AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESEARCH ARTICLE
Host-Microbe Biology

Exploration of the Biosynthetic Potential of the *Populus* Microbiome

Patricia M. Blair,[a] Miriam L. Land,[a] Marek J. Piatek,[a] Daniel A. Jacobson,[a] Tse-Yuan S. Lu,[a] Mitchel J. Doktycz,[a] Dale A. Pelletier[a]

[a]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

ABSTRACT Natural products (NPs) isolated from bacteria have dramatically advanced human society, especially in medicine and agriculture. The rapidity and ease of genome sequencing have enabled bioinformatics-guided NP discovery and characterization. As a result, NP potential and diversity within a complex community, such as the microbiome of a plant, are rapidly expanding areas of scientific exploration. Here, we assess biosynthetic diversity in the *Populus* microbiome by analyzing both bacterial isolate genomes and metagenome samples. We utilize the fully sequenced genomes of isolates from the *Populus* root microbiome to characterize a subset of organisms for NP potential. The more than 3,400 individual gene clusters identified in 339 bacterial isolates, including 173 newly sequenced organisms, were diverse across NP types and distinct from known NP clusters. The ribosomally synthesized and posttranslationally modified peptides were both widespread and divergent from previously characterized molecules. Lactones and siderophores were prevalent in the genomes, suggesting a high level of communication and pressure to compete for resources. We then consider the overall bacterial diversity and NP variety of metagenome samples compared to the sequenced isolate collection and other plant microbiomes. The sequenced collection, curated to reflect the phylogenetic diversity of the *Populus* microbiome, also reflects the overall NP diversity trends seen in the metagenomic samples. In our study, only about 1% of all clusters from sequenced isolates were positively matched to a previously characterized gene cluster, suggesting a great opportunity for the discovery of novel NPs involved in communication and control in the *Populus* root microbiome.

IMPORTANCE The plant root microbiome is one of the most diverse and abundant biological communities known. Plant-associated bacteria can have a profound effect on plant growth and development, and especially on protection from disease and environmental stress. These organisms are also known to be a rich source of antibiotic and antifungal drugs. In order to better understand the ways bacterial communities influence plant health, we evaluated the diversity and uniqueness of the natural product gene clusters in bacteria isolated from poplar trees. The complex molecule clusters are abundant, and the majority are unique, suggesting a great potential to discover new molecules that could not only affect plant health but also could have applications as antibiotic agents.

KEYWORDS biosynthetic gene clusters, natural product biosynthesis, plant-microbe interactions, quorum sensing, rhizosphere-inhabiting microbes, siderophores

A diverse array of microbial organisms living in the rhizosphere confer beneficial attributes to the host plant (1–4). Microbial symbionts facilitate nutrient exchange, environmental stress tolerance, pathogen resistance, and modulation of the host immune response. *Populus*, the first woody plant to be sequenced (5) and have its microbiome characterized (6–8), is a key model system for studying the plant micro-

Received 11 April 2018 **Accepted** 17 August 2018 **Published** 2 October 2018

Citation Blair PM, Land ML, Piatek MJ, Jacobson DA, Lu T-YS, Doktycz MJ, Pelletier DA. 2018. Exploration of the biosynthetic potential of the *Populus* microbiome. mSystems 3: e00045-18. https://doi.org/10.1128/mSystems.00045-18.

Editor Pieter C. Dorrestein, University of California, San Diego

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Mitchel J. Doktycz, doktyczmj@ornl.gov, or Dale A. Pelletier, pelletierda@ornl.gov.

Plant root-associated bacteria have the potential to make thousands of unique natural products, which could have profound effects on microbiome stability and plant health

biome and its effect on nutrient cycling, plant health, and ecological diversity (4). A deeper understanding of how the microbiome of *Populus* utilizes secondary metabolism for intra- and interspecies communication to maximize plant growth and development will be critical for realizing the effective production of sustainable fuel sources and for understanding the cycling of materials through ecosystems.

While the root microbiome can depend on environmental conditions such as soil type, season, climate, and water availability (6, 8, 9), there is clear evidence for rich diversity within the *Populus* microbiome (8, 10). Rhizosphere and endosphere microbial communities are distinct from each other, indicating different selection processes for survival within these compartments (6, 7, 11). Soil type and host plant genotype dramatically influence microbial population, so it follows that the small-molecule clusters can also widely vary between microbiome samples (12). For example, potato plant microbial natural product (NP) diversity was found to change as the plants matured, reflecting changes in the microbial composition and the role of the microbiome during growth and development (13). This strongly indicates that different interactions and functions are required for bacteria to compete in these distinct compartments and across the development of a plant, and thus, the possible NP production will differ, leading to increased NP diversity (14). While soil microbiome analyses have surveyed the diversity of nonribosomal peptide (NRP) or polyketide (PK) clusters (13, 15), little is known about the extent of biosynthetic gene cluster (BGC) diversity across all known NP types within a plant-associated microbial community.

Bacteria not only interact with one another using chemical signaling (16), but they also interact directly and indirectly with the host plant through chemical signaling (17), nutrient exchange (18, 19), induction of immune response pathways (20), and promotion of root hairs or lateral root growth (21). For example, beneficial *Pseudomonas* microorganisms are of particular importance in protecting the host from potential pathogens through the induction of the plant immune response (22). Complex NPs with antibiotic properties have also been shown to protect the plant from pathogens and to stabilize the microbiome (23–26). Bioinformatic mining of the genomes of soil-dwelling bacteria has revealed that bacteria have the ability to produce many more NPs than have been discovered through traditional activity-based isolation methods (27, 28). In particular, special focus has been directed toward the actinomycetes because of the prevalence of BGCs in their genomes (29) and known production of bioactive NPs, including the widely successful antibiotics tetracycline, chloramphenicol, erythromycin, and rifamycin (30, 31). The types of NPs that bacteria have the potential to synthesize, the situations under which these compounds are produced, and the mechanisms by which NPs influence microbe-microbe and microbe-plant interactions (32) are questions of great importance in developing a comprehensive understanding of the plant microbiome.

A variety of computational programs, such as antiSMASH (33), BAGEL (34), and PRISM (35), have enabled the prediction of BGCs from bacterial genomes. Although BGCs are recognized on the basis of homology to NP biosynthetic pathways that have already been identified, newer programs are venturing into putative cluster identification, and predictions are continually being refined as the products of novel clusters are identified and characterized (33). In particular, the ribosomally synthesized and post-translationally modified peptide (RiPP) natural products benefit from genome mining efforts. The structures and masses of these NPs, derived from precursor peptides encoded in the genome, can be predicted with increasing accuracy on the basis of the presence of certain identifiable biosynthetic enzymes that install posttranslational modifications onto the core region of the precursor peptide. While many RiPPs have been identified as having antibiotic, anticancer, or antifungal activity *in vitro*, the actual *in vivo* role is, in most cases, undetermined (36).

Here, we analyze metagenome samples from the roots of field-grown *Populus* for genes involved in secondary metabolism. Further, we investigate the sequenced genomes of 339 bacterial strains, including 173 newly sequenced genomes, isolated from the root endosphere and rhizosphere of *Populus deltoides* and *Populus trichocarpa* in

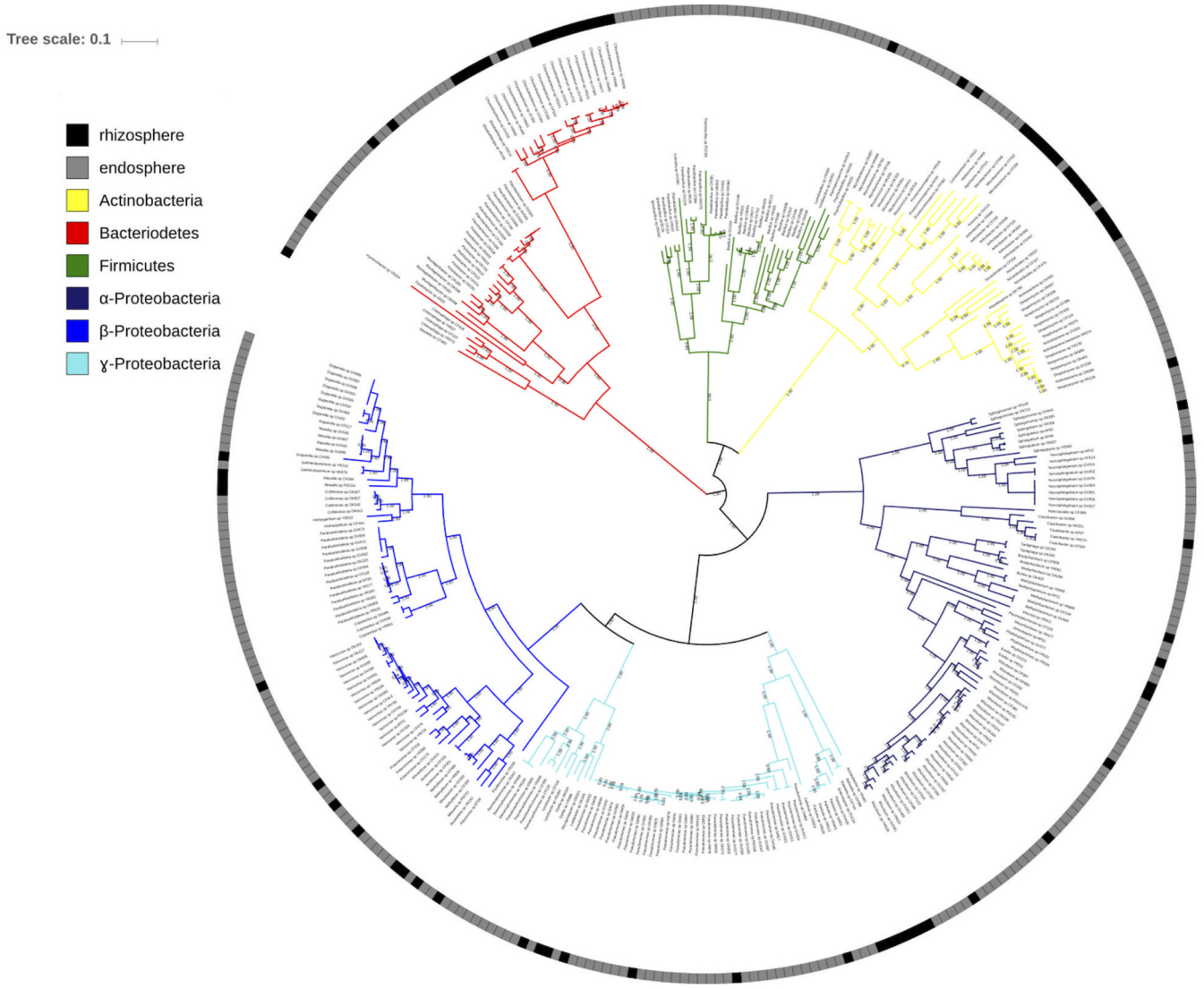

FIG 1 Phylogenetic tree of genome-sequenced bacterial strains from the *Populus* microbiome. The tree was inferred from approximately maximum likelihood approach with FastTree v2.1.9. Branch support values are noted on each branch, and branches are colored by phylum or class as follows: yellow, *Actinobacteria*; red, *Bacteroidetes*; green, *Firmicutes*; dark blue, *Alphaproteobacteria*; blue, *Betaproteobacteria*; light blue, *Gammaproteobacteria*. The isolation compartment is noted with color strip as indicated: black, rhizosphere isolate; gray, endosphere isolate.

