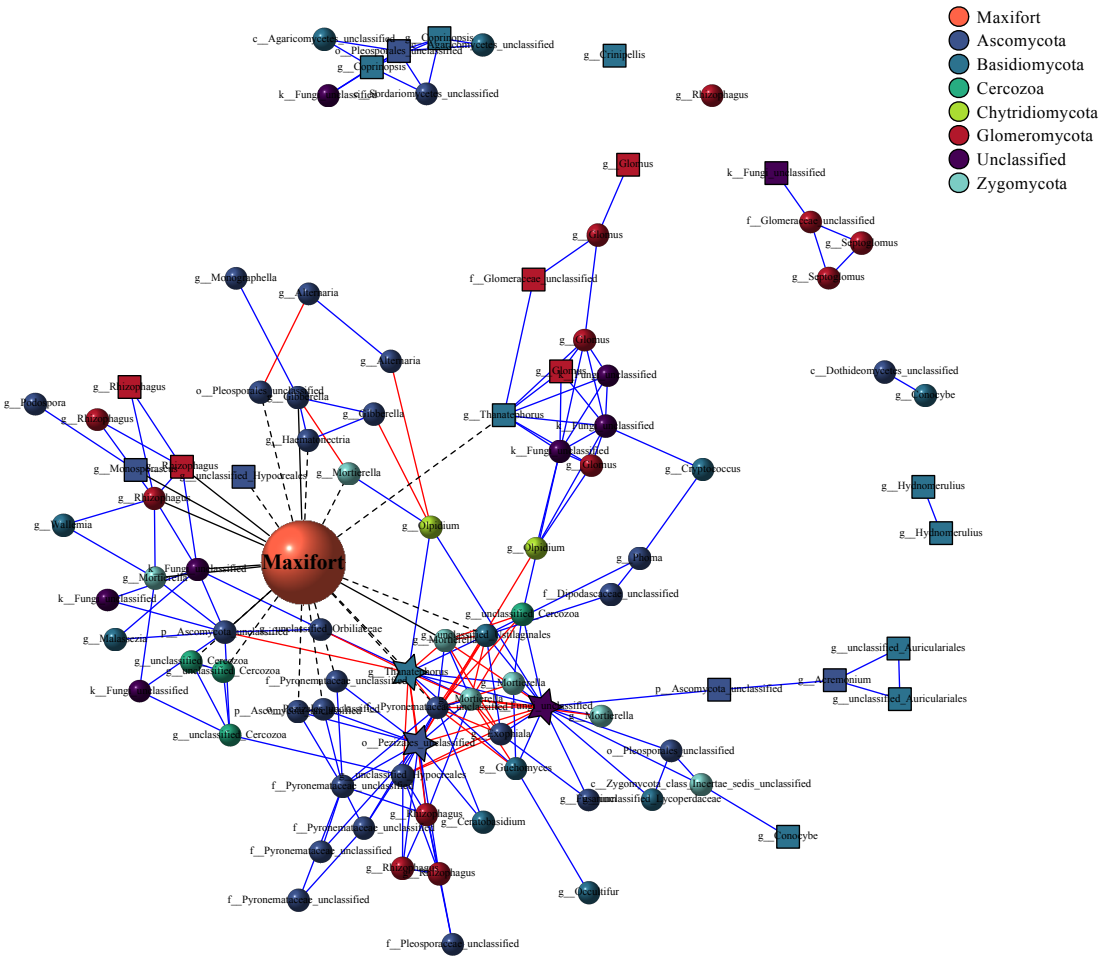


FIG 3 Phenotype-OTU network analysis (PhONA) of endosphere fungal taxa for BHN589 grafted on Maxifort. Node color indicates the phylum, except that the tomato-color node represents yield associated with the rootstock. Nodes connected to the rootstock yield node with black links are taxa that were predictive of rootstock yield, where dotted and solid lines indicate negative and positive associations (Pearson's correlation coefficient) with the yield node, respectively. Red and blue links represent negative and positive associations, respectively, between OTUs. Nodes are labeled with the lowest available taxonomic information. Square-shaped and star-shaped nodes represent DAOTUs and key nodes detected in role analyses, respectively.

