Microbiome Networks: A Systems Framework for Identifying Candidate Microbial Assemblages for Disease Management

R. Poudel, A. Jumpponen, D. C. Schlatter, T. C. Paulitz, B. B. McSpadden Gardener, L. L. Kinkel, and K. A. Garrett

First and seventh authors: Plant Pathology Department, Institute for Sustainable Food Systems, and Emerging Pathogens Institute, University of Florida, Gainesville 32611-0680; second author: Division of Biology and Ecological Genomics Institute, Kansas State University, Manhattan 66506; third and fourth authors: U.S. Department of Agriculture-Agriculture Research Service, Wheat Health, Genetics, and Quality Research Unit, Washington State University, Pullman, WA 99164; fifth author: Department of Plant Pathology, The Ohio State University-OARDC, Wooster 44691; and sixth author: Department of Plant Pathology, University of Minnesota, St. Paul 55108.
Current address of fifth author: Sustane Natural Fertilizers, Cannon Falls, MN 55009.

ABSTRACT

Poudel, R., Jumpponen, A., Schlatter, D. C., Paulitz, T. C., McSpadden Gardener, B. B., Kinkel, L. L., and Garrett, K. A. 2016. Microbiome networks: A systems framework for identifying candidate microbial assemblages for disease management. Phytopathology 106:1083-1096.

Network models of soil and plant microbiomes provide new opportunities for enhancing disease management, but also challenges for interpretation. We present a framework for interpreting microbiome networks, illustrating how observed network structures can be used to generate testable hypotheses about candidate microbes affecting plant health. The framework includes four types of network analyses. "General network analysis" identifies candidate taxa for maintaining an existing microbial community. "Host-focused analysis" includes a node representing a plant response such as yield, identifying taxa with direct or indirect associations with that node. "Pathogen-focused analysis" identifies taxa with direct or indirect associations with taxa known a priori as pathogens. "Disease-focused analysis" identifies taxa associated with disease. Positive direct or indirect associations with desirable outcomes, or negative associations with undesirable outcomes, indicate candidate taxa. Network analysis provides characterization not only of taxa with *direct* associations with important outcomes such as disease suppression, biofertilization, or expression of plant host resistance, but also taxa with *indirect* associations via their association with other key taxa. We illustrate the interpretation of network structure with analyses of microbiomes in the oak phyllosphere, and in wheat rhizosphere and bulk soil associated with the presence or absence of infection by *Rhizoctonia solani*.

Additional keywords: biocontrol, networks, phytobiome, *Quercus macrocarpa*, *Triticum aestivum*.

Individual plants are complex systems whose productivity is a function of interactions with and among diverse organisms associated with them, as well as with their abiotic environment. These complex networks of interaction have profound effects on plant health, stress tolerance, growth dynamics, and yield. Plant microbiome research represents the most recent development of fields of inquiry previously referred to as soil microbiology, plant microbial ecology, or plant–microbe interactions. As the tools of these research areas change, so do the foci of investigations. Nearly 20 years ago, plant pathologists first started using the emerging molecular tools of microbial ecology to more fully characterize the uncultured fraction of microorganisms associated with plants and soils, to gain insights previously unavailable to a science heavily reliant on culture-dependent methods. Prior to low cost, high-throughput sequencing, molecular profiling methods, such as ARISA (Robleto et al. 1998), DGGE (Postma et al. 2000), and T-RFLP (McSpadden Gardener and Weller 2001), were used to characterize the relative abundance and diversity of microorganisms that could affect plant health. While relatively crude, such approaches proved useful for discovering new species of biocontrol agents that suppress pathogens (e.g., Benitez and Gardener 2009). With the rise of new sequencing technologies, the depth and complexity of microbial community profiling has greatly expanded. However, applications of microbiome studies have continued to focus on the identification of functionally important microbial populations based on correlations and/or differential abundances of molecular markers (Debenport et al. 2015; Mazzola et al. 2015). The limitation of that approach is that it generally considers each group of markers in isolation, ignoring the likely connections that occur among the populations co-inhabiting the phytosphere. Network analysis offers a means to evaluate not only direct interactions, but also potential chains of indirect interactions among community members. Network models offer a new perspective on microbial communities associated with plants and soils, and they are quickly becoming a common feature of microbiome analyses.

As sequencing technologies evolve, the large number of microbial operational taxonomic units (OTUs) recovered from microbiome studies presents a challenge for analysts of microbial community structure and function. The most common approaches to evaluating microbial communities have been estimation of diversity, and multivariate statistical approaches such as non-metric multidimensional scaling or principal coordinate analysis to classify sample sets by some discriminating variable, such as location, soil type, or plant health status. The two data sets included in new network analysis case studies in this paper were previously analyzed in papers that illustrate this general approach (Jumpponen and Jones 2010; Yin et al. 2013). These approaches provide information about overall community composition, but they fail to capture the potential interactions among the OTUs. Understanding interactions among microbial populations, their plant hosts, and the abiotic environment can open up new opportunities to better manage microbial communities crucial to plant health.

Network models can portray the members of a microbial community along with inference about their interactions. In

microbiome network models, a node typically represents a taxonomic group (such as a named species or sequence-defined OTU). A link is defined between two nodes if they have some type of relationship, most commonly a positive correlation in frequency. These measures of relationship represent microbial co-occurrence and may capture associations based on ecological properties of soil (Barberan et al. 2012; van Bruggen and Semenov 2000), the methane cycling potential of roots (Edwards et al. 2015), or shared bacterial communities among plant species (Bakker et al. 2014). Similarly, networks of host co-occurrence may also be useful for anticipating sharing of pathogens among hosts (e.g., Cox et al. 2013), and the resulting ecological and evolutionary effects on pathogens and other microbes. Further, microbiome network models can be used to evaluate the dynamic structure of microbial communities based on network topologies, providing a quantitative characterization of a complex community. With advances in technologies for microbial characterization, network analysis of microbiomes sampled at very frequent intervals may allow prediction of causal relationships between taxa even based on relative abundance and co-association. Such analyses at frequent intervals could also aid in defining the resilient or background microbial communities in contrast to taxa that fluctuate greatly as a function of time or respond to experimental treatments.

The implication of network topologies has been described for many types of systems (Barberan et al. 2012; Chadès et al. 2011; Edwards et al. 2015; Faust and Raes 2012; Faust et al. 2012; Shaw and Pautasso 2014; Zhou et al. 2010). Many of these applications use network topologies to infer the functions, structures, and dynamics of interacting entities, in social science (Barrat et al. 2004; Germanos et al. 2015), food webs (Kefi et al. 2012), molecular biology (MacNeil and Walhout 2011), engineering (Barrat et al. 2004), and ecology (Sole and Montoya 2001). For example, topological features (Box 1) such as mean node degree indicate the mean number of interactions among the nodes in a system, where highly connected nodes are referred to as hub nodes. Hub nodes are useful in targeting key locations for epidemic surveillance (Grange et al. 2014; Hernandez Nopsa et al. 2015), allocating resources for

---

BOX 1: KEY TRAITS OF NETWORKS

Nodes: Nodes, or vertices, are the individual entities that are the building blocks of a network. Depending on the system, microbes, proteins, genes, or agents can be represented as nodes. In our case studies, nodes are OTUs and, in some cases, host or pathogen response variables.

Links: Links, or edges, are connections between nodes in a network, which define the relation/interaction between the nodes. In our examples, links represent the positive or negatives associations between taxon frequencies.

Node degree: A node's degree is the number of links between itself and other nodes. The frequency distribution of node degree provides information about the nature of a network, and has often been used to infer network robustness.

Module: A set of nodes connected to each other by many links, while connected by few links to nodes of other groups.

Clustering coefficient: The clustering coefficient is a ratio of the number of links between the neighbors of a node, and the maximum number of links that could possibly exist between its neighbors. The clustering coefficient of a node is always a number between 0 and 1.

Betweenness centrality: The number of times a node acts as a bridge along the shortest path between two other nodes. The betweenness centrality of a node reflects the function of a node to connect subnetworks.

---

effective conservation of endangered species (Grange et al. 2014), and developing strategies for immunization (Bai et al. 2007; Wang et al. 2009). The number of modules in a network is another topological feature (e.g., Bakker et al. 2014), along with many other potential characterizations (Box 1) (Barberan et al. 2012; Zhou et al. 2010). Thus, applying knowledge about network models, developed across multiple disciplines, to the analysis of microbiome data and associated relevant variables (such as plant, soil, and water measurements) offers an opportunity to enhance microbiome-based disease management and biofertilization of crops.

Network models provide tools for understanding microbiomes as a system, and how microbiome structure can influence crop health. When pairs of candidate biocontrol taxa, selected because of promising performance as individual biocontrol agents, were deployed together, the combination was often less effective for biocontrol than use of an individual taxon (Xu et al. 2011). Designing useful synthetic communities based on microbiomes provides a direct way to harness the benefits of microbiome research (Knight 2015). However, given that microbial communities are often hyperdiverse (Jumpponen and Jones 2009) and that understanding of the functionality of the majority of taxa in real production systems is limited, designing a synthetic community model will be challenging. To address the challenge, many researchers have used host phenotypes to define the nature of microbiomes. Preferential selection of microbes by plant hosts, a feature long ago established using culture-dependent methods (Curl and Trulove 1986), has been thoroughly demonstrated for multiple host types (Edwards et al. 2015; Lundberg et al. 2012; McKenzie et al. 2012; Peiffer et al. 2013) across multiple biomes (Faust et al. 2015). Such associations have been extended to connect microbiome and biochemical data, allowing researchers to suggest linkages at both the ecological and biochemical levels. For example, recent work compared the host influence on the soil microbial assemblage in a phytohormone mutant and a related wild type of *Arabidopsis thaliana* (Lebeis et al. 2015). A synthetic community was established based on an enriched pool of microbes that were only able to colonize the wild type and not the mutant. In a further step, work by Panke-Buisse et al. (2015) reported successive enrichment of the microbiome by harvesting the soil associated with specific plant phenotypes over multiple generations, showing how microbiomes can be selected to modify plant traits and physiology. Although the rationale for selecting subsets of taxa raises multiple questions, host phenotype-based selection of microbes followed by an enrichment assay is promising. These results support the idea that synthetic communities of select microbial agents can be deployed, with the potential to successfully colonize hosts, providing new avenues for modulating the phytobiome with the aim to enhance crop yields and/or sustainability (Berendsen et al. 2012). Analysis of microbiome networks offers an approach for selecting subsets of candidate microbes for testing in synthetic communities.

