



# Elucidating the Diversity and Potential Function of Nonribosomal Peptide and Polyketide Biosynthetic Gene Clusters in the Root Microbiome

Barak Dror,<sup>a,b</sup> Zongqiang Wang,<sup>c</sup> Sean F. Brady,<sup>c</sup> Edouard Jurkevitch,<sup>b</sup>  Eddie Cytryn<sup>a</sup>

<sup>a</sup>Institute of Soil, Water and Environmental Sciences, Agricultural Research Organization, Rishon-LeZion, Israel

<sup>b</sup>Department of Plant Pathology and Microbiology, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel

<sup>c</sup>Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, New York, New York, USA

**ABSTRACT** Polyketides (PKs) and nonribosomal peptides (NRPs) are two microbial secondary metabolite (SM) families known for their variety of functions, including antimicrobials, siderophores, and others. Despite their involvement in bacterium-bacterium and bacterium-plant interactions, root-associated SMs are largely unexplored due to the limited cultivability of bacteria. Here, we analyzed the diversity and expression of SM-encoding biosynthetic gene clusters (BGCs) in root microbiomes by culture-independent amplicon sequencing, shotgun metagenomics, and metatranscriptomics. Roots (tomato and lettuce) harbored distinct compositions of nonribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs) relative to the adjacent bulk soil, and specific BGC markers were both enriched and highly expressed in the root microbiomes. While several of the highly abundant and expressed sequences were remotely associated with known BGCs, the low similarity to characterized genes suggests their potential novelty. Low-similarity genes were screened against a large set of soil-derived cosmid libraries, from which five whole BGCs of unknown function were retrieved. Three clusters were taxonomically affiliated with *Actinobacteria*, while the remaining were not associated with known bacteria. One *Streptomyces*-derived BGC was predicted to encode a polyene with potential antifungal activity, while the others were too novel to predict chemical structure. Screening against a suite of metagenomic data sets revealed higher abundances of retrieved clusters in roots and soil samples. In contrast, they were almost completely absent in aquatic and gut environments, supporting the notion that they might play an important role in root ecosystems. Overall, our results indicate that root microbiomes harbor a specific assemblage of undiscovered SMs.

**IMPORTANCE** We identified distinct secondary-metabolite-encoding genes that are enriched (relative to adjacent bulk soil) and expressed in root ecosystems yet almost completely absent in human gut and aquatic environments. Several of the genes were distantly related to genes encoding antimicrobials and siderophores, and their high sequence variability relative to known sequences suggests that they may encode novel metabolites and may have unique ecological functions. This study demonstrates that plant roots harbor a diverse array of unique secondary-metabolite-encoding genes that are highly enriched and expressed in the root ecosystem. The secondary metabolites encoded by these genes might assist the bacteria that produce them in colonization and persistence in the root environment. To explore this hypothesis, future investigations should assess their potential role in interbacterial and bacterium-plant interactions.

**KEYWORDS** soil microbiome, root microbiome, polyketides, nonribosomal peptides, secondary metabolites, plant-microbe interactions, secondary metabolism

**Citation** Dror B, Wang Z, Brady SF, Jurkevitch E, Cytryn E. 2020. Elucidating the diversity and potential function of nonribosomal peptide and polyketide biosynthetic gene clusters in the root microbiome. *mSystems* 5:e00866-20. <https://doi.org/10.1128/mSystems.00866-20>.

**Editor** Marnix Medema, Wageningen University

**Copyright** © 2020 Dror et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Eddie Cytryn, [eddie@volcani.agri.gov.il](mailto:eddie@volcani.agri.gov.il).

**Received** 9 September 2020

**Accepted** 2 December 2020

**Published** 22 December 2020

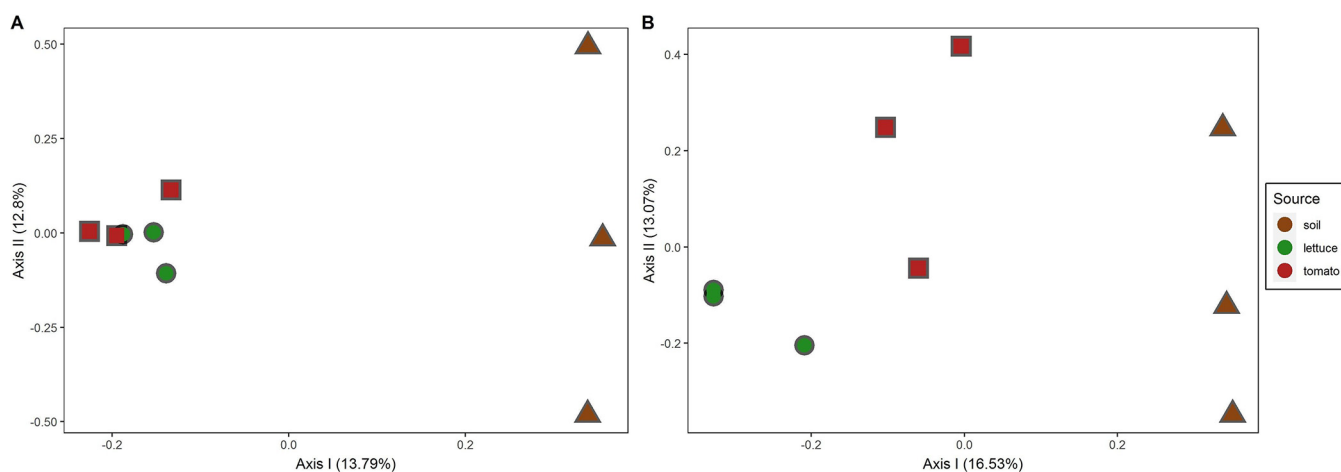
Soil is an extremely diverse ecosystem that contains a myriad of micro- and macroorganisms, including nematodes, arthropods, fungi, and bacteria. The rhizosphere is a narrow region of soil directly influenced by root exudates and mucilage (1, 2). This “hot spot” of organic matter and nutrients “enriches” a specific fraction of the soil microbial community known as the root microbiome, which is significantly different than the surrounding soil microbiome (3). Over the past 2 decades, several studies have linked specific constituents of the root microbiome to enhanced plant growth and development and inhibition of soilborne plant pathogens (4) by direct antagonism and/or induced systemic resistance (5). These functions are often facilitated by the vast array of secondary metabolites (SMs) produced by root-associated bacteria, which play a key role in inter- and intraspecies interactions (6, 7).

Many important soil and root-associated bacterial SMs are nonribosomal peptides (NRPs) or polyketides (PKs), produced by nonribosomal peptide synthetases (NRPSs) or polyketide synthases (PKSs), respectively. These are encoded on large biosynthetic gene clusters (BGCs) that often exceed 50,000 bp (8). Enzymatic complexes in these families follow a similar biosynthetic logic wherein molecules are assembled in an iterative building process using conserved domains that are organized in modules (9, 10). NRPSs and PKSs are responsible for the synthesis of a wide array of siderophores, toxins, pigments, and antimicrobial compounds (11) that are believed to play pivotal roles in bacterial adaptation to soil and rhizosphere ecosystems and in plant health and development (12). Despite their ecological (rhizosphere competence) and translational (biocontrol agents and novel antimicrobial compounds for plant protection) importance, little is known about the occurrence, diversity, and dynamics of NRPSs and PKSs in root ecosystems or their role in intra- and intermicrobial and plant-bacterium interactions.

A major challenge in exploring the role and function of SMs in soil stems from the fact that the majority of bacteria cannot be cultivated using conventional methods, making it difficult to study these bacteria and the diversity, expression, and function of the metabolites that they produce (13). Despite the progress made in culturing techniques, our capacity to isolate soil and root-associated bacteria is highly constrained, primarily because it is challenging to mimic the natural conditions required for growing these bacteria (14). Furthermore, many bacterial BGCs are silent under laboratory conditions, and therefore, the metabolites that they encode are extremely challenging to isolate (15).

To circumvent the above-mentioned barriers, a myriad of culture-independent sequencing-based and omics tools have been developed to reveal the scope and composition of soil-derived BGCs encoding NRPSs and PKSs (16, 17) and to infer the chemical composition and structure of the metabolites produced by these synthetases (18, 19). For instance, amplicon sequencing-based approaches have been developed to target short fragments within adenylation (AD) (in NRPS) and ketosynthase (KS) (in PKS) domains. These amplicons can be used to ascertain the diversity and abundance of bacterial BGCs in complex environments as both AD and KS domains are important (in concert with other components) for the assembly, and thus the identity and activity, of the synthesized metabolites (20, 21). To date, a few studies have explored the diversity and composition of bacterial SM-encoding BGCs in soil, demonstrating the vast genetic diversity and novelty of NRPS and PKS genes (22, 23). However, little is known regarding the distribution of these gene families in the root microbiome, and their functional role in this complex community remains an enigma (24).

This study proposes a unique approach to analyze the diversity and potential functions of NRPSs and PKSs in the root, specifically focused on elucidating (i) the composition and diversity of NRPS- and PKS-encoding genes in the root environment relative to adjacent bulk soil, (ii) NRPS and PKS composition and expression in the root as a function of plant type, (iii) the sequence and inferred SM structures of whole bacterial BGCs that are highly abundant or expressed in root environments, and (iv) the occurrence of root-enriched bacterial BGCs in other ecosystems.