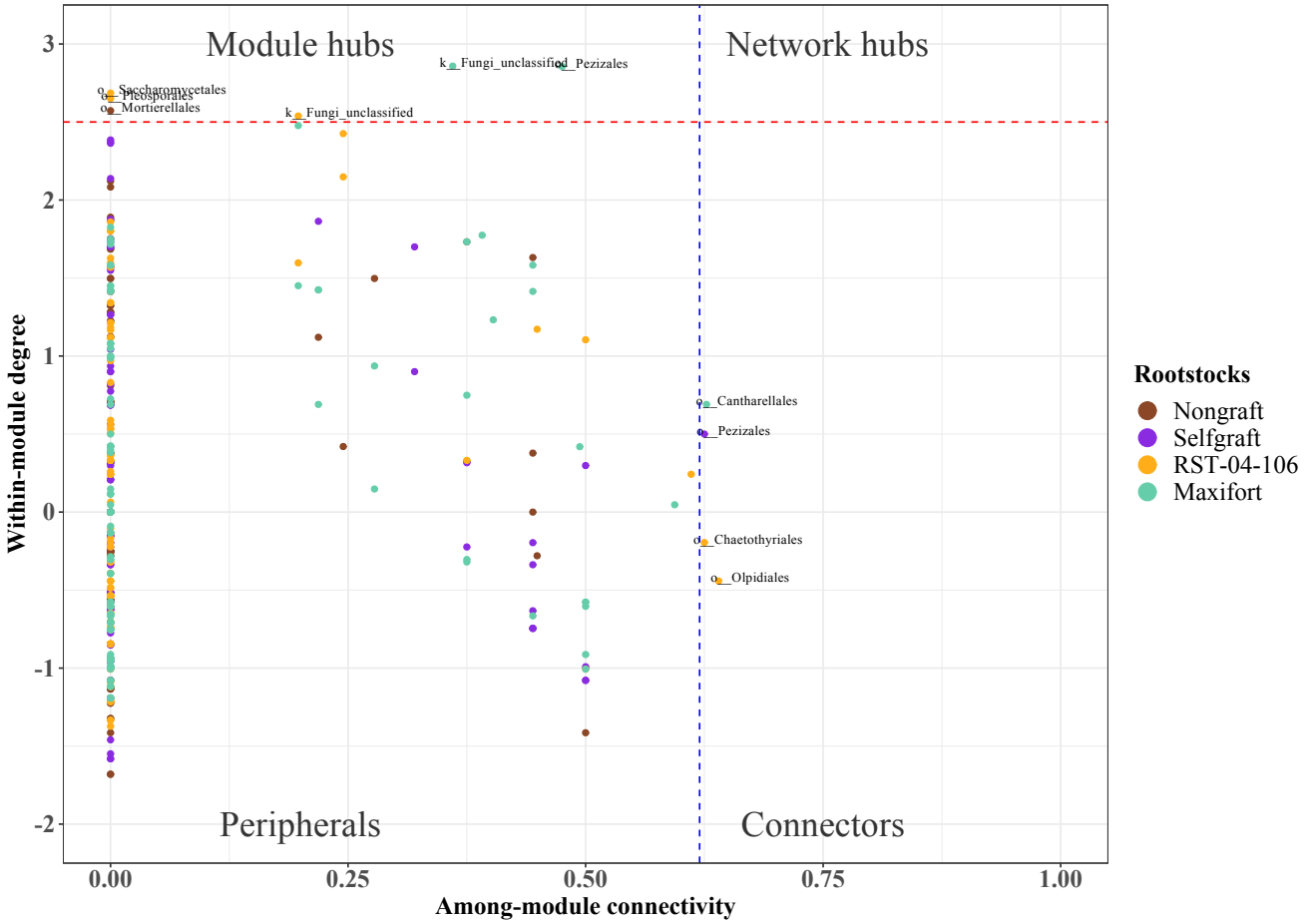


FIG 4 Partitioning of endosphere fungal OTUs according to their network roles. Nodes were divided into four categories based on within-module degree and among-module connectivity. The blue dashed line represents a threshold value (0.62) for among-module connectivity, and the red dashed line represents a threshold value (2.5) for within-module degree. Nodes were categorized as peripherals, connectors, module hubs, and network hubs. Node color indicates rootstock treatment.

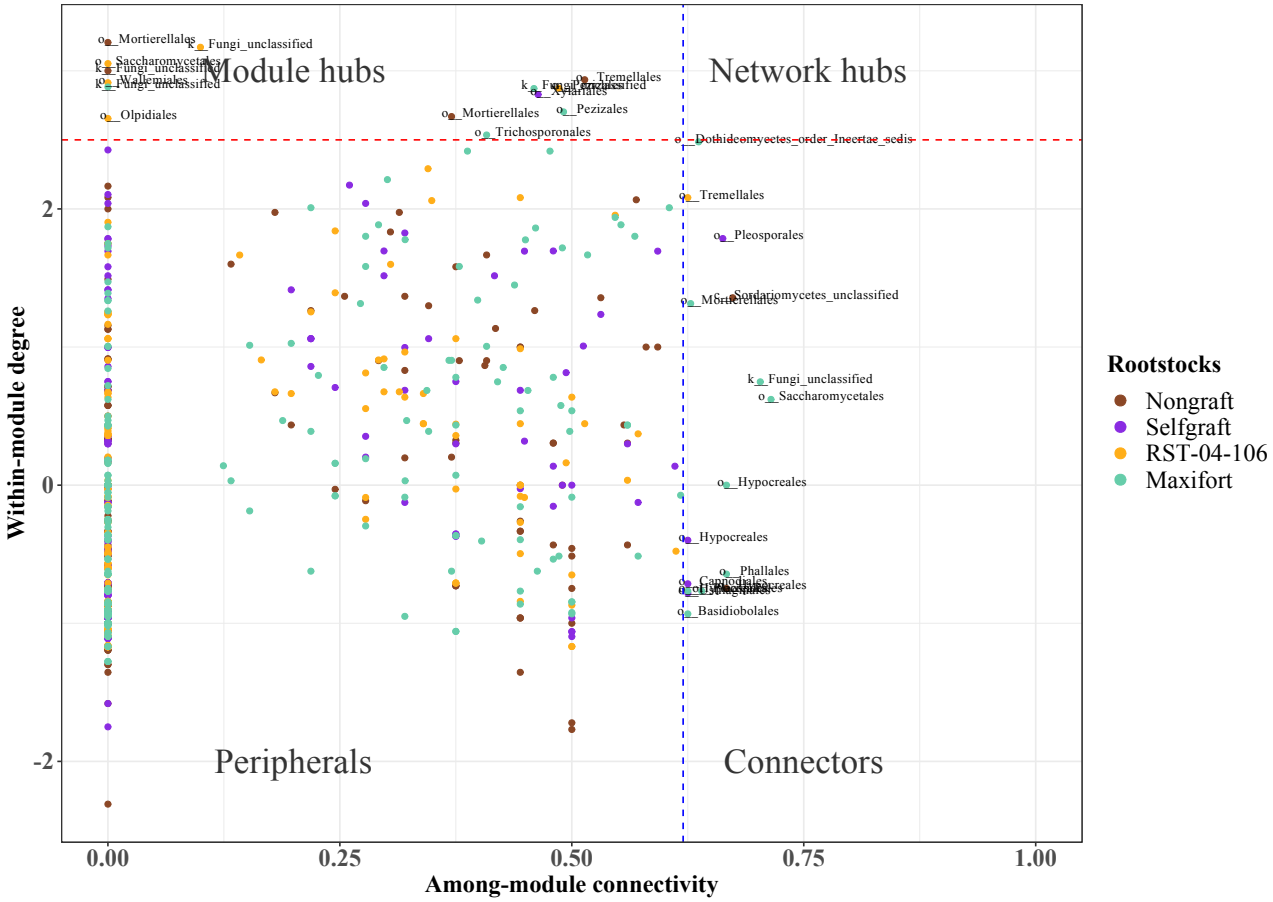
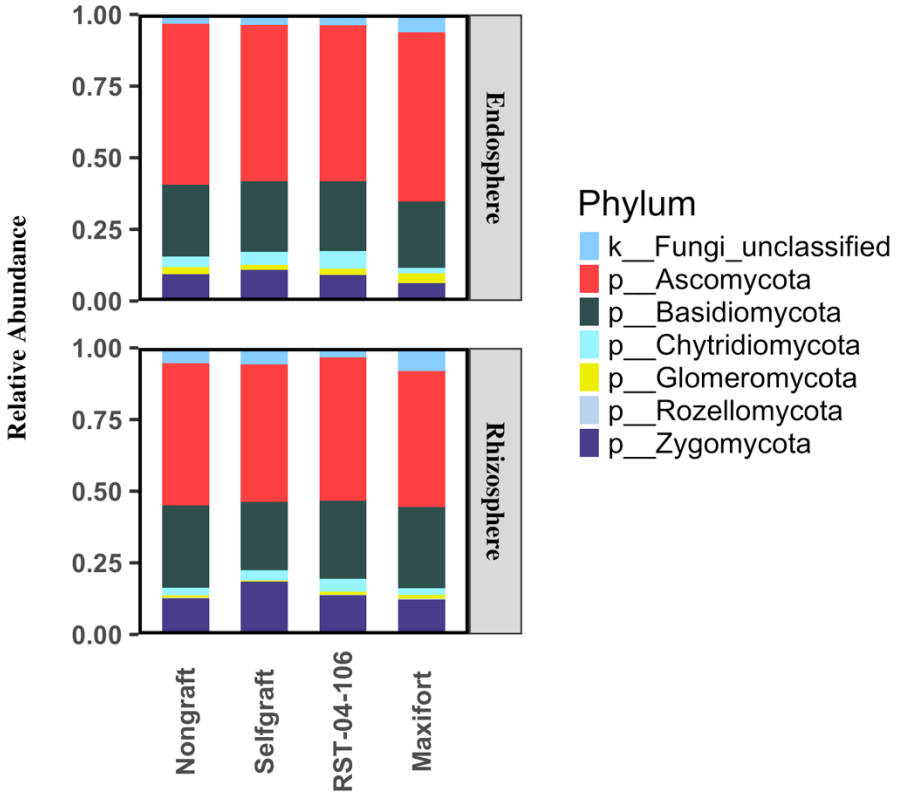