In this paper, we present a framework for selecting candidate sets of functionally important plant-associated microbes based on network analysis of microbiomes. This framework serves as a general guide for interpreting microbiome networks, taking into account which types of information about the system are and are not typically available. We illustrate how the observed structure of networks can be used to explore the possible biological roles of microbes in plant disease development, biological control, and biofertilization, as well as microbial taxa that may be targets for selection in crop breeding. The main objective of this paper is to develop this framework for evaluating the OTUs of microbiome networks that are likely to be important for disease management, and to illustrate the application of this framework to the phyllosphere microbiome of *Quercus macrocarpa* and the soil and rhizosphere microbiomes associated with the presence or absence of infection of wheat by *Rhizoctonia solani.* Although there are many limitations to interpreting microbial networks with confidence based on current technologies, this exploratory framework

outlines steps to identify candidate taxa in a microbiome-based strategy for disease management and crop improvement.

FRAMEWORK

Despite the utility of network models for capturing and displaying the complexity of microbial interactions, there are a number of challenges for their interpretation and for translating these analyses into disease management strategies. One challenge is based on the fact that co-occurrence of two taxa may result from a number of different biological and/or methodological processes. Several methods for detecting association, such as correlation or proportionality, have been used, often with filters based on statistical cut-offs and P value criteria (Bakker et al. 2014; Barberan et al. 2012; Edwards et al. 2015), or correlation measures combined with random matrix theory to define the threshold of association (Zhou et al. 2010). Other methods include those in SparCC, designed to reduce the spurious correlation associated with compositional bias (Friedman and Alm 2012; Kurtz et al. 2015), and local similarity analysis, which evaluates the temporal association of species and environmental variables (Ruan et al. 2006). Each method used to define an association provides a different perspective; thus, careful interpretation should take into account the caveats of each mathematical approach, as well as the characteristics of the data type evaluated.

In this framework, we first discuss general considerations for interpreting microbiome networks, such as methods for defining whether links exist, the question of whether links represent biological interactions or shared niches, and compositional bias. Second, we discuss four scenarios for interpreting microbiome networks to identify useful taxa: general analyses of network structure, host-focused analyses, pathogen-focused analyses, and disease-focused analyses. In each scenario, the goal is to identify candidate taxa that may have particularly important roles in plant health, and thus serve as useful components of a synthetic community to support plant health. Two case studies illustrate the application of these analyses.

General considerations for interpreting microbiome networks. A natural question about links in a network of OTUs is whether they represent biological interactions, or environmental preferences detected because of the sampling scheme. We illustrate how detected links may indicate a mixture of biological interactions and niche preferences, when networks are evaluated with links based on proportionality or correlation (Box 2). For purposes of this illustration, the biological interactions and habitat preferences of each taxon are known a priori (in contrast to most microbiome studies). The choice of which environmental conditions to sample, and at what scale, yields different network structures (as do different choices about definitions of association). For example, no link is detected between D and E when sampling is solely from diseased plants, because D and E both have consistently too low a frequency to be detected in these samples. However, D and E's shared niche produces a positive link when sampling includes both diseased and healthy plants, because D and E both have an undetectable frequency in diseased plants and both have a consistently high frequency in healthy plants (Box 2). Thus, cautious interpretation of networks is essential, keeping in mind the different mechanisms by which a link can be generated from data.

The majority of published studies of microbiome networks appear to focus on positive correlations among taxa. One important reason studies may avoid negative correlations is to avoid the problem of compositional bias. Using current technologies, researchers typically evaluate the *frequency* of an OTU, the number of sequences for the OTU divided by the total number of sequences obtained in the sample, rather than absolute abundance. As a result, if a common taxon increases in abundance in a sample, there will be an observed decrease in frequencies for many other taxa, even if their (unknown) abundance remains unchanged. Interpreting negative correlations is more problematic because of this compositional bias. Yet, these negative correlations are key to understanding biological control and disease suppression. In this framework, we take the approach that negative correlations are important for identifying candidate taxa, where some proportion of taxa will not "pan out" in further evaluation for disease management, because they were originally identified only due to compositional bias or because they consistently occupied different niches than pathogens.

Two other important issues for interpretation are directionality and the distinction between correlation and partial correlation. Networks are directed if links are associated with a direction. For example, taxon A might benefit taxon B, while taxon B has no effect on taxon A, and this could be indicated in a network by an arrow leading from A to B, without any reciprocal arrow from B to A. When networks are constructed based on correlation, the directionality will typically be lost, and there will simply be evidence for a positive association between A and B. Bayesian network analysis has the potential to detect more information about directionality in relationships among taxa (Darwiche 2009). The ability of Bayesian network analysis to deal with conditional probabilities also provides the potential to better discriminate cases where A affects B only through effects on C compared with A having a direct effect on B. However, for current microbiome data, the number of responses (OTU frequencies, etc.) is typically much greater than the number of samples, and this makes fitting Bayesian networks impractical for complete microbiome networks until larger sample sizes are routinely available.

Finally, it is also important to keep in mind that the most important taxa for plant health may appear not to have any links to other taxa, nor to have correlations with plant or pathogen performance. In observational studies, the absence of any links or correlations can occur if the sampling distribution of frequencies for a taxon is too limited, such that there is not enough range in frequency to evaluate associations. For example, if a taxon is always present in samples at a consistently high frequency, no associations will be detected, even if the absence of that taxon would have major effects.

General network analysis. In many analyses, the evaluation of a soil or plant-associated community is exploratory, and not closely related to specific plant outcomes or involving known pathogens. In this case, there may be interest in maintaining the existing microbial community, under the assumption that it is a desirable community. What are candidate taxa that may be important to the maintenance of structure and function of a community? In a network where links represent *positive* associations (Box 3 and Fig. 1A), some taxa may have limited importance from the standpoint of maintaining network structure. Guimera and Amaral (2005) described an approach to categorize nodes into functional groups based on the modular structure of networks, applying this approach to metabolic networks for different organisms. Peripheral taxa have few links to other taxa. Modular hub taxa are highly connected within a module. Connector taxa provide a link among multiple modules. Network hub taxa are highly connected both in general and within a module. From the standpoint of this general structure, the taxa that may be most important for maintaining a network are network hubs, connectors, and, perhaps secondarily, modular hubs. In addition to the key nodes, negative associations may also have a role in community resilience (Shade et al. 2012), despite the difficulty in assessing negative associations using current technologies.

Modular hubs may be community components that facilitate the stable occurrence of many other taxa. These other taxa may possess attributes that we find desirable for either growth promotion or integrated pest management, in some cases. The hubs themselves need not be the organisms with desirable traits. Rather, the modular hubs may be the foundation species—or keystones—whose presence supports the establishment and growth of other taxa that may possess the desirable attributes. Module hubs may be important candidate taxa for synthetic communities in experiments designed to evaluate plant responses or to maintain plant productivity. Recent

studies (Agler et al. 2016) showed that plant responses to other microbes may be through a hub node, where a hub node maintained microbial diversity. One of the hub nodes observed was a pathogen, illustrating both the importance of the network structure for community stability, and the fact that community stability will not always be desirable for plant health.

Host-focused network analysis. In other analyses, a plant variable (or variables) such as yield will be part of the data set, and can be represented as a node in the network of variables. Analyses to assess potential biofertilizer components would exemplify this application, as well as identification of "subclinical" parasites, plant antagonists, or previously unidentified pathogens. In this case, the general network analysis may still be of interest, but particular focus is likely to be on the taxa that are directly or indirectly linked to the plant variable of interest. Approaches analogous to the analysis of
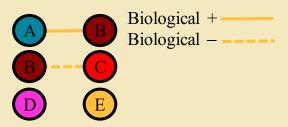
quantitative trait loci may be useful for identifying important taxa, particularly for direct effects (Garrett et al. 2012). In a network representing only positive associations, all those taxa that are directly or indirectly linked to a desirable plant variable may be candidate taxa. However, as the path between a taxon and the host response node runs through more and more other taxa, the probability that some links are artifacts increases. In a network representing both positive and negative associations, star network motifs offer a simple interpretation for candidate taxa, but motifs such as lines are more complicated (Fig. 1B).

Pathogen-focused network analysis. When information about the frequency of one or more pathogens is available, the analysis is similar to the host-focused analysis, except that generally higher values of the pathogen response variables will be undesirable. Thus, in a network representing only positive associations with the
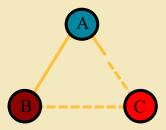

BOX 2

Illustration of how biological interactions and niche preferences may be detected when networks are evaluated with links based on measures of association such as correlation or proportionality. Nodes A to E represent five hypothetical taxa, and ii to iv represent three hypothetical sampling scenarios.

  (i) Underlying biological interactions and niche preferences, which will typically not be known. Taxa A, B, and C are abundant enough when hosts are diseased to be detected in samples, and Taxa D and E are abundant enough when hosts are healthy to be detected in samples. A and B have a positive biological interaction, B and C have a negative biological interaction, while D and E lack a biological interaction.

  (ii) Estimated links when sampling diseased hosts. A and B have a positive association, and B and C have a negative association based on biological interactions; A and C may have a negative link if considering simple correlation, or no link for conditional correlation. (D and E are not detected in samples.)

  (iii) Estimated links when sampling healthy hosts. D and E have no link because only a single type of environment is being sampled and there is no biological interaction. (A, B, and C are not detected in samples). Sampling in a single environment could tend to result in D and E always being present at high frequency, such that there would be no correlation to detect across samples.