**FIG 1** Principal-coordinate analysis (PCoA) of lettuce root, tomato root, and soil (from tomato lysimeters) samples using the Bray-Curtis similarity index. (A) Analysis of KS (from PKS) domain amplicons (roots versus soil ANOSIM statistic  $R=0.332$  [ $P < 0.05$ ]; tomato versus lettuce versus soil ANOSIM  $R=0.432$  [not significant (NS)]). (B) Analysis of AD (NRPS) domain amplicons (roots versus soil ANOSIM statistic  $R=0.308$  [ $P < 0.01$ ]; tomato versus lettuce versus soil ANOSIM  $R=0.6872$  [ $P < 0.01$ ]).

## RESULTS

**Composition and diversity of NRPS and PKS genes in roots versus bulk soil.** To determine the composition and diversity of NRPSs and PKSs in tomato and lettuce root samples relative to bulk soil (previous studies targeting this controlled lysimeter system showed that bulk soils from tomato and lettuce microbiomes were almost identical, and therefore, only tomato soil was analyzed here), we applied a previously described amplicon sequencing approach to amplify the conserved adenylation (AD) and ketosynthase (KS) domains of NRPSs and PKSs, respectively (23). Overall, sequencing yielded totals of 1,850,442 and 2,174,020 raw KS and AD reads with average read lengths of 280 bp and 235 bp, respectively (see Table S1 in the supplemental material). Further filtering steps using QIIME2 and DADA2 denoising methods resulted in 2,980 and 3,269 nonredundant KS and AD domain sequences, respectively.

We observed significantly greater diversity of both AD and KS domains in the bulk tomato soil than in the adjacent roots (Fig. S1A and C). In contrast, no difference in diversity was observed between tomato and lettuce roots for either of the SM-encoding domains (Fig. S1B and D). To assess differences in AD and KS domain diversity between samples, a principal-coordinate analysis (PCoA) and analysis of similarity (ANOSIM) using the Bray-Curtis similarity index were performed (Fig. 1A and B). The KS and AD domain profiles of the roots (from both tomato and lettuce) formed distinct clusters, which were significantly different from those of the adjacent bulk soil ( $R=0.332$  [ $P < 0.05$ ] for PKS;  $R=0.308$  [ $P < 0.01$ ] for NRPS).

To explore the potential novelty of root-associated SM-encoding genes, the amplified AD and KS domain sequences from the root and soil samples were first aligned against the MIBiG (Minimum Information about a Biosynthetic Gene Cluster) repository (25) using blastp, with a  $>50\%$  amino acid sequence identity cutoff, and then grouped according to their identity to the MIBiG reference genes (Table 1). On average, more than 25% of the AD and almost 13% of the KS domain sequences in the root environment showed less than 50% amino acid identity with genes found in the MIBiG database (characterized as “unassigned”), whereas fewer than 1% and 6% of the AD and KS sequences, respectively, shared over 85% similarity to the reference MIBiG genes. These results demonstrate the profusion of potentially novel SM-encoding genes in both root and soil environments.

**Pinpointing predicted root-enriched NRPs and PKSs.** As SMs are known to play critical roles in bacterium-bacterium and bacterium-plant interactions, we were interested in the associated metabolites synthesized by BGCs whose AD or KS domains were highly abundant and enriched in tomato or lettuce roots relative to bulk soil. To

**TABLE 1** Abundance of NRPS (AD) and PKS (KS) amplicons based on levels of similarity to MIBiG NRPS and PKS reference genes in the tomato and cucumber root and tomato soil samples<sup>a</sup>

Gene	Identity (%)	% of sequences (total no. of hits)			
		Tomato	Lettuce	Avg (roots)	Soil
AD	Unassigned	24.1 (225)	28.6 (202)	26.10 (427)	14.0 (234)
	50–70	56.2 (523)	53.9 (381)	55.2 (904)	66.8 (1,114)
	70–85	18.7 (174)	16.2 (115)	17.6 (289)	18.3 (305)
	85–95	0.8 (8)	1.1 (8)	0.97 (16)	0.72 (12)
	95–100	0.0	0.0	0.0	0.06 (1)
KS	Unassigned	10.2 (108)	14.4 (202)	12.6 (310)	5.0 (17)
	50–70	46.2 (486)	46.5 (652)	46.4 (1,138)	56.2 (190)
	70–85	37.5 (394)	32.1 (451)	34.4 (845)	33.7 (114)
	85–95	5.1 (54)	6.0 (85)	5.6 (139)	4.7 (16)
	95–100	0.7 (8)	0.7 (11)	0.7 (19)	0.2 (1)

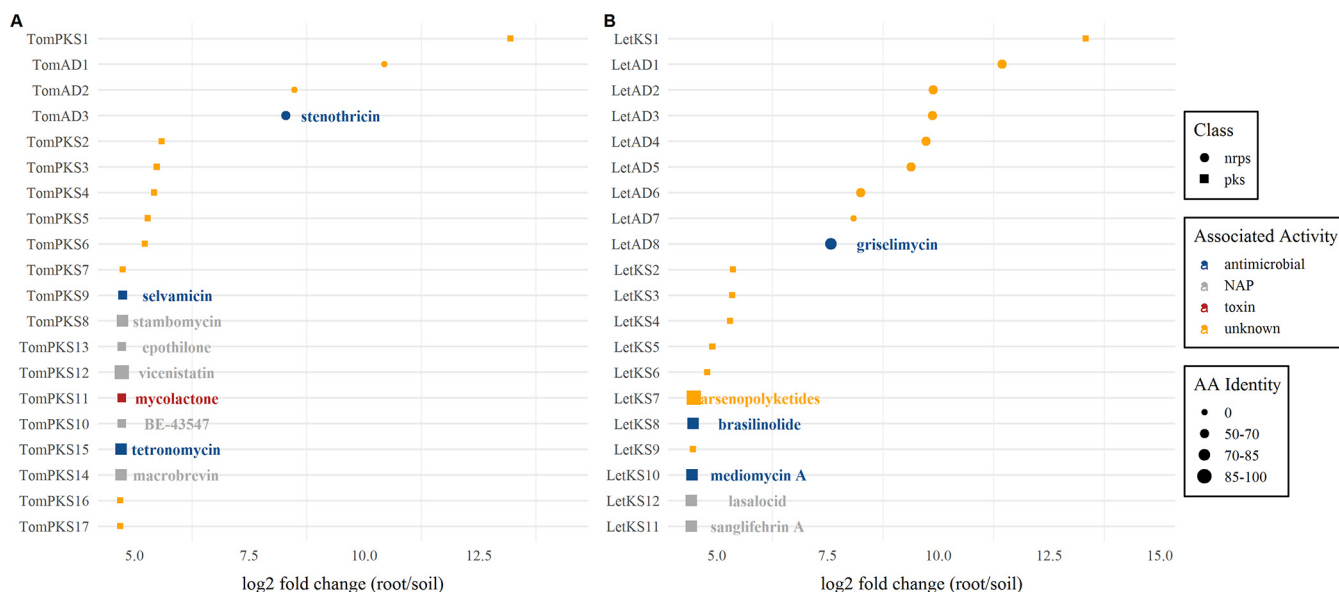
<sup>a</sup>Identity groups were determined as unassigned (<50% amino acid sequence identity) and 50 to 70%, 70 to 85%, 85 to 95%, and 95 to 100% sequence identity. Numbers outside parentheses indicate the percentages of sequences associated with each group, and numbers inside parentheses indicate the total number of hits for each group.

do so, the MIBiG-aligned amplicons were annotated to the corresponding BGC-associated metabolites, using a cutoff E value of  $<10^{-40}$ . Sequences that did not meet these criteria were defined as “unknown.” Previous analyses have shown that compared to reference KS or AD domain sequences, amplicons with E values of  $<10^{-40}$  are likely derived from the same BGC family as the reference sequence and thus may be inferred to encode a similar function (26–29).

A differential abundance analysis using DESeq2 of the top 20 highly abundant AD and KS amplicons revealed that 55% (11/20) and 70% (14/20) of the amplicons in both of the plant root samples (tomato and lettuce, respectively) (adjusted *P* value of  $<0.1$ ;  $\log_2$  fold change of  $>5$ ) were not associated with known BGCs; thus, their associated metabolites cannot be inferred (Fig. 2). Nonetheless, several root-enriched AD and KS domain sequences (9 in tomato and 6 in lettuce) were above these threshold values and thus can be considered congeners to known metabolites. These included the non-ribosomal peptides stenothricin (30) and griselimycin (31), whose BGC NRPS analogues were highly abundant in the tomato and lettuce roots, respectively, and were less profuse in the bulk soil. While we cannot determine the actual role of the metabolites potentially encoded by these enriched BGCs, both stenothricin and griselimycin are known for their antimicrobial activity.