FIG 5 Partitioning of rhizosphere fungal OTUs according to their network roles. Nodes were divided into four categories based on within-module degree and among-module connectivity. The blue dashed line represents a threshold value (0.62) for among-module connectivity, and the red dashed line represents a threshold value (2.5) for within-module degree. Nodes were categorized as peripherals, connectors, module hubs, and network hubs. Node color indicates rootstock treatment.

53



54

55 **FIG S1** Relative abundance of endosphere and rhizosphere fungi at the phylum level recovered

56 from four tomato rootstock treatments: nongraft BHN589, selfgraft BHN589, and BHN589

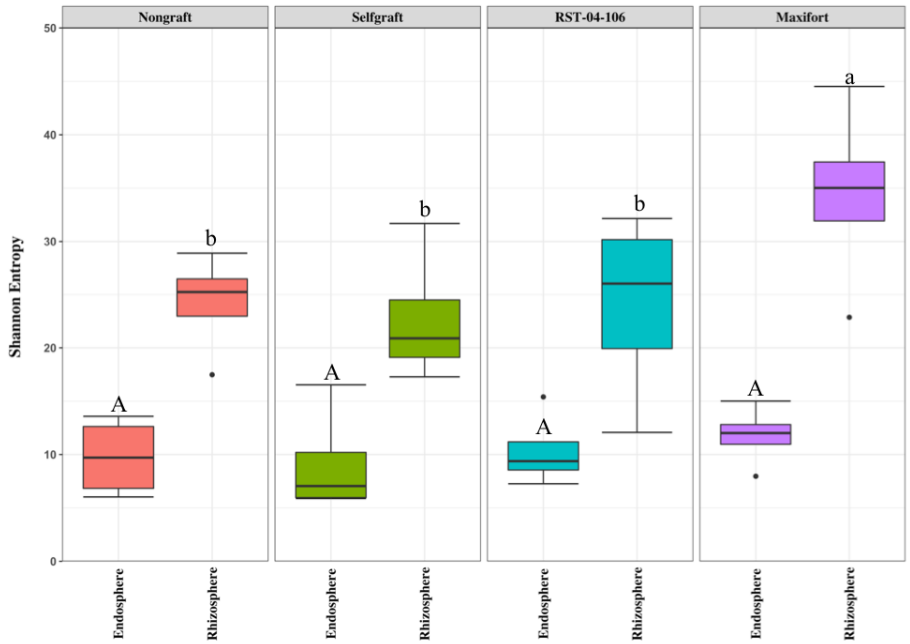
57 grafted on two hybrid rootstocks (RST-04-106 and Maxifort). Each individual bar represents a

58 rootstock treatment, and the colored area within the bar represents the relative abundance of the

59 corresponding phylum.

60

(A)


Nongraft
Selfgraft
RST-04-106
Maxifort
Shannon Entropy
0
10
20
30
40
50
A
b
A
b
A
b
A
a
Endosphere
Rhizosphere
Endosphere
Rhizosphere
Endosphere
Rhizosphere
Endosphere
Rhizosphere

(B)