order to connect the diversity observed in the metagenome samples to the biosynthetic potential of individual organisms. We determine the key organisms that have the greatest biosynthetic richness and classify clusters on the basis of NP family to better understand community structure in the rhizosphere. We next evaluate distribution of certain NP classes across strains. The biosynthetic potential is assessed for novel compounds and better-known signaling molecules such as siderophores and quorum-sensing homoserine lactones. Finally, we examine the most prevalent family characterized, the RiPPs, for broadly distributed and genus-specific NP clusters. These examples of NPs identified from genome mining reveal the diversity of compounds that can be discovered within the *Populus* rhizosphere.

RESULTS

Phylogenomic analysis of bacterial BGCs. We constructed a phylogenetic tree utilizing the genomes of 339 diverse bacterial strains isolated from the *Populus* microbiome (Fig. 1). Organisms were selected for genome sequencing on the basis of ease of cultivation, representation of the overall phylogenetic diversity of the *Populus*

TABLE 1 Percentage composition of rhizosphere and endosphere metagenomes for the four most common bacterial phyla, and the representation of those phyla in the sequence collection[a]

Phylum and class	% of metagenome	% of sequenced collection	Genome size (Mb)[b]	No. of BGCs	Avg no. of BGCs per genome[b]	% of genome dedicated to BGCs[b]
Actinobacteria	18.70	13.86				
Actinobacteria	14.62	13.86	6.96 ± 2.84	994	21.15 ± 20.85	9.80 ± 4.71
Bacteroidetes	3.01	14.16				
Chitinophagia	0.53	1.47	8.74 ± 0.71	112	22.40 ± 8.93	7.51 ± 9.59
Cytophagia	0.96	0.59	6.35 ± 3.02	19	9.50 ± 7.78	3.65 ± 2.69
Flavobacteriia	0.55	8.55	5.00 ± 0.77	327	11.28 ± 5.99	4.74 ± 2.14
Sphingobacteriia	0.56	3.54	6.48 ± 0.93	95	7.92 ± 4.44	3.18 ± 2.14
Firmicutes	2.58	10.32				
Bacilli	1.35	10.32	5.88 ± 1.15	281	8.03 ± 4.43	5.12 ± 2.16
Proteobacteria	34.56	61.65				
Alphaproteobacteria	17.69	22.42	6.10 ± 1.13	474	6.24 ± 2.92	4.94 ± 2.05
Betaproteobacteria	6.93	21.24	7.06 ± 1.59	605	8.40 ± 4.54	4.13 ± 2.47
Gammaproteobacteria	5.43	17.99	5.82 ± 0.93	502	8.23 ± 4.78	4.56 ± 1.99

[a]Genome size, number of BGCs, average number of BGCs per genome, and percentage of genome dedicated to secondary metabolism are detailed for each class within the four major bacterial phyla, *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*.
[b]Values are means ± standard deviations.

microbiome (6), and potential for use in controlled laboratory/greenhouse studies (21, 37) (see Fig. S1A in the supplemental material). The genomes sequenced, assembled, and annotated in collaboration with the Joint Genome Institute (JGI) as of July 2017 were downloaded from the JGI Integrated Microbial Genomes and Microbiomes website (https://img.jgi.doe.gov/) for further analysis (38). Of the genomes analyzed, 173 are being reported for the first time (Table S1). These bacteria were isolated from wild *Populus deltoides* and *Populus trichocarpa* root samples collected as noted (Table S1) from root surfaces (rhizosphere) or surface-sterilized and ground roots (endosphere isolates).

Bacterial genome size and gene count are linearly related (39); in the sequenced isolates, the number of detected genes was compared to genome size, and no outliers were detected, revealing that genomes were well assembled and annotated (Fig. S1B). The sequenced genomes were subsequently mined for biosynthetic gene clusters (BGCs) using antiSMASH 3.0 (33). Bioinformatic analysis of the collection identified a total of 3,409 BGCs from more than 35 natural product (NP) families (Table S3). Members of *Actinobacteria* harbored the greatest number of BGCs per genome (21.15 ± 20.85 BGCs/genome), but the large standard deviation reflects the fact that the phylum contains the cluster-rich *Streptomyces* genus (45.29 ± 17.81 BGCs/genome) but also genera with few clusters (for example, *Arthrobacter* with 4.43 ± 1.62 BGCs/genome, *Microbacterium* with 3.00 ± 0.00 BGCs/genome, *Nocardioides* with 3.50 ± 1.91 BGCs/genome, and *Promicromonospora* with 4.67 ± 1.53 BGCs/genome) (Table 1 and Fig. S2A). In contrast, *Proteobacteria* and *Firmicutes* harbored many fewer clusters and were less biosynthetically rich (7.56 ± 4.21 and 8.03 ± 4.43 BGCs per genome). Though there is rough scaling between genome size and number of BGCs, when normalized to genome size, *Actinobacteria* microorganisms dedicate a larger portion of the genome to secondary metabolism than all other phyla (Table 1 and Fig. S2B) (40). In contrast, *Paraburkholderia*, Gram-negative bacteria commonly found in close symbiotic relationship with both plants and fungi, dedicate a disproportionately small amount of genomic space to NP biosynthesis (Fig. S2B).

BGCs were grouped into categories on the basis of similar biosynthetic pathways, including nonribosomal peptides (NRPs), polyketides (PKs), ribosomally synthesized and posttranslationally modified peptides (RiPPs), terpenes, saccharides, fatty acids, lactones, siderophores, and hybrids. The most abundant BGCs in the *Populus* microbiome collection encode RiPPs (606 clusters from 97 genomes [Fig. 2]). Clusters from NP families known to produce antibiotics (NRPs, PKs, and RiPPs in particular) were some of

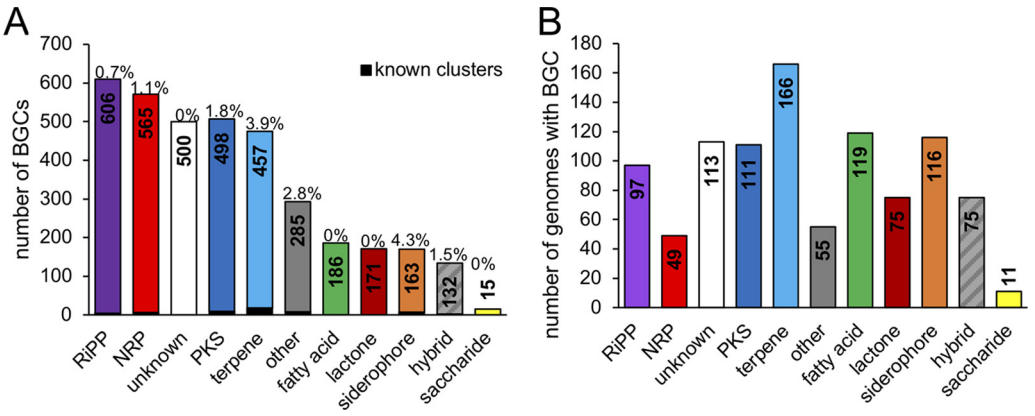

FIG 2 Diversity of natural product cluster classes identified in the sequenced bacterial isolates from the *Populus* microbiome. (A) Total number of gene clusters identified in the collection from each natural product class. The total number of clusters is indicated in boldface type at the top of each bar. Clusters with >85% sequence similarity to a known cluster are indicated as black bars in each class, with the percentage of the class indicated above each bar. (B) Total number of genomes containing clusters of each natural product class. The total number of genomes containing natural products in that class is listed at the top of each bar.

the more abundant cluster types, highlighting the importance of the ability to produce metabolites for bacterial competition. Similarly, terpenes, historically studied in plants and mammals and now known to be widespread in bacteria as well (41), were present in the largest number of genomes, with 49% of the isolates carrying a gene encoding a terpene synthase. Bacterial terpenes are implicated in interkingdom signaling, as the volatile compounds have recently been shown to elicit profound effects on plants (42, 43). Siderophores, discussed in more detail below, were represented in more than one-third of the sequenced isolates, highlighting the importance of nutrient acquisition in the soil. Intriguingly, a total of 500 previously unclassified clusters were identified from more than 30% of organisms in the sequenced collection (Fig. 2). The antiSMASH algorithm identifies clusters using enzymes that perform less common biosynthetic transformations in an attempt to capture possible NPs that fall into previously unexplored families (33).

