Elucidating the diversity and potential function of nonribosomal peptide and polyketide biosynthetic gene clusters in the root microbiome

Barak Dror[a,b], Zongqiang Wang[c], Sean F. Brady[c], Edouard Jurkevitch[b], Eddie Cytryn[a]#

[a] Institute of Soil, Water and Environmental Sciences, Agricultural Research Organization, Rishon-Lezion, Israel.

[b] Department of Plant Pathology and Microbiology, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel.

[c] Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, New York, New York, USA.

#Address correspondence to Eddie Cytryn, eddie@volcani.agro.gov.il

Abstract

Polyketides (PK) and nonribosomal peptides (NRP) are two microbial secondary metabolite (SM) families largely known for their variety of functions, including antimicrobials, siderophores and others. Despite their involvement in bacteria-bacteria and bacteria-plant interactions, root-associated SM are largely unexplored due to the limited cultivability of bacteria. Here, we analyzed the diversity, composition and expression of SM-encoding biosynthetic gene clusters (BGC) in root microbiomes by culture-independent amplicon sequencing, shotgun metagenomics and metatranscriptomics. Roots (tomato and lettuce) harbored distinct compositions of nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) relative to the adjacent bulk soil, and specific BGCs were both enriched and highly expressed in the root microbiomes. While several of the highly abundant and expressed sequences were remotely associated with BGCs encoding antimicrobials, the low similarity to characterized genes suggests their potential novelty. A subset of these genes were screened against a large set of soil-derived cosmid libraries, from which five whole NRPS/PKS BGCs of unknown function were retrieved. Three clusters were taxonomically affiliated with *Actinobacteria*, while the remaining two were associated with hosts of unknown phylogeny. One *Streptomyces*-derived BGC was predicted to encode for a polyene with potential antifungal activity, while the others were too novel to predict. Screening against a suite of metagenomic datasets revealed that all clusters were profuse in soil and roots, and almost completely absent in aquatic and gut environments, supporting the notion that they play a specific role in root ecosystems. Overall, our results indicate that root microbiomes harbor a specific assemblage of yet undiscovered SM.

Importance

We identified distinct secondary metabolite (polyketide and nonribosomal peptide) 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 for antimicrobials and siderophores, although the high sequence variability suggests that they encode for 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 are undoubtedly required for bacterial colonization and persistence in the root environment and are most likely associated with inter-bacterial and bacterial-plant interactions.

Keywords: Soil microbiome, root microbiome, polyketides, nonribosomal peptides, secondary metabolites, plant-microbe interactions

Introduction

Soil is an extremely diverse ecosystem that contains a myriad of micro- and macro-organisms, 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 "recruits" a specific fraction of the soil microbial community known as the root microbiome, which is significantly more abundant and less diverse than the surrounding soil microbiome (3). Over the past two decades, several studies have linked specific constituents of the root microbiome to enhanced plant growth and development and inhibition of soilborne plant pathogens (4), either by direct antagonism, and/or induced systemic resistance (5). These functions are often facilitated by the vast array of secondary metabolites (SM) produced by root-associated bacteria, which play a key role in inter- and intra-species interactions (6, 7).

Many important soil and root-associated bacterial SM are nonribosomal peptides (NRP), or polyketides (PK), produced by nonribosomal peptide synthetases (NRPS) and polyketide synthases (PKS), respectively. These are encoded on large biosynthetic gene clusters (BGC) 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). NRPS and PKS are responsible for the synthesis of a wide array of siderophores, toxins, pigments and antimicrobial compounds (11) that are believed to play a pivotal role 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 NRPS and PKS in root ecosystems, or their role in intra- and inter microbial and plant-bacterial interactions.

A major challenge in exploring the role and function of SM 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 they produce (13). Despite 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 BGC are not expressed using conventional *in-vitro* growth conditions and therefore the metabolites that they encode cannot be isolated (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 BGC encoding NRPS and PKS (16, 17), and to infer the chemical composition and structure of the metabolites produced by these synthases (18, 19). For instance, a novel amplicon sequencing-based approach was developed that targets short fragments within adenylation (AD- in NRPS) and ketosynthase (KS- in PKS) domains. These amplicons can be targeted to ascertain the diversity and abundance of bacterial BGC in complex environments, and help uncover the potential function of these gene clusters (20). To date, a handful of studies have explored the diversity and composition of bacterial SM-encoding BGC in soil, demonstrating the vast genetic diversity and novelty of NRPS and PKS genes (21, 22). However, little is known regarding the distribution and of these gene families in the root microbiome, and their functional role in the complex community interactions in root ecosystems remains an enigma (23).

This study proposes a novel approach to analyze the diversity and potential functions of NRPS and PKS 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 and bacterial phylogeny; (iii) the sequence and inferred SM structure of whole bacterial BGC that are highly abundant or expressed in root environments; and (iv) the occurrence of root-enriched bacterial BGC in other ecosystems.

Results

1. Composition and diversity of NRPS and PKS in roots vs. bulk soil

To determine the composition of root-associated NRPS/PKS and adjacent soil in lettuce and tomato, we applied a previously described amplicon sequencing approach to amplify the conserved adenylation (AD) and ketosynthase (KS) domains of NRPS/PKS, respectively (22). Overall, sequencing yielded a total of 1,850,442 and 2,174,020 raw KS and AD reads with average read lengths of 280 bp and 235 bp, respectively. Further filtering using QIIME2 and DADA2 denoising methods, resulted in 2,980 and 3,269 non-redundant KS and AD sequences, respectively.

We observed significantly higher diversity of both AD and KS domains in the bulk tomato soil compared to the adjacent roots (**Fig. S1- A and C**). In contrast, no difference in diversity was observed between tomato and lettuce roots for either of the SM encoding domains (**Fig. S1- B and D**). To assess differences in AD and KS domain diversity between samples, principal coordinate analysis (PCoA) and analysis of similarity (ANOSIM) using the Bray-Curtis similarity index (Fig. **1A** and **1B**) was performed. The KS and AD domain profiles of the roots (from both tomato and lettuce) formed distinct clusters, which were significantly different from the adjacent bulk tomato soil ($R=0.332$, $p< 0.05$ for PKS, $R=0.308$, $p<0.01$ for NRPS).

The phylogeny of the bacteria harboring the identified AD and KS sequences in the root and soil samples was inferred using Kaiju (24) (**Fig. 2A**). At the class level, we observed high abundance of Actinobacteria, specifically among bacteria harboring KS domains (50%, 56% and 63% of the lettuce, tomato and soil microbial communities, respectively). Higher abundance of Gammaproteobacteria was observed for the bacteria harboring AD domain sequences, especially in the lettuce roots (encompassing 25% of the sequences). Soil also had twice the number of Deltaproteobacteria-associated AD and KS domain sequences (13.8% and 11.24% compared with 6.3 and 6.5 and 8.1 and 5.3% in lettuce and tomato, respectively), and as high as 25% (soil AD) and 56% (tomato KS) of the sequences could not be assigned to known bacterial taxa. Zooming in to the order level, we saw higher abundance of Rhizobiales in the root-associated samples, whereas higher Myxococcales abundance was observed in the soil samples (**Fig. 2B**).

2. <u>Associated activity of AD/KS domains encoding root-enriched NRP and PK</u>

As SM are known to play critical roles in bacteria-bacteria and bacteria-plant interactions, we were interested in the associated activity of metabolites synthesized by gene clusters whose AD/KS domains were highly enriched in tomato and lettuce roots relative to bulk tomato soil. Using DESEQ2, AD and KS domain amplicons that were significantly more abundant in lettuce and tomato roots than in soil (adjusted p-Value<0.1) were detected. These amplicons were then aligned against the MIBiG repository (blastp, >50% amino acid sequence identity) and annotated based on corresponding BGC-associated metabolites, by using a cut-off e-Value of below 10-40. Previous analyses have shown that when compared to reference KS or AD domain sequences, amplicons with e-Values of <10-40 are likely derived from the same gene cluster or at least the same family of gene clusters as the reference sequence, and thus may be inferred to have similar function (25, 26).