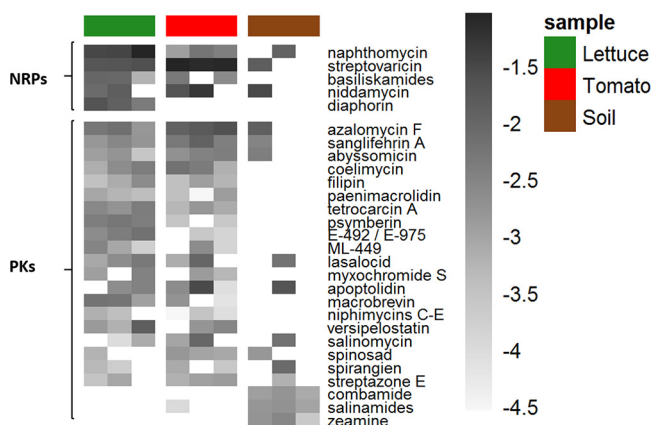
Next, we calculated the relative abundances of amplicons that were associated with known metabolites (based on the criteria described above) in the different root and soil samples (Fig. S2). To pinpoint associated BGCs that may play a role in adaptation to the root environment, we focused our analysis on inferred metabolites that were present in at least four of the root samples (tomato and lettuce) and in no more than one soil sample (Fig. 3). In addition, to identify BGCs specifically relevant to soil, we also selected inferred metabolites that were present in all three soil samples and in no more than one root-associated sample. For NRPs, we found BGCs associated with four metabolites that were highly abundant in both of the root samples (e.g., the *Streptomyces*-derived antibiotic macrolide family streptovaricin). For PKs, we again found several highly abundant *Streptomyces*-derived inferred metabolites, among others. These included lasalocid, sanglifehrin A, and azalomycin A. Interestingly, amplicons associated with the two former metabolites were also found to be highly enriched in roots relative to soil for lettuce (Fig. 2B). Overall, we found that 26 associated metabolites were present in at least one of the root-associated samples and completely absent in the soil samples, e.g., diaphorin (in lettuce) and basiliskamides (present in both tomato and lettuce) (Fig. S2).

Due to the potential biases associated with the above-described PCR-based

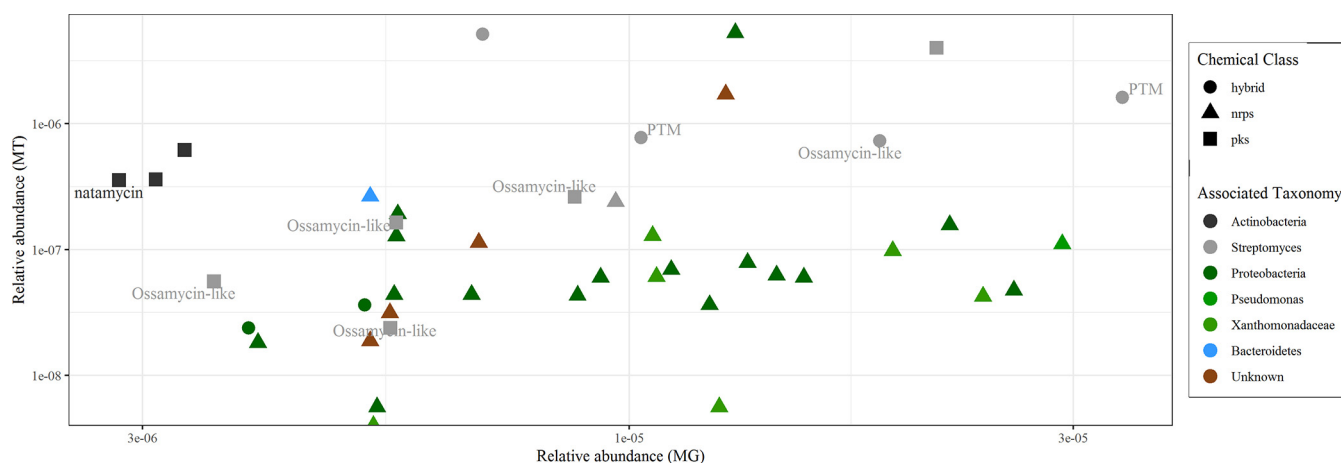


**FIG 2** Differential abundances of the top 20 highly abundant NRPS and PKS amplicons in tomato (A) and lettuce (B) root microbial communities relative to bulk tomato soil. DESeq2 was used to calculate  $\log_2$  fold changes and test for significance (adjusted  $P$  value [false discovery rate {FDR}] of  $<0.1$ ). Associated metabolites (based on the MIBiG repository) are presented only for sequences with  $E$  values of  $<10^{-40}$  and  $>50\%$  amino acid (AA) identity. Identity bars represent amino acid identity to the MIBiG reference sequence (0, not identified). NAP, non-antimicrobial pharmaceutical (e.g. anti-tumor, anti-inflammatory, etc.).

approach, we analyzed previously reported (32) shotgun metagenomes of the same tomato and lettuce root samples ( $n=3$  each). Assembled open reading frame (ORF) sequences from the metagenomes identified using Prodigal were aligned against the MIBiG database, generating a list of the 50 most abundant NRPSs and PKSs in each of the root data sets (representing the normalized abundance within samples by plotting the coefficient of variance [CV] for each gene) (Fig. S3A and B). In addition, as gene clusters are often silent or expressed under very specific conditions, we evaluated gene expression in parallel to gene occurrence to uncover active BGCs with ecological importance in the highly dynamic root ecosystem. Thus, in parallel to the shotgun metagenome analysis described above, we applied a similar analytical approach using the previously collected shotgun metatranscriptomes (32) to identify NRPSs and PKSs with enhanced expression in lettuce and/or tomato root microbiomes. Interestingly, 60% (30/50) and 46% (23/50) of the AD and KS domains that were highly abundant in the tomato and lettuce root samples, respectively, were highly expressed as well (Fig. S3).



**FIG 3** Relative abundances of NRPS (top)- and PKS (bottom)-associated metabolites based on MIBiG annotations. Only AD and KS amplicon hits to reference MIBiG sequences with over  $>50\%$  amino acid identity and an  $E$  value of  $<10^{-40}$  were included. Counts were normalized and  $\log_{10}$  transformed. Only associated metabolites that were present in  $>2$  different samples are shown.



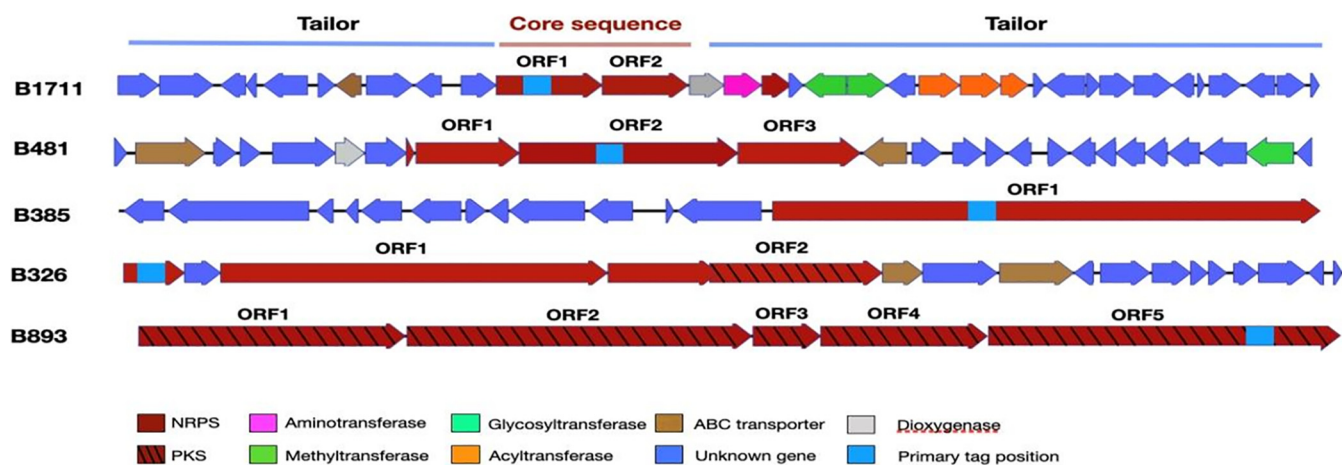
**FIG 4** Relative abundances and taxonomy of highly abundant NRPS/PKSs in tomato and lettuce root metagenomes (MG) and metatranscriptomes (MT). Shapes represent BGC-associated chemical classes. Taxonomy was assigned using MEGAN 5.10 and the lowest-common-ancestor (LCA) classification algorithm. Only low-CV sequences were chosen.

Next, we filtered the highly abundant hits based on their CV values ( $<50$ ) in order to analyze sequences with lower dispersion levels within tomato and lettuce root-associated samples, followed by taxonomic annotation using MEGAN. The resulting 42 sequences were clustered into two main phyla: *Actinobacteria* (13/42) and *Proteobacteria* (25/42). Several sequences could not be assigned a taxonomic affiliation, and one was assigned to the *Bacteroidetes* phylum (Fig. 4). While we could not infer the associated metabolites synthesized by most of these highly abundant sequences (including all of those assigned to the phylum *Proteobacteria*), suggesting their potential novelty, we managed to annotate several of the *Actinobacteria*-associated BGCs. These were distantly associated with ossamycin (5 hits) and polycyclic tetramate macrolactams (PTMs) (2 hits, including the most highly abundant NRPS/PKS-related sequence and the 5th most expressed). Ossamycin is a known antifungal and cytotoxic macrocyclic polyketide originally isolated from soil-associated *Streptomyces hygrosopicus* subsp. *ossamyceticus* (33, 34), while PTMs are a family of biologically important metabolites, including HSAF (heat-stable antifungal factor), ikarugamycin, and clifednamides, generally associated with different isolates of *Actinobacteria* and *Gammaproteobacteria* (35).