In order to compare the sequenced isolates from different compartments, we divided results by endosphere and rhizosphere (Fig. S3). Normalizing to the number of genomes within each compartment, endophytic bacteria tend to have slightly more BGCs than rhizosphere isolates (Fig. S3B and C), containing on average more RiPP, NRP, terpene, hybrid, and less common NP clusters. This observation, and the enrichment of homoserine lactone (HSL) and butyrolactone BGCs in endophytic genomes, is in agreement with the metagenomic data, highlighting the importance of quorum sensing in the more confined endosphere. However, siderophore clusters were actually overrepresented in the rhizosphere isolates compared to endophytic isolates (44). This could be the result of bias in the selection of organisms for sequencing and may not reflect an overall trend in iron acquisition in the rhizosphere versus endosphere. Nonetheless, as nutritional availability and microbiome structure differ between compartments, it is not surprising that endophytic bacteria and rhizosphere isolates would be capable of producing different types of NPs.

The most abundant gene cluster type found in the collection encodes RiPPs, of which the majority are bacteriocins and lanthipeptides. However, these clusters are present within only 29% of the collection, while almost half of the sequenced organisms (49%) contain terpene clusters (Fig. 2B). Previous studies have focused on the iteratively constructed NRP and PK NPs, whose enzymatic machinery is diverse even within the microbiota of different deep-sea sponges or urban park soils, not just from other characterized enzymatic domains of similar function but also within the microbiome itself (45, 46). A systematic survey of NPs in the genomes of human gut microbiota showed a vast number of previously uncharacterized gene cluster families,

TABLE 2 Gene cluster product and NP class for clusters with >85% sequence similarity to a characterized BGC

NP name	NP class	No. of BGCs
Alkylresorcinol	PK	7
Carotenoid	Terpene	5
Paeninodin	RiPP	4
Petrobactin	Other	4
Ectoine	Other	4
Albaflavenone	Terpene	4
2-Methylisoborneol	Terpene	3
Hopene	Terpene	3
Flexirubin	Terpene	2
Desferrioxamine B	Siderophore	2
Mycobactin	Siderophore	2
Pyochelin	NRP	2
Zwittermycin A	Hybrid	1
Polymyxin	NRP	1
Mangotoxin	Other	1
Putisolvin	NRP	1
Vicibactin	Siderophore	1
Heterobactin	Siderophore	1
Enterocin	PK	1
Coelichelin	NRP	1
Antimycin	Hybrid	1
Naringenin	PK	1
Bacillibactin	NRP	1
Isorenieratene	Terpene	1
Total		54

along with a vast diversity of BGCs belonging to known families (47, 48). We thus looked at the homology of BGCs identified in the *Populus* microbiome to determine the extent of overlap with previously characterized clusters.

Presence of known BGCs. Importantly, within each NP family, there exists the possibility of widespread structural diversity. Using MiBIG (49), a repository of gene clusters and their experimentally confirmed products, BGCs with >85% similarity to already characterized clusters were identified. Surprisingly, only 54 of the 3,409 clusters (1.6%) matched a known BGC (Fig. 2A and Table 2). The most prevalent known clusters are siderophores and include desferrioxamine B and petrobactin (Fig. S4C and D). While RiPP clusters are the most common BGC in the collection of isolates, only one cluster matched an already-described compound: the lassopeptide paeninodin, which was identified in four genomes (50). The limitation of this analysis is that similarity analysis is performed only with MiBIG-curated clusters, which do not comprise all BGCs positively connected to a NP. However, the diversity of clusters present in the collection, as described below, suggests that many of the isolates harbor unique gene clusters capable of producing previously uncharacterized NPs.

RiPP natural products are diverse and are unique to the soil microbiome. RiPPs are the most prevalent cluster type found in the sequenced collection (Fig. 2). A total of 136 lanthipeptide-producing clusters were identified in addition to 9 lanthipeptide hybrid clusters from 73 of the isolates, representing 23% of all RiPP clusters. The lanthipeptide precursor peptides (LanA proteins) from the collection were compared to known sequences in both GenBank and Uniprot (Fig. 3A and Fig. S5). The lanthipeptide clusters are widely distributed in the *Populus* collection, as they are found in all four represented phyla. The *Populus* microbiome-derived LanA proteins clustered together in most cases, but these clusters were distributed throughout the tree, suggesting that the lanthipeptides represented in the current collection of *Populus* isolates cover new sequence space in the NP family. Lanthipeptides often act as antibiotics, disrupting cell membranes; these molecules could be important in the plant microbiome to protect against pathogens and to keep certain bacterial populations in check.

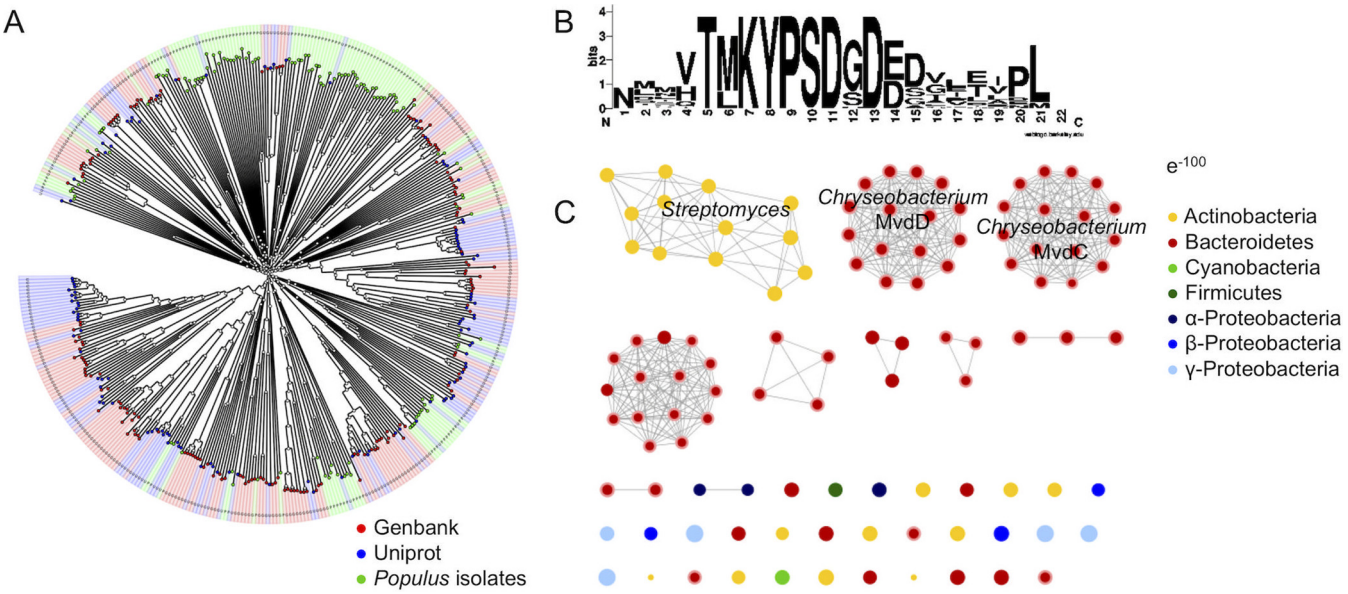

FIG 3 Ribosomally produced and posttranslationally modified (RiPP) clusters from the *Populus* microbiome sequenced isolates. (A) Tree of LanA precursor peptide sequences. The length of the branch represents distance in amino acid sequence diversity. GenBank LanA (red), Uniprot LanA (blue), and sequenced *Populus* isolates (green) are indicated. Genome names are included in Fig. S4 in the supplemental material. (B) WebLogo of the core region of the *Chryseobacterium* microviridin precursor peptides shows conservation of residues required for lactone and lactam formation. (C) Sequence similarity network (SSN) of all ATP-grasp enzymes in the sequenced *Populus* isolates, The nodes, color coded by phylum or class as indicated, represent individual genes and are connected with edges based on an E value of −100, and sized on the basis of amino acid sequence length. The phylum or class is indicated in color as follows: yellow, *Actinobacteria*; red, *Bacteroidetes*; green, *Firmicutes*; dark blue, *Alphaproteobacteria*; blue, *Betaproteobacteria*; light blue, *Gammaproteobacteria*.

Unlike lanthipeptides, one class of RiPP was found to be present in only one bacterial genus in our sequenced bacterial strains. Interestingly, the microviridin cluster was located within 16 of the 18 sequenced *Chryseobacterium* isolates, but not within any other genus in the collection. Microviridins are RiPPs with a cagelike structure formed as a result of two dedicated ATP-grasp ligases, which form two lactone rings on Thr/Asp and Ser/Glu and a lactam ring from Lys/Glu residues within a conserved TXKYPSDXD/E motif in the core region of the MvdA precursor peptide (Fig. 3B) (51, 52). While microviridins have thus far been isolated only from *Cyanobacteria*, the *in vitro* reconstitution of the pathway has enabled the study of microviridins from other bacteria (53, 54). Homology searching of the MdnC ATP-grasp ligase in the NCBI database showed that the microviridin biosynthetic machinery is not confined to cyanobacteria but is also present in *Proteobacteria* and *Bacteroidetes*, in particular *Chryseobacterium* (55, 56). *In vitro* experimentation has revealed that microviridins act as protease inhibitors (54), but the natural function in the microbiome of *Populus* remains to be determined.

The GC content of the microviridin BGC, compared to the entire *Chryseobacterium* genome, does not indicate that the cluster was obtained through horizontal gene transfer from cyanobacteria (cluster percent GC 36.04 ± 1.16; organism percent GC 37.47 ± 1.23) (57). Extensive searches did not yield microviridin-like precursor peptides within other genomes in the collection, and although other putative ATP-grasp enzymes were identified on the basis of homology to MvdC from *Planktothrix agardhii* (Fig. 3C), no other clusters contained both ATP-grasp ligases and a precursor meeting the requirements of microviridins (58). It should be noted that ATP-grasp proteins from other Pfams do not appear in BGCs, as they function instead in primary metabolic pathways (51).