Most of the highly enriched AD and KS domain amplicons in both of the plant roots, were not associated with characterized BGC, thus their potential activity cannot be speculated (**Fig. 3**). Nonetheless, several root-enriched AD and KS domains sequences were associated with SM with known antimicrobial activity, *e.g*. stenothricin (NRP, identified in tomato roots (27)) and griselmycin (NRP identified in lettuce roots (28)). In contrast, some of the root-enriched KS domains encoded for polyketides characterized as non-antimicrobial pharmaceuticals (NAP) such as anti-tumor and anti-inflammatory compounds, with no clear function in the root ecosystem (for instance the polyketides vicenistatin (29) or BE-43547 (30), identified in tomato roots, are only known for anti-tumor activity).

Due to the potential biases associated with the above PCR-based approach, we analyzed previously published (31) shotgun metagenomes of the same tomato and lettuce root samples. Assembled AD and KS domains from metagenomes were aligned against the MIBiG database, generating a list of the fifty most abundant NRPS/PKS-related genes in each of the root datasets (accounting for normalized abundance within-samples by plotting the coefficient of variance (CV) per gene, **Fig. 4A** and **B**). As in the amplicon sequencing analysis, many of the AD and KS genes in both plants were associated with antimicrobials and NAPs. Highly abundant domains in tomato roots (**Fig. 4A**) encoded for metabolites associated with antimicrobial, (e.g. streptazone E (I)), antifungal (heat-stable antifungal factor- HSAF, II and ECO-02301 (III)), and insecticidal (melingmycin, V) activity (32-34). In addition, several genes were linked to NAPs such as nemadectin (linked to nematocidal activity (35)) and the immunosuppressant meridamycin (IV and VI, respectively, (36)). In lettuce (**Fig. 4B**), we also found several AD and KS contigs that encoded for metabolites with associated antimicrobial and insecticidal activity, namely the

insecticide orfamide A (I and VI, (37)), the antimicrobial sessilin (III- V, (38)) and again HSAF (II).

Gene clusters are often silent or expressed under very specific conditions. Therefore evaluating gene expression in parallel to gene occurrence may uncover active BGCs and associated metabolites with ecological importance in the highly dynamic root ecosystem. Thus, in parallel to the shotgun metagenome analysis described above we applied shotgun meta-transcriptomics to identify NRPS/PKS gene clusters with enhanced expression in lettuce and tomato root microbiomes. Interestingly, many of the AD and KS domains that were highly abundant in the tomato and lettuce root microbial communities were highly expressed as well (**Fig. 4**). These included AD and KS contigs associated with HSAF in tomato and orfamide and sessilin in lettuce. Not only were these abundant, but they also displayed rather low CV values, indicating that they were present in all of the replicates tested. We also analyzed the AD and KS domains that were substantially more abundant in the meta-transcriptome compared with the metagenome datasets in order to pinpoint highly expressed SM encoding genes that are rare (**Fig. S2**). While we found only few sequences that were significantly expressed (relative to their presence, one in tomato and nine in lettuce), these were found to be associated with antimicrobial or NAP activity, notably the meroterpenoid antimicrobial LL-Z1272beta (39) in tomato and the bioactive cyclic depsipeptide polyoxypeptin (40) in lettuce.

Finally, to evaluate the extent to which the amplicon sequencing method was able to detect NRPS/PKS-associated metabolites in the targeted samples relative to the PCR-independent shotgun-metagenomic analyses, we analyzed the distribution of total MIBiG-associated genes in all four datasets (lettuce and tomato NRPS/PKS amplicons, tomato and lettuce metagenome, **Fig. S3**). In general, approximately 35% and 40% of the MBiG-characterized genes were found it

both amplicon sequences and metagenomes of the lettuce and tomato roots, respectively. In contrast, approximately 61% and 30% of the tomato and lettuce genes, respectively, were only detected in the shotgun metagenomic datasets. Less than 8% of the NRPS/PKS-related genes were found only within the amplicon sequencing datasets (28 and 35 genes in tomato and lettuce, 3.5% and 4.33%, respectively) and less than a fifth (139 genes, or 17.2%) were common to all four culture-independent datasets.

3. Potential novelty of root-associated NRPS/PKS sequences

To explore the potential novelty of root-associated SM producing genes, the amplified AD and KS domain sequences from the root and soil samples were grouped according to the amino acid identity to MIBiG reference genes (**Table 1**). Results demonstrate the high potential novelty of SM in both root and soil environments. Almost 40% of the KS and AD domain sequences in the root environment showed less than 50% amino acid identity with characterized genes in the MIBiG database (characterized as 'unassigned') whereas less than 6.5% of the sequences shared over 85% similarity to the reference MIBiG genes.

Using Kaiju, we evaluated the phylogenetic composition of the unassigned sequences, as these have the highest probability of encoding for yet unknown metabolites (**Fig. 5**). Depending upon the environment (soil vs. roots), 50-75% of the MIBiG-unassigned sequences could not be assigned to any specific bacterial taxa, further implying the novelty of these sequences hosts. The rest of the sequences in lettuce roots were primarily identified as being affiliated with Gammaproteobacteria, whereas in the tomato roots and soil microbiomes they were primarily

associated with Actinobacteria. We also found that ~10% of the unassigned PKS-associated sequences in soil belonged to the Epsilonproteobacteria class.

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

The identification of candidate NRPS/PKS-related sequences that were either root-enriched, abundant and/or highly expressed, encouraged us to extrapolate these sequences whole BGC. We screened these candidates against a large set of previously collated soil and rhizosphere cosmid libraries using the bioinformatic platform eSNaPD (41)(see Materials and Methods for full pipeline). Five cosmid library targets showed low e-Values and high nucleotide sequence identity (>75%) to candidate AD or KS sequences. Sequencing and annotation of the metagenomic insert captured in each cosmid revealed an overall of three NRPS and two hybrid NRPS/PKS gene clusters (Fig. 6). Based on gene content and sequence identity, none of the identified gene clusters were predicted to be identical to BGCs previously associated with a known metabolite. The NRPS and PKS open reading frames (ORF) of two recovered clones (B326 and B385) were not affiliated with any known bacterial taxa (<50% nucleic acid identity to the nr database), while the other three clones were related to genes from Actinobacteria (Table 2).

The five apparently novel BGCs recovered from the metagenomic libraries were targeted due to their abundance in tomato and lettuce root metagenome and metatranscriptome datasets, suggesting a link to root ecosystems. To test this hypothesis, we qualitatively assessed the presence 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) datasets. Our analysis clearly demonstrated that the recovered BGCs are highly

ubiquitous in terrestrial (soil and root) ecosystems (Fig. 7). Interestingly, not a single sequence was found in any of the gut microbiome communities analyzed, and very few were detected in aquatic environments.

Of the cosmids recovered from metagenomic libraries, clone B481 was nearly identical to an uncharacterized NRPS BGC found in the genome of *Streptomyces cyaneogriseus* (Fig. 8a). The only potential mode of action we could glean from the recovered gene clusters arises from clone B893, which is related to an uncharacterized PKS gene cluster found in the genome of *Saccharothrix saharansis*, a filamentous Actinobacteria isolated from desert soil. A detailed bioinformatic analysis of its PKS domains indicates the gene cluster likely encodes for an extended polyene substructure (Fig. 8b). 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. 8b). While polyene substructures like this are seen in a number of natural products (42) they are most commonly seen in polyene antifungal agents (e.g., Nystatin, Filipin and Pimaricin), suggesting that this BGC encodes for a currently unknown antifungal compound. This is further supported by the fact that B893 was extracted based on a candidate sequence from tomato associated with the novel polyene antifungal ECO-0230, identified in the soil bacterium *Streptomyces aizunensis* NRRL B-11277 (**Fig. 4A**)(34)

Discussion

NRP and PK produced by root-associated microbial communities play an important role in plant root ecosystems (43, 44). Several studies have identified and characterized BGCs and/or associated metabolites in prominent plant growth promoting and biocontrol agents originally isolated from plant roots (45, 46). 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 before to understand SM diversity and distribution in bulk soil and roots (17, 25, 47, 48), but NRPS/PKS have not been comprehensively explored in root ecosystems. Furthermore, to the best of our knowledge, this is the first study to explore expression of bacterial NRPS/PKS encoding genes in root ecosystems.