Finally, to evaluate the extent to which the amplicon sequencing method was able to detect NRPSs and PKSs in the targeted samples relative to the PCR-independent shotgun metagenomic analyses, we analyzed the distribution of total MIBiG-associated genes in all four data sets (lettuce and tomato NRPS and PKS amplicons and tomato and lettuce metagenomes) (Fig. S4). In general, approximately 34% and 25% of the MIBiG-characterized genes were found in both amplicon sequences and metagenomes of the lettuce and tomato roots, respectively. In contrast, approximately 33% and 20% of the tomato and lettuce genes, respectively, were detected only in the shotgun metagenomic data sets. Fewer than 4% and 5% of the NRPSs and PKSs were found only within the tomato and lettuce amplicon sequencing data sets (28 and 35 genes in tomato and lettuce, respectively), and fewer than one-fifth (139 genes or 17.2%) were common to all four culture-independent data sets.

#### Extraction and environmental distribution of whole SM-encoding gene clusters.

The identification of NRPSs and PKSs that were either enriched in roots relative to adjacent bulk soil or abundant and/or highly expressed in lettuce and/or tomato roots encouraged us to capture whole BGCs associated with these sequences in order to shed light on their phylogenetic affiliation and potentially infer their function and chemical structure. This was achieved by screening the root-associated NRPS and PKS candidate sequences identified in this study against a large set of previously collated soil and rhizosphere cosmid libraries using the bioinformatic platform eSNaPD (29)



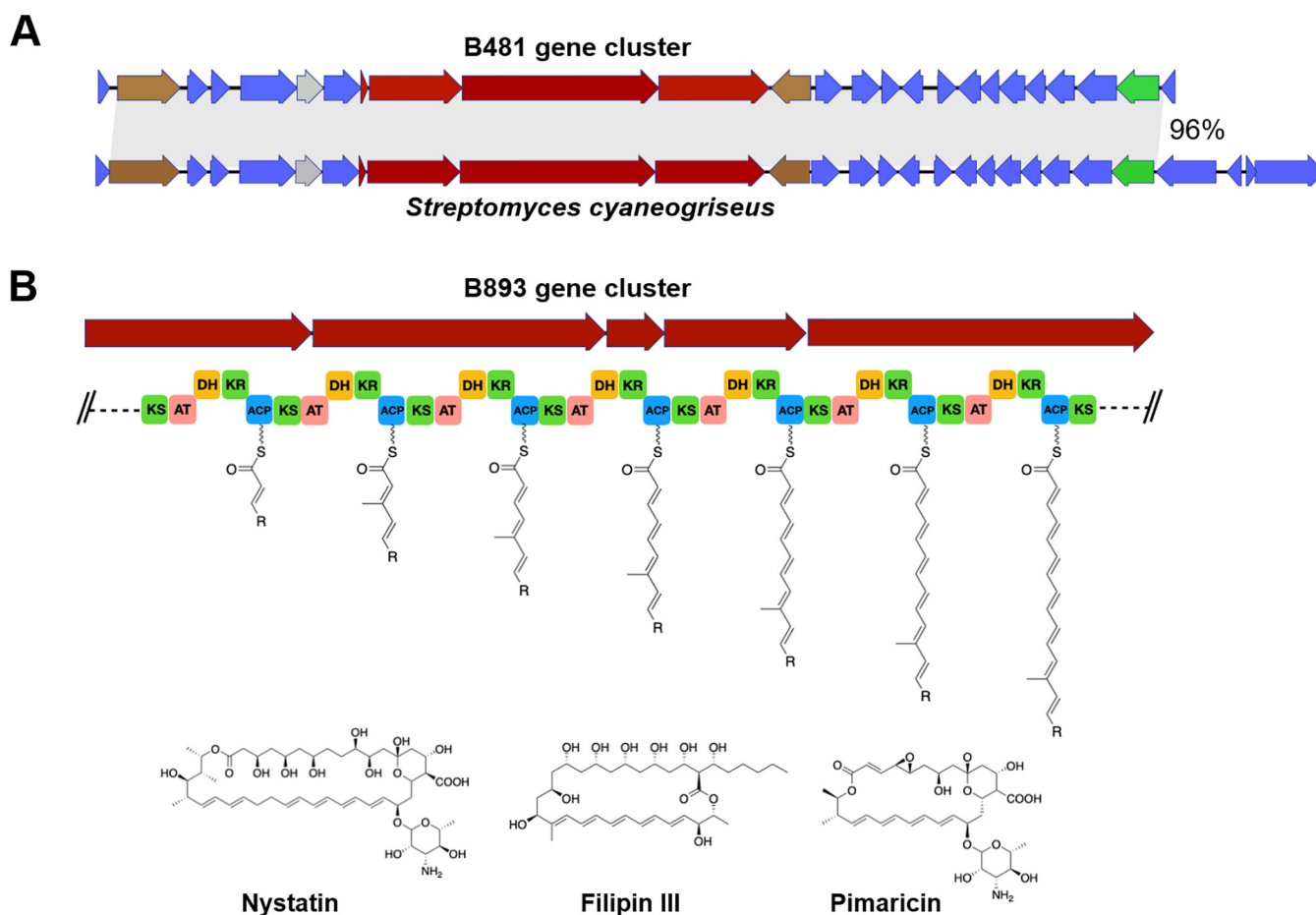
**FIG 5** Overview of retrieved (from archived soil cosmid libraries) biosynthetic gene clusters containing AD and KS domain contigs that were abundant and/or highly expressed in tomato and cucumber root microbiomes.

(see Materials and Methods for the full pipeline). Five cosmid library targets showed low E values and high nucleotide sequence identities (>75%) to candidate NRPSs or PKSs. Sequencing and annotation of the metagenomic insert captured in each cosmid revealed three NRPS and two hybrid NRPS/PKS gene clusters (Fig. 5). Based on gene content and sequence identity, the identified gene clusters were not identical to any BGCs associated with known metabolites. The NRPS and PKS ORFs of two recovered clones (B326 and B385) were not affiliated with any known bacterial taxa (<50% nucleic acid identity to the NCBI database), while the other three clones were related to genes from *Actinobacteria* (Table 2). Of the cosmids recovered from the metagenomic libraries, clone B481 was nearly identical to an uncharacterized NRPS BGC found in the genome of *Streptomyces cyaneogriseus* (Fig. 6A). The only predicted chemical structure that we could infer from the recovered BGCs was for clone B893, which was related to an uncharacterized PKS gene cluster found in the genome of *Saccharothrix saharensis*, a filamentous actinobacterium isolated from desert soil. A detailed bioinformatic analysis of its PKS domains revealed that the gene cluster likely encodes an extended polyene substructure (Fig. 6B). The seven PKS modules captured on the clone all contain dehydratase (DH) and ketoreduction (KR) domains, indicating that each module introduces a double bond into the polyketide backbone (Fig. 6B). While

**TABLE 2** Predicted taxonomy and source data set of recovered cosmids<sup>a</sup>

Gene cluster	Domain (metabolite)	Coverage (%)	Identity (%)	GenBank accession no.	Source data set	Affiliated organism	Taxon
1711	B1711-1 (NRPS)	84	58.92	<a href="#">WP_136726459.1</a>	Lettuce	<i>Streptomyces</i> sp. NEAU-C40	<i>Actinobacteria</i>
	B1711-2 (NRPS)	97	69.36	<a href="#">WP_136726460.1</a>	Lettuce	<i>Streptomyces</i> sp. NEAU-C40	<i>Actinobacteria</i>
481	B481-1 (NRPS)	99	92.97	<a href="#">WP_052808563.1</a>	Tomato	<i>Streptomyces cyaneogriseus</i>	<i>Actinobacteria</i>
	B481-2 (NRPS)	99	96.60	<a href="#">WP_052808562.1</a>	Tomato	<i>Streptomyces cyaneogriseus</i>	<i>Actinobacteria</i>
	B481-3 (NRPS)	99	93.30	<a href="#">WP_044379030.1</a>	Tomato	<i>Streptomyces cyaneogriseus</i>	<i>Actinobacteria</i>
385	B385 (NRPS)	UA	UA		Lettuce	UA	
326	B326-1 (NRPS)	UA	UA		Lettuce	UA	
	B326-2 (PKS)	UA	UA		Lettuce	UA	
893	B893-1 (PKS)	98	90.90	<a href="#">WP_141974662.1</a>	Tomato	<i>Saccharothrix saharensis</i>	<i>Actinobacteria</i>
	B893-2 (PKS)	99	89.02	<a href="#">WP_141974663.1</a>	Tomato	<i>Saccharothrix saharensis</i>	<i>Actinobacteria</i>
	B893-3 (PKS)	98	92.31	<a href="#">TQM77832.1</a>	Tomato	<i>Saccharothrix saharensis</i>	<i>Actinobacteria</i>
	B893-4 (PKS)	100	87.62	<a href="#">WP_141974664.1</a>	Tomato	<i>Saccharothrix saharensis</i>	<i>Actinobacteria</i>
	B893-5 (PKS)	97	89.78	<a href="#">WP_141974665.1</a>	Tomato	<i>Saccharothrix saharensis</i>	<i>Actinobacteria</i>

<sup>a</sup>UA, unassigned. Only affiliated organisms with >50% identity are shown.



**FIG 6** (A) Comparison of the B481 cluster and the related BGC in *S. cyaneogriseus*. The gray area indicates the aligned region. (B) Structure prediction for PKS modules present on clone B893. Examples of known polyene antifungal agents are shown at the bottom. KS, ketosynthase; AT, acyltransferase; DH, dehydratase; KR, ketoreductase; ACP, acyl carrier protein.