Chryseobacterium isolates from the *Populus* rhizosphere have one to four MvdA genes per cluster, and the core regions of these precursors are typically longer than those of previously identified MvdA sequences (18 residues on average versus 12 to 14 amino acids in cyanobacterial microviridins) (Fig. 3B). The *Chryseobacterium* isolates

from *P. deltoides* have class III precursor peptides, in which a single leader region is fused to a single core region (55). Within a BGC, the individual precursor sequences differ, indicating leniency in peptide recognition of the ATP-grasp ligases. Only 50% of the clusters contained an *N*-acetyltransferase (Table S4), unlike microviridin clusters from cyanobacteria; however, this modification has been shown to not be necessary for protease inhibition (59). Additionally, the presence of a resistance protein (PF00903) in 81% of the clusters suggests that the microviridins may have antibacterial activity related to protease inhibition and the resistance protein may be required for immunity of the native producer (Table S4). It remains to be determined whether the microviridin clusters, as well as other secondary metabolite BGCs from other organisms in the *Populus* microbiome, are transcribed and translated under ecologically relevant conditions and whether the NPs are involved in intra- and interspecies communication.

Lactone clusters highlight the importance of quorum sensing. Quorum sensing (QS) is a vital communication method in the bacterial realm which relies on the production of signaling molecules that convey information about cell density and result in the triggering of various cellular processes. QS regulates biofilm formation, secondary metabolite production, and many other cellular processes. One of the most well-described QS systems is the homoserine lactone cluster, which is comprised of the signaling molecule synthase and the signal receptor (60, 61). HSL clusters are identified in antiSMASH using hidden Markov models (HMMs) based on the Pfams for the synthetase (33). While originally implicated in intraspecies communication, recent studies have also shown that QS molecules can be detected by nonproducing bacteria and even the host plant, whereby plant metabolism, the plant immune response, and root development can be modulated in response to the HSLs (16, 60, 62, 63). Accordingly, the metagenomic analysis of the *P. deltoides* microbiome shows an enrichment of HSLs in the endosphere (Fig. 4B).

Schaefer et al. determined that, of 129 proteobacterial isolates from *P. deltoides* tested, 40% demonstrated HSL activity (16). Secondary metabolite prediction in the sequenced organisms encountered 131 HSL and 40 butyrolactone gene clusters, most prevalent in the *Rhizobium* and *Streptomyces* genera, respectively (Fig. S6A and E). Interestingly, a LuxR gene was found in *Streptomyces* sp. strain OK807, which, while not adjacent to a LuxI gene, had high sequence similarity to proteobacterial sequences (Fig. S6E) (64). Homology searches with the *Streptomyces* sp. OK807 LuxR yield other LuxR genes in *Streptomyces* and do not appear to be the result of horizontal gene transfer from *Proteobacteria* (data not shown). Further investigation will be required in order to determine the signal molecule the *Streptomyces* LuxR detects and what effect LuxR activation has on the organism.

While HSL clusters are known to be prevalent in *Proteobacteria*, LuxR genes are not always adjacent to an HSL synthase gene (Fig. S6D) (16, 65, 66). Indeed, a total of 436 LuxR proteins were found in the 339 sequenced isolates, compared to only 131 HSL synthase proteins. Of the 163 isolates with genes encoding a LuxR, 63 did not contain a LuxI homolog (Fig. S6E), indicating that nonproducing bacteria may be able to eavesdrop on chemical signaling of cell density to regulate their own metabolic pathways (67). Recent evidence suggests that many of these LuxR solos regulate other BGCs, as is the case for the NRP-polyketide synthase (PKS) bactobolin and the β-lactam carbapenem (64). Alternatively, some LuxR solos may be capable of detecting tree-produced signaling molecules. The OryR homolog PipR in *Pseudomonas* sp. strain GM79 was found to recognize an as-yet unknown metabolite in *Populus* leaf macerates, suggesting that LuxR-type proteins can even be involved in interkingdom signaling (63). While the presence of PipR homologs in other *Proteobacteria* (Fig. S6D) (16) does not necessarily indicate a novel NP, plant-derived molecules may also be utilized in the root microbiome to control bacterial cellular processes such as NP biosynthesis or biofilm formation.

Siderophore clusters suggest interspecies competition. Not only is signaling important in a complex environment, bacteria must also compete with one another for

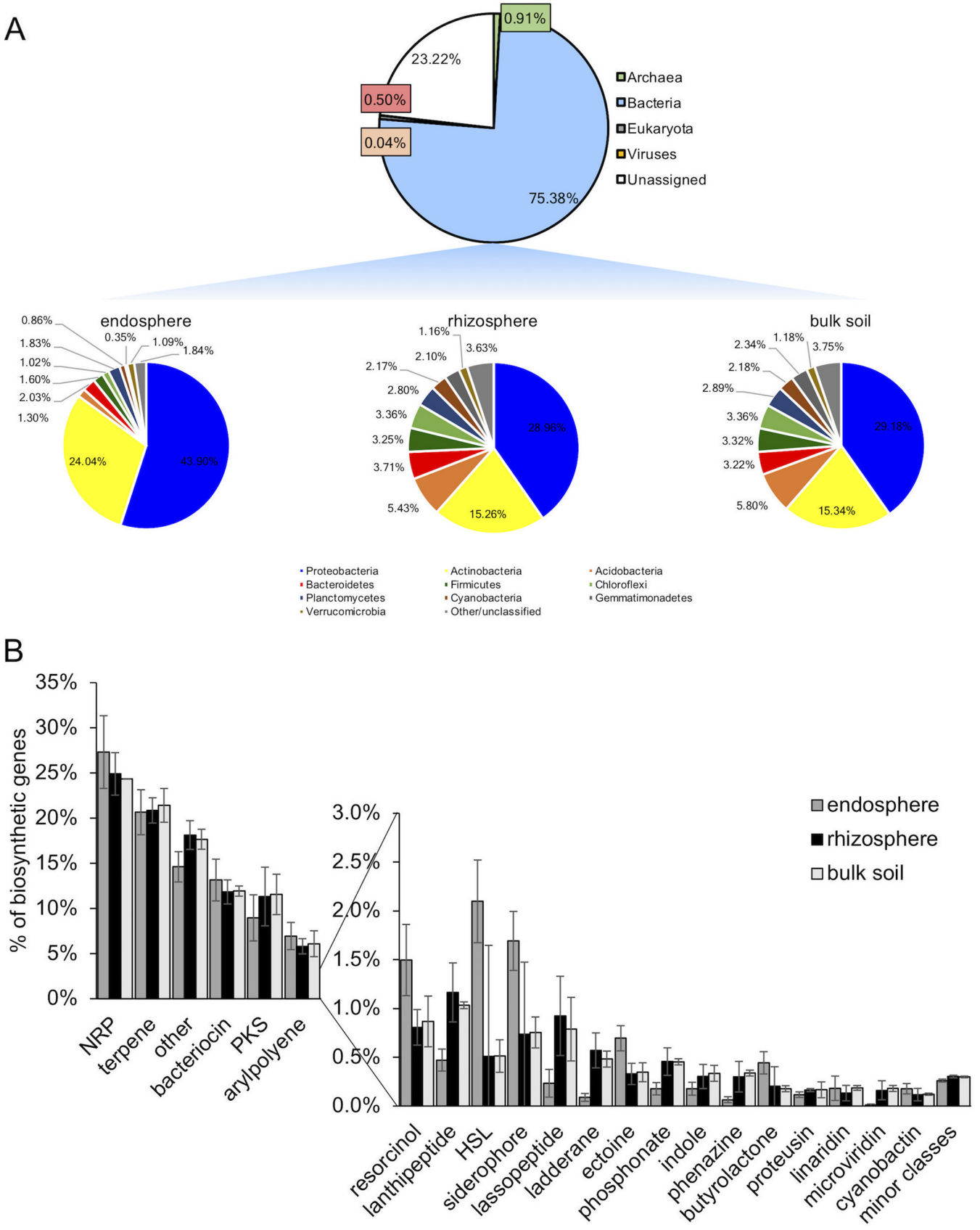

FIG 4 Natural product biosynthetic analysis of the metagenome of *P. deltoides*. (A) Composition of the metagenome by kingdom and the phylogenetic distribution of the bacterial component within the bulk soil, rhizosphere, and soil, based on BLAST phylogeny. (B) Composition of natural product clusters

(Continued on next page)

nutrient acquisition. Siderophores are secondary metabolites that primarily chelate iron, a metal essential for cellular metabolism and growth. Siderophores can be biosynthesized from a variety of pathways, including NRP (enterobactin [68]), RiPP (microcin M [69]), and NRP-independent pathways (desferrioxamine B [70]). AntiSMASH uses PF04183.5, the IucA_IucC family involved in aerobactin biosynthesis (71), to identify siderophore-producing clusters from a biosynthetic pathway exclusive to iron-chelating molecules; thus, the number of siderophore-producing clusters may be dramatically underestimated (33). Indeed, clusters with 100% similarity to the NRP-produced coelichelin and pyochelin were identified in *Pseudomonas* sp. strain GV058 and *Pseudomonas* sp. strain GV077, and *Streptomyces atratus* sp. OK807, respectively (Table 2).

Despite the limitations of estimating the overall number of possible siderophore-producing clusters, siderophore BGCs were widely distributed across genera in our sequenced collection (Fig. S4A and B). Of the 163 gene clusters identified as containing siderophore-producing enzyme genes, 10 had >90% similarity to clusters with a product described in the NP database MiBIG (49), including desferrioxamine B and petrobactin (Fig. S4C and D). These NRP-independent pathways were found in bacteria from 40 different genera. The percent occurrence of siderophore clusters in the sequenced collection does not reflect the enrichment of the cluster observed in the endosphere metagenomic analysis; only 37% of the sequenced isolates contained a siderophore cluster, while 45.33% of sequenced rhizosphere isolates had the potential to make an NRP-independent siderophore. Bacteria compete for limited resources in the plant microbiome, and the presence of a siderophore cluster could be a competitive advantage for the survival of certain species, both within the root and in the rhizosphere. The prevalence of siderophore gene clusters across multiple biosynthetic pathways in the sequenced collection is evidence for resource competition in the *Populus* microbiome as well as the importance of siderophores in plant root colonization (44).