  (iv) Estimated links when sampling both diseased and healthy hosts. A and B have a strong positive link, based on both shared niche and biological interaction; A and C, and B and C, have a positive association because of a shared niche, but a negative influence from biological interaction: thus, whether the estimated correlation is positive or negative depends on the relative strength of biological interactions and niche preference; A, B, and C all have negative links with D and E based on niche; D and E have a detectable positive association because they occur at too low a frequency to be detected in diseased hosts, and both occur at high frequency in healthy hosts.

i) Biological interactions and niches
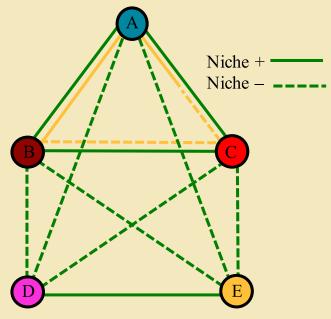
ii) Sampling under diseased conditions

iii) Sampling under healthy conditions

iv) Sampling across conditions

pathogen response variable, there will be no candidate taxa (Fig. 1B, in a case where the green node is a pathogen fitness variable, in a positive star). Instead, the candidate taxa will show up in a network representing both positive and negative associations. This type of analysis may be most important when studying pathogens that can persist in the absence of a particular host (or when host data are simply lacking).

Disease-focused network analysis. System interpretation is strengthened when both host and pathogen information are available, and i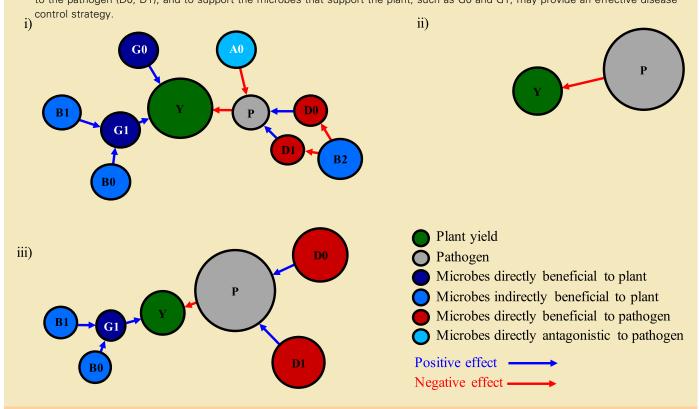deally information about disease. Candidate taxa are those with positive associations with desirable host responses and negative associations with pathogen frequency or level of disease. We illustrate example scenarios for healthy and diseased host states, where the interactions among a pathogen and other microbes are highlighted (Box 3). In this case we consider the diseased state to be associated with reduced plant yield or performance, while the healthy state is associated with higher yield and performance. This diagram indicates how diseased or healthy states are often a result not only of a pathogen's direct effects, but also of the indirect

---

BOX 3

Plant yield "Y" (green node) represented as a function of microbial interactions when a pathogen taxon "P" (gray node) is introduced. Dark blue nodes represent microbes that are directly beneficial to the plant. Medium blue nodes represent microbes that are indirectly beneficial to the plants via their effects on other microbes. Light blue nodes represent microbes antagonistic to the pathogen, whereas red nodes represent microbes that are beneficial to the pathogen. The size of the node represents the abundance of that taxon. Blue links indicate a positive effect and red links indicate a negative effect, where the direction of the arrow indicates the direction of the effect. i, ii, and iii represent three hypothetical scenarios reflecting plant states.

 (i) **Scenario with high yield, reflecting a relatively healthy state.** The plant is in a relatively healthy state as a result of microbial interactions. Although a pathogen is present in the system, it is suppressed due to the activity of beneficial and antagonistic microbes. The network indicates the topological function of the microbes, which may provide some guidelines in selecting microbes for biofertilization. For example, one microbe directly benefits the plant (G1) and functions as a connector between indirectly beneficial microbes B0 and B1 and the plant. Similarly, indirectly beneficial microbe B2 suppresses microbes (D0 and D1) that are directly beneficial to the pathogen.

(ii) **Scenario with one pathogen causing a disease in a plant.** There is a severe decrease in plant yield Y due to pathogen P. Finding a single antagonistic microbe specific to a single pathogen species would have been considered a reasonable strategy in early approaches to biocontrol. However, selecting only a single biocontrol agent fails to incorporate the potential roles of the whole microbial community in disease etiology, and thus may be less successful and sustainable. This figure describes a scenario in which a single pathogen species causes a disease in isolation, which may prove to be true for few pathogen species as microbiomes are better understood.

(iii) **Scenario with low yield reflecting a diseased state.** In this case, considering the whole network of microbial interactions suggests that the diseased state is not only a function of a single pathogen, but results from a perturbation in the community. Compared to the first scenario, some of the beneficial taxa (A0, G0, and B2) are no longer present, the abundance of the pathogen (P) and pathogen associated taxa (D0 and D1) is higher, and the abundance of other beneficial taxa (B0, B1, and G1) is lower. In this scenario, a more effective approach to designing biocontrol agents may be not only to target the pathogen, but also to target the friends of the pathogen on one hand, while supporting the enemies of the pathogen on the other hand–applying the old saying "the enemy of my enemy is my friend". In this hypothetical scenario, a microbial assemblage consisting of potential taxa to inhibit both the pathogen (P) and the microbes that are directly beneficial to the pathogen (D0, D1), and to support the microbes that support the plant, such as G0 and G1, may provide an effective disease control strategy.

effects of microbial interactions at a community level. While there may be scenarios where a single pathogen causes a disease without the influence of other microbial taxa, as microbial communities are better understood this may prove to be the exception rather than the rule. Identification of candidate taxa from this framework, for disease management or biofertilization, is based on the whole set of microbial interactions and their network topologies.

CASE STUDY METHODS

We illustrate the framework for identifying candidate taxa using observed microbiome data, in some cases in combination with hypothetical data for the sake of illustration. The microbiome data are typical current data sets in the sense that they include the relative frequency of a large number of OTUs, where there is some taxonomic information associated with most OTUs but little or no information about function.

Oak phyllosphere fungal communities. *Sampling and molecular processing.* We evaluated data previously described in Jumpponen and Jones (2010), by reanalyzing the sequence library using MOTHUR (Schloss et al. 2009). In these published analyses, the authors evaluated seasonal dynamics of fungal communities

present in the leaves of *Quercus macrocarpa* Michx. (bur oak) in isolated stands located in urban and nonurban sites in and outside of the city of Manhattan, Kansas. The study identified distinct temporal dynamics in the fungal communities. The methods, including detailed stand location information and the sampling strategy, have been described previously. Briefly, leaf samples were collected from 70 individual oak trees, half from urban areas and half from rural sites. The previous publication described responses in fungal diversity and community composition in the context of temporal dynamics and land use (with no network analysis). In the current analyses, we applied the framework described above and treated each tree sample as representing an independent experimental unit to build a fungal co-occurrence network.

The details of sample processing have been described in Jumpponen and Jones (2009, 2010). Briefly, the total DNA was extracted from leaf disks (MOBIO UltraClean Soil DNA Isolation Kit) in a FastPrep tissue homogenizer. The DNA extracts were PCR-amplified with ITS1F and ITS4 primers (Gardes and Bruns 1993) combining platform specific A-primer and B-primer for 454-GS-FLX sequencing of the Internal Transcribed Spacer 2 (ITS2) region. The primers included sample-specific Molecular Identifier (MID) tags that allowed sample separation postsequencing.
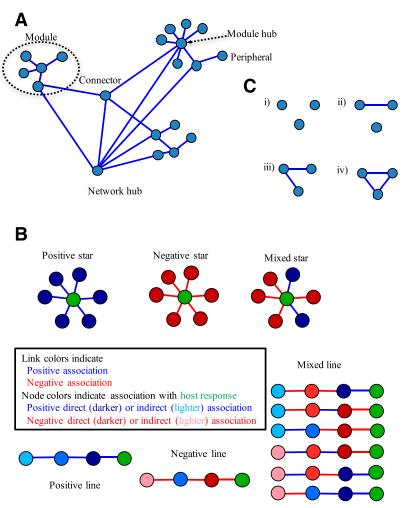

Fig. 1. Illustration of potential network structures and interpretations. **A,** A network with links that represent positive associations, illustrating nodes that can be classified as a peripheral, a modular hub, a connector, and a network hub. **B,** Network motifs that may be observed in plant-focused network analyses, and the candidate taxa identified in them. Green nodes represent a desirable plant response variable, blue nodes indicate microbial taxa with positive direct or indirect associations with the plant response, and red nodes indicate microbial taxa with negative direct or indirect associations with the plant response. The darkest shaded blue or red nodes indicate direct associations with the plant response ("first degree association"), while lighter shaded blue or red nodes indicate indirect associations with the plant response ("second degree or third degree associations"). (Note that if, instead of desirable host responses, the green nodes were to represent an undesirable response such as pathogen frequency, then the candidate taxa would be those with a negative association that are red). **C,** Triad analyses representing possible subnetwork structures among sets of three taxa (motifs). Considering (i) no co-associations within the triad, (ii) one pair of operational taxonomic units (OTUs) co-associated within the triad, (iii) two pairs co-associated within the triad, or (iv) all three OTUs positively co-associated within the triad. Blue links represent positive association.