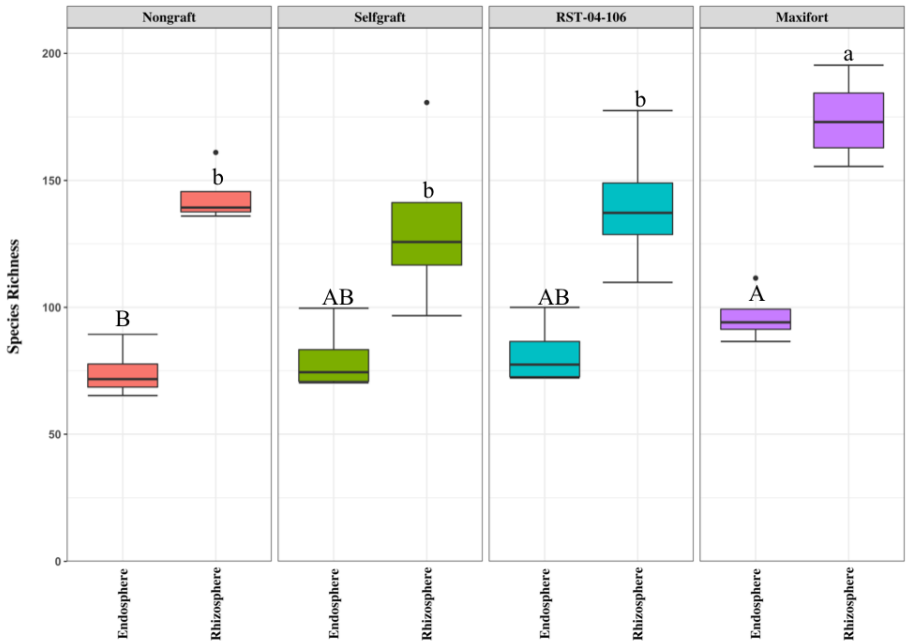


FIG S2 Comparison of overall fungal diversity (A) and richness (B) associated with tomato rootstock genotypes and controls, evaluated in the endosphere and rhizosphere. The plot is divided by the four tomato rootstock treatments: nongraft BHN589, selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (RST-04-106 and Maxifort). Shannon entropy and species richness, measures of community diversity, were both higher for Maxifort ($p < 0.005$) compared to the self-graft and RST-04-106 in the rhizosphere, while there was not evidence for a difference in Shannon entropy in the endosphere ($p = 0.634$). Treatment means were separated using the "difflsmeans" function as specified in the lmerTest package in R. Tests for boxplots sharing a letter or letter case type had $p > 0.05$.

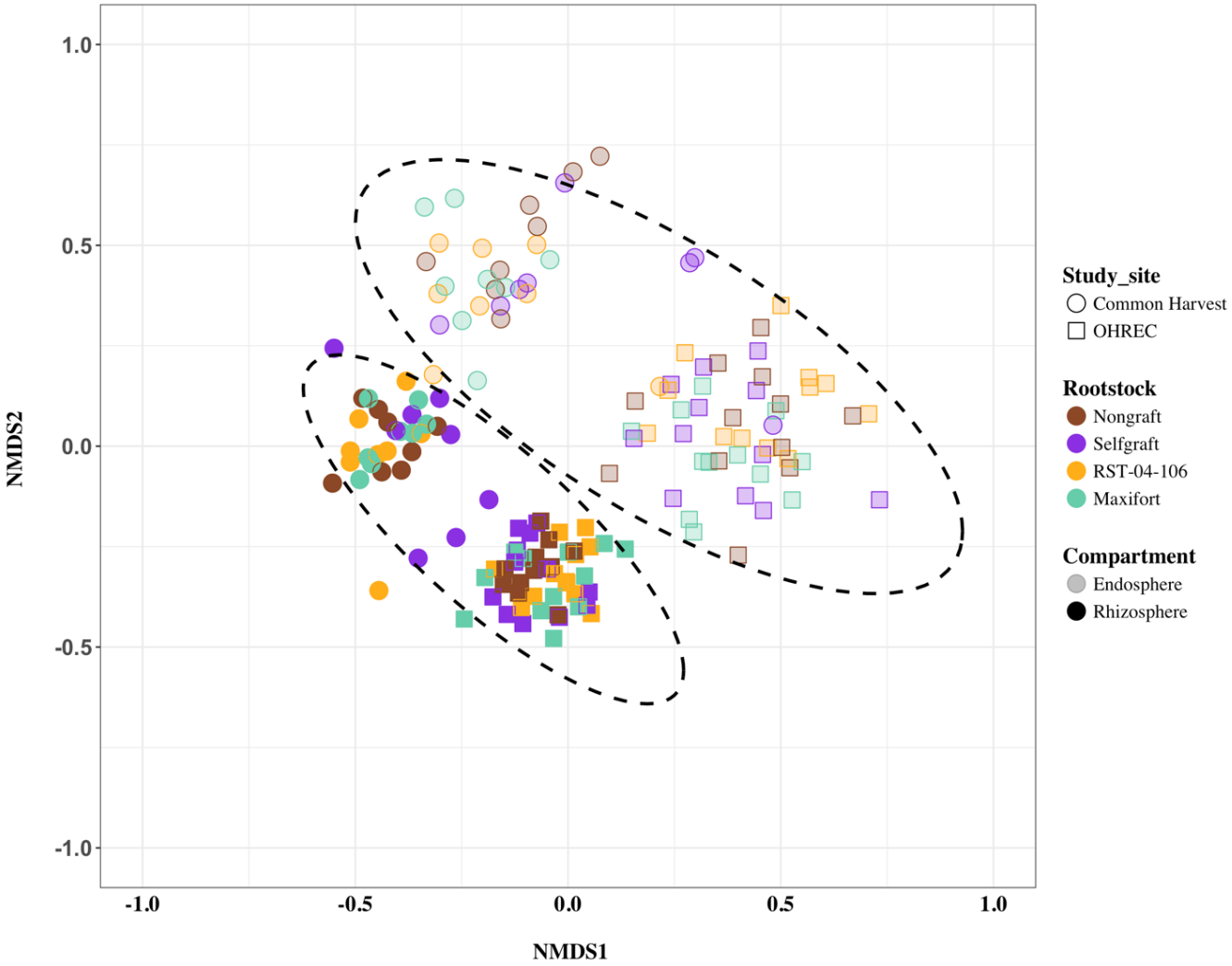


FIG S3 Non-metric multidimensional scaling (NMDS) ordination plot of samples labeled by tomato rootstock (nongraft BHN589, selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (RST-04-106 and Maxifort)), compartment (endosphere or rhizosphere), and study site, based on the Bray-Curtis dissimilarity distance matrix of fungal OTUs. Color indicates rootstock treatment, shape indicates study site, and size indicates compartment. Ellipses surrounding the samples indicate the 95% CI of the endosphere and rhizosphere sample centroids.