Analysis of the bacterial component of the *Populus* metagenome. In order to determine how well the sequenced isolates reflected the overall soil microbiome, we then surveyed the metagenome of *Populus deltoides* rhizosphere, endosphere, and bulk soil samples to gain a broad perspective of secondary metabolism diversity in a plant-associated microbial community. In comparison to the human gut microbiome, which is dominated by *Firmicutes* and *Bacteroidetes*, the *Populus* root microbiome comprises a greater level of bacterial diversity at the phylum level, with *Actinobacteria*, *Bacteroidetes*, *Firmicutes*, and *Proteobacteria* dominating the bacterial fraction (Fig. 4A) (72, 73). The bacterial distribution we observed is in agreement with previous metagenomics analyses of maize (74), *Arabidopsis* (75), and *Populus tremula* × *Populus alba* field-grown trees (11). Similar to these previous plant microbiome experiments, which showed that *Proteobacteria* make up a greater percentage of the bacterial community moving from bulk soil to the endosphere (11), the component of *Proteobacteria* dramatically increased from 29.2% of bulk soil bacterial reads to 43.9% in the endosphere (alphaproteobacteria increased most significantly, from 12.9% to 25.7%). *Actinobacteria* also increased from 15.3% to 24.0%, as seen in previous analyses (Fig. 4A) (76). Differences in plant host species, genotype, soil, and environment can have dramatic effects on the composition of the microbiome (77). In spite of the differences, the most prevalent phyla in our samples matched those of other studies: *Proteobacteria*, *Firmicutes*, *Bacteroidetes*, and *Actinobacteria* dominated the bacterial metagenome counts (78), and this bacterial diversity is reflected in the genome-sequenced isolate collection as well (Table 1). We expect that the increased diversity of species found in soil samples compared to the human gut suggests that the plant root microbiome

FIG 4 Legend (Continued)
identified in the metagenome of *P. deltoides*, separated into endosphere, rhizosphere, and bulk soil components. Clusters were grouped by natural product family and divided by the total number of clusters identified to give relative prevalence of cluster type within each compartment.

could similarly have greater diversity in natural product biosynthetic pathways. Thus, the assembled *Populus* metagenomes were analyzed by antiSMASH (33) for the presence of genes associated with secondary metabolite biosynthesis. Samples were compiled and separated on the basis of location of sample extraction: endosphere, rhizosphere, or bulk soil (Fig. 4B).

Because of the short sequence reads common to metagenomic samples that lead to fragmented genes, the numbers of biosynthetic genes identified per sample were low compared to overall data set size (4,825 $\pm$ 1,214 biosynthetic genes in the endosphere, 9,046 $\pm$ 1,744 biosynthetic genes in the rhizosphere, and 7,545 $\pm$ 2,552 biosynthetic genes in bulk soil). Raw sequence counts were highest in the rhizosphere; however, raw counts for bacterial contigs were also highest in the rhizosphere versus endosphere and bulk soil (79). Thus, we normalized the data in order to compare the relative prevalence of NP type within all identified genes in a compartment, thereby accounting for different metagenome sample sizes (Fig. 4B).

Across all compartments, NP types commonly associated with production of antibiotics, including the NRP, PK, and RiPP families, were the most prevalent. Terpenes, odoriferous and volatile hydrocarbons constructed from isoprenoid building blocks, were also prevalent in all compartments (41). These trends reflect the prevalence observed within the sequenced isolates.

The quorum-sensing homoserine lactones and butyrolactones are enriched in the endosphere in both sequenced isolates and metagenomic data, suggesting a great importance of intraspecies communication within the root. While bacteria in the rhizosphere also rely on quorum-sensing molecules for regulation of certain pathways, the environmental pressures in the endosphere may enhance the need for population-based transcriptional control. Signaling provides an advantage to bacteria in the endosphere as they compete to colonize and acquire nutrients for survival; HSLs allow bacteria to appropriately respond to nutrient availability and the presence of other organisms in order to efficiently colonize and survive.

The endosphere is also enriched in siderophore biosynthetic genes compared to the rhizosphere and bulk soil, indicating that nutritional availability varies across the compartments. Bacteria in the endosphere may need to compete more for iron than those organisms exposed to soil. Biosynthetic genes connected to less common processes in NP biosynthesis (condensation of multiple complex moieties, for example), classified as "other" were most common in the rhizosphere and bulk soil samples, perhaps reflecting the larger percentage of unclassified bacteria and thus, NPs with less common or less well-characterized biosynthetic pathways in these compartments (Table S2).

DISCUSSION

In this study, we investigated the natural product (NP) biosynthetic potential of the *Populus* microbiome by analyzing metagenomic data as well as a set of sequenced bacterial isolates taken from the *Populus* rhizosphere and endosphere. Metagenomic analysis reveals the overall potential of the microbiome, including the estimated 95% of bacteria that are unculturable by current techniques (2), to produce secondary metabolites. Just as the diversity of bacterial isolates is dramatically increased compared to the well-studied human gut microbiome, NP diversity may also be enhanced in the *Populus* microbiome. The most abundant NP genes identified include RiPPs, NRPs, terpenes, and those of unknown class. The metagenome samples showed that the endophytic samples are enriched in siderophores and HSLs, but siderophores were found to be more prevalent in rhizosphere isolates when analyzing fully sequenced genomes. As the bacterial population is also distinct in the endosphere versus the rhizosphere, the molecular arsenal with which organisms arm themselves in these distinct spheres must also differ (44).

Metagenomic data are suggestive of the overall potential of the microbiome to make a vast array of NPs, from the most abundant NRP class to less abundant hybrids and putative secondary metabolite clusters with no closely related NP type. However,

metagenomic analysis has its limitations. The libraries contain many short sequence reads, leading to a lack of genomic context for the identified enzymes. This makes connecting biosynthetic genes to the organism producing the product extremely challenging and inhibits further investigation of clusters. In order to enable an analysis of the distribution of NP classes across NP-producing organisms as well as the structure and function of individual NPs in the microbiome, the cluster should be positively correlated to a specific genome.

Bioinformatic analysis of the genomes of 339 bacterial strains isolated from the *Populus* microbiome revealed that the NP cluster distribution clearly favors *Actinobacteria* but also that the most prevalent members are not necessarily the most prolific in terms of NP biosynthesis. Nonetheless, less abundant members of the microbiome may still have a large impact on overall community structure by producing molecules that affect intra- and interspecies communication.

However, the presence of a BGC in a genome does not indicate that a product is synthesized; many BGCs are cryptic, and biosynthetic potential is not a reflection of an organism's impact on community structure (80). Endophytic bacteria harbor more BGCs on average than rhizosphere isolates, perhaps because of a greater need to compete for space and nutrients within a confined and crowded environment. The presence of BGCs involved in nutrient acquisition, such as siderophore clusters for iron uptake, reveal that competition for resources likely dictates community structure. Thus, certain NPs may also serve as antibiotics in order to regulate population levels within the community. While certain classes of NPs are expected in the plant microbiome, such as quorum-sensing (QS) molecules and siderophores, there are a vast number of other NP gene clusters within the genomes of these soil-dwelling bacteria. The products of these BGCs may have significant roles in the shaping of the microbiome and thus the overall health of the host plant. We investigated the diversity of other NP clusters within the 339 sequenced genomes and discovered a RiPP cluster present in only one genus of bacteria. Microviridin BGCs are found only in *Chryseobacterium* spp., suggesting a niche role for the NP. In contrast to the microviridins, other RiPP clusters, such as those implicated in lanthipeptide biosynthesis, are broadly distributed across bacterial genera. While an abundance of clusters exists for compounds already known to influence community dynamics, such as siderophores and HSLs, only 1.6% of clusters have an NP already described, and 500 clusters have no discernible NP classification, suggesting that a wealth of novel chemical diversity exists in the *Populus* rhizosphere. Determining the production, and subsequently the structure and function, of these bioinformatically predicted molecules will be important for understanding the multifaceted communication networks and community structuring taking place in the *Populus* microbiome.

MATERIALS AND METHODS

Metagenome sampling. Samples were collected from an experimental cultivar trial in Blount County in Tennessee at a site managed by the University of Tennessee Institute of Agriculture (UTIA) – East Tennessee Research and Education Center (ETREC) (35°50′N, 83°57′W) in August 2014. Metagenome samples were collected from the bulk soil, rhizosphere, and endosphere of *Populus deltoides* roots, and DNA was extracted in the laboratory as described previously (81). DNA extraction was performed using the MoBio PowerSoil DNA isolation kit (MoBio Laboratories, Inc., Carlsbad, CA, USA) after bead-beating frozen samples (3 min in frozen blocks using one 5-mm steel bead per sample) to pulverize roots.

Strain isolation. Root samples from *Populus deltoides* and *Populus trichocarpa* were collected from mature trees and were processed for strain isolation as described previously (6, 22, 82–84). Caney Fork River samples were from native *P. deltoides* in the Buffalo Valley Recreation Area in DeKalb County, TN (36˚6′N, 85˚50′W). Clatskanie samples were from common garden-grown *P. trichocarpa* in the Columbia River Valley in Oregon (123°40′W, 46°6′N). Corvallis samples were from common garden-grown *P. trichocarpa* in the Columbia River Valley in Oregon (44°35′N, 123°11′W). Oak Ridge samples were from native *P. deltoides* on the Oak Ridge Reservation, TN, or *P. trichocarpa* grown in Corvallis, OR, soil in a greenhouse in Oak Ridge, TN (35°55′N, 84°19′W). Yadkin River samples were collected from native *P. deltoides* in eastern North Carolina (35˚43′N, 80˚24′W).

To collect rhizosphere isolates, root washes were serially plated on R2A agar (Thermo Fisher). Endosphere strains were isolated by surface sterilizing roots (wash five times in sterile H_2O, incubate in 95% ethanol [EtOH] for 30 s, incubate in 5% NaOCl for 3 min, wash six times in sterile H_2O), pulverizing roots with a sterile mortar and pestle in 10 mM $MgSO_4$ and serially plating dilutions on R2A agar. Colonies were picked and restreaked a minimum of three times on R2A agar. Strains were identified by 16S rDNA

PCR amplification using primers 8F (F stands for forward) (AGAGTTTGATCCTGGCTCAG) and 1429R (R stands for reverse) (GGTTACCTTGTTACGACTT), followed by Sanger sequencing and analysis prior to complete genome sequencing.