Bioinformatics and sequence processing. The sequence data from the raw file (.sff) were curated using the MOTHUR pipeline (v.1.35.1; Schloss et al. 2009) following the steps as outlined in 454 SOP (http://www.mothur.org/wiki/454_SOP). Briefly, the fasta, qual, and flow files were extracted from .sff files, denoised, and any sequences shorter than 200 base pairs (bp) or containing ambiguous base calls or more than eight homopolymers or missing MIDs were removed from the library. The chimeric sequences were identified using UCHIME (Edgar et al. 2011) and removed. The sequences were assigned a taxonomy string using a Bayesian classifier (Wang et al. 2007) against the UNITE plus INSD nonredundant ITS database (Koljalg et al. 2013). Uncorrected pairwise distances between DNA sequences were used to cluster the sequences into OTUs at 97% similarity using the nearest-neighbor joining algorithm. Finally, consensus taxon affinities were assigned to the clustered OTUs. The number of sequences binning to each OTU and sample was recorded in a count matrix with columns as OTUs and rows as samples in the R programming environment (R Core Team 2015). To minimize the inflation of rare OTUs in the community analysis, we included only those OTUs with sequence count greater than 10 (Brown et al. 2015; Oliver et al. 2015), and that occurred in at least five of the individual samples.

Network construction. To minimize the effects of compositional bias, we used the SparCC method. This method evaluates the variance of the log-ratio for transformed data, rather than the relative abundance, to infer pairwise relations (Friedman and Alm 2012). Estimation of an average pairwise relation was based on 20 iterations, where the algorithm assumes the absence of a large number of co-occurring OTUs with strong correlations. Further, the OTU count data were resampled 100 times, and were used to generate randomized correlation tables. The randomized correlation matrix was then used to calculate bootstrapped P values for each pairwise correlation. The resulting correlation matrix was used to define links between OTUs in the network models. Two OTUs were considered to be linked if their absolute pairwise correlation was greater than 0.25 and there was strong evidence for the association ($P < 0.001$).

The resulting co-occurrence network consisted of 171 OTUs represented at the genus level, referred to here as nodes. Of these, 54 were linked nodes and the remaining were unlinked nodes based on the thresholds defined above, with a total of 102 links. A total of 78% of the links represented co-abundance, or positive associations (blue links in Fig. 2) between OTU pairs, whereas 22% represented co-exclusion, or negative associations (red links). We performed a "general network analysis" in this co-occurrence network to define the basic topologies of the microbial communities.

Rhizoctonia bare-patch bacterial communities: Background on role in disease suppression. In a long-term cropping system study in Washington State, suppression of *Rhizoctonia* patch of wheat was observed over a 15-year period, documented in Schillinger and Paulitz (2014). This disease, caused by *Rhizoctonia solani* AG-8, produces bare patches in the field where the wheat is severely stunted (Paulitz et al. 2002). It was first described in Australia, and found in the Pacific Northwest of the United States in the mid-1980s. The pathogen infects the seminal and crown roots of wheat and barley, but can also attack broadleaf crops. Unlike most soilborne diseases, the spatial distribution in the field can be easily observed. The disease can be a problem in no-till or direct seeding, especially when tillage is stopped during the conversion from conventional to no-till. In the cropping system rotation study which started in 1996 (Cook et al. 2002), patches were observed three years after the plots were converted to no-till. Over the next 10 years, the patches were mapped with a high-resolution GPS. New patches appeared from one year to the next, some patches persisted over a number of years, but other patches disappeared (Schillinger and Paulitz 2014). In addition, the patch area started to decline in year 8 and declined to almost zero by year 11. Before the decline

was complete, in 2008, 2009, and 2010, we sampled from replicated plots in the continuous wheat rotation. We sampled bulk and rhizosphere soil from diseased patches, adjacent healthy areas, and patches that had disappeared. DNA was extracted and amplified with the V3 region of the bacterial 16S ribosomal RNA gene, and pyrosequenced (454) as described by Yin et al. (2013). We hypothesized that bacterial communities in the diseased areas and the healthy areas of the field would differ. The initial microbiome analyses (Yin et al. 2013) sought to identify the full spectrum of bacteria that could be associated with bare patch disease suppression. The new network analysis of these data, presented here for the first time, was aimed at determining the relationships among these bacterial populations. Although many of the trends previously identified by analysis of variance were observed, network analysis provides a more detailed view of the relationships among the communities and identifies a set of candidate taxa based on interactions that could potentially be used for disease management.

Using a generalized linear model of the relative frequencies of OTUs in samples from healthy and diseased plants, we observed bacteria in the class *Acidobacteria* and the genus *Gemmatimonas* in higher frequencies in the rhizosphere of healthy plants outside the diseased patches, whereas *Dyella* and *Acidobacteria* subgroup Gp7 occurred at higher frequencies in recovered patches. In contrast, *Chitinophaga*, *Pedobacter*, *Oxalobacteriaceae* (*Duganella* and *Massilia*), and *Chyseobacterium* were more frequent in the rhizosphere of diseased plants within the patches. These trends were verified in soils using qPCR. We also cycled soils in the greenhouse by inoculating them with *R. solani* AG-8 and analyzing bacteria in rhizosphere and bulk soils. We verified that the same taxa from diseased rhizospheres accumulated in the greenhouse. Finally, we isolated one of these bacteria, *Chryseobacterium soldanellicola* and demonstrated its biocontrol activity against *R. solani* in the greenhouse, completing Koch's postulates of biological control, going beyond simple correlative studies (Yin et al. 2013).

Bioinformatic and sequence processing. Bacterial community data (V3 region of the 16S) from the 2008 samples described in Yin et al. (2013) were reanalyzed to generate an OTU table for co-occurrence network construction. Briefly, sequences were removed if they were shorter than 100 bp in length, had more than 2 bp primer mismatches, contained more than 1 bp barcode mismatches, had greater than 2 ambiguous bases, or had a homopolymer run greater than 7 bp long. The resulting 142,851 high-quality sequences were binned into OTUs at 97% similarity using an open-reference with subsampling strategy using the pick_open_reference_otus.py pipeline in QIIME (Caporaso et al. 2010) with default settings and using the greengenes_13_8 release as a reference. Chimeric sequences were identified with ChimeraSlayer and removed, along with any nonbacterial OTUs or those identified as mitochondria or chloroplasts. After singleton OTUs were discarded the OTU table was rarefied by randomly subsampling sequences at the lowest count available for a single sample.

Network construction. Because microbial community data are sparse (have high numbers of OTUs with '0' counts), the rarefied OTU table was filtered to remove any OTUs with a total abundance of less than 10 sequences across all samples, and OTUs present in less than 5 of the individual samples. The resulting data set consisted of 1371 OTUs. The microbiome network was constructed using SparCC to identify links, where 133 OTUs were involved in either coabundance (positive associations represented by blue links in Fig. 3) or coexclusion (negative associations represented by red links) based on a threshold of $P > 0.001$ and an absolute pairwise correlation of >0.60. The resulting correlations among OTUs were graphed as a network using the igraph package (Csardi and Nepusz 2006) in R and visualized in Gephi (Bastian et al. 2009).

Applying the network analysis framework. *General network analysis: Modularity in the oak phyllosphere.* A module is a set of nodes with a large number of inter-connecting links, but few

links to nodes of other groups. Modules were detected using the netcarto function in the rnetcarto package in R (Doulcier and Stouffer 2015). This function is based on simulated annealing optimization, identifying modules by maximizing the network modularity. Simulated annealing randomly selects the local solution and thus requires no prior input about the anticipated number and size of a module. At each iteration, assignment of OTUs into a module and role varies due to the model stochasticity. Thus the evaluation was performed in 200 iterations, and the consensus result from the 200 iterations was evaluated. The nodes belonging to specific modules were assigned roles in the network structure as described by Olesen et al. (2007), with the slight modifications described in Guimera and Amaral (2005). The observed nodes were classified as peripherals, connectors, module hubs, or network hubs based on within-module degree and among-module connectivity. Both positive and negative links were included in the analysis, reflecting the potential for both types of association to influence stability. Nodes with within-module degree greater than 2.5 were classified as hubs and less than 2.5 as non-hubs. Based on the level of among-module connectivity, the hub nodes were further designated as network hubs (among-module connectivity >0.62) and module hubs (among-module connectivity <0.62), whereas the nonhubs were designated as connectors (among-module connectivity >0.62) or peripherals (among-module connectivity <0.62).

Host-focused network analysis. We illustrate a hypothetical host-focused analysis by inserting a hypothetical host response node ("photosynthetic rate") in the oak phyllosphere example, because the original data did not include a host response. The links to this hypothetical node were selected simply to illustrate the type of relationship that might be observed. Most microbiome network analyses have not included a host response node, but including a host response node helps visualization for selection of candidate taxa for contexts such as biofertilization.

Pathogen-focused network analysis. As in the host-focused network case, we illustrate a hypothetical pathogen-focused analysis by designating a certain node as a pathogen in the oak phyllosphere, for the sake of discussion.

Disease-focused network analysis. The *Rhizoctonia* system was used to illustrate a disease-focused network analysis where a node's direct or indirect association with the presence or absence of disease was the focus of analysis. Taxa linked through positive or negative association with the presence of disease were identified as described above. This data set included information about the presence or absence of disease, but not the level of disease. A node representing "disease" was introduced, where positive (blue) or negative (red) links to that node indicate association with the presence or absence of disease.

CASE STUDY RESULTS

General network analysis: Oak phyllosphere. The co-occurrence network of the oak phyllosphere fungal community (Fig. 2A) had modularity 0.132. On average only 32% of the total links were between the modules, and the majority of the links were within the modules (68%). On average each module consisted of 7.14 ± 1.7 (mean $\pm$ SE) nodes, whereas the number of nodes in each module ranged from 2 to 13. In total 7 modules with nodes belonging to both Ascomycota and Basidiomycota were observed. The entire fungal network had an average path length of 3.04 with average node degree 1.89.

Each of the nodes that occurred in a module was assigned a role based on its connectivity with other nodes in the network. The majority (91%) of the linked nodes were peripheral, and further classified as ultra- and non-ultra-peripheral nodes based on among-module connectivity. A total of 65% of the peripheral nodes were ultra-peripheral (among-module connectivity = 0), and the remaining 35% were non-ultra-peripheral (0.62 > among-module connectivity >0), indicating that these nodes were linked within their own module but not with other modules. Only one node, belonging to the genus *Aureobasidium* (order Dothidiales), which made up 11.4% of the total number of sequences and was the second most abundant taxon, occurred as a connector for the whole network. The connector node represents a node with a high network betweenness value, and can thus function as a bridge between modules. The phyllosphere fungal community lacked a network hub (for the set of thresholds evaluated here). Such a node would have had high within-module degree and among-module-connectivity, and could represent a supergeneralist taxon.