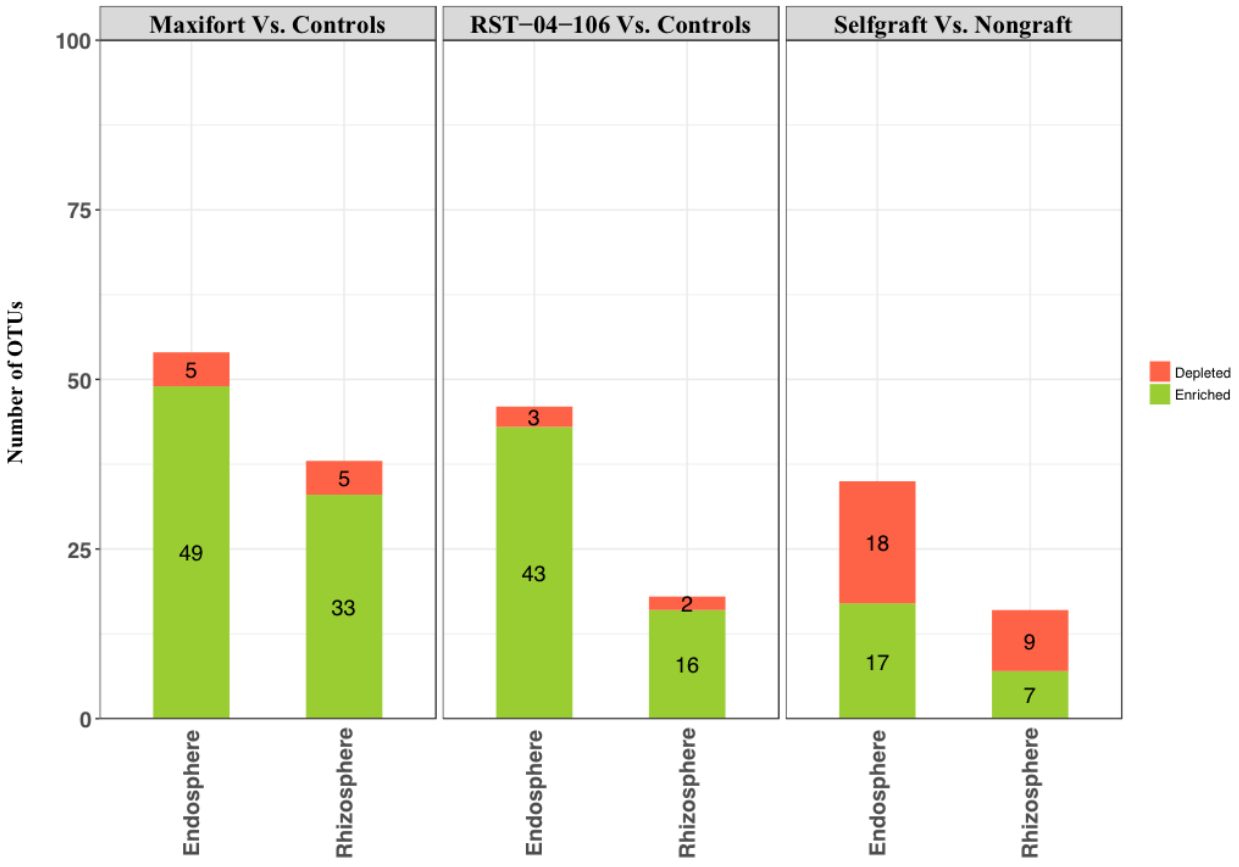
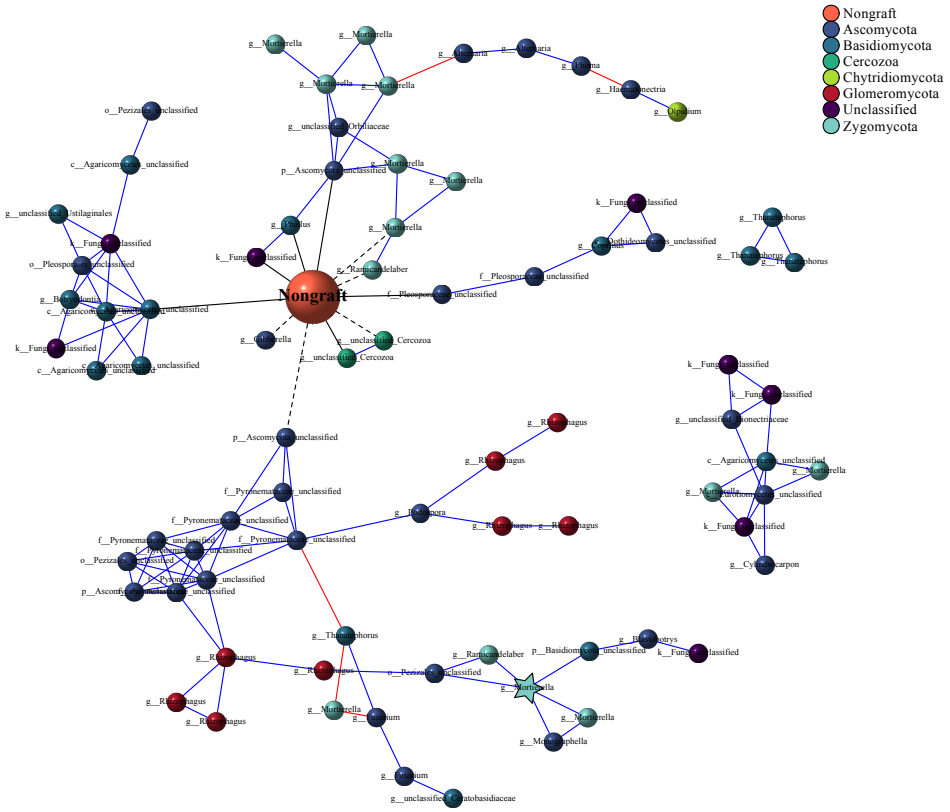
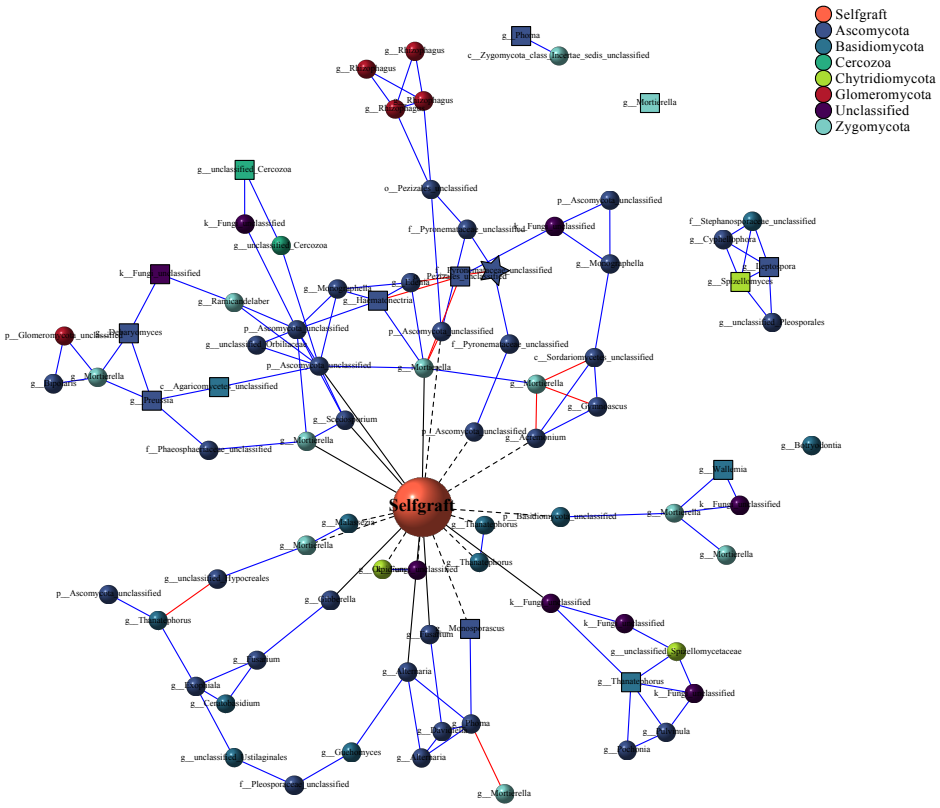