Genome and metagenome sequences. Metagenome and genome sequencing of the isolates was performed at the Joint Genome Institute (JGI, Walnut Creek, CA, USA) (http://jgi.doe.gov/) as described previously (82, 83). Isolates were sequenced on next-generation sequencing platforms, most commonly Illumina HiSeq technology (see Table S2 in the supplemental material). Sequenced genomic DNA was assembled as noted (Table S2) and annotated using the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4) (85). The metagenome samples analyzed were downloaded from Integrated Microbial Genomes & Microbiomes (IMG) and were subsequently analyzed by antiSMASH 3.0. Raw gene counts were normalized to metagenome size: metagenome no. 3300006177, 2,569 Mb and 3,453 biosynthetic genes (BGs), metagenome no. 3300006048, 3,449 Mb and 6,407 BGs; metagenome no. 3300006051, 3,169 Mb and 4,797 BGs; metagenome no. 3300006038, 3,467 Mb and 4,643 BGs; metagenome no. 3300006049, 2,091 Mb and 5,608 BGs; metagenome no. 3300009100, 4,243 Mb and 7,633 BGs; metagenome no. 3300006969, 1,993 Mb and 7,509 BGs; metagenome no. 3300006853, 4,672 Mb and 10,246 BGs; metagenome no. 3300006845, 5,908 Mb and 4,414 BGs; metagenome no. 3300009147, 4,951 Mb and 10,051 BGs; metagenome no. 3300006844, 5,764 Mb and 8,085 BGs; metagenome no. 3300006880, 4,651 Mb and 9,824 BGs; metagenome no. 3300006846, 4,343 Mb and 7,954 BGs; metagenome no. 3300006847, 5,274 Mb and 11,629 BGs.

Gene cluster identification and classification. Genomes were downloaded from JGI in FASTA format (July 2017) and were submitted for analysis using antiSMASH 3.0 (33). Metagenome samples were additionally analyzed using BAGEL3 (34). Individual clusters and groups of similar clusters were further analyzed using PRISM (35) (http://magarveylab.ca) and RODEO (86) (http://ripprodeo.org) using FASTA files or protein accession numbers, respectively.

Metagenome analysis. Metagenomes from five *P. deltoides* trees were downloaded from JGI-IMG in FASTA format (June 2017) and were submitted for analysis using antiSMASH 3.0 (33). Phylogeny of rhizosphere, endosphere, and bulk soil metagenomes was determined on the IMG/MER platform (http://img.jgi.doe.gov) using the best BLAST hits of protein-coding genes of the assembled samples at 30% cumulative percent identity.

GC content analysis. The percent GC of nucleotide sequences for biosynthetic gene clusters was calculated using the output files for predicted clusters in antiSMASH and compared to NCBI Taxonomy records of organism percent GC content. For microviridin cluster percent GC, the antiSMASH-defined cluster nucleotide sequence percent GC was compared to the organism percent GC.

Phylogenetic tree generation. The phylogenetic tree was generated using genomes downloaded from KBase (https://kbase.us). FastTree v.2.1.9 (87) was used to determine the approximately maximum likelihood phylogenies for the isolates, and the resulting tree based on nearest-neighbor interchanges was visualized in interactive tree of life (iTOL) (http://itol.embl.de/) (88).

Proteins with LanA predictions were identified by the presence of either a leader/code peptide or a LD_lanti_pre domain using antiSMASH 3.0. There were 73 genomes with one or more predictions and a total of 136 unique sequences. The search term "lantibiotic" was used to identify confirmed lantibiotic proteins in GenBank and Uniprot. These known proteins were used to provide a context for the PMI newly predicted proteins. A total of 417 unique sequences from the three sources were aligned with ClustalX, and a neighbor-joining tree was constructed with a bootstrap value of 1,000. The dendrogram of proteins was colored by source to confirm that the new sequences include both proteins similar to known lantibiotic proteins and novel sequences.

LuxR homologs were identified using Pfam03472 in the IMG database (16). From this set of 444 proteins, only those with a domain matching Pfam00196 (LuxR-type DNA binding helix-turn-helix [HTH] domain) were selected, leaving 436 proteins. LuxI proteins were identified in IMG by searching for COG3916 within the sequenced genomes and selecting proteins adjacent to a LuxR homolog. LuxR protein sequences were exported from JGI-IMG and aligned using MUSCLE3.8.31 (89). The phylogenetic tree data were exported, and the tree was visualized using iTOL (http://itol.embl.de/) (88).

Sequence similarity network generation. Enzyme function initiative-enzyme similarity tool (EFI-EST) (http://efi.igb.illinois.edu/efi-est/) (90) was used to generate the sequence similarity network (SSN) for siderophore synthase genes and were visualized in Organic layout in Cytoscape v. 3.5.1 (91). Nodes were annotated using the JGI gene set list generated from the output of the search for siderophore synthase genes, which was accomplished by searching for COG4264 hits in the sequenced genomes.

Sequence logo generation. WebLogo (http://weblogo.berkeley.edu/) (92) was used to generate a logo for the core sequence of microviridins after first aligning with MUSCLE 3.8.31 (89).

Data availability. The DNA sequence data sets supporting the conclusions of this article are available in the JGI Integrated Microbial Genomes & Microbiomes repository (https://img.jgi.doe.gov/) using the genome identifiers (IDs) provided in Table S1. The IMG accession numbers for the metagenome samples are as follows: 3300006177 (endosphere), 3300006048 (endosphere), 3300006051 (endosphere), 3300006038 (endosphere), 3300006049 (bulk soil), 3300009100 (bulk soil), 3300006969 (bulk soil), 3300006853 (bulk soil), 3300006845 (bulk soil), 3300009147 (rhizosphere), 3300006844 (rhizosphere), 3300006880 (rhizosphere), 3300006846 (rhizosphere), and 3300006847 (rhizosphere).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00045-18.

FIG S1, PDF file, 0.1 MB.
FIG S2, PDF file, 0.1 MB.
FIG S3, PDF file, 0.02 MB.
FIG S4, PDF file, 0.2 MB.
FIG S5, PDF file, 0.2 MB.
FIG S6, PDF file, 0.3 MB.
TABLE S1, XLSX file, 0.03 MB.
TABLE S2, XLSX file, 0.01 MB.
TABLE S3, XLSX file, 0.01 MB.
TABLE S4, XLSX file, 0.01 MB.

ACKNOWLEDGMENTS

This article has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy (DOE).

This research was funded by the U.S. DOE Office of Biological and Environmental Research, Genomic Science Program as part of the Plant Microbe Interfaces Scientific Focus Area (http://pmi.ornl.gov). Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725. The sequencing work was conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231. The sequencing work was conducted using JGI CSP1429 under project manager Nichole Shapiro.

P.M.B. designed, performed, and analyzed experiments and wrote the article. M.L.L., D.A.J., and M.J.P. conducted bioinformatic data collection. T.-Y.S.L. isolated strains and genomic DNA (gDNA). D.A.P. and M.J.D. designed experiments and edited the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. Berendsen RL, Pieterse CM, Bakker PA. 2012. The rhizosphere microbiome and plant health. Trends Plant Sci 17:478–486. https://doi.org/10.1016/j.tplants.2012.04.001.
2. Mendes R, Garbeva P, Raaijmakers JM. 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. FEMS Microbiol Rev 37:634–663. https://doi.org/10.1111/1574-6976.12028.
3. Bardgett RD, van der Putten WH. 2014. Belowground biodiversity and ecosystem functioning. Nature 515:505–511. https://doi.org/10.1038/nature13855.
4. Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. 2013. Going back to the roots: the microbial ecology of the rhizosphere. Nat Rev Microbiol 11:789–799. https://doi.org/10.1038/nrmicro3109.
5. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen G-L, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, dePamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, et al. 2006. The genome of black cottonwood, *Populus trichocarpa*. Science 313:1596–1604. https://doi.org/10.1126/science.1128691.
6. Gottel NR, Castro HF, Kerley M, Yang Z, Pelletier DA, Podar M, Karpinets T, Uberbacher E, Tuskan GA, Vilgalys R, Doktycz MJ, Schadt CW. 2011. Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. Appl Environ Microbiol 77:5934–5944. https://doi.org/10.1128/AEM.05255-11.
7. Utturkar SM, Cude WN, Robeson MS, Yang ZK, Klingeman DM, Land ML, Allman SL, Lu TYS, Brown SD, Schadt CW, Podar M, Doktycz MJ, Pelletier DA. 2016. Enrichment of root endophytic bacteria from *Populus deltoides* and single-cell-genomics analysis. Appl Environ Microbiol 82:5698–5708. https://doi.org/10.1128/AEM.01285-16.
8. Shakya M, Gottel N, Castro H, Yang ZK, Gunter L, Labbe J, Muchero W, Bonito G, Vilgalys R, Tuskan G, Podar M, Schadt CW. 2013. A multifactor analysis of fungal and bacterial community structure in the root microbiome of mature *Populus deltoides* trees. PLoS One 8:e76382. https://doi.org/10.1371/journal.pone.0076382.
9. Castro HF, Classen AT, Austin EE, Norby RJ, Schadt CW. 2010. Soil microbial community responses to multiple experimental climate change drivers. Appl Environ Microbiol 76:999–1007. https://doi.org/10.1128/AEM.02874-09.
10. Bonito G, Reynolds H, Robeson MS, Nelson J, Hodkinson BP, Tuskan G, Schadt CW, Vilgalys R. 2014. Plant host and soil origin influence fungal and bacterial assemblages in the roots of woody plants. Mol Ecol 23:3356–3370. https://doi.org/10.1111/mec.12821.
11. Beckers B, Op De Beeck M, Weyens N, Boerjan W, Vangronsveld J. 2017. Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. Microbiome 5:25. https://doi.org/10.1186/s40168-017-0241-2.
12. Reddy BVB, Kallifidas D, Kim JH, Charlop-Powers Z, Feng Z, Brady SF. 2012. Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. Appl Environ Microbiol 78:3744–3752. https://doi.org/10.1128/AEM.00102-12.
13. Aleti G, Nikolić B, Brader G, Pandey RV, Antonielli L, Pfeiffer S, Oswald A, Sessitsch A. 2017. Secondary metabolite genes encoded by potato rhizosphere microbiomes in the Andean highlands are diverse and vary with sampling site and vegetation stage. Sci Rep 7:2330. https://doi.org/10.1038/s41598-017-02314-x.
14. Hacquard S, Schadt CW. 2015. Towards a holistic understanding of the beneficial interactions across the *Populus* microbiome. New Phytol 205:1424–1430. https://doi.org/10.1111/nph.13133.
15. Charlop-Powers Z, Owen JG, Reddy BVB, Ternei MA, Guimarães DO, de Frias UA, Pupo MT, Seepe P, Feng Z, Brady SF. 2015. Global biogeographic sampling of bacterial secondary metabolism. Elife 4:e05048. https://doi.org/10.7554/eLife.05048.
16. Schaefer AL, Lappala CR, Morlen RP, Pelletier DA, Lu TY, Lankford PK, Harwood CS, Greenberg EP. 2013. LuxR- and LuxI-type quorum-sensing circuits are prevalent in members of the *Populus deltoides* microbiome. Appl Environ Microbiol 79:5745–5752. https://doi.org/10.1128/AEM.01417-13.