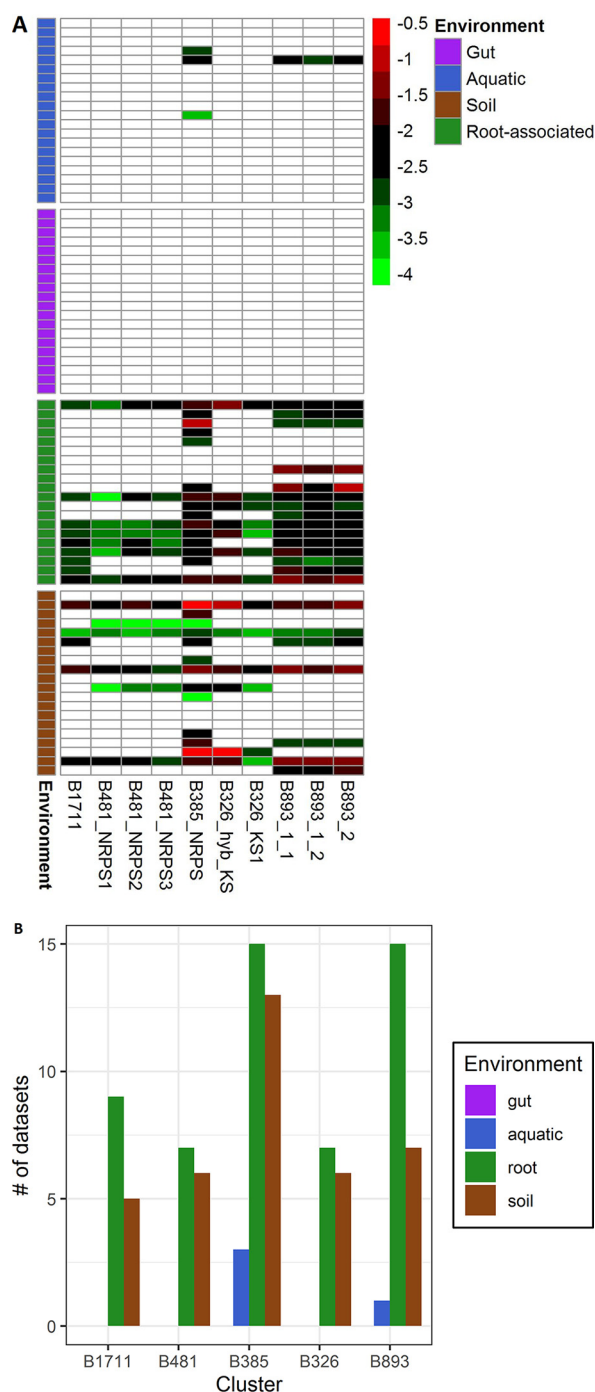
polyene substructures like this are seen in a number of natural products (36, 37), they are most commonly seen in polyene antifungal agents, including many that are derived from *Streptomyces* species (e.g., cyphomycin, nystatin, filipin, and pimaricin). This may suggest that the BGC encodes an antifungal compound.

The five recovered BGCs were initially targeted due to their abundance in tomato and lettuce root data sets, suggesting a link to root ecosystems. To test this hypothesis, we assessed the abundances of the five BGCs in a large collection of publicly available shotgun metagenomes (20 metagenomes from each environment) from four distinct environments, targeting gut (animal and human), aquatic (freshwater and marine), soil (different soil types), and root-associated (various plant species) data sets. Our analysis demonstrated that the recovered BGCs are ubiquitous in most of the queried root samples and in some of the soil samples (Fig. 7A). Only clone B893 showed a significantly higher abundance in the root samples than in the soil samples ( $P < 0.05$  by a Wilcoxon test) (Fig. S5). B893 was found in 16/20 of the root-associated data sets that we examined (compared with 8/20 soil data sets) (Fig. 7B). In contrast, none of the five BGCs were found in any of the gut microbiome communities analyzed, and very few were detected in the aquatic environments (Fig. 7B).

## DISCUSSION

NRPs and PKs produced by root-associated microbial communities play an important role in plant root ecosystems (38, 39). Several studies have identified and characterized BGCs and/or associated metabolites in prominent plant-growth-promoting and





**FIG 7** (A) Relative abundances of clone-recovered AD and KS domains in shotgun metagenomes ( $n=20$ ) from 4 environmental ecosystems: gut (animal and human), aquatic (freshwater and marine), soil, and root associated. Analysis was conducted using the IMG Blastn feature with an E value of  $<10^{-5}$  and an identity of  $>85\%$ . Counts were normalized with *rpoB* and *gyrB* for each data set. (B) Presence/absence of each recovered gene cluster in the different environments. Only gene clusters where all AD/KS domains were found are considered present.

biocontrol agents originally isolated from plant roots (40, 41). However, the large fraction of uncultivated bacteria in root ecosystems and the limitations associated with culturing bacteria encouraged us to examine the composition of genes encoding NRPS and PKS in root environments using culture-independent approaches. These methods have been applied previously to understand SM-encoding gene diversity and

distribution in bulk soil (17, 26, 42, 43), but NRPSs and PKSs have not been comprehensively explored in root ecosystems. Furthermore, to the best of our knowledge, this is the first study to explore the expression of secondary-metabolite-associated genes in root ecosystems.

Our results demonstrate distinct compositions of NRPSs and PKSs in plant (tomato and lettuce) roots relative to adjacent bulk soil, that these genes in root microbial communities are less diverse than those found in soil microbiomes, and that a fraction of these NRPSs and PKSs are highly expressed in root ecosystems. This is consistent with several previous studies showing that the phylogeny and functionality of root-associated microbial communities are significantly different from those in adjacent soil (44–46). It is well established that plants interact with soil bacterial communities through the secretion of root exudates (47, 48), resulting in the selection of specific microbial populations from the soil microbiome. This appears to be the case for the recruitment of plant-growth-promoting rhizobacteria (PGPR), which are known to harbor specific SM-encoding genes (4, 49, 50). Thus, while at this point, we cannot infer the actual function of the highly abundant and expressed NRPSs and PKSs in the root environment, we can infer that they likely play a role in various processes, e.g., competition and root colonization (51, 52).

While most of the detected SM-associated genes were too novel to link to any known metabolites, a few of the highly abundant, expressed, and/or root-enriched NRPSs and PKSs were associated with known metabolites. Azalomycin F, for instance, found to be associated with sequences in both of the plant microbiome samples, is a polyketide with reported antifungal activity against a variety of phytopathogens, which is produced by different soil- and root-associated *Streptomyces* isolates (53, 54). Diaphorin, associated with sequences found only in the tomato root microbiome, is a pederin analogue known to be produced by the psyllid *Diaphorina citri* endosymbiont "*Candidatus Proffotella armatura*" (*Betaproteobacteria*), with potential antifungal and cytotoxic activity (55). In this regard, the described culture-independent approaches are a promising platform for identifying novel BGCs and elucidating their roles in soil ecosystems and within the framework of drug discovery efforts, despite their current limitations (16, 17).

A large fraction of the highly abundant and highly expressed NRPSs and PKSs identified in this study were not assigned, or had low identity (50 to 70%), to previously characterized genes in the MIBiG repository. Recently, a previously unidentified hybrid NRPS/PKS gene cluster was found to be essential for *Rhizoctonia* suppression by an endophyte *Flavobacterium* sp. (56), highlighting the vast amount of root-associated SMs with unidentified functional roles, which undoubtedly play a pivotal role in bacterium-plant interactions.

We screened a large set of soil cosmid libraries with candidate sequences from our amplicon sequencing and metagenomic analyses that were enriched (relative to bulk soil) and/or highly abundant and expressed in tomato and lettuce roots, taking advantage of a unique culture-independent platform capable of extracting and analyzing long NRPSs and PKSs (27, 57). Five clones containing uncharacterized gene clusters with no known function, including two that were not associated with any known taxonomic group, were identified. The fact that all of these BGCs were rather common in various root-associated environmental metagenomes but rare or completely absent in other environments suggests their potential importance in these habitats. While we can only speculate as to their synthesized metabolites' actual activity, they were associated with bacterial groups well known for their antimicrobial capacity. *Saccharothrix saharensis* (*Actinobacteria*), for instance, which contains a BGC closely related to clone B893, is a soil-dwelling bacterium known to produce an array of antimicrobials (58). This BGC likely encodes a polyene substructure, often seen in antifungal agents, as it is capable of directly disrupting the fungal membrane (59). Of particular interest is clone B481, associated with SM-associated genes from *Streptomyces cyaneogriseus*, known

for its ability to produce the biopesticide nemadectin (60). We speculate that this BGC may be associated with bacterium-fungus competition in the root ecosystems.

At the broader level, our results emphasize the need to look beyond basic descriptive diversity and composition information regarding the SM capacity of microbial communities. The pipeline adopted in this study, where potentially important NRPSs and PKSs are first identified, followed by the extraction of BGCs from cosmid libraries in order to identify potentially novel BGCs, has the advantage of being resource-efficient while spawning deeper knowledge regarding potentially important gene clusters. Future studies will focus on expressing these cosmid library BGCs in suitable hosts, enabling us to characterize their encoded metabolites and test their *in vitro* and *in planta* activities against various phytopathogens. Our results coincide with studies conducted in other plants, showcasing the as-yet-unexplored diversity of NRPSs and PKSs (43, 61).