Our results demonstrate that bacterial communities with specific NRPS/PKS encoding genes are enriched in plant (tomato and lettuce) roots relative to adjacent bulk soil; that NRPS/PKS-related genes in root microbial communities are less diverse than those found in soil microbiomes; and that a fraction of these NRPS/PKS-associated genes 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 (49-51). It is well established that plants interact with soil bacterial communities through secretion of root exudates (52, 53), resulting in selection of specific microbial populations from the soil microbiome. This appears to be the case in the recruitment of plant-growth promoting rhizobacteria (PGPR), which are known to harbor specific SM encoding genes (54-56).

We inferred the associated activity of root NRPS and PKS by creating a custom repository, where each MIBiG alignment was literature surveyed for its experimentally determined *in-vitro* or *in-vivo* activity. Highly abundant and/or root-enriched NRPS/PKS genes were linked to a diverse array of biological activities, including various antibacterials, antifungals and siderophores, among others. For example, sequences associated with sessilin and orfamide

were both highly abundant in lettuce metagenome and meta-trancriptome datasets. These two NRPS-synthesized cyclic lipopeptides, produced by different Pseudomonas strains, were previously shown to have antifungal and insecticidal activity (37, 57), and were found to play a role in the biocontrol of the soilborne fungal pathogen Rhizoctonia, the causative agent of root rot in beans (58). In tomato roots we found sequences that were inferred to encode for additional predicted antifungals, including the recently identified polyketide Selvamicin (59) and HSAF (33), the latter was also found to be highly abundant in lettuce.

Our ability to identify and pinpoint sequences predicted to encode for NRP and PK that are both highly abundant (shotgun metagenome datasets) and highly expressed (metatranscriptome datasets) in root microbiomes, supports their potential involvement in various processes in the root environment, including rhizosphere competence and soil suppressivness (60, 61). In this regard, several of the dominant NRPS/PKS sequences in the root microbiomes were associated with metabolites characterized by us as non-antimicrobial pharmaceuticals (NAPs), and we can assume that these metabolites play unique ecological functions in the root environment, *e.g* inter- and/or intra-species communication (6). Recently, a previously unidentified hybrid NRPS/PKS gene cluster was found to be essential for Rhizoctonia suppression by an endophyte *Flavobacterium* (62), highlighting the vast amount of yet unidentified functional traits that are involved in plant-bacterial interactions.

Culture-independent approaches are a promising platform for identification of novel BGCs and for elucidating their roles in soil ecosystems and within the framework of drug discovery efforts (16, 17). Indeed, most of highly abundant and highly expressed NRPS- and PKS-associated sequences identified in this study were not assigned to or had low identity (50-70%) when compared to previously characterized genes in the MIBiG repository. This again

highlights the huge untapped chemical diversity of SM in root microbiomes, as was also reflected by the fact that these MIBiG-unassigned sequences were also unassigned taxonomically.

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 NRPS and PKS sequences (26, 63). 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 gene clusters were profuse in various soil- and root-associated environmental metagenomes, but rare or completely absent in other environments, highlights their importance in these habitats. While we can only speculate as to their synthesized metabolite's actual activity, they were associated with bacterial groups well-known for their antimicrobial capacity. *Saccharohrix saharansis (*Actinobacteria*)*, for instance, which contained SM-associated genes closely related to clone B893, is a soil-dwelling bacteria known to produce an array of antimicrobials (64). This gene cluster likely encodes for a polyene substructure, often seen in antifungal agents, as it capable of directly targeting the fungal membrane (65). Of particular interest is clone B481, associated with SM-associated genes from *Streptomyces cyaneogrisues*, known for its ability to produce the biopesticide Nemadectin (66). We speculate that it may be associated with bacterial-fungal competition in the root ecosystems.

At the broader level, our results emphasize the need to look beyond basic descriptive diversity and composition information regarding secondary metabolites capacity of microbial

communities. The unique pipeline adopted in this study, where potentially important AD and KS sequences are first identified and pinpointed, followed by extraction of BGC from cosmid libraries in order to identify whole potentially novel gene clusters, has the advantages of being resource-efficient, which spawning deeper knowledge regarding generated gene clusters. Future studies will focus on expressing these cosmid library gene clusters in suitable hosts will enable us to isolate and characterize their encoded products and test their activity both *in-vitro* and *in-planta*. In general, our results are on par with different studies conducted in other plants, showcasing the yet unexplored diversity of NRPS/PKS by exploring AD and KS domains using amplicon sequencing or PCR cloning of root samples (48, 67).

Overall, this study indicates that the root microbiome harbors a unique, diverse and potentially novel array of secondary metabolites, 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 SM in the root microbiome (*e.g.*, plant health and growth stage and presence of phytopathogens). This will undoubtedly help expose the ecological role of SM in root ecosystems and provide a platform for drug discovery and novel and environmentally sustainable compounds for plant protection.

Materials and Methods

Amplicon sequencing, shotgun metagenomic and meta-transcriptomics analyses

Tomato and lettuce root and soil samples were collected as previously described (68). Briefly, tomato (*Solanum lycopersicum*-Heinz 4107) and lettuce (*Lactuca sativa*-Romaine-Assaph) seedling were planted and grown for 42 days in a random block lysimeters experiment, at

Lachish agricultural research station in Kiryat-Gat, Israel. Analysis of each sample type (soil, tomato roots and lettuce roots) was based on replicate samples from three different lysimeters. Each replicate consisted of a composite sample collected from 2-4 plants or 3 soil samples, taken from the distant edges of the lysimeters and away from plant roots. Soil samples were collected, frozen in liquid nitrogen on site, and stored at −80 °C for until further analysis. Root samples were collected intact, and soil particles were removed by shaking. The roots were then dried and frozen in liquid nitrogen on site and kept at −80 °C until processed. in this study, extracted DNA was used as template for NRPS/PKS PCR amplification using degenerated primers (A3F/A7R for AD NRPS and degKS2F/degKS2R for KS PKS domains, as previously reported (22). Resulted barcoded libraries were pooled and sequenced on an Iluumina MiSeq, employing V2 chemistry at the University of Illinoi Chicago Sequencing Core (UICSQC). A total of 18 samples were sequenced, which included sampling location (tomato soil vs. tomato and lettuce roots) and SM family (AD domains for NRPS and KS domains for PKS) with three replicates for each treatment. Resulting sequences were processed and demultiplexed using QIIME2 pipeline and the integrated DADA2 method (69). Exact Sequence Variants (ESVs) represented by fewer than 3 sequences were removed from downstream analyses. Raw amplicon sequences data is available via the MG-RAST data depository, under mgm4862150.3. In addition, shotgun metagenome and metatranscriptome datasets of lettuce and tomato roots, previously generated and analyzed from same samples were also used for NRPS/PKS identification as described below (31). Shotgun sequences data is available via the NCBI Sequence Read Archive (SRA) data depository, with the project number PRJNA602301.

<u>Identification of NRPS/PKS sequences</u>

Fig. S4 present a conceptual description of applied pipeline.

NRPS/PKS. For chemical diversity analysis of NRP/PK gene clusters within samples, the different datasets (ESVs generated via amplicon sequencing and metagenome/meta-transcriptome assembled genes) were aligned against the Minimum Information about a Biosynthetic Gene cluster repository (MIBiG, version August 6th, 2018). Only core NRPS/PKS (AD and KS domains) genes were included in the analysis. Alignment was performed by using diamond blastx command line, with >50% amino acid sequence identity.

Phylogenetic taxonomy. Taxonomy of both ESVs from the amplicon sequencing analysis and of assembled genes identified as NRPS/PKS-associated was determined using Kaiju (against the nr database)(24).

Functional annotation of aligned NRPS/PKS sequences

To associate NRPS/PKS hit with its potential activity (only for hits with $<10^{-20}$ e-Value), we performed a manual literature survey per MIBiG-identified chemical function. Only studies demonstrating experimental function of the metabolites were used. The survey was done by searching the corresponding metabolite name using Google Scholar and determining the biological function as it was experimentally tested. Based on these activity annotations, different categories were set- antimicrobial (including antifungal and antibacterial activity), insecticidal, non-antimicrobial pharmaceutical (NAP, antitumor, antiviral, anti-inflammatory, etc.), siderophore and pigment. The full curated repository is publicly available via the Cytryn lab website (https://cytrynlab.wixsite.com/cytrynlab/databases).