FIG S4 Number of DAOTUs in a contrast analysis, evaluated for the endosphere and rhizosphere compartments for four tomato rootstock treatments: nongraft and selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (Maxifort and RST-04-106). The green color in each bar represents the number of enriched taxa, and the red color represents the number of depleted taxa. The number of differentially changed taxa was greater for the endosphere than for the rhizosphere. Among the contrast pairs, hybrid rootstocks had a greater number of enriched taxa compared to depleted taxa. However, the number of depleted taxa was higher compared to enriched taxa in the controls. Among the treatments, Maxifort had the highest number of DAOTUs in both compartments.

A)


Nongraft
Ascomycota
Basidiomycota
Cercozoa
Chytridiomycota
Glomeromycota
Unclassified
Zygomycota
Nongraft

B)


Selfgraft
Ascomycota
Basidiomycota
Cercozoa
Chytridiomycota
Glomeromycota
Unclassified
Zygomycota
Selfgraft

C)

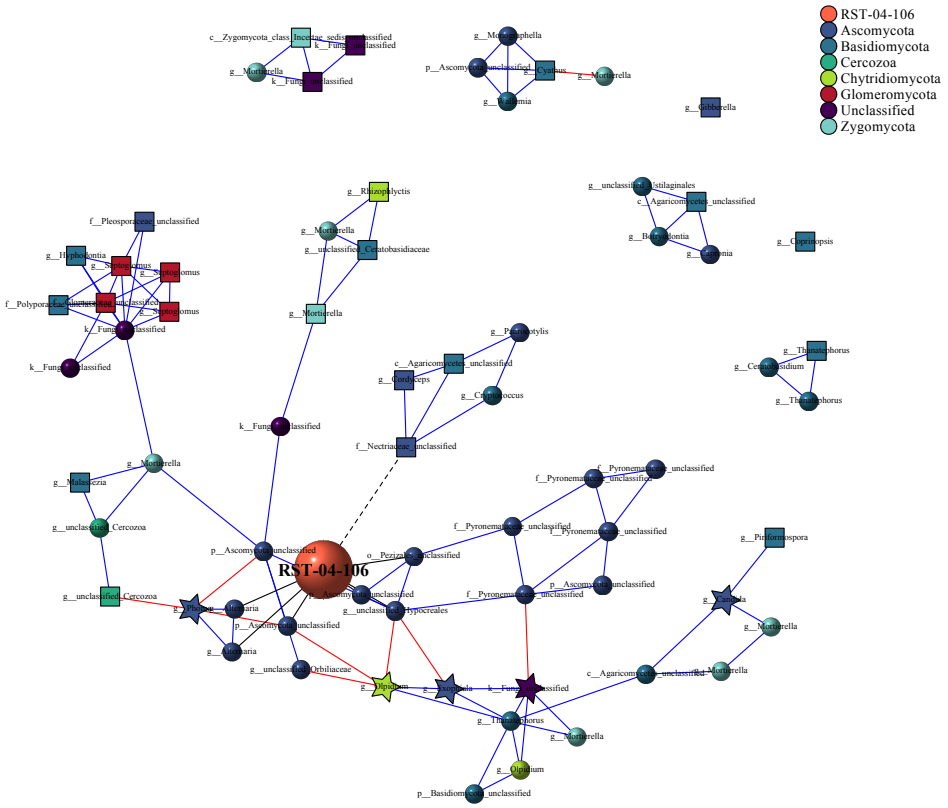
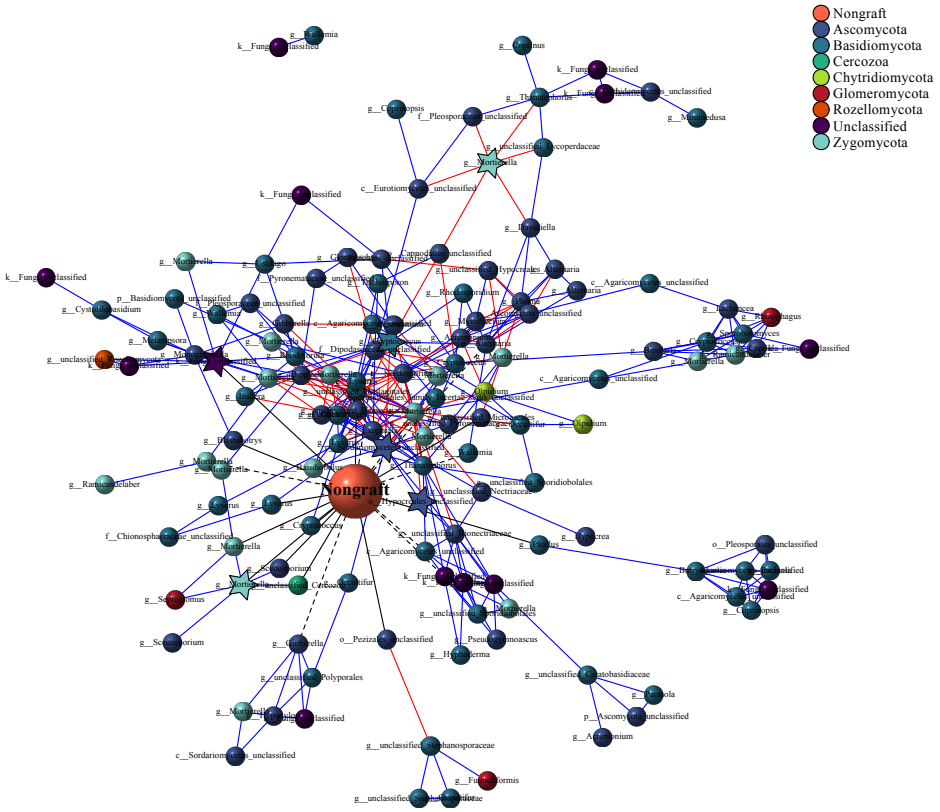
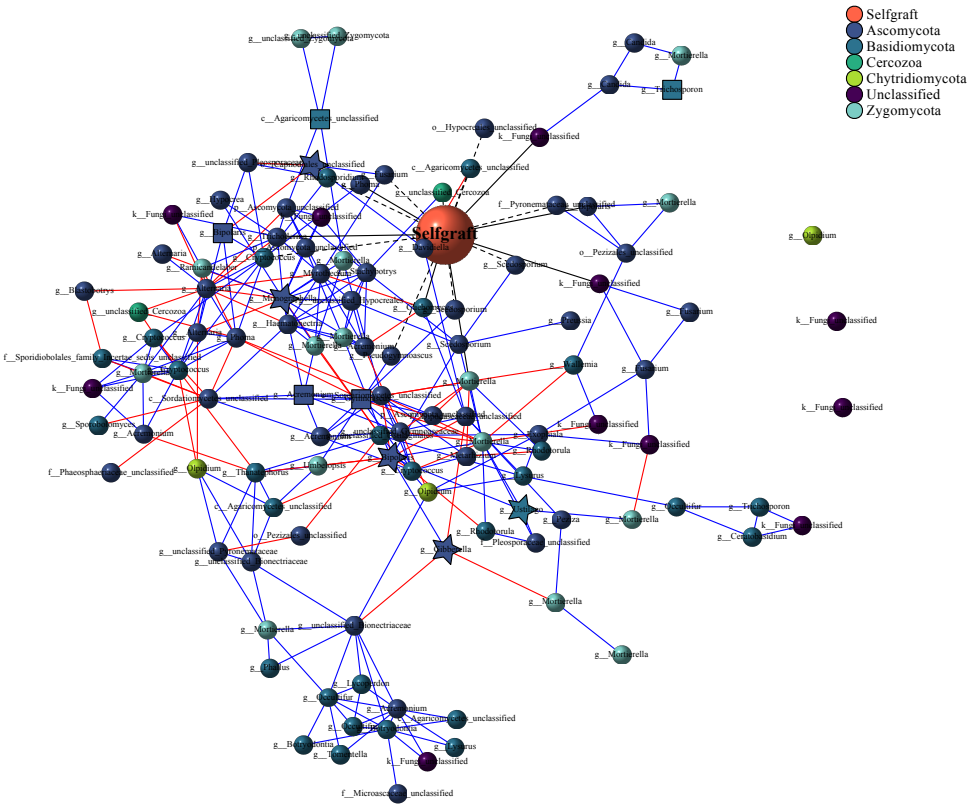