17. Venturi V, Fuqua C. 2013. Chemical signaling between plants and plant-pathogenic bacteria. Annu Rev Phytopathol 51:17–37. https://doi.org/10.1146/annurev-phyto-082712-102239.
18. Kahn ML, Kraus J, Somerville JE. 1985. A model of nutrient exchange in the Rhizobium-legume symbiosis, p 193–199. *In* Evans HJ, Bottomley PJ, Newton WE (ed), Nitrogen fixation research progress. Proceedings of the 6th International Symposium on Nitrogen Fixation, Corvallis, OR 97221, August 4–10, 1985. Martinus Nijhoff Publishers, Dordrecht, The Netherlands.
19. Timm CM, Campbell AG, Utturkar SM, Jun S-R, Parales RE, Tan WA, Robeson MS, Lu T-YS, Jawdy S, Brown SD, Ussery DW, Schadt CW, Tuskan GA, Doktycz MJ, Weston DJ, Pelletier DA. 2015. Metabolic functions of *Pseudomonas fluorescens* strains from *Populus deltoides* depend on rhizosphere or endosphere isolation compartment. Front Microbiol 6:1118. https://doi.org/10.3389/fmicb.2015.01118.
20. Chisholm ST, Coaker G, Day B, Staskawicz BJ. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124:803–814. https://doi.org/10.1016/j.cell.2006.02.008.
21. Timm CM, Pelletier DA, Jawdy SS, Gunter LE, Henning JA, Engle N, Aufrecht J, Gee E, Nookaew I, Yang Z, Lu TY, Tschaplinski TJ, Doktycz MJ, Tuskan GA, Weston DJ. 2016. Two poplar-associated bacterial isolates induce additive favorable responses in a constructed plant-microbiome system. Front Plant Sci 7:497. https://doi.org/10.3389/fpls.2016.00497.
22. Weston DJ, Pelletier DA, Morrell-Falvey JL, Tschaplinski TJ, Jawdy SS, Lu TY, Allen SM, Melton SJ, Martin MZ, Schadt CW, Karve AA, Chen JG, Yang X, Doktycz MJ, Tuskan GA. 2012. *Pseudomonas fluorescens* induces strain-dependent and strain-independent host plant responses in defense networks, primary metabolism, photosynthesis, and fitness. Mol Plant Microbe Interact 25:765–778. https://doi.org/10.1094/MPMI-09-11-0253.
23. Ratcliff WC, Denison RF. 2011. Alternative actions for antibiotics. Science 332:547–548. https://doi.org/10.1126/science.1205970.
24. Yim G, Wang HH, Davies J. 2007. Antibiotics as signalling molecules. Philos Trans R Soc Lond B Biol Sci 362:1195–1200. https://doi.org/10.1098/rstb.2007.2044.
25. Fajardo A, Martinez JL. 2008. Antibiotics as signals that trigger specific bacterial responses. Curr Opin Microbiol 11:161–167. https://doi.org/10.1016/j.mib.2008.02.006.
26. Aminov RI. 2013. Biotic acts of antibiotics. Front Microbiol 4:241. https://doi.org/10.3389/fmicb.2013.00241.
27. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147. https://doi.org/10.1038/417141a.
28. Milshteyn A, Schneider JS, Brady SF. 2014. Mining the metabiome: identifying novel natural products from microbial communities. Chem Biol 21:1211–1223. https://doi.org/10.1016/j.chembiol.2014.08.006.
29. Behie SW, Bonet B, Zacharia VM, McClung DJ, Traxler MF. 2017. Molecules to ecosystems: Actinomycete natural products *in situ*. Front Microbiol 7:2149. https://doi.org/10.3389/fmicb.2016.02149.
30. Berdy J. 2005. Bioactive microbial metabolites. J Antibiot 58:1–26. https://doi.org/10.1038/ja.2005.1.
31. Newman DJ, Cragg GM. 2016. Natural products as sources of new drugs from 1981 to 2014. J Nat Prod 79:629–661. https://doi.org/10.1021/acs.jnatprod.5b01055.
32. Singh BK, Millard P, Whiteley AS, Murrell JC. 2004. Unravelling rhizosphere-microbial interactions: opportunities and limitations. Trends Microbiol 12:386–393. https://doi.org/10.1016/j.tim.2004.06.008.
33. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237–W243. https://doi.org/10.1093/nar/gkv437.
34. van Heel AJ, de Jong A, Montalbán-López M, Kok J, Kuipers OP. 2013. BAGEL3: automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. Nucleic Acids Res 41:W448–W453. https://doi.org/10.1093/nar/gkt391.
35. Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster ALH, Wyatt MA, Magarvey NA. 2015. Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Res 43:9645–9662. https://doi.org/10.1093/nar/gkv1012.
36. Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian K-D, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJT, Rebuffat S, Ross RP, Sahl H-G, Schmidt EW, Selsted ME, Severinov K, et al. 2013. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep 30:108–160. https://doi.org/10.1039/C2NP20085F.
37. Henning JA, Weston DJ, Pelletier DA, Timm CM, Jawdy SS, Classen AT. 2016. Root bacterial endophytes alter plant phenotype, but not physiology. PeerJ 4:e2606. https://doi.org/10.7717/peerj.2606.
38. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG: the Integrated Microbial Genomes database and comparative analysis system. Nucleic Acids Res 40:D115–D122. https://doi.org/10.1093/nar/gkr1044.
39. Konstantinidis KT, Tiedje JM. 2004. Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci U S A 101:3160–3165. https://doi.org/10.1073/pnas.0308653100.
40. Viaene T, Langendries S, Beirinckx S, Maes M, Goormachtig S. 2016. Streptomyces as a plant's best friend? FEMS Microbiol Ecol 92:fiw119. https://doi.org/10.1093/femsec/fiw119.
41. Yamada Y, Kuzuyama T, Komatsu M, Shin-ya K, Omura S, Cane DE, Ikeda H. 2015. Terpene synthases are widely distributed in bacteria. Proc Natl Acad Sci U S A 112:857–862. https://doi.org/10.1073/pnas.1422108112.
42. Schulz-Bohm K, Martín-Sánchez L, Garbeva P. 2017. Microbial volatiles: small molecules with an important role in intra- and inter-kingdom interactions. Front Microbiol 8:2484. https://doi.org/10.3389/fmicb.2017.02484.
43. Tahir HA, Gu Q, Wu H, Raza W, Hanif A, Wu L, Colman MV, Gao X. 2017. Plant growth promotion by volatile organic compounds produced by *Bacillus subtilis* SYST2. Front Microbiol 8:171. https://doi.org/10.3389/fmicb.2017.00171.
44. Compant S, Clément C, Sessitsch A. 2010. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: their role, colonization, mechanisms involved and prospects for utilization. Soil Biol Biochem 42:669–678. https://doi.org/10.1016/j.soilbio.2009.11.024.
45. Charlop-Powers Z, Pregitzer CC, Lemetre C, Ternei MA, Maniko J, Hover BM, Calle PY, McGuire KL, Garbarino J, Forgione HM, Charlop-Powers S, Brady SF. 2016. Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. Proc Natl Acad Sci U S A 113:14811–14816. https://doi.org/10.1073/pnas.1615581113.
46. Borchert E, Jackson SA, O'Gara F, Dobson ADW. 2016. Diversity of natural product biosynthetic genes in the microbiome of the deep sea sponges *Inflatella pellicula*, *Poecillastra compressa*, and *Stelletta normani*. Front Microbiol 7:1027.
47. Donia MS, Cimermancic P, Schulze CJ, Wieland Brown LC, Martin J, Mitreva M, Clardy J, Linington RG, Fischbach MA. 2014. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. Cell 158:1402–1414. https://doi.org/10.1016/j.cell.2014.08.032.
48. Donia MS, Fischbach MA. 2015. Small molecules from the human microbiota. Science 349:1254766. https://doi.org/10.1126/science.1254766.
49. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann H-W, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, Yim G, Yu F, Xie Y, Aigle B, Apel AK, Balibar CJ, Balskus EP, Barona-Gómez F, Bechthold A, Bode HB, Borriss R, Brady SF, Brakhage AA, Caffrey P, Cheng Y-Q, Clardy J, Cox RJ, De Mot R, Donadio S, Donia MS, van der Donk WA, Dorrestein PC, Doyle S, Driessen AJM, Ehling-Schulz M, Entian K-D, Fischbach MA, Gerwick L, Gerwick WH, Gross H, et al. 2015. Minimum information about a biosynthetic gene cluster. Nat Chem Biol 11:625. https://doi.org/10.1038/nchembio.1890.