Soil libraries amplicons generation screening

Metagenomic libraries were constructed as previously reported (70). 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 Eco EC100 (Lucigen). Each library was expanded to contain 20 x106 unique cosmid clones with ~30-45 kb eDNA (environmental DNA) inserts and then arrayed into 768 sub-pool (2×384-well plates) containing ~25,000 unique cosmid clones per well. Each sub-pool 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 NRPS and PKS, two set of degenerate primers (AD & KS) were applied to amplify the conserved region in Adenylation and Keto-synthase domains in the biosynthetic gene cluster: AD (forward) 5'-SATBTAYACSTCVGGHWCSAC-3' and (reverse) 5'-CCANRTCNCCBGTSYKGTACA-3'; KS (forward) 5'-TGYTCSDSSTCGCTSGTSGCS-3' (reverse) 5'-GTNCCSGTSCCRTGBGCYTCS-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 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 a 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 (71) to generate the 95% identity centroid sequences (NPST).

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 in the degenerate primers (26). Specific primers with Tm value ~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 whole-cell PCR strategy was used (17). Briefly, library glycerol stocks that contain target hits from eSNaPD analysis were inoculated into 3 ml LB broth (Kanamycin & Chloramphenicol) and grow overnight at 37 °C to confluence. The overnight cultured cells were diluted to 2000 CFU ml-1 calculated by OD600nm. 384 well plates were inoculated with 50 ul of the resulting dilution (600 CFU/well) with an Eppendorf epMotion 5075 liquid hander, grown overnight and screened using real-time PCR with touchdown PCR program, to identify wells containing target clones. Target positive wells were diluted to a concentration of ~5 CFU ml-1 and the process repeated to identify new wells containing target clones. Five clone pools were then plated on solid agar plate and target single clones were identified by clone PCR.

Analysis of Recovered Gene Clusters

Recovered single cosmid clones were miniprepped by QIAprep kit and sequenced using Miseq technology. 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 (72). 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. Putative function and source organism of genes in the BGC were assigned based on the closest characterized gene in NCBI. KnownClusterBlast in antiSMASH 5.0 (73) was utilized to analyze the relationship between known characterized gene clusters and recovered BGC. The structure prediction of adenylation domain and keto-synthase domain in BGC was given by the antiSMASH prediction which employ three prediction algorithms, NPRSPredictor2, Stachelhaus code, and SVM prediction. These predicted

building blocks were then utilized to predict final structure combined with known characterized BGC in cultured bacteria.

Recovered clones 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 humans feces (Gut), aquatic, soil and root-associated. We selected 20 shotgun metagenomes from each of these ecosystems and performed blastn search using the IMG workstation tool. Additional filtering was performed based on eValue threshold (10^{-5}) and identity (>85%). Hits that matched this filtering step were considered as present in the queried dataset. Presence/Absence heatmap created using pheatmap R package (74). Selected metagenome datasets are given in table S1.

Statistical Analyses

Alpha (Simpson index) and beta (Bray-Curtis) diversity indices across environments (bulk soil vs. roots and soil vs. tomato vs. lettuce) were calculated using R package vegan (75). Variation of NRPS/PKS-associated genes was visualized by Principal Coordinate Analysis (PCoA) using same R package. To obtain this figure, we performed ordination on ESV count table (constructed by qiime2) using Bray-Curtis distances, followed by plotting using R ggplot2 package. Difference significance between groups was determined using vegan ANOSIM (Analysis of similarities) analysis.

Enrichment of ESVs between soil and tomato and lettuce roots and of NRPS/PKS-related sequences between roots shotgun metagenomes and metatranscriptomes was determined using DESeq2 (76). Only sequences with adjusted P-valu <0.1 were chosen.

Acknowledgments

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

Funding

This manuscript was supported by the US-Israel Binational Agricultural Research and Development Fund (BARD) Grant #: IS-5177-19F.

Literature Cited

1. Lynch J, Whipps J. 1990. Substrate flow in the rhizosphere. Plant and soil 129:1-10.
2. Pinton R, Varanini Z, Nannipieri P. 2007. The rhizosphere: biochemistry and organic substances at the soil-plant interface. CRC press.
3. 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.
4. Mendes R, Kruijt M, De Bruijn I, Dekkers E, van der Voort M, Schneider JH, Piceno YM, DeSantis TZ, Andersen GL, Bakker PA. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science:1203980.
5. Vacheron J, Desbrosses G, Bouffaud M-L, Touraine B, Moënne-Loccoz Y, Muller D, Legendre L, Wisniewski-Dyé F, Prigent-Combaret C. 2013. Plant growth-promoting rhizobacteria and root system functioning. Frontiers in plant science 4:356.
6. 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 in microbiology 25:280-292.
7. Philippot L, Raaijmakers JM, Lemanceau P, Van Der Putten WH. 2013. Going back to the roots: the microbial ecology of the rhizosphere. Nature Reviews Microbiology 11:789.
8. Reddy BVB, Kallifidas D, Kim JH, Charlop-Powers Z, Feng Z, Brady SF. 2012. Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. Applied and environmental microbiology:AEM. 00102-12.
9. Payne JA, Schoppet M, Hansen MH, Cryle MJ. 2017. Diversity of nature's assembly lines–recent discoveries in non-ribosomal peptide synthesis. Molecular BioSystems 13:9-22.
10. Hertweck C. 2009. The biosynthetic logic of polyketide diversity. Angewandte Chemie International Edition 48:4688-4716.
11. Newman DJ, Cragg GM. 2016. Natural products as sources of new drugs from 1981 to 2014. Journal of natural products 79:629-661.
12. Chowdhury SP, Hartmann A, Gao X, Borriss R. 2015. Biocontrol mechanism by root-associated Bacillus amyloliquefaciens FZB42–a review. Frontiers in microbiology 6:780.
13. Pham VH, Kim J. 2012. Cultivation of unculturable soil bacteria. Trends in biotechnology 30:475-484.
14. Alain K, Querellou J. 2009. Cultivating the uncultured: limits, advances and future challenges. Extremophiles 13:583-594.
15. Abrudan MI, Smakman F, Grimbergen AJ, Westhoff S, Miller EL, van Wezel GP, Rozen DE. 2015. Socially mediated induction and suppression of antibiosis during bacterial coexistence. Proceedings of the National Academy of Sciences 112:11054-11059.