FIG S5 Phenotype-OTU network analysis (PhONA) of **endosphere** fungal taxa for tomato rootstock treatments: (A) nongraft and (B) selfgraft BHN589, and BHN589 grafted on (C) RST-04-106.Node color indicates the phylum, except that the tomato-color node represents yield associated with the rootstock. Nodes connected to the rootstock yield node with black links are taxa that were predictive of rootstock yield, where dotted and solid lines indicate negative and positive associations with the yield node, respectively. Red and blue links represent negative and positive associations, respectively, between OTUs. Nodes are labeled with the lowest available taxonomic information. Square-shaped and star-shaped nodes represent DAOTUs and key nodes detected in role analyses, respectively.

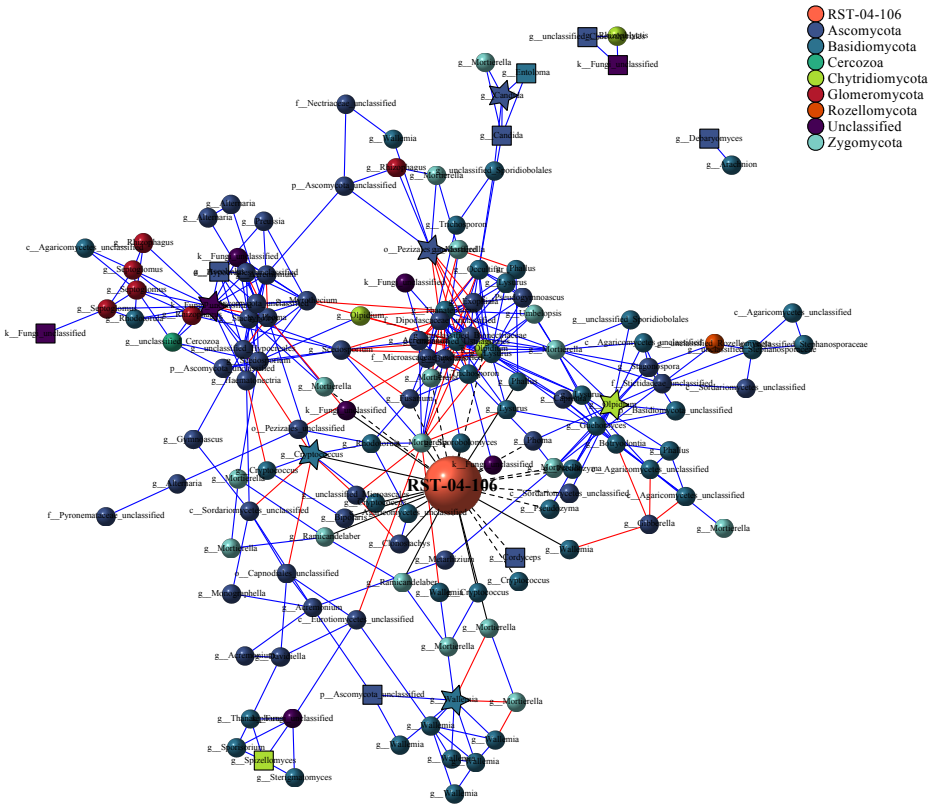
A)


Nongraft
Ascomycota
Basidiomycota
Cercozoa
Chytridiomycota
Glomeromycota
Rozellomycota
Unclassified
Zygomycota
Nongraft

B)