50. Zhu S, Hegemann JD, Fage CD, Zimmermann M, Xie X, Linne U, Marahiel MA. 2016. Insights into the unique phosphorylation of the lasso peptide paeninodin. J Biol Chem 291:13662–13678. https://doi.org/10.1074/jbc.M116.722108.
51. Li K, Condurso HL, Li G, Ding Y, Bruner SD. 2016. Structural basis for precursor protein-directed ribosomal peptide macrocyclization. Nat Chem Biol 12:973–979. https://doi.org/10.1038/nchembio.2200.
52. Weiz AR, Ishida K, Quitterer F, Meyer S, Kehr JC, Muller KM, Groll M, Hertweck C, Dittmann E. 2014. Harnessing the evolvability of tricyclic microviridins to dissect protease-inhibitor interactions. Angew Chem Int Ed Engl 53:3735–3738. https://doi.org/10.1002/anie.201309721.
53. Reyna-González E, Schmid B, Petras D, Süssmuth RD, Dittmann E. 2016. Leader peptide-free *in vitro* reconstitution of microviridin biosynthesis enables design of synthetic protease-targeted libraries. Angew Chem Int Ed Engl 55:9398–9401. https://doi.org/10.1002/anie.201604345.
54. Ziemert N, Ishida K, Liaimer A, Hertweck C, Dittmann E. 2008. Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. Angew Chem Int Ed Engl 47:7756–7759. https://doi.org/10.1002/anie.200802730.
55. Ahmed MN, Reyna-González E, Schmid B, Wiebach V, Süssmuth RD, Dittmann E, Fewer DP. 2017. Phylogenomic analysis of the microviridin biosynthetic pathway coupled with targeted chemo-enzymatic synthesis yields potent protease inhibitors. ACS Chem Biol 12:1538–1546. https://doi.org/10.1021/acschembio.7b00124.
56. Gatte-Picchi D, Weiz A, Ishida K, Hertweck C, Dittmann E. 2014. Functional analysis of environmental DNA-derived microviridins provides new insights into the diversity of the tricyclic peptide family. Appl Environ Microbiol 80:1380–1387. https://doi.org/10.1128/AEM.03502-13.
57. Garcia-Vallve S, Romeu A, Palau J. 2000. Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Res 10:1719–1725. https://doi.org/10.1101/gr.130000.
58. Iyer LM, Abhiman S, Maxwell Burroughs A, Aravind L. 2009. Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of novel metabolites and peptide modifications of proteins. Mol Biosyst 5:1636–1660. https://doi.org/10.1039/b917682a.
59. Ziemert N, Ishida K, Weiz A, Hertweck C, Dittmann E. 2010. Exploiting the natural diversity of microviridin gene clusters for discovery of novel tricyclic depsipeptides. Appl Environ Microbiol 76:3568–3574. https://doi.org/10.1128/AEM.02858-09.
60. Gerdt JP, Wittenwyler DM, Combs JB, Boursier ME, Brummond JW, Xu H, Blackwell HE. 2017. Chemical interrogation of LuxR-type quorum sensing receptors reveals new insights into receptor selectivity and the potential for interspecies bacterial signaling. ACS Chem Biol 12:2457–2464. https://doi.org/10.1021/acschembio.7b00458.
61. Fuqua C, Parsek MR, Greenberg EP. 2001. Regulation of gene expression by cell-to-cell communication: acyl-homoserine lactone quorum sensing. Annu Rev Genet 35:439–468. https://doi.org/10.1146/annurev.genet.35.102401.090913.
62. Kan J, Fang R, Jia Y. 2017. Interkingdom signaling in plant-microbe interactions. Sci China Life Sci 60:785–796. https://doi.org/10.1007/s11427-017-9092-3.
63. Schaefer AL, Oda Y, Coutinho BG, Pelletier DA, Weiburg J, Venturi V, Greenberg EP, Harwood CS. 2016. A LuxR homolog in a cottonwood tree endophyte that activates gene expression in response to a plant signal or specific peptides. mBio 7:e01101-16. https://doi.org/10.1128/mBio.01101-16.
64. Brotherton CA, Medema MH, Greenberg EP. 2018. luxR homolog-linked biosynthetic gene clusters in Proteobacteria. mSystems 3:e00208-17. https://doi.org/10.1128/mSystems.00208-17.
65. Case RJ, Labbate M, Kjelleberg S. 2008. AHL-driven quorum-sensing circuits: their frequency and function among the Proteobacteria. ISME J 2:345. https://doi.org/10.1038/ismej.2008.13.
66. Subramoni S, Venturi V. 2009. PpoR is a conserved unpaired LuxR solo of *Pseudomonas putida* which binds N-acyl homoserine lactones. BMC Microbiol 9:125. https://doi.org/10.1186/1471-2180-9-125.
67. Lee J-H, Lequette Y, Greenberg EP. 2006. Activity of purified QscR, a *Pseudomonas aeruginosa* orphan quorum-sensing transcription factor. Mol Microbiol 59:602–609. https://doi.org/10.1111/j.1365-2958.2005.04960.x.
68. Gehring AM, Bradley KA, Walsh CT. 1997. Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. Biochemistry 36:8495–8503. https://doi.org/10.1021/bi970453p.
69. Vassiliadis G, Destoumieux-Garzón D, Lombard C, Rebuffat S, Peduzzi J. 2010. Isolation and characterization of two members of the siderophore-microcin family, microcins M and H47. Antimicrob Agents Chemother 54:288–297. https://doi.org/10.1128/AAC.00744-09.
70. Challis GL. 2005. A widely distributed bacterial pathway for siderophore biosynthesis independent of nonribosomal peptide synthetases. Chembiochem 6:601–611. https://doi.org/10.1002/cbic.200400283.
71. Tanabe T, Funahashi T, Nakao H, Miyoshi S, Shinoda S, Yamamoto S. 2003. Identification and characterization of genes required for biosynthesis and transport of the siderophore vibrioferrin in *Vibrio parahaemolyticus*. J Bacteriol 185:6938–6949. https://doi.org/10.1128/JB.185.23.6938-6949.2003.
72. Hacquard S, Garrido-Oter R, González A, Spaepen S, Ackermann G, Lebeis S, McHardy AC, Dangl JL, Knight R, Ley R, Schulze-Lefert P. 2015. Microbiota and host nutrition across plant and animal kingdoms. Cell Host Microbe 17:603–616. https://doi.org/10.1016/j.chom.2015.04.009.
73. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, Gill SR, Nelson KE, Relman DA. 2005. Diversity of the human intestinal microbial flora. Science 308:1635–1638. https://doi.org/10.1126/science.1110591.
74. Niu B, Paulson JN, Zheng X, Kolter R. 2017. Simplified and representative bacterial community of maize roots. Proc Natl Acad Sci U S A 114:e2450–e2459. https://doi.org/10.1073/pnas.1616148114.
75. Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S, Tremblay J, Engelbrektson A, Kunin V, Del Rio TG, Edgar RC, Eickhorst T, Ley RE, Hugenholtz P, Tringe SG, Dangl JL. 2012. Defining the core *Arabidopsis thaliana* root microbiome. Nature 488:86–90. https://doi.org/10.1038/nature11237.
76. Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ. 2018. Assembly and ecological function of the root microbiome across angiosperm plant species. Proc Natl Acad Sci U S A 115:E1157–E1165. https://doi.org/10.1073/pnas.1717617115.
77. Ofek-Lalzar M, Sela N, Goldman-Voronov M, Green SJ, Hadar Y, Minz D. 2014. Niche and host-associated functional signatures of the root surface microbiome. Nat Commun 5:4950. https://doi.org/10.1038/ncomms5950.
78. Finkel OM, Castrillo G, Herrera Paredes S, Salas González I, Dangl JL. 2017. Understanding and exploiting plant beneficial microbes. Curr Opin Plant Biol 38:155–163. https://doi.org/10.1016/j.pbi.2017.04.018.
79. Buyer JS, Roberts DP, Russek-Cohen E. 2002. Soil and plant effects on microbial community structure. Can J Microbiol 48:955–964. https://doi.org/10.1139/w02-095.
80. Rutledge PJ, Challis GL. 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol 13:509–523. https://doi.org/10.1038/nrmicro3496.
81. Cregger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, Schadt CW. 2018. The *Populus* holobiont: dissecting the effects of plant niches and genotype on the microbiome. Microbiome 6:31. https://doi.org/10.1186/s40168-018-0413-8.
82. Brown SD, Utturkar SM, Klingeman DM, Johnson CM, Martin SL, Land ML, Lu TY, Schadt CW, Doktycz MJ, Pelletier DA. 2012. Twenty-one genome sequences from *Pseudomonas* species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of *Populus deltoides*. J Bacteriol 194:5991–5993. https://doi.org/10.1128/JB.01243-12.
83. Levy A, Salas Gonzalez I, Mittelviefhaus M, Clingenpeel S, Herrera Paredes S, Miao J, Wang K, Devescovi G, Stillman K, Monteiro F, Rangel Alvarez B, Lundberg DS, Lu TY, Lebeis S, Jin Z, McDonald M, Klein AP, Feltcher ME, Rio TG, Grant SR, Doty SL, Ley RE, Zhao B, Venturi V, Pelletier DA, Vorholt JA, Tringe SG, Woyke T, Dangl JL. 2018. Genomic features of bacterial adaptation to plants. Nat Genet 50:138–150. https://doi.org/10.1038/s41588-017-0012-9.
84. Brown SD, Klingeman DM, Lu T-YS, Johnson CM, Utturkar SM, Land ML, Schadt CW, Doktycz MJ, Pelletier DA. 2012. Draft genome sequence of *Rhizobium* sp. strain PDO1-076, a bacterium isolated from *Populus deltoides*. J Bacteriol 194:2383–2384. https://doi.org/10.1128/JB.00198-12.
85. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen IM, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). Stand Genomic Sci 10:86. https://doi.org/10.1186/s40793-015-0077-y.
86. Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai HC, Zakai UI, Mitchell DA. 2017. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nat Chem Biol 13:470–478. https://doi.org/10.1038/nchembio.2319.

87. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
88. Letunic I, Bork P. 2011. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39:W475–W478. https://doi.org/10.1093/nar/gkr201.
89. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
90. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. 2015. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. Biochim Biophys Acta 1854:1019–1037. https://doi.org/10.1016/j.bbapap.2015.04.015.
91. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303.
92. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188–1190. https://doi.org/10.1101/gr.849004.