509 16. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. 2018. Novel soil bacteria
510 possess diverse genes for secondary metabolite biosynthesis. Nature 558:440.
511 17. Hover BM, Kim S-H, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, Maniko J, Estrela AB,
512 Molina H, Park S. 2018. Culture-independent discovery of the malacidins as calcium-dependent
513 antibiotics with activity against multidrug-resistant Gram-positive pathogens. Nature
514 microbiology 3:415.
515 18. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos EL,
516 Kim HU, Nave M. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster
517 boundary identification. Nucleic acids research 45:W36-W41.
518 19. Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. 2017. PRISM 3: expanded prediction of
519 natural product chemical structures from microbial genomes. Nucleic acids research 45:W49-
520 W54.
521 20. Lemetre C, Maniko J, Charlop-Powers Z, Sparrow B, Lowe AJ, Brady SF. 2017. Bacterial natural
522 product biosynthetic domain composition in soil correlates with changes in latitude on a
523 continent-wide scale. Proceedings of the National Academy of Sciences:201710262.
524 21. Amos GC, Borsetto C, Laskaris P, Krsek M, Berry AE, Newsham KK, Calvo-Bado L, Pearce DA,
525 Vallin C, Wellington EM. 2015. Designing and implementing an assay for the detection of rare
526 and divergent NRPS and PKS clones in European, Antarctic and Cuban soils. PloS one
527 10:e0138327.
528 22. Charlop-Powers Z, Pregitzer CC, Lemetre C, Ternei MA, Maniko J, Hover BM, Calle PY, McGuire
529 KL, Garbarino J, Forgione HM. 2016. Urban park soil microbiomes are a rich reservoir of natural
530 product biosynthetic diversity. Proceedings of the national academy of sciences 113:14811-
531 14816.
532 23. Dror B, Jurkevitch E, Cytryn E. 2020. State-of-the-art methodologies to identify antimicrobial
533 secondary metabolites in soil bacterial communities-A review. Soil Biology and
534 Biochemistry:107838.
535 24. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics
536 with Kaiju. Nature communications 7:1-9.
537 25. Wu C, Shang Z, Lemetre C, Ternei MA, Brady SF. 2019. Cadasides, calcium-dependent acidic
538 lipopeptides from the soil metagenome that are active against multidrug-resistant bacteria.
539 Journal of the American Chemical Society 141:3910-3919.
540 26. Owen JG, Charlop-Powers Z, Smith AG, Ternei MA, Calle PY, Reddy BVB, Montiel D, Brady SF.
541 2015. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted
542 discovery of epoxyketone proteasome inhibitors. Proceedings of the National Academy of
543 Sciences 112:4221-4226.
544 27. Liu W-T, Lamsa A, Wong WR, Boudreau PD, Kersten R, Peng Y, Moree WJ, Duggan BM, Moore
545 BS, Gerwick WH. 2014. MS/MS-based networking and peptidogenomics guided genome mining
546 revealed the stenothricin gene cluster in Streptomyces roseosporus. The Journal of antibiotics
547 67:99-104.
548 28. Terlain B, Thomas J. 1971. Structure of griselimycin, polypeptide antibiotic extracted
549 Streptomyces cultures. I. Identification of the products liberated by hydrolysis. Bulletin de la
550 Societe chimique de France 6:2349-2356.
551 29. Shindo K, KAMISHOHARA M, ODAGAWA A, MATSUOKA M, KAWAI H. 1993. Vicenistatin, a novel
552 20-membered macrocyclic lactam antitumor antibiotic. The Journal of antibiotics 46:1076-1081.
553 30. Igarashi Y, Shimasaki R, Miyanaga S, Oku N, Onaka H, Sakurai H, Saiki I, Kitani S, Nihira T,
554 Wimonsiravude W. 2010. Rakicidin D, an inhibitor of tumor cell invasion from marine-derived
555 Streptomyces sp. The Journal of antibiotics 63:563-565.

556 31. Zolti A, Green SJ, Sela N, Hadar Y, Minz D. 2019. The microbiome as a biosensor: functional
557 profiles elucidate hidden stress in hosts. bioRxiv:752261.
558 32. Ohno S, Katsuyama Y, Tajima Y, Izumikawa M, Takagi M, Fujie M, Satoh N, Shin - ya K, Ohnishi Y.
559 2015. Identification and characterization of the streptazone E biosynthetic gene cluster in
560 Streptomyces sp. MSC090213JE08. ChemBioChem 16:2385-2391.
561 33. Yu F, Zaleta-Rivera K, Zhu X, Huffman J, Millet JC, Harris SD, Yuen G, Li X-C, Du L. 2007. Structure
562 and biosynthesis of heat-stable antifungal factor (HSAF), a broad-spectrum antimycotic with a
563 novel mode of action. Antimicrobial agents and chemotherapy 51:64-72.
564 34. McAlpine JB, Bachmann BO, Piraee M, Tremblay S, Alarco A-M, Zazopoulos E, Farnet CM. 2005.
565 Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel
566 antifungal agent, as an example. Journal of natural products 68:493-496.
567 35. Doscher M, Wood I, Pankavich J, Ricks C. 1989. Efficacy of nemadectin, a new broad-spectrum
568 endectocide, against natural infections of canine gastrointestinal helminths. Veterinary
569 parasitology 34:255-259.
570 36. Salituro GM, Zink DL, Dahl A, Nielsen J, Wu E, Huang L, Kastner C, Dumont FJ. 1995.
571 Meridamycin: a novel nonimmunosuppressive FKBP12 ligand from Streptomyces hygroscopicus.
572 Tetrahedron letters 36:997-1000.
573 37. Jang JY, Yang SY, Kim YC, Lee CW, Park MS, Kim JC, Kim IS. 2013. Identification of orfamide A as
574 an insecticidal metabolite produced by Pseudomonas protegens F6. Journal of agricultural and
575 food chemistry 61:6786-6791.
576 38. D'aes J, Kieu NP, Léclère V, Tokarski C, Olorunleke FE, De Maeyer K, Jacques P, Höfte M, Ongena
577 M. 2014. To settle or to move? The interplay between two classes of cyclic lipopeptides in the
578 biocontrol strain P seudomonas CMR 12a. Environmental microbiology 16:2282-2300.
579 39. Mogi T, Ui H, Shiomi K, Ōmura S, Miyoshi H, Kita K. 2009. Antibiotics LL-Z1272 identified as novel
580 inhibitors discriminating bacterial and mitochondrial quinol oxidases. Biochimica et Biophysica
581 Acta (BBA)-Bioenergetics 1787:129-133.
582 40. Umezawa K, Nakazawa K, Uemura T, Ikeda Y, Kondo S, Naganawa H, Kinoshita N, Hashizume H,
583 Hamada M, Takeuchi T. 1998. Polyoxypeptin isolated from Streptomyces: a bioactive cyclic
584 depsipeptide containing the novel amino acid 3-hydroxy-3-methylproline. Tetrahedron letters
585 39:1389-1392.
586 41. Reddy BVB, Milshteyn A, Charlop-Powers Z, Brady SF. 2014. eSNaPD: a versatile, web-based
587 bioinformatics platform for surveying and mining natural product biosynthetic diversity from
588 metagenomes. Chemistry & biology 21:1023-1033.
589 42. Hamilton-Miller J. 1973. Chemistry and biology of the polyene macrolide antibiotics.
590 Bacteriological reviews 37:166.
591 43. Wang H, Fewer DP, Holm L, Rouhiainen L, Sivonen K. 2014. Atlas of nonribosomal peptide and
592 polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes.
593 Proceedings of the National Academy of Sciences 111:9259-9264.
594 44. Vannier N, Agler M, Hacquard S. 2019. Microbiota-mediated disease resistance in plants. PLoS
595 pathogens 15.
596 45. Loper JE, Gross H. 2007. Genomic analysis of antifungal metabolite production by Pseudomonas
597 fluorescens Pf-5, p 265-278, New Perspectives and Approaches in Plant Growth-Promoting
598 Rhizobacteria Research. Springer.
599 46. Tran H, Ficke A, Asiimwe T, Höfte M, Raaijmakers JM. 2007. Role of the cyclic lipopeptide
600 massetolide A in biological control of Phytophthora infestans and in colonization of tomato
601 plants by Pseudomonas fluorescens. New Phytologist 175:731-742.

47. 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.
48. 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. Scientific reports 7:2330.
49. Uroz S, Buée M, Murat C, Frey‐Klett P, Martin F. 2010. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. Environmental Microbiology Reports 2:281-288.
50. 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. Microbial biotechnology 6:36-44.
51. 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. The ISME journal 11:56.
52. Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, Cho H, Karaoz U, Loqué D, Bowen BP. 2018. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nature microbiology 3:470-480.
53. 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.
54. 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. Ecological Engineering 107:8-32.
55. Wilkinson SW, Magerøy 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. Annual review of phytopathology 57:505-529.
56. Mendes R, Kruijt M, De Bruijn I, Dekkers E, van der Voort M, Schneider JH, Piceno YM, DeSantis TZ, Andersen GL, Bakker PA. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science 332:1097-1100.
57. Olorunleke FE, Hua GKH, Kieu NP, Ma Z, Höfte M. 2015. Interplay between orfamides, sessilins and phenazines in the control of R hizoctonia diseases by P seudomonas sp. CMR 12a. Environmental microbiology reports 7:774-781.
58. Hua GKH, Höfte M. 2015. The involvement of phenazines and cyclic lipopeptide sessilin in biocontrol of Rhizoctonia root rot on bean (Phaseolus vulgaris) by Pseudomonas sp. CMR12a is influenced by substrate composition. Plant and soil 388:243-253.
59. Van Arnam EB, Ruzzini AC, Sit CS, Horn H, Pinto-Tomás AA, Currie CR, Clardy J. 2016. Selvamicin, an atypical antifungal polyene from two alternative genomic contexts. Proceedings of the National Academy of Sciences 113:12940-12945.
60. Schlatter D, Kinkel L, Thomashow L, Weller D, Paulitz T. 2017. Disease suppressive soils: new insights from the soil microbiome. Phytopathology 107:1284-1297.
61. 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. Applied soil ecology 99:141-149.
62. Carrión VJ, Perez-Jaramillo J, Cordovez V, Tracanna V, De Hollander M, Ruiz-Buck D, Mendes LW, van Ijcken WF, Gomez-Exposito R, Elsayed SS. 2019. Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. Science 366:606-612.