Overall, this study indicates that the root microbiome harbors a unique, diverse, and potentially novel array of SM-synthesizing genes, which are significantly different from those in the bulk soil microbiome. To enhance our current understanding, future research should focus on identifying additional factors shaping the occurrence and expression of SMs in the root microbiome (e.g., plant health, the presence of phytopathogens, and plant growth). This will undoubtedly help expose the ecological role of SMs in root ecosystems and provide a platform for drug discovery and novel and environmentally sustainable compounds for plant protection.

## MATERIALS AND METHODS

Figure S6 in the supplemental material presents a conceptual description of the pipeline applied in this study.

**Amplicon sequencing, shotgun metagenomics, and metatranscriptomics analyses.** Tomato soil and root samples and lettuce root samples were collected as previously described (62). Briefly, tomato (*Solanum lycopersicum* Heinz 4107) and lettuce (*Lactuca sativa* [romaine] Assaph) seedlings were planted and grown for 42 days in a random-block lysimeter experiment at the Lachish agricultural research station in Kiryat-Gat, Israel. Each sample type (soil, tomato roots, and lettuce roots) was analyzed using triplicate samples from three different lysimeters (thus,  $n = 3$  for each sample type). Each triplicate consisted of a composite sample collected from 2 to 4 plants or 3 soil samples, taken from the distant edges of the lysimeters and away from plant roots. As the same soil was used throughout the experiment, and given that soil samples were collected at a sufficient distance from growing plants, bulk soil from the tomato lysimeters served as a reference point for both tomato and lettuce soils. A previous study showed that they harbored almost identical microbial communities (32). Soil samples were collected, frozen in liquid nitrogen on-site, and stored at  $-80^{\circ}\text{C}$  until further analysis. Root samples were collected intact, and soil particles were removed by shaking and briefly rinsing. The roots were then lightly dried, immediately frozen in liquid nitrogen on-site, and kept at  $-80^{\circ}\text{C}$  until processed. In this study, extracted DNA was used as the template for NRPS and PKS PCR amplification using degenerate primers (A3F/A7R for AD-NRPS and degKS2F/degKS2R for KS-PKS domains, as previously reported [23]). The resulting barcoded libraries were pooled and sequenced on an Illumina MiSeq instrument, employing V2 chemistry, at the University of Illinois—Chicago Sequencing Core (UICSQC). A total of 18 samples were sequenced, which included sampling location (tomato soil versus tomato and lettuce roots) and SM family (AD and KS domains), with three replicates for each treatment. The resulting sequences were processed and demultiplexed using the QIIME2 pipeline and the integrated DADA2 method (63). Exact sequence variants (ESVs) represented by fewer than 3 sequences were removed from downstream analyses. Raw amplicon sequences are available via the MG-RAST data repository under accession number [mgm4862150.3](https://mgm4862150.3). In addition, shotgun metagenome and metatranscriptome data sets of lettuce and tomato roots ( $n = 3$  for each data set type [hence, 6 for tomato and 6 for lettuce]) previously generated and analyzed from the same samples were also used for NRPS and PKS identification as described below (32). Shotgun sequence data are available via the NCBI Sequence Read Archive (SRA) data repository under project accession number [PRJNA602301](https://PRJNA602301).

**Identification and annotation of NRPSs and PKSs.** For chemical diversity analysis of NRPS and PKS gene clusters, the different data sets (ESVs generated via amplicon sequencing and metagenome/metatranscriptome-assembled genes) were aligned against the Minimum Information about a Biosynthetic Gene Cluster (MIBiG) repository (version 6 August 2018). Only core NRPS and PKS genes (AD and KS domains) were included in the analysis. Alignment was performed using the diamond blastx command line, with  $>50\%$  amino acid sequence identity. To associate each NRPS or PKS hit with its potentially synthesized metabolite, an E value of  $<10^{-40}$  was used as a cutoff. Hits that did not pass this threshold were regarded as “unknown.”

For taxonomic annotation, sequences were aligned against the nonredundant (nr) BLAST NCBI database, followed by lowest-common-ancestor (LCA) classification using the MEGAN 6.15 Ultimate edition by taking the top 10% of hits and filtering for a minimum score of 50 and a maximum E value of 0.01

(64). Conversion of gene identifiers to taxonomic path was done using the mapping files provided by MEGAN as of October 2016.

**Soil library amplicon generation screening.** Metagenomic libraries were constructed as previously reported (65). Briefly, in each library, crude environmental DNA (eDNA) was extracted directly from field-collected soil, gel purified, blunt ended, ligated into cosmid pWEB::TNC (Epicenter), packaged into lambda phage, and transferred into *E. coli* EC100 (Lucigen). Each library was expanded to contain  $20 \times 10^6$  unique cosmid clones with ~30- to 45-kb eDNA inserts and then arrayed into 768 subpools (two 384-well plates) containing ~25,000 unique cosmid clones per well. Each subpool was then stored as a glycerol stock for the clone recovery of interesting hits and as cosmid DNA to facilitate PCR-based screening. To generate an amplicon sequence database of NRPSs and PKSs, the following two sets of degenerate primers (AD and KS) were applied to amplify the conserved regions in the adenylation and ketosynthase domains in the biosynthetic gene cluster: AD forward primer 5'-SATBTAYACSTCVGGHWCASAC-3' and reverse primer 5'-CCANRTCNCBBGTSYKGTACA-3', and KS forward primer 5'-TGYTCSDSSTCGCTSGTSGCS-3' and reverse primer 5'-GTNCCSGTSCCRGTBGCYTC-3'. The 5' ends of the primers were augmented with MiSeq sequencing adapters followed by unique 8-bp barcode sequences identifying the soil metagenome from which they were amplified. Amplicons were pooled as collections of 96 samples and cleaned using magnetic beads. Cleaned, pooled amplicons were used as the template in a second PCR. Prior to sequencing, all PCR amplicons were quantified by gel electrophoresis and mixed in an equal molar ratio. The resulting pool was fluorometrically quantified with HS D1000 ScreenTape and sequenced on an Illumina MiSeq instrument. Reads were debarcoded and trimmed to 240 bp. The reads from each sample were clustered using UCLUST (66) to generate the 95% tags.

**Recovery of BGC clones from metagenomic library pools.** The library well locations for target AD or KS domains were identified using well-specific barcodes incorporated into the degenerate primers (27). Specific primers with melting temperature ( $T_m$ ) values of ~60°C (18 to 20 bp) were designed to amplify each unique conserved sequence of interest. To recover the single cosmid clone from each library subpool, a serial dilution of the whole-cell PCR strategy was used (17). Briefly, library glycerol stocks that contained target hits from eSNaPD analysis were inoculated into 3 ml LB broth (kanamycin and chloramphenicol) and grown overnight at 37°C to confluence. The cells cultured overnight were diluted to 2,000 CFU ml<sup>-1</sup>, calculated by the optical density at 600 nm (OD<sub>600</sub>). The 384-well plates were inoculated with 50 µl of the resulting diluent (600 CFU/well) with an Eppendorf epMotion 5075 liquid handler, grown overnight, and screened using real-time PCR with a touchdown PCR program to identify wells containing target clones. Target-positive wells were diluted to a concentration of ~5 CFU ml<sup>-1</sup>, and the process was repeated to identify new wells containing target clones. Five clone pools were then plated on a solid-agar plate, and target single clones were identified by clone PCR.

**Analysis of recovered gene clusters.** Recovered single-cosmid clones were miniprepmed by using a QIAprep kit and sequenced using MiSeq technology. The M13-40FOR and T7 universal primers were utilized to sequence both ends of the insert sequences. Reads, amplicons, and end sequences were assembled together to generate constant contigs using Newbler 2.6 (67). Fully assembled contigs were then analyzed using an in-house annotation script consisting of open reading frame (ORF) prediction with MetaGeneMark, HMM Scan, and BLAST search. The annotation script was developed using Python and is available at the GitHub open-source repository ([https://github.com/brady-lab-rockefeller/gene\\_annotation](https://github.com/brady-lab-rockefeller/gene_annotation)). Putative functions and source organisms of genes in the BGC were assigned based on the closest characterized gene in the NCBI database. KnownClusterBlast in antiSMASH 5.0 (68) was utilized to analyze the relationship between known characterized gene clusters and recovered BGCs. The structure prediction of the adenylation domains and ketosynthase domains in BGCs was given by the antiSMASH prediction, which employs three prediction algorithms, NPRSPredictor2, Stachelhaus code, and SVM (support-vector machine) prediction. These predicted building blocks were then utilized to predict a final structure combined with known characterized BGCs in cultured bacteria.

**Recovered clone search in environmental shotgun metagenomes.** AD and KS domains from all five recovered gene clusters were searched against shotgun metagenomes from four different environments: animal and human feces (gut), aquatic, soil, and root associated. We selected 20 Illumina-sequenced shotgun metagenomes from each of these ecosystems using the JGI IMG/MER advanced search option, followed by a Blastn search using the IMG website online tool. Additional filtering was performed based on an E value threshold ( $10^{-40}$ ) and identity (>85%). For each hit, counts were normalized using *gyrB* and *rpoB* housekeeping gene counts (obtained via the IMG/MER platform). A relative-abundance heat map was created using the pheatmap R package (69). For gene clusters with more than one AD/KS domain, results are shown only for data sets that contained all cluster-belonging domains. A gene cluster presence/absence plot was created using the ggplot2 R package. Further information regarding selected metagenome data sets is shown in Table S2.