Selfgraft
Ascomycota
Basidiomycota
Cercozoa
Chytridiomycota
Unclassified
Zygomycota
Selfgraft

C)


RST-04-106
Ascomycota
Basidiomycota
Cercozoa
Chytridiomycota
Glomeromycota
Rozellomycota
Unclassified
Zygomycota

D)

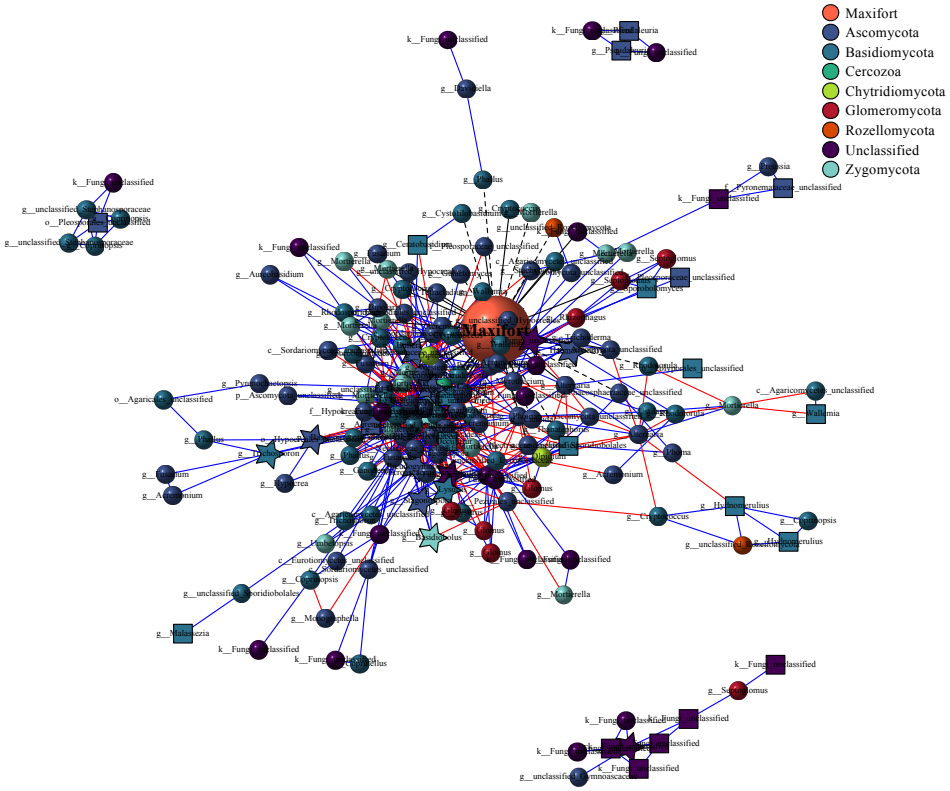


FIG S6 Phenotype-OTU network analysis (PhONA) of **rhizosphere** fungal taxa for tomato rootstock treatments: (A) nongraft and (B) selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks ((C) Maxifort and (D) RST-04-106). Node color indicates the phylum, except that the tomato-color node represents yield associated with the rootstock. Nodes connected to the rootstock yield node with black links are taxa that were predictive of rootstock yield, where dotted and solid lines indicate negative and positive associations with the yield node, respectively. Red and blue links represent negative and positive associations, respectively, between OTUs. Nodes are labeled with the lowest available taxonomic information. Square-shaped and star-shaped nodes represent DAOTUs and key nodes detected in role analyses, respectively.

TABLE S1 Sites included in the study, their soil type, and geographic coordinates.

Study sites	Location	Latitude	Longitude	Soil type
Olathe Horticulture Research and Extension Center (OHREC)	Johnson County, KS	38.88N	94.99W	Chase silt loam
Common Harvest	Douglas County, KS	38.96N	95.20W	Eudora-Kimo complex

143 **TABLE S2** Results of the multivariate permutational analysis of variance (PERMANOVA) for
144 fungal taxon abundance data. Permutation was based on the Bray-Curtis distance matrix
145 generated for root associated fungal communities at the OTU level from four tomato rootstock
146 treatments: nongraft and selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks
147 (Maxifort and RST-04-106) (1000 permutations). P values < 0.05 are in bold.

Factor	Sum of Squares	% Explained	P value
Rootstocks	1.24	2.07	**< 0.01**
Compartment	5.36	8.92	**< 0.001**
Study_Site	5.01	8.34	**< 0.001**
Year	3.23	5.38	**< 0.001**
Rootstocks:Compartment	0.58	0.97	0.972
Rootstocks:Study_Site	1.39	2.32	**< 0.01**
Compartment:Study_Site	1.59	2.65	**< 0.001**
Rootstocks:Year	1.00	1.66	0.111
Compartment:Year	1.00	1.66	**< 0.001**
Study_Site:Year	1.46	2.43	**< 0.001**
Rootstocks:Compartment:Study_Site	0.55	0.91	0.998
Rootstocks:Compartment:Year	0.49	0.82	1.000
Rootstocks:Study_Site:Year	1.22	2.03	**< 0.01**
Compartment:Study_Site:Year	0.60	1.00	**< 0.01**
Rootstocks:Compartment:Study_Site:Year	0.60	1.00	0.989
Residuals	34.75	57.86	
Total	60.07	100.01	

148

149 **TABLE S3** Network attributes and links observed in the fungal association networks for four tomato rootstock treatments: nongraft

150 and selfgraft BHN589, and BHN589 grafted on two hybrid rootstocks (Maxifort and RST-04-106) in each of rhizosphere and

151 endosphere compartments.

152

Compartment	Rootstocks	Node	Edge	Node Degree	Density	Modules (SA)	Negative Edge	Positive Edge	Negative/Positive Edge
Endosphere	Nongraft	605	155	0.26	0.0004	18	5	150	0.03
	Selfgraft	600	145	0.24	0.0004	20	9	136	0.07
	RST-04-106	584	156	0.27	0.0005	17	10	146	0.07
	Maxifort	700	221	0.32	0.0005	20	36	185	0.19
Rhizosphere	Nongraft	950	381	0.40	0.0004	17	82	299	0.27
	Selfgraft	995	312	0.31	0.0003	19	67	245	0.27
	RST-04-106	991	397	0.40	0.0004	23	58	339	0.17
	Maxifort	1175	785	0.67	0.0006	18	289	496	0.58

153