649 63. Owen JG, Reddy BVB, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH, Brady SF. 2013. Mapping
650 gene clusters within arrayed metagenomic libraries to expand the structural diversity of
651 biomedically relevant natural products. Proceedings of the National Academy of Sciences
652 110:11797-11802.
653 64. Lamari L, Zitouni A, Boudjella H, Badji B, Sabaou N, Lebrihi A, Lefebvre G, Seguin E, Tillequin F.
654 2002. New Dithiolopyrrolone Antibiotics from Saccharothrix sp. SA 233. The Journal of
655 antibiotics 55:696-701.
656 65. te Welscher YM, van Leeuwen MR, de Kruijff B, Dijksterhuis J, Breukink E. 2012. Polyene
657 antibiotic that inhibits membrane transport proteins. Proceedings of the National Academy of
658 Sciences 109:11156-11159.
659 66. Wang H, Li C, Zhang B, He H, Jin P, Wang J, Zhang J, Wang X, Xiang W. 2015. Complete genome
660 sequence of Streptomyces cyaneogriseus ssp. noncyanogenus, the thermotolerant producer of
661 commercial antibiotics nemadectin. Journal of biotechnology 204:1-2.
662 67. Zhao B, Gao Z, Shao Y, Yan J, Hu Y, Yu J, Liu Q, Chen F. 2012. Diversity analysis of type I
663 ketosynthase in rhizosphere soil of cucumber. Journal of basic microbiology 52:224-231.
664 68. Zolti A, Green SJ, Mordechay EB, Hadar Y, Minz D. 2019. Root microbiome response to treated
665 wastewater irrigation. Science of The Total Environment 655:899-907.
666 69. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ,
667 Arumugam M, Asnicar F. 2018. QIIME 2: Reproducible, interactive, scalable, and extensible
668 microbiome data science. PeerJ Preprints,
669 70. Brady SF. 2007. Construction of soil environmental DNA cosmid libraries and screening for
670 clones that produce biologically active small molecules. Nature protocols 2:1297.
671 71. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics
672 26:2460-2461.
673 72. Zhang T, Luo Y, Chen Y, Li X, Yu J. 2012. BIGrat: a repeat resolver for pyrosequencing-based re-
674 sequencing with Newbler. BMC research notes 5:567.
675 73. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH
676 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic acids research
677 47:W81-W87.
678 74. Kolde R, Kolde MR. 2015. Package 'pheatmap'. R Package 1.
679 75. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M. 2007. The
680 vegan package. Community ecology package 10:631-637.
681 76. Love M, Anders S, Huber M. 2014. Differential gene expression analysis based on the negative
682 binomial distribution. Genome Biol 15:550.

683

684

685

686

687

Figures and Tables

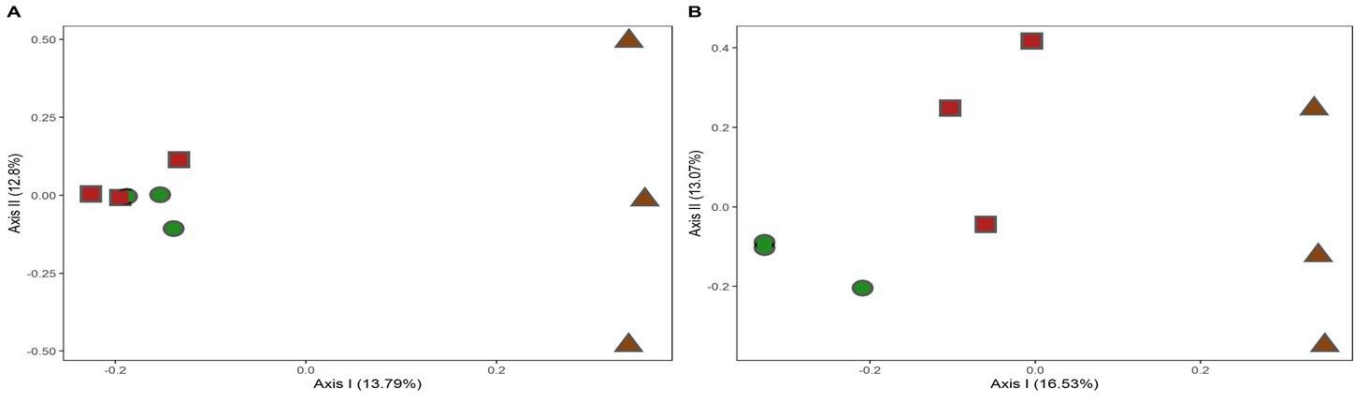

Figure 1. Principal coordinate analysis (PCoA) of lettuce (green circles), tomato (red squares) and (tomato) soil (brown triangle) samples using the Bray-Curtis similarity index. (A) Analysis of KS (from PKS) domain amplicons (Roots vs. soil ANOSIM statistic R=0.332, p-Value<0.05, tomato vs. lettuce vs. soil ANOSIM R=0.432, NS). (B) Analysis of AD (NRPS) domain amplicons (Roots vs. soil ANOSIM statistic R- 0.308, p-Value<0.01, tomato vs. lettuce vs. soil ANOSIM R=0.6872, p-Value<0.01).

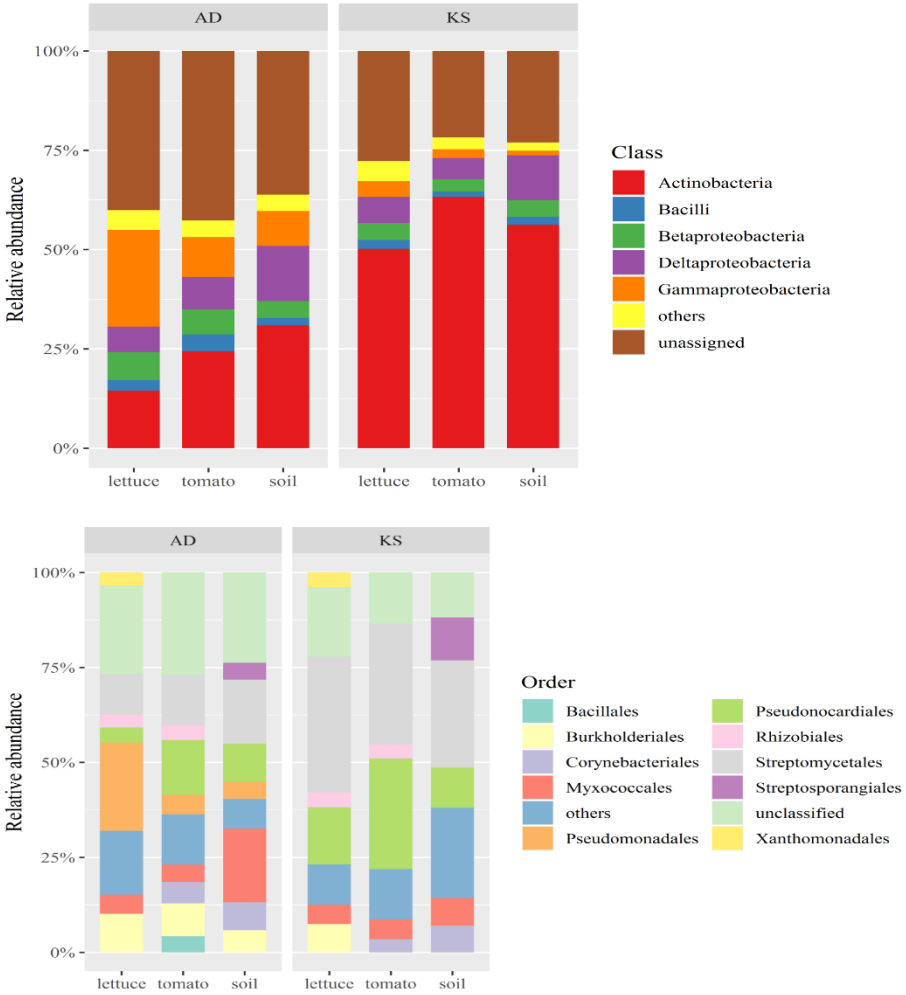

Figure 2. Taxonomic inference of bacteria harboring AD/KS domains in tomato, lettuce and bulk soil samples. (A) Class level. (B) Order level. Bars represent the relative abundance of each bacterial class. N=3 per sample. Taxonomy assignment was done using Kaiju (against the nr database).