Host-focused analysis: Oak phyllosphere. In this hypothetical scenario (Fig. 2C), 11 nodes are positively and directly associated with the host response variable (photosynthetic rate), whereas 3 nodes are directly negatively associated. The majority of the other nodes are also indirectly associated with photosynthetic rate, where those with second degree associations (through one other OTU) or third degree associations (through two other OTUs) are also indicated (Fig. 2C). The differences in node degree among the nodes with direct links to the host response may reflect the relative importance of the nodes in maintaining the current network. Those taxa with the most direct positive links to the host response would be the most likely candidate taxa for such applications as biofertilization, potentially along with those taxa with the most important roles as hubs and bridges for candidate taxa. Hub nodes may be key in mediating interactions among other microbes, stabilization of microbial diversity and host and microbe interactions (Agler et al. 2016; van der Heijden and Hartmann 2016). Similarly, exploring strategies to control the nodes with direct negative associations to host response variables may support disease management or crop improvements. In addition to direct (first degree) associations, exploring second, third, or higher degree associations could support more comprehensive or community-level management methods. Higher degree associations present challenges for interpretation, because of the potential artifacts as lines become longer, and the potential for mixed types of associations (negative association with the host response node through one line, and positive through another line; Figs. 1B and 2).

Pathogen-focused analysis: Oak phyllosphere. This hypothetical scenario (Fig. 2D) illustrates a case with taxa both positively and negatively associated with a pathogen. The taxa that have direct or indirect negative associations with the pathogen are potential candidates for biocontrol agents. For example, those taxa with positive associations with other taxa that have negative associations with the pathogen are also potentially beneficial, although the longer the chain of links is, the higher the uncertainty becomes. These positive and negative associations with the pathogen node could result from niche modulation by the pathogen or other microbes, potentially through effects on host defense mechanisms altering the local niche and the potential for colonization by other microbes (Chou et al. 2000; Mukhtar et al. 2011; Voegele et al. 2005). Nodes colored dark, medium, and light blue represent first degree, second degree, and third degree neighbors for the pathogen node (Fig. 2D), respectively. The node degree of the potential biocontrol agents varied among the linked nodes, as indicated by node color density (Fig. 2A). Understanding the role of nodes in the network could further help in selecting effective biocontrol agents. For example, if a taxon is not only negatively associated with the pathogen, but also is a bridge node to a positively linked module, this could suggest a key role. Finding key nodes and hub nodes supports prediction of ecological roles of nodes/taxa, although the rationale for selecting such taxa becomes complex when key nodes or hub nodes represent pathogens (Agler et al. 2016). Thus, when selecting candidate taxa, it may be possible to target groups that affect pathogens through multiple pathways, but preserve other microbial interactions, with the aim of providing an effective pool of microbes for disease management.

Disease-focused analysis: *Rhizoctonia* system. The modules observed were phylogenetically diverse and consisted of Gram-positive and Gram-negative taxa belonging to many distinct families. Although taxa with a direct negative or positive association with disease may be detected without the use of network analyses, this analysis offers insight into the interrelationships among the diverse array of taxa coexisting in disease-prone or healthy soil communities. Specifically, potentially plant-beneficial taxa tended to co-associate in two distinct modules, whereas taxa associated with diseased patches co-associated to form a single module. Together, this suggests that taxa associated with diseased or healthy plant states respond to similar environmental conditions or act together as cohesive consortia to impact plant health. However, many of the detected associations could result from sampling (due to temporal and spatial resolution), PCR, and statistical bias, so that follow-up experiments are needed to test candidate taxa.

Three large interconnected modules were observed, along with three smaller unlinked modules (Fig. 3). In this network, the OTU frequencies in the rhizosphere are indicated by the size of the nodes—larger diameter circles indicate taxa more abundant in the rhizosphere. The color of the nodes indicates the association with the diseased (orange nodes) or healthy (purple nodes) state.

The large module at the bottom of the figure consists of rhizosphere taxa associated with diseased plants (patches). This includes families such as Enterobacteriaceae, Pseudomonadaceae, Chitinophagaceae, Caulobacteraceae and Flavobacteriaceae. Flavobacteriaceae would include the genus *Chryseobacterium*, which Yin et al. (2013) isolated and showed to have biocontrol activity. In general, this module associated with diseased roots consists of copiotrophic bacteria. This module is negatively linked to an Actinomycetales OTU associated with the healthy state and N1423W (a member of the Gemmatimonadetes). The other large module in the upper part of the network consists of taxa more associated with the healthy state and the bulk soil, and includes Gemmatimonadaceae, Acidobacteriaceae, and the Actinobacteria families Koribacteraceae, Gaiellaceae and Solirubrobacteriaceae.
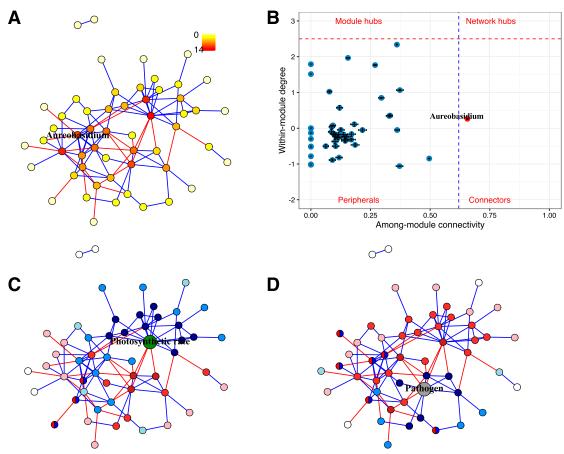

Fig. 2. Illustration of network analyses for the oak phyllosphere fungal communities. In the network diagrams, blue and red links represent positive and negative associations, respectively. Network layouts are based on the Fruchterman-Reingold algorithm in the igraph package, where each node represents an OTU and two nodes are connected if their frequencies are significantly associated. (Node size does not reflect taxon frequency.) **A,** General network analysis of the co-occurrence of oak phyllosphere fungal communities. Node color indicates the node degree for each OTU. The OTU categorized as *Aureobasidium*, the strongest connector node, is indicated. **B,** Partitioning of 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, as described by Guimera et al. (2005). The standard error bar at each node is based on 200 iterations. (Note that both positive and negative associations were included in this analysis, in contrast to many applications.) **C,** A hypothetical host-focused analysis based on the oak phyllosphere network (A), where, for illustrative purposes, a hypothetical host response variable (photosynthetic rate) is represented as a green node. Blue nodes have a positive association with the host response, where dark blue nodes have a direct link as first degree neighbors, medium blue nodes have an indirect link through one other taxon as second degree neighbors, and light blue nodes have an indirect link through two other taxa as third degree neighbors. Other taxa with longer shortest paths to the host response (fourth degree neighbors, etc.) may also be of interest as candidates, although there will be lower confidence in their role. Similarly dark red, medium red, and light red nodes represent taxa that are negatively associated with the photosynthetic rate, as first, second and third degree neighbors, respectively. Nodes colored half red and half blue have a mixed association with the host response node, having both positive and negative higher degree associations. **D,** A hypothetical pathogen-focused analysis based on the oak phyllosphere network (A), where, for illustrative purposes, a hypothetical pathogen is represented as a grey node. For the pathogen-focused scenario, red nodes have a direct or indirect negative association with the pathogen, and may be candidate taxa for biocontrol. Strategies to limit taxa with positive associations (blue nodes) might also help to reduce pathogen populations.

These tend to be slower growing oligotrophic bacteria. This module is linked to a submodule consisting of rhizosphere families: Burkholderiaceae, Microbacteriaceae and Sphingobacteriaceae, some of which were associated with healthy plants. This analysis also identified the presence of three smaller unlinked modules. One was represented by rhizosphere-predominant Sphingomonadaceae associated with diseased roots (orange nodes). Another consisted of Actinobacteria more predominant in the bulk soil such as Micromonosporaceae and Pseudonocardiaceae, which were associated with healthy roots.
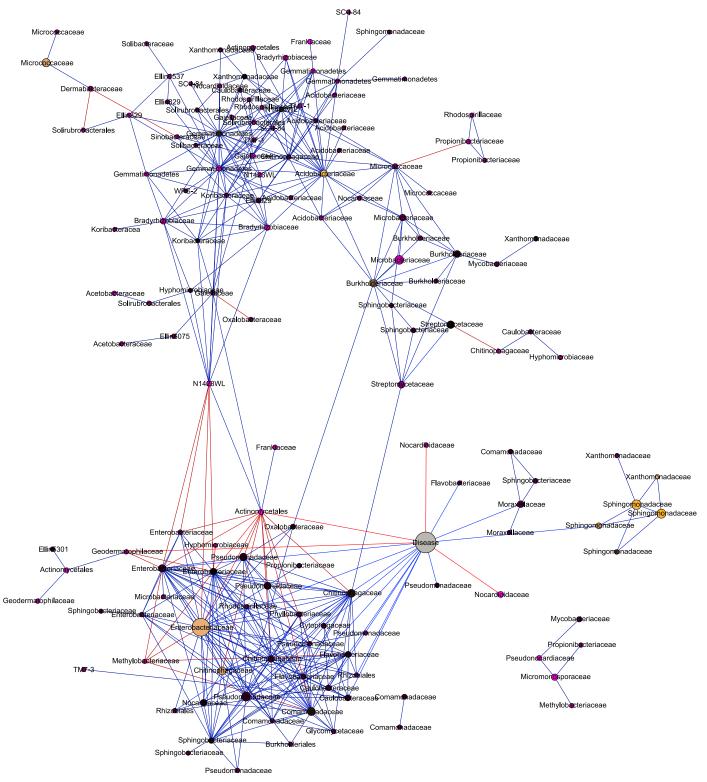

Fig. 3. Disease-focused analysis of soil and rhizosphere taxa positively or negatively associated with infection of wheat by *Rhizoctonia solani* (diseased patches in the field). The size of nodes (operational taxonomic units [OTUs]) is scaled by their frequency in the rhizosphere, where larger diameter circles indicate a greater abundance in the rhizosphere. The color of the nodes is scaled by the association of that OTU with the disease (brighter orange) or healthy (brighter purple) state. Darker nodes have little difference in relative abundance between healthy and diseased states. The disease state is represented by a large gray node. Relationships among OTU nodes were assessed using SparCC methods and those between OTUs and the disease state were evaluated using GLMs. Positive relationships are indicated by blue links and negative relationships are indicated by red links. OTU taxon assignments to the family level are indicated for each node. Where OTUs could not be confidently assigned to the family level, higher taxonomic assignments (e.g., order, class, or phylum) are indicated.