**Statistical analyses.** Alpha (Simpson index) and beta (Bray-Curtis) diversity indices across environments (bulk soil versus roots and soil versus tomato versus lettuce) were calculated using the R package vegan (70). Variation in NRPS/PKS-associated genes was visualized by principal-coordinate analysis (PCoA) using the same R package. To obtain this figure, we performed ordination on an ESV count table (constructed by QIIME2) using Bray-Curtis distances, followed by plotting using the R ggplot2 package. Difference significance between groups was determined using vegan ANOSIM (analysis of similarity).

Enrichment of ESVs between soil and tomato and lettuce roots and of NRPS/PKS-related sequences between root shotgun metagenomes and metatranscriptomes was determined using DESeq2 (71). Only sequences with a corrected-adjusted *P* value of <0.1 (Wald test *P* values corrected for multiple testing by the Benjamini-Hochberg method [72]) were chosen.

**Data availability.** Raw AD and KS amplicons from the tomato root, lettuce root, and tomato bulk soil microbial communities sequenced in this study are available via the MG-RAST data repository under accession number [mgm4862150.3](https://doi.org/10.1093/mgm/4862150.3). Previously sequenced shotgun metagenome and metatranscriptome data sets (32) are available via the NCBI SRA data repository under project accession number [PRJNA602301](https://doi.org/10.1093/prj/na602301).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, TIF file, 0.1 MB.

**FIG S2**, TIF file, 1.3 MB.

**FIG S3**, TIF file, 1.3 MB.

**FIG S4**, TIF file, 0.2 MB.

**FIG S5**, TIF file, 0.2 MB.

**FIG S6**, TIF file, 0.4 MB.

**TABLE S1**, CSV file, 0.01 MB.

**TABLE S2**, CSV file, 0.01 MB.

## ACKNOWLEDGMENTS

We thank Avihai Zolti, Asaf Levi, Jonathan Friedman, and Noa Sela for providing excellent technical and methodological help.

The manuscript was supported by U.S.-Israel Binational Agricultural Research and Development Fund (BARD) grant number IS-5177-19F.

## REFERENCES

- Lynch J, Whipps J. 1990. Substrate flow in the rhizosphere. *Plant Soil* 129:1–10. <https://doi.org/10.1007/BF00011685>.
- Pinton R, Varanini Z, Nannipieri P (ed). 2007. *The rhizosphere: biochemistry and organic substances at the soil-plant interface*, 2nd ed. CRC Press, Boca Raton, FL.
- Huang X-F, Chaparro JM, Reardon KF, Zhang R, Shen Q, Vivanco JM. 2014. Rhizosphere interactions: root exudates, microbes, and microbial communities. *Botany* 92:267–275. <https://doi.org/10.1139/cjb-2013-0225>.
- Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM, Piceno YM, DeSantis TZ, Andersen GL, Bakker PAHM, Raaijmakers JM. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* 332:1097–1100. <https://doi.org/10.1126/science.1203980>.
- Vacheron J, Desbrosses G, Bouffaud M-L, Touraine B, Moëne-Loccoz Y, Muller D, Legendre L, Wisniewski-Dyé F, Prigent-Combaret C. 2013. Plant growth-promoting rhizobacteria and root system functioning. *Front Plant Sci* 4:356. <https://doi.org/10.3389/fpls.2013.00356>.
- Tyc O, Song C, Dickschat JS, Vos M, Garbeva P. 2017. The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. *Trends Microbiol* 25:280–292. <https://doi.org/10.1016/j.tim.2016.12.002>.
- 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>.
- 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>.
- Payne JAE, Schoppet M, Hansen MH, Cryle MJ. 2017. Diversity of nature's assembly lines—recent discoveries in non-ribosomal peptide synthesis. *Mol Biosyst* 13:9–22. <https://doi.org/10.1039/c6mb00675b>.
- Hertweck C. 2009. The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* 48:4688–4716. <https://doi.org/10.1002/anie.200806121>.
- 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>.
- Chowdhury SP, Hartmann A, Gao X, Borriss R. 2015. Biocontrol mechanism by root-associated *Bacillus amyloliquefaciens* FZB42—a review. *Front Microbiol* 6:780. <https://doi.org/10.3389/fmicb.2015.00780>.
- Pham VH, Kim J. 2012. Cultivation of unculturable soil bacteria. *Trends Biotechnol* 30:475–484. <https://doi.org/10.1016/j.tibtech.2012.05.007>.
- Alain K, Querellou J. 2009. Cultivating the uncultured: limits, advances and future challenges. *Extremophiles* 13:583–594. <https://doi.org/10.1007/s00792-009-0261-3>.
- Abrudan MI, Smakman F, Grimbergen AJ, Westhoff S, Miller EL, van Wezel GP, Rozen DE. 2015. Socially mediated induction and suppression of anti-biosis during bacterial coexistence. *Proc Natl Acad Sci U S A* 112:11054–11059. <https://doi.org/10.1073/pnas.1504076112>.
- Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. 2018. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* 558:440–444. <https://doi.org/10.1038/s41586-018-0207-y>.
- Hover BM, Kim S-H, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, Maniko J, Estrela AB, Molina H, Park S, Perlin DS, Brady SF. 2018. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nat Microbiol* 3:415–422. <https://doi.org/10.1038/s41564-018-0110-1>.
- Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.
- Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. 2017. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res* 45:W49–W54. <https://doi.org/10.1093/nar/gkx320>.
- Lemetre C, Maniko J, Charlop-Powers Z, Sparrow B, Lowe AJ, Brady SF. 2017. Bacterial natural product biosynthetic domain composition in soil correlates with changes in latitude on a continent-wide scale. *Proc Natl Acad Sci U S A* 114:11615–11620. <https://doi.org/10.1073/pnas.1710262114>.
- Chevrette MG, Gutiérrez-García K, Selem-Mojica N, Aguilar-Martínez C, Yañez-Olvera A, Ramos-Aboites HE, Hoskisson PA, Barona-Gómez F. 2020. Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat Prod Rep* 37:566–599. <https://doi.org/10.1039/c9np00048h>.
- Amos GC, Borsetto C, Laskaris P, Krsek M, Berry AE, Newsham KK, Calvo-Bado L, Pearce DA, Vallin C, Wellington EM. 2015. Designing and implementing an assay for the detection of rare and divergent NRPS and PKS clones in European, Antarctic and Cuban soils. *PLoS One* 10:e0138327. <https://doi.org/10.1371/journal.pone.0138327>.
- 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>.
24. Dror B, Jurkevitch E, Cytryn E. 2020. State-of-the-art methodologies to identify antimicrobial secondary metabolites in soil bacterial communities—a review. *Soil Biol Biochem* 147:107838. <https://doi.org/10.1016/j.soilbio.2020.107838>.
  25. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooff JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG, Weber T, Medema MH. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res* 48:D454–D458. <https://doi.org/10.1093/nar/gkz882>.
  26. Wu C, Shang Z, Lemetre C, Ternei MA, Brady SF. 2019. Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug-resistant bacteria. *J Am Chem Soc* 141:3910–3919. <https://doi.org/10.1021/jacs.8b12087>.
  27. Owen JG, Charlop-Powers Z, Smith AG, Ternei MA, Calle PY, Reddy BVB, Montiel D, Brady SF. 2015. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc Natl Acad Sci U S A* 112:4221–4226. <https://doi.org/10.1073/pnas.1501124112>.
  28. Peek J, Lilic M, Montiel D, Milshteyn A, Woodworth I, Biggins JB, Ternei MA, Calle PY, Danziger M, Warrier T, Saito K, Braffman N, Fay A, Glickman MS, Darst SA, Campbell EA, Brady SF. 2018. Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nat Commun* 9:4147. <https://doi.org/10.1038/s41467-018-06587-2>.
  29. Reddy BVB, Milshteyn A, Charlop-Powers Z, Brady SF. 2014. eSNAPd: a versatile, Web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. *Chem Biol* 21:1023–1033. <https://doi.org/10.1016/j.chembiol.2014.06.007>.
  30. Liu W-T, Lamsa A, Wong WR, Boudreau PD, Kersten R, Peng Y, Moree WJ, Duggan BM, Moore BS, Gerwick WH, Linington RG, Pogliano G, Dorrestein PC. 2014. MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in *Streptomyces roseosporus*. *J Antibiot (Tokyo)* 67:99–104. <https://doi.org/10.1038/ja.2013.99>.
  31. Terlain B, Thomas J. 1971. Structure of griselimycin, polypeptide antibiotic extracted *Streptomyces* cultures. I. Identification of the products liberated by hydrolysis. *Bull Soc Chim Fr* 6:2349–2356.
  32. Zolti A, Green SJ, Sela N, Hadar Y, Minz D. 2019. The microbiome as a biosensor: functional profiles elucidate hidden stress in hosts. *bioRxiv* 752261. <https://doi.org/10.1101/752261>.
  33. Bilyk O, Samborsky M, Leadlay PF. 2019. The biosynthetic pathway to ossamycin, a macrocyclic polyketide bearing a spiroacetal moiety. *PLoS One* 14:e0215958. <https://doi.org/10.1371/journal.pone.0215958>.
  34. Schmitz H, Jubinski SD, Hooper IR, Crook KE, Jr, Price KE, Lein J. 1965. Ossamycin, a new cytotoxic agent. *J Antibiot (Tokyo)* 18:82–88.
  35. Blodgett JA, Oh D-C, Cao S, Currie CR, Kolter R, Clardy J. 2010. Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proc Natl Acad Sci U S A* 107:11692–11697. <https://doi.org/10.1073/pnas.1001513107>.
  36. Hamilton-Miller J. 1973. Chemistry and biology of the polyene macrolide antibiotics. *Bacteriol Rev* 37:166–196. <https://doi.org/10.1128/BR.37.3.166-196.1973>.
  37. Chevrette MG, Carlson CM, Ortega HE, Thomas C, Ananiev GE, Barns KJ, Book AJ, Cagnazzo J, Carlos C, Flanigan W, Grubbs KJ, Horn HA, Hoffmann FM, Klassen JL, Knack JJ, Lewin GR, McDonald BR, Muller L, Melo WGP, Pinto-Tomás AA, Schmitz A, Wendt-Pienkowski E, Wildman S, Zhao M, Zhang F, Bugni TS, Andes DR, Pupo MT, Currie CR. 2019. The antimicrobial potential of *Streptomyces* from insect microbiomes. *Nat Commun* 10:516. <https://doi.org/10.1038/s41467-019-08438-0>.
  38. Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K. 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc Natl Acad Sci U S A* 111:9259–9264. <https://doi.org/10.1073/pnas.1401734111>.
  39. Vannier N, Agler M, Hacquard S. 2019. Microbiota-mediated disease resistance in plants. *PLoS Pathog* 15:e1007740. <https://doi.org/10.1371/journal.ppat.1007740>.
  40. Loper JE, Gross H. 2007. Genomic analysis of antifungal metabolite production by *Pseudomonas fluorescens* Pf-5, p 265–278. In Lemanceau P, Bakker P, Raaijmakers J, Bloemberg G, Höfte M, Cooke BM (ed), *New perspectives and approaches in plant growth-promoting rhizobacteria research*. Springer, Dordrecht, The Netherlands.
  41. Tran H, Ficke A, Asimwe T, Höfte M, Raaijmakers JM. 2007. Role of the cyclic lipopeptide massetolide A in biological control of *Phytophthora infestans* and in colonization of tomato plants by *Pseudomonas fluorescens*. *New Phytol* 175:731–742. <https://doi.org/10.1111/j.1469-8137.2007.02138.x>.
  42. Charlop-Powers Z, Owen JG, Reddy BVB, Ternei MA, Guimaraes 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>.
  43. 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>.
  44. Uroz S, Buée M, Murat C, Frey-Klett P, Martin F. 2010. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ Microbiol Rep* 2:281–288. <https://doi.org/10.1111/j.1758-2229.2009.00117.x>.
  45. García-Salamanca A, Molina-Henares MA, van Dillewijn P, Solano J, Pizarro-Tobías P, Roca A, Duque E, Ramos JL. 2013. Bacterial diversity in the rhizosphere of maize and the surrounding carbonate-rich bulk soil. *Microb Biotechnol* 6:36–44. <https://doi.org/10.1111/j.1751-7915.2012.00358.x>.
  46. Yan Y, Kuramae EE, de Hollander M, Klinkhamer PG, van Veen JA. 2017. Functional traits dominate the diversity-related selection of bacterial communities in the rhizosphere. *ISME J* 11:56–66. <https://doi.org/10.1038/ismej.2016.108>.
  47. Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, Cho H, Karaoz U, Loqué D, Bowen BP, Firestone MK, Northen TR, Brodie EL. 2018. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat Microbiol* 3:470–480. <https://doi.org/10.1038/s41564-018-0129-3>.
  48. Yang Y, Wang N, Guo X, Zhang Y, Ye B. 2017. Comparative analysis of bacterial community structure in the rhizosphere of maize by high-throughput pyrosequencing. *PLoS One* 12:e0178425. <https://doi.org/10.1371/journal.pone.0178425>.
  49. Meena VS, Meena SK, Verma JP, Kumar A, Aeron A, Mishra PK, Bisht JK, Pattanayak A, Naveed M, Dotaniya M. 2017. Plant beneficial rhizospheric microorganism (PBRM) strategies to improve nutrients use efficiency: a review. *Ecol Eng* 107:8–32. <https://doi.org/10.1016/j.ecoleng.2017.06.058>.
  50. Wilkinson SW, Mageroy MH, López Sánchez A, Smith LM, Furci L, Cotton TA, Krokene P, Ton J. 2019. Surviving in a hostile world: plant strategies to resist pests and diseases. *Annu Rev Phytopathol* 57:505–529. <https://doi.org/10.1146/annurev-phyto-082718-095959>.
  51. Schlatter D, Kinkel L, Thomashow L, Weller D, Paulitz T. 2017. Disease suppressive soils: new insights from the soil microbiome. *Phytopathology* 107:1284–1297. <https://doi.org/10.1094/PHYTO-03-17-0111-RVW>.
  52. Bach E, dos Santos Seger GD, de Carvalho Fernandes G, Lisboa BB, Passaglia LMP. 2016. Evaluation of biological control and rhizosphere competence of plant growth promoting bacteria. *Appl Soil Ecol* 99:141–149. <https://doi.org/10.1016/j.apsoil.2015.11.002>.
  53. Cheng J, Yang SH, Palaniyandi SA, Han JS, Yoon T-M, Kim T-J, Suh J-W. 2010. Azalomycin F complex is an antifungal substance produced by *Streptomyces malaysiensis* MJM1968 isolated from agricultural soil. *J Korean Soc Appl Biol Chem* 53:545–552. <https://doi.org/10.3839/jksabc.2010.084>.
  54. Liu C, Zhuang X, Yu Z, Wang Z, Wang Y, Guo X, Xiang W, Huang S. 2019. Community structures and antifungal activity of root-associated endophytic actinobacteria of healthy and diseased soybean. *Microorganisms* 7:243. <https://doi.org/10.3390/microorganisms7080243>.
  55. Yamada T, Hamada M, Floreancig P, Nakabachi A. 2019. Diaphorin, a polyketide synthesized by an intracellular symbiont of the Asian citrus psyllid, is potentially harmful for biological control agents. *PLoS One* 14:e0216319. <https://doi.org/10.1371/journal.pone.0216319>.
  56. Carrión VJ, Perez-Jaramillo J, Cordovez V, Tracanna V, de Hollander M, Ruiz-Buck D, Mendes LW, van Ijcken WFJ, Gomez-Exposito R, Elsayed SS, Mohanraju P, Arifah A, van der Oost J, Paulson JN, Mendes R, van Wezel GP, Medema MH, Raaijmakers JM. 2019. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* 366:606–612. <https://doi.org/10.1126/science.aaw9285>.
  57. Owen JG, Reddy BVB, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH, Brady SF. 2013. Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc Natl Acad Sci U S A* 110:11797–11802. <https://doi.org/10.1073/pnas.1222159110>.
  58. Lamari L, Zitouni A, Boudjella H, Badji B, Sabaou N, Lebrihi A, Lefebvre G, Seguin E, Tillequin F. 2002. New dithioliopyrrolone antibiotics from