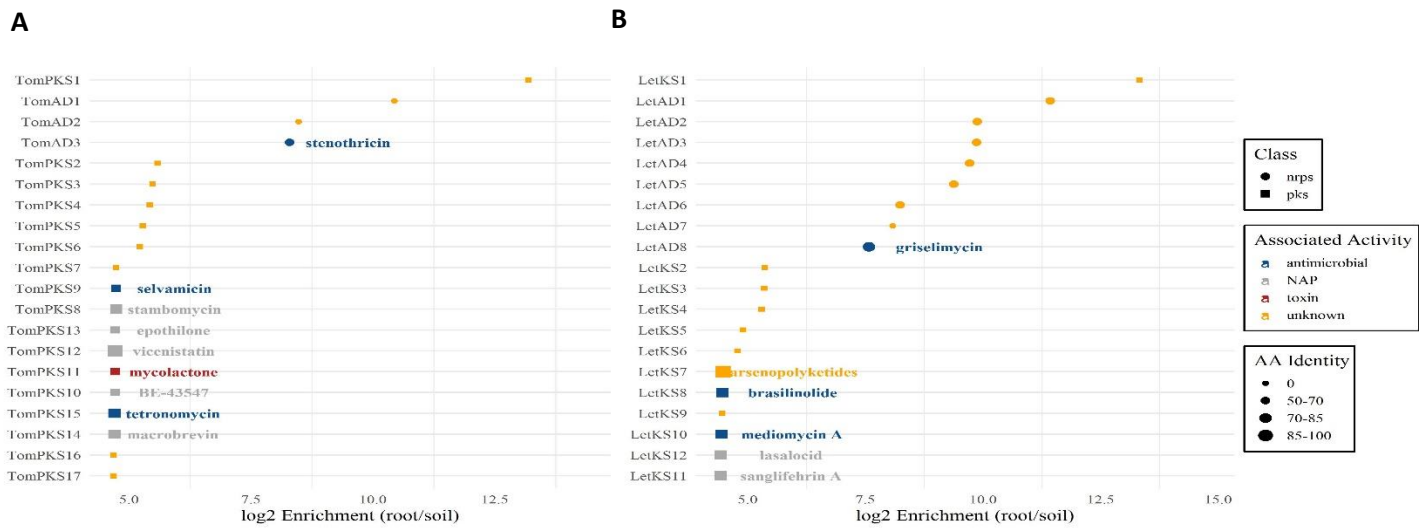

Figure 3. Root enriched NRPS/PKS amplicons relative to bulk soil in tomato (A) and lettuce (B). Only the top 20 sequences with adjusted Pvalue <0.1 are shown. Associated metabolites (based on the MIBiG repository) are presented only for sequences with e-Value<10-40. Identity bars represent amino acid identity to MIBiG reference sequence (0-not identified). Associated activity was determined based on experimentally verified activity of related metabolite.

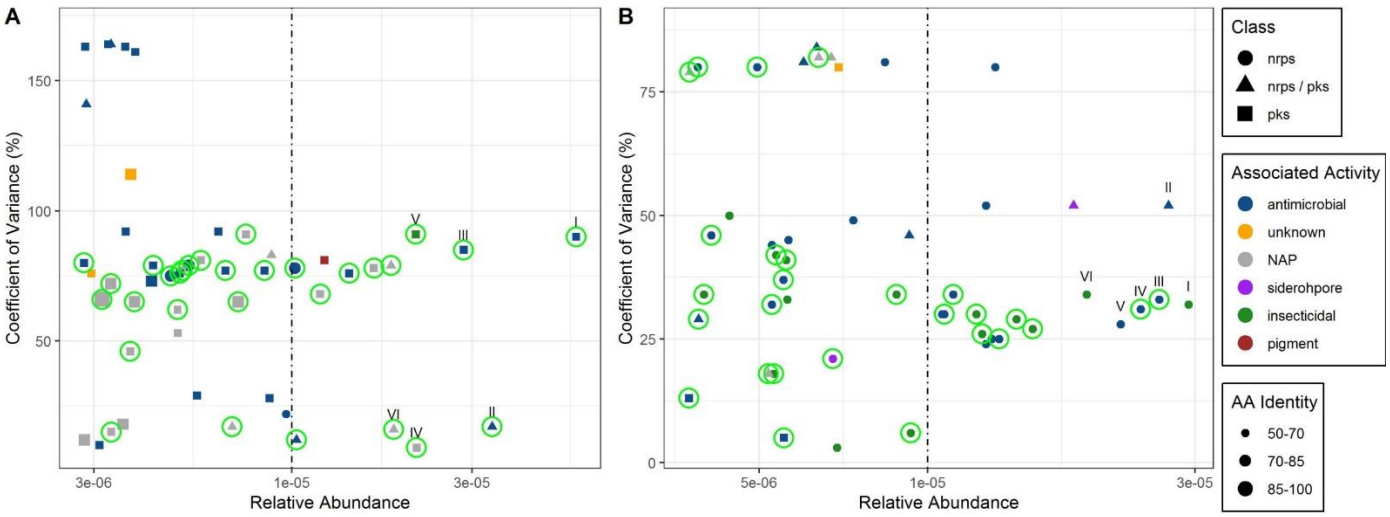

Figure 4. Top 50 NRPS/PKS-related sequences in tomato (A) and lettuce (B) root microbiomes extrapolated from shotgun metagenomic data. Associated activity (based on the MIBiG repository) are presented only for sequences with e-Values<10-40. Identity bars represent amino

acid identity to MIBiG reference sequence (0-not identified). Relative abundance represent the mean of three repeats per sample. CV (coefficient of variance) was determined as STD/Mean per sequence. Green-circled points represent sequences that are highly abundant in the corresponding meta-transcriptome datasets per sample type. Roman numbers represent associated metabolites (A: I- Streptazone E , II- HSAF, III-ECO-02301, IV-nemadectin, V-melingmycin, VI-meridamycin; B: I-orfamide,II-HSAF, III-V-sessilin, VI-orfamide). Red letters represent sequences that were later identified in the cosmid libraries (A-B893, B- B385, C- B326).

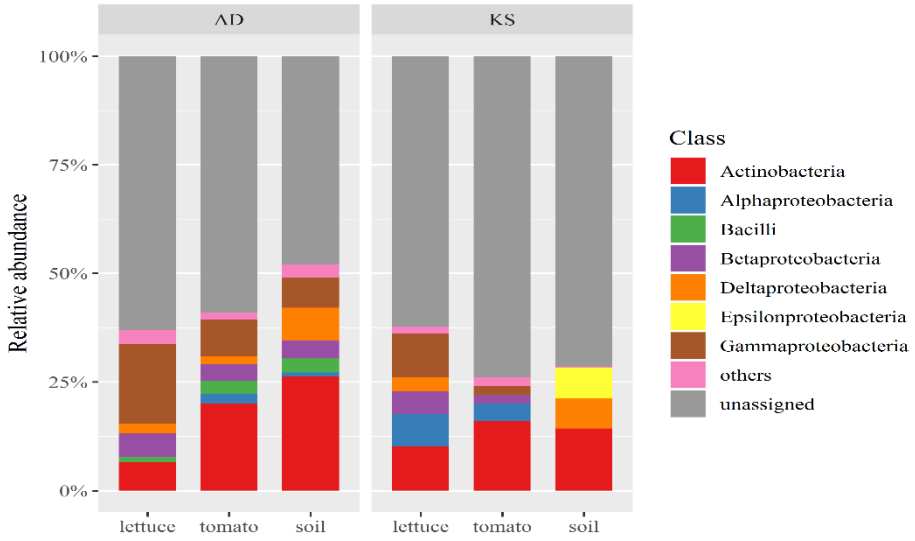

Figure 5. Predicted taxonomy of MIBiG-unassigned (< 50% amino acid sequence identity) AD and KS amplicons based on Kaiju comparisons (to the nr database).