DISCUSSION

The four types of network analysis in this framework reveal different aspects of microbiome networks, and different candidate taxa for further evaluation. The general network analysis in the oak phyllosphere identified candidate taxa for maintaining microbial communities based on their high number of links within and among modules. The illustration of the host-focused analysis identified candidate taxa that had direct positive or negative associations to a desired host response, and others that were only a limited number of links away from the host response variable. The illustration of the pathogen-focused analysis shows how it can be used to find candidate taxa that have a direct negative link to a known pathogen, and others that have positive links to support the potential pathogen antagonists. The disease-focused analysis identified candidate taxa positively associated with the absence of infection by *Rhizoctonia solani*, and others positively linked to these. In the host-, pathogen-, and disease-focused analyses, candidate taxa with direct links to the response variables can be considered first degree neighbors, which could represent either candidate taxa for biofertilization or target candidates for disease management. These could be discovered in a simple analysis of associations with the response variables. Including the network analysis makes it possible to evaluate the candidates that support these, the second degree neighbors, third degree neighbors, and so on. Inclusion of these taxa in synthetic microbial assemblages may increase the stability of biocontrol or biofertilization efforts.

For analyses where there are not obvious response variables aside from OTU frequencies, the general network analysis offers a means to identify candidate taxa that are potentially important for maintaining the network as a whole. The majority of the nodes in the oak phyllosphere data were categorized as peripheral nodes, with one node categorized as a connector. The modular structure of co-occurring OTUs suggests diversity in species roles and functionality (Montoya et al. 2015), as well as in their environmental preferences. Modularity may suggest niche overlap or common resource requirements among the co-occurring microbiome taxa, or any of a number of phenomena in different types of ecological networks. For example, modularity may indicate predators and prey that frequently co-occur in food webs (Kondoh 2008), a convergence in functional traits among unrelated species (Bascompte et al. 2006) in pollinator-plant networks (Olesen et al. 2007), and the nonrandom and repeated association of plants and visiting insects (Dupont and Olesen 2009). The modules detected in our analysis consisted of both Ascomycota and Basidiomycota, suggesting their common co-occurrence in this system, although some patterns might be influenced by higher primer specificity toward the Ascomycota (Monard et al. 2013).

The structure of modules in a network may be important for system stability and functionality (May 1972; Scheffer et al. 2012; Thebault and Fontaine 2010), but this has been difficult to test for many systems (Sinha 2005). The majority of the oak phyllosphere nodes were categorized as peripheral nodes, connected only with other member nodes within a module, which is sometimes interpreted as evidence for specialization. Definitions for specialists and generalists differ from study to study. For example, Barberan et al. (2012) and Collins and Glenn (1997) described specialist or satellite species based on species occurrence across samples. Based on network structure, networks may include generalists observed as network hubs and connectors, and specialists observed as peripheral nodes. Nodes that are linked among the modules may reflect a wider range of habitat and resource use, although the most generalist taxa will appear more consistently across samples in observational studies, and thus may not be detected in networks of co-association. Network-based categorization of taxa as generalists or specialists may have evolutionary significance in some cases. Connector nodes in metabolite networks may be conserved metabolites that are involved in essential biochemical pathways and conserved across multiples species (Guimera and Amaral 2005). It is possible that connector and hub species may represent multifunctional taxa.

A pathogen-focused network analysis can reveal how second degree, third degree, and higher associations with pathogen abundance can influence disease risk. For some systems, such as disease-suppressive soils, these associations may be key to management outcomes. In the future, as technologies for characterizing microbiomes become common and less expensive, network analyses may become a standard procedure to identify key taxa for managing many plant pathogens. Pathogen-focused network analysis could also be useful for surveying new microbes from natural habitats with potential for identifying novel antibiotics or chemicals.

The results of the network analysis of the *Rhizoctonia* system support the hypotheses posed in Yin et al. (2013) that taxa closely associated with the diseased patches were copiotrophic rhizosphere bacteria adapted to large amounts of root exudates coming from diseased roots. The destruction of roots by *Rhizoctonia* infection probably releases not only more exudates, but other cell wall and cytoplasm components that can be used by these bacteria, which include genera such as *Flavobacterium* and *Pseudomonas*, previously implicated in biological control. *Chitinophaga* has been found as one of the most abundant taxa in wheat soils in other studies, and may be digesting chitin from the fungal biomass. In the healthy/recovered patches, this large food base is not available, so these areas are dominated by oligotrophs more adapted to small nutrient fluxes in the bulk soil. However, unlike previous analyses, the network analysis identified three other smaller modules or groups - a rhizosphere predominant group linked to the larger bulk soil/healthy module (which was not linked to the rhizosphere/diseased module), a smaller module consisting of rhizosphere Sphingomonadaceae associated with diseased roots, and an Actinobacteria module of bulk soil families associated with healthy roots. Interestingly, OTUs in the same family can be associated with healthy or diseased roots, indicating a tremendous amount of functional diversity within families. For example, within the Pseudomonadaceae the ecological niche and antibiotic phenotypes of *Pseudomonas* species can vary widely, as found in a study in the same area of Washington State (Parejko et al. 2012). Thus, it is not surprising that one taxon could be associated with both disease states. Many copiotrophs may build up on leaky roots, but we hypothesize that only those producing antifungal compounds or that have some other antagonistic function will suppress disease. Detailed functional studies of microbial interactions will be a crucial next step to testing hypotheses from network analyses about candidate taxa that may play important roles in suppressing or facilitating plant disease. As microbiome networks are better understood, it may be possible to evaluate and predict the shifts over time in taxon frequency driven by biological interactions, shifts that ultimately lead to disease-suppressive soils.

The primary goal of our framework is to use microbiome network studies to provide guidance for selecting a useful number of candidate taxa for biological control, biofertilization, and microbe-associated crop breeding. Using current technologies, there remain challenges in interpretations for these networks. One important source of potential error in interpretation of microbiome networks is compositional bias, where an increase in abundance of one taxon may lead to reduced frequencies for other taxa, even if the abundance of the other taxa remains unchanged. As technologies to evaluate abundance improve, this problem may ultimately be removed. Next-generation sequencing (NGS) platforms have been evolving rapidly in the past decade. For example, single molecule sequencing tools such as the Oxford Nanopore Technologies MinION (https://www.nanoporetech.com/) have great promise for producing data that can potentially not only characterize communities, but also provide reasonable estimates of the relative abundance of organisms residing in complex environmental samples. Proposed strategies include direct amplicon preparations following Oxford Nanopore methods (Kilianski et al. 2015) to obtain long nanopore reads from low input template DNAs or "Intramolecular-ligated Nanopore Consensus Sequencing" (Li et al. 2016) similar to

PacBio sequencing of circularized templates to reduce sequencing errors (Lou et al. 2013; Travers et al. 2010). These reports are encouraging as they provide proof of concept, paving the way for single molecule sequencing that might provide abundance data for improved network analyses.

In the case studies, we chose different correlation thresholds for the two different data sets, while using the same P value threshold ($P < 0.001$). The choice of thresholds was made to create networks to identify a practical number of candidate taxa that could be evaluated further in follow-up experiments. The choice of thresholds will have important effects on which candidate taxa are selected. As more attention is paid to the arbitrary aspects of defining thresholds for statistical significance (Madden et al. 2015), alternatives include the potential to use sensitivity analysis to identify those taxa that are candidates across a range of threshold values. Thresholds are also used to define OTU size based on the sequence count, as well as in defining the pairwise association between two OTUs, and in defining the level of significance based on P values. Sensitivity analysis could help to capture the relation between OTUs across a range of thresholds, as opposed to a single set of thresholds, potentially providing results that are more robust.

Selection of candidate taxa involves a number of decisions that may decrease the chance of artifacts and identification of spurious associations. In our analysis, we chose a method that reduces compositional bias. We also used criteria to select OTUs based on a sequence count greater than 10, and presence in at least five of the replicates. Relatively rare taxa are likely to be associated with lower statistical power in tests of associations, and, from a practical standpoint, may be difficult as candidates in follow-up experiments. Still, many candidate taxa will be identified as a consequence of statistical error and biases associated with PCR and sequencing.

Follow-up experiments can be designed for evaluating candidates based on concepts similar to Koch's postulates, where candidate assemblages are deployed as a complete group and as targeted subsets, to evaluate the performance of the assemblage and the importance of the role of individual taxa to the assemblage performance. If the goal of follow-up experiments is pragmatic development of successful bio-control, a limited number of assemblages might be evaluated, based on consistent performance in sensitivity analyses. If the goal is to better understand and map the microbiome network structure, experimental assemblages could be designed to test network-based hypotheses about causality in relationships among taxa and outcomes for the host or disease. Because some important taxa may be present at consistent levels in observational studies, and thus "invisible" in network analyses, it may be important to include experiments in which their abundance is altered to have a more complete understanding of the network. In the future, with advanced and less expensive sequencing technology supporting the analysis of taxon *abundance* rather than merely *frequency*, network analyses at multiple time points with a higher number of replicates will contribute to better understanding of the dynamics of microbial co-occurrence and factors contributing to community resilience (Shade et al. 2012). Such analyses would also contribute to understanding shifts in microbial communities over time, or as a function of experimental treatments, and could contribute to understanding directionality of interactions among taxa, and the functional roles of taxa.