- Saccharothrix sp. SA 233. *J Antibiot (Tokyo)* 55:696–701. <https://doi.org/10.7164/antibiotics.55.696>.
59. te Welscher YM, van Leeuwen MR, de Kruijff B, Dijksterhuis J, Breukink E. 2012. Polyene antibiotic that inhibits membrane transport proteins. *Proc Natl Acad Sci U S A* 109:11156–11159. <https://doi.org/10.1073/pnas.1203375109>.
60. Wang H, Li C, Zhang B, He H, Jin P, Wang J, Zhang J, Wang X, Xiang W. 2015. Complete genome sequence of *Streptomyces cyaneogriseus* ssp. noncyanogenus, the thermotolerant producer of commercial antibiotics nemadectin. *J Biotechnol* 204:1–2. <https://doi.org/10.1016/j.jbiotec.2015.03.012>.
61. Zhao B, Gao Z, Shao Y, Yan J, Hu Y, Yu J, Liu Q, Chen F. 2012. Diversity analysis of type I ketosynthase in rhizosphere soil of cucumber. *J Basic Microbiol* 52:224–231. <https://doi.org/10.1002/jobm.201000455>.
62. Zolti A, Green SJ, Mordechay EB, Hadar Y, Minz D. 2019. Root microbiome response to treated wastewater irrigation. *Sci Total Environ* 655:899–907. <https://doi.org/10.1016/j.scitotenv.2018.11.251>.
63. Bolyen E, Rideout JR, Dillon MR, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852–857. <https://doi.org/10.1038/s41587-019-0209-9>.
64. Huson DH, Beier S, Flade I, Górska A, El-Hadidi M, Mitra S, Ruscheweyh H-J, Tappu R. 2016. MEGAN community edition—interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 12:e1004957. <https://doi.org/10.1371/journal.pcbi.1004957>.
65. Brady SF. 2007. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat Protoc* 2:1297–1305. <https://doi.org/10.1038/nprot.2007.195>.
66. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
67. Zhang T, Luo Y, Chen Y, Li X, Yu J. 2012. BIGrat: a repeat resolver for pyrosequencing-based re-sequencing with Newbler. *BMC Res Notes* 5:567. <https://doi.org/10.1186/1756-0500-5-567>.
68. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 47:W81–W87. <https://doi.org/10.1093/nar/gkz310>.
69. Kolde R, Kolde MR. 2015. Package ‘pheatmap’. R package 1.
70. Oksanen J, Kindt R, Legendre P, O’Hara B, Stevens MHH, Oksanen MJ, MASS Suggests. 2007. The vegan package. *Community Ecol Package* 10:631–637.
71. Love M, Anders S, Huber M. 2014. Differential gene expression analysis based on the negative binomial distribution. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
72. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57:289–300.