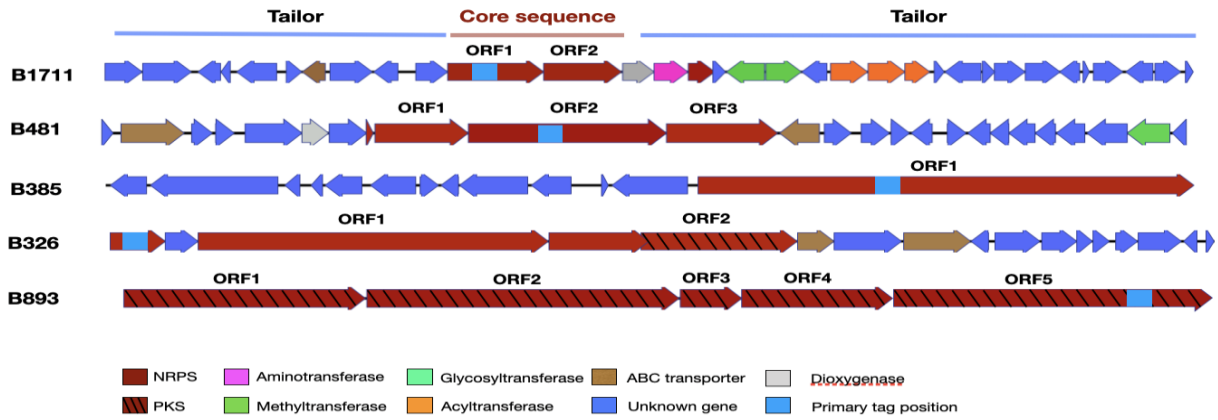

Figure 6. Overview of retrieved (from archived soil cosmid libraries) biosynthetic gene clusters containing AD and KS domain contigs that were abundant and/or highly expressed in tomato and cucumber root microbiomes.

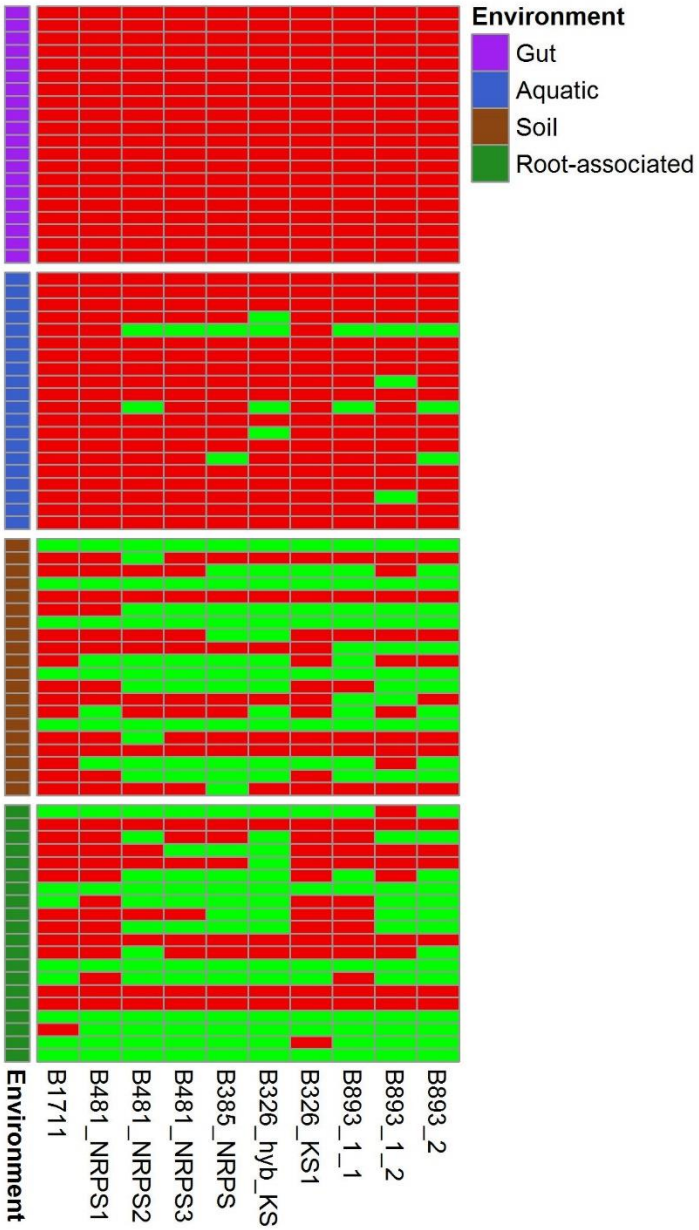

Figure 7. Presence/absence of extracted clones AD and KS domains in shotgun metagenomes from 4 environmental ecosystems- gut (animal and human), aquatic (freshwater and marine), soil and root-associated. N= 20,21,17 and 20, respectively. Analysis was conducted using IMG BlastN feature with eValue<10-5 and identity >85%.

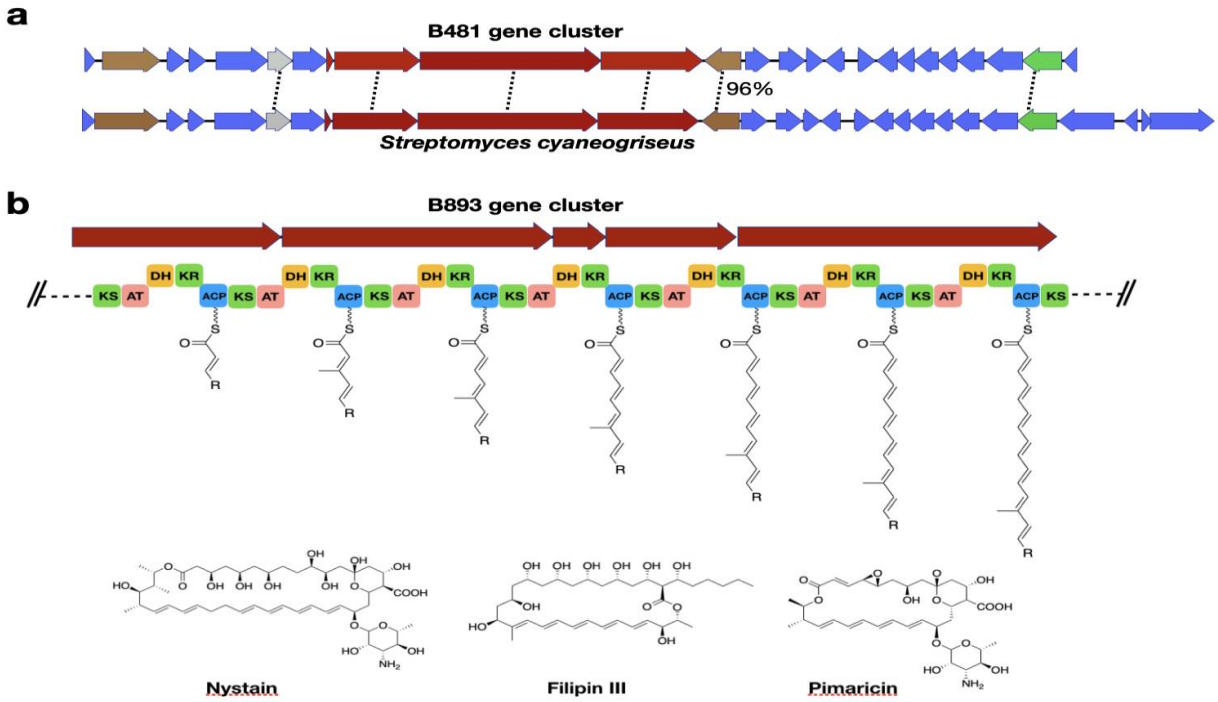

Figure 8. (a) Comparison of B481 cluster and the related