Another approach to identifying key relationships among taxa, beyond pair-wise associations, is to consider patterns of association within clusters of OTUs such as triads (sets of three OTUs) across communities, and specifically the patterns of relative abundance of distinct cluster motifs. Thus, as an example, rather than considering networks defined by the presence of significant pairwise co-associations, triad analyses focus on characterizing subnetwork structures among sets of three taxa (motifs). Considering only positive co-associations, there are four possible triad motifs: (i) no co-associations within the triad; (ii) one pair of OTUs co-associated within the triad; (iii) two pairs co-associated within the triad; or (iv) all three OTUs positively co-associated within the triad (Fig. 1C). Including both positive and negative co-associations or directed interactions expands the number of possible motifs, and can provide more detailed insights into the ecology of interactions within complex networks (Schlatter and Kinkel 2014; Tran et al. 2013). Quantifying the relative abundance of distinct triad (or larger) motifs within communities, and among communities having different disease or disease-suppressive states, offers the potential to capture higher-order structure within the microbiome and to generate hypotheses about the roles of distinct network motifs in generating community functional activities. Alternatively, characterizing the patterns of representation for individual OTUs in different triad motifs can provide useful information on the role of specific OTUs within the community.

The networks that are a focus of this paper are based on two-dimensional adjacency matrices, where entries in the matrices indicate information about pair-wise interactions. As substantially greater data volumes become available in typical microbiome studies, it may become possible to evaluate adjacency matrices with three or more dimensions, including information about associations between taxa that are conditional on the abundance of other taxa, and/or conditional on levels of environmental variables. With more data, inclusion of more environmental variable nodes along with taxon nodes in microbiome networks, evaluated using Bayesian network methods (Darwiche 2009), may also make it feasible to discriminate between microbial associations based on shared niches versus associations based on biological interactions. Future greater data availability may also provide the opportunity to evaluate linked *spatial* networks of microbial community movement and spatial networks of signaling molecules (Garrett 2012). A strength of many types of network analysis is the use of matrix multiplication to predict future levels of each network component given current values (or given scenarios where some values are enhanced). In the context of microbiome networks, the effects of the abundance of a candidate microbial assemblage (or a targeted change in its abundance) on the abundance of other taxa, or on host or pathogen response variables, could be predicted over time. This type of analysis would be most useful if the associations among nodes were based on known biological interactions, and the dynamic nature of these associations in response to factors such as climate change is understood. With current technological limitations, matrix multiplication would propagate the error resulting when associations are based merely on niche preferences or compositional bias. However, in the future, with greater amounts of higher quality data, analyses of network-wide responses to changes in the abundance of particular taxa may become a standard tool for understanding microbiomes.

ACKNOWLEDGMENTS

LITERATURE CITED

Agler, M. T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S. T., Weigel, D., and Kemen, E. M. 2016. Microbial hub taxa link host and abiotic factors to plant microbiome variation. PLoS Biol. 14:e1002352.

Bai, W. H., Zhou, T., and Wang, B. H. 2007. Immunization of susceptible-infected model on scale-free networks. Physica A: Statistical Mechanics and its Applications 384:656-662.

Bakker, M. G., Schlatter, D. C., Otto-Hanson, L., and Kinkel, L. L. 2014. Diffuse symbioses: Roles of plant–plant, plant–microbe and microbe–microbe interactions in structuring the soil microbiome. Mol. Ecol. 23:1571-1583.

Barberan, A., Bates, S. T., Casamayor, E. O., and Fierer, N. 2012. Using network analysis to explore co-occurrence patterns in soil microbial communities. ISME J. 6:343-351.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. 2004. The architecture of complex weighted networks. Proc. Natl. Acad. Sci. USA 101:3747-3752.

Bascompte, J., Jordano, P., and Olesen, J. M. 2006. Asymmetric co-evolutionary networks facilitate biodiversity maintenance. Science 312:431-433.

Bastian, M., Heymann, S., and Jacomy, M. 2009. Gephi: An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Benitez, M. S., and Gardener, B. B. 2009. Linking sequence to function in soil bacteria: Sequence-directed isolation of novel bacteria contributing to soilborne plant disease suppression. Appl. Environ. Microbiol. 75:915-924.

Berendsen, R. L., Pieterse, C. M. J., and Bakker, P. A. H. M. 2012. The rhizosphere microbiome and plant health. Trends Plant Sci. 17:478-486.

Brown, S. P., Veach, A. M., Rigdon-Huss, A. R., Grond, K., Lickteig, S. K., Lothamer, K., Oliver, A. K., and Jumpponen, A. 2015. Scraping the bottom of the barrel: Are rare high throughput sequences artifacts? Fungal Ecol. 13:221-225.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J., and Knight, R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7:335-336.

Chadès, I., Martin, T. G., Nicol, S., Burgman, M. A., Possingham, H. P., and Buckley, Y. M. 2011. General rules for managing and surveying networks of pests, diseases, and endangered species. Proc. Natl. Acad. Sci. USA 108:8323-8328.

Chou, H. M., Bundock, N., Rolfe, S. A., and Scholes, J. D. 2000. Infection of *Arabidopsis thaliana* leaves with *Albugo candida* (white blister rust) causes a reprogramming of host metabolism. Mol. Plant Pathol. 1:99-113.

Collins, S. L., and Glenn, S. M. 1997. Effects of organismal and distance scaling on analysis of species distribution and abundance. Ecol. Appl. 7:543-551.

Cook, R. J., Schillinger, W. F., and Christensen, N. W. 2002. Rhizoctonia root rot and take-all of wheat in diverse direct-seed spring cropping systems. Can. J. Plant Pathol. 24:349-358.

Cox, C. M., Bockus, W. W., Holt, R. D., Fang, L., and Garrett, K. A. 2013. Spatial connectedness of plant species: Potential links for apparent competition via plant diseases. Plant Pathol. 62:1195-1428.

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

Curl, E. A., and Trulove, B. 1986. The rhizosphere. Adv. Agric. Sci. 15:290.

Darwiche, A. 2009. Modeling and Reasoning with Bayesian Networks. Cambridge University Press.

Debenport, S. J., Assigbetse, K., Bayala, R., Chapuis-Lardy, L., Dick, R. P., and McSpadden Gardener, B. B. 2015. Association of shifting populations in the root zone microbiome of millet with enhanced crop productivity in the Sahel region (Africa). Appl. Environ. Microbiol. 81:2841-2851.

Doulcier, G., and Stouffer, D. 2015. Rnetcarto: Fast network modularity and roles computation by simulated annealing. R package version 0.2.4.

Dupont, Y. L., and Olesen, J. M. 2009. Ecological modules and roles of species in heathland plant-insect flower visitor networks. J. Anim. Ecol. 78:346-353.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194-2200.

Edwards, J., Johnson, C., Santos-Medellin, C., Lurie, E., Podishetty, N. K., Bhatnagar, S., Eisen, J. A., and Sundaresan, V. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. Proc. Natl. Acad. Sci. USA 112:E911-920.

Faust, K., Lima-Mendez, G., Lerat, J. S., Sathirapongsasuti, J. F., Knight, R., Huttenhower, C., Lenaerts, T., and Raes, J. 2015. Cross-biome comparison of microbial association networks. Front. Microbiol. 6:1200.

Faust, K., and Raes, J. 2012. Microbial interactions: From networks to models. Nat. Rev. Microbiol. 10:538-550.

Faust, K., Sathirapongsasuti, J. F., Izard, J., Segata, N., Gevers, D., Raes, J., and Huttenhower, C. 2012. Microbial co-occurrence relationships in the human microbiome. PLoS Comput. Biol. 8:e1002606.

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

Gardes, M., and Bruns, T. D. 1993. ITS primers with enhanced specificity for basidiomycetes–application to the identification of mycorrhizae and rusts. Mol. Ecol. 2:113-118.

Garrett, K. A. 2012. Information networks for disease: Commonalities in human management networks and within-host signalling networks. Eur. J. Plant Pathol. 133:75-88.

Garrett, K. A., Jumpponen, A., Toomajian, C., and Gomez-Montano, L. 2012. Climate change and plant health: Designing research spillover from plant genomics for understanding the role of microbial communities. Can. J. Plant Pathol. 34:349-361.

Germanos, R., Deacon, R., and Mooney-Somers, J. 2015. The social and cultural significance of women's sexual identities should guide health promotion. LGBT Health 2:162-168.

Grange, Z. L., Van Andel, M., French, N. P., and Gartrell, B. D. 2014. Network analysis of translocated Takahe populations to identify disease surveillance targets. Conserv. Biol. 28:518-528.

Guimera, R., and Amaral, L. A. N. 2005. Functional cartography of complex metabolic networks. Nature 433:895-900.

Hernandez Nopsa, J. F., Daglish, G. J., Hagstrum, D. W., Leslie, J. F., Phillips, T. W., Scoglio, C., Thomas-Sharma, S., Walter, G. H., and Garrett, K. A. 2015. Ecological networks in stored grain: Key postharvest nodes for emerging pests, pathogens, and mycotoxins. BioScience 65:985-1002.

Jumpponen, A., and Jones, K. L. 2009. Massively parallel 454 sequencing indicates hyperdiverse fungal communities in temperate *Quercus macrocarpa* phyllosphere. New Phytol. 184:438-448.

Jumpponen, A., and Jones, K. L. 2010. Seasonally dynamic fungal communities in the *Quercus macrocarpa* phyllosphere differ between urban and nonurban environments. New Phytol. 186:496-513.

Kefi, S., Berlow, E. L., Wieters, E. A., Navarrete, S. A., Petchey, O. L., Wood, S. A., Boit, A., Joppa, L. N., Lafferty, K. D., Williams, R. J., Martinez, N. D., Menge, B. A., Blanchette, C. A., Iles, A. C., and Brose, U. 2012. More than a meal... integrating non-feeding interactions into food webs. Ecol. Lett. 15:291-300.

Kilianski, A., Haas, J. L., Corriveau, E. J., Liem, A. T., Willis, K. L., Kadavy, D. R., Rosenzweig, C. N., and Minot, S. S. 2015. Bacterial and viral identification and differentiation by amplicon sequencing on the MinION nanopore sequencer. Gigascience 4:12.

Knight, R. 2015. Why microbiome treatments could pay off soon. Nature 518:S5.

Koljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., Callaghan, T. M., Douglas, B., Drenkhan, T., Eberhardt, U., Duenas, M., Grebenc, T., Griffith, G. W., Hartmann, M., Kirk, P. M., Kohout, P., Larsson, E., Lindahl, B. D., Luecking, R., Martin, M. P., Matheny, P. B., Nguyen, N. H., Niskanen, T., Oja, J., Peay, K. G., Peintner, U., Peterson, M., Poldmaa, K., Saag, L., Saar, I., Schuessler, A., Scott, J. A., Senes, C., Smith, M. E., Suija, A., Taylor, D. L., Telleria, M. T., Weiss, M., and Larsson, K. H. 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol. Ecol. 22:5271-5277.

Kondoh, M. 2008. Building trophic modules into a persistent food web. Proc. Natl. Acad. Sci. USA 105:16631-16635.

Kurtz, Z. D., Muller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. 2015. Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput. Biol. 11:e1004226.

Lebeis, S. L., Paredes, S. H., Lundberg, D. S., Breakfield, N., Gehring, J., McDonald, M., Malfatti, S., del Rio, T. G., Jones, C. D., Tringe, S. G., and Dangl, J. L. 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science 349:860-864.

Li, C., Chng, K. R., Boey, J. H. E., Ng, H. Q. A., Wilm, A., and Nagarajan, N. 2016. INC-Seq: Accurate single molecule reads using nanopore sequencing. bioRxiv. doi: http://dx.doi.org/10.1101/038042

Lou, D. I., Hussmann, J. A., McBee, R. M., Acevedo, A., Andino, R., Press, W. H., and Sawyer, S. L. 2013. High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. Proc. Natl. Acad. Sci. USA 110:19872-19877.

Lundberg, D. S., Lebeis, S. L., Paredes, S. H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., del Rio, T. G., Edgar, R. C., Eickhorst, T., Ley, R. E., Hugenholtz, P., Tringe, S. G., and Dangl, J. L. 2012. Defining the core *Arabidopsis thaliana* root microbiome. Nature 488:86-90.

MacNeil, L. T., and Walhout, A. J. 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. Genome Res. 21:645-657.

Madden, L. V., Shah, D. A., and Esker, P. D. 2015. Does the *P* value have a future in plant pathology? Phytopathology 105:1400-1407.

May, R. M. 1972. Will a large complex system be stable? Nature 238:413-414.

Mazzola, M., Hewavitharana, S. S., and Strauss, S. L. 2015. Brassica seed meal soil amendments transform the rhizosphere microbiome and improve apple production through resistance to pathogen reinfestation. Phytopathology 105:460-469.

McKenzie, V. J., Bowers, R. M., Fierer, N., Knight, R., and Lauber, C. L. 2012. Co-habiting amphibian species harbor unique skin bacterial communities in wild populations. ISME J. 6:588-596.

McSpadden Gardener, B. B., and Weller, D. M. 2001. Changes in populations of rhizosphere bacteria associated with take-all disease of wheat. Appl. Environ. Microbiol. 67:4414-4425.

Monard, C., Gantner, S., and Stenlid, J. 2013. Utilizing ITS1 and ITS2 to study environmental fungal diversity using pyrosequencing. FEMS Microbiol. Ecol. 84:165-175.

Montoya, D., Yallop, M. L., and Memmott, J. 2015. Functional group diversity increases with modularity in complex food webs. Nat. Commun. 6:7379.

Mukhtar, M. S., Carvunis, A. R., Dreze, M., Epple, P., Steinbrenner, J., Moore, J., Tasan, M., Galli, M., Hao, T., Nishimura, M. T., Pevzner, S. J., Donovan, S. E., Ghamsari, L., Santhanam, B., Romero, V., Poulin, M. M., Gebreab, F., Gutierrez, B. J., Tam, S., Monachello, D., Boxem, M., Harbort, C. J., McDonald, N., Gai, L., Chen, H., He, Y., European Union Effectoromics, C., Vandenhaute, J., Roth, F. P., Hill, D. E., Ecker, J. R., Vidal, M., Beynon, J., Braun, P., and Dangl, J. L. 2011. Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science 333:596-601.

Olesen, J. M., Bascompte, J., Dupont, Y. L., and Jordano, P. 2007. The modularity of pollination networks. Proc. Natl. Acad. Sci. USA 104:19891-19896.

Oliver, A. K., Brown, S. P., Callaham, M. A., and Jumpponen, A. 2015. Polymerase matters: Non-proofreading enzymes inflate fungal community richness estimates by up to 15%. Fungal Ecol. 15:86-89.

Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., and Kao-Kniffin, J. 2015. Selection on soil microbiomes reveals reproducible impacts on plant function. ISME J. 9:980-989.

Parejko, J. A., Mavrodi, D. V., Mavrodi, O. V., Weller, D. M., and Thomashow, L. S. 2012. Population structure and diversity of phenazine-1-carboxylic acid producing fluorescent *Pseudomonas* spp. from dryland cereal fields of central Washington State (USA). Microbial Ecol. 64:226-241.

Paulitz, T. C., Smiley, R. W., and Cook, R. J. 2002. Insights into the prevalence and management of soilborne cereal pathogens under direct seeding in the Pacific Northwest, USA. Can. J. Plant Pathol. 24:416-428.

Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangl, J. L., Buckler, E. S., and Ley, R. E. 2013. Diversity and heritability of the maize rhizosphere microbiome under field conditions. Proc. Natl. Acad. Sci. USA 110:6548-6553.

Postma, J., Willemsen-de Klein, M. J., and van Elsas, J. D. 2000. Effect of the indigenous microflora on the development of root and crown rot caused by *Pythium aphanidermatum* in cucumber grown on rockwool. Phytopathology 90:125-133.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www. R-project.org

Robleto, E. A., Borneman, J., and Triplett, E. W. 1998. Effects of bacterial antibiotic production on rhizosphere microbial communities from a culture-independent perspective. Appl. Environ. Microbiol. 64:5020-5022.

Ruan, Q., Dutta, D., Schwalbach, M. S., Steele, J. A., Fuhrman, J. A., and Sun, F. 2006. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 22: 2532-2538.

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., van de Koppel, J., van de Leemput, I. A., Levin, S. A., van Nes, E. H., Pascual, M., and Vandermeer, J. 2012. Anticipating critical transitions. Science 338:344-348.

Schillinger, W. F., and Paulitz, T. C. 2014. Natural suppression of Rhizoctonia bare patch in a long-term no-till cropping systems experiment. Plant Dis. 98:389-394.

Schlatter, D. C., and Kinkel, L. L. 2014. Antibiotics: Conflict and communication in microbial communities. Microbe 9:282-288.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., and Weber, C. F. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75:7537-7541.

Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., Burgmann, H., Huber, D. H., Langenheder, S., Lennon, J. T., Martiny, J. B., Matulich, K. L., Schmidt, T. M., and Handelsman, J. 2012. Fundamentals of microbial community resistance and resilience. Front. Microbiol. 3:417.

Shaw, M. W., and Pautasso, M. 2014. Networks and plant disease management: Concepts and applications. Annu. Rev. Phytopathol. 52:477-493.

Sinha, S. 2005. Complexity vs. stability in small-world networks. Physica A: Statistical Mechanics and its Applications 346:147-153.

Sole, R. V., and Montoya, J. M. 2001. Complexity and fragility in ecological networks. Proc. R. Soc. B-Biological Sci. 268:2039-2045.

Thebault, E., and Fontaine, C. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science 329:853-856.

Tran, N. H., Choi, K. P., and Zhang, L. 2013. Counting motifs in the human interactome. Nat. Commun. 4:2241.

Travers, K. J., Chin, C. S., Rank, D. R., Eid, J. S., and Turner, S. W. 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 38:e159.

van Bruggen, A. H. C., and Semenov, A. M. 2000. In search of biological indicators for soil health and disease suppression. Appl. Soil Ecol. 15:13-24.

van der Heijden, M. G., and Hartmann, M. 2016. Networking in the plant microbiome. PLoS Biol. 14:e1002378.

Voegele, R. T., Hahn, M., Lohaus, G., Link, T., Heiser, I., and Mendgen, K. 2005. Possible roles for mannitol and mannitol dehydrogenase in the biotrophic plant pathogen *Uromyces fabae*. Plant Physiol. 137:190-198.

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73:5261-5267.

Wang, Y. B., Xiao, G. X., Hu, J., Cheng, T. H., and Wang, L. S. 2009. Imperfect targeted immunization in scale-free networks. Physica A: Statistical Mechanics and its Applications 388:2535-2546.

Xu, X.-M., Jeffries, P., Pautasso, M., and Jeger, M. J. 2011. Combined use of biocontrol agents to manage plant diseases in theory and practice. Phytopathology 101:1024-1031.

Yin, C., Hulbert, S. H., Schroeder, K. L., Mavrodi, O., Mavrodi, D., Dhingra, A., Schillinger, W. F., and Paulitz, T. C. 2013. Role of bacterial communities in the natural suppression of *Rhizoctonia solani* bare patch disease of wheat (*Triticum aestivum* L.). Appl. Environ. Microbiol. 79:7428-7438.

Zhou, J., Deng, Y., Luo, F., He, Z., Tu, Q., and Zhi, X. 2010. Functional molecular ecological networks. MBio 1:e00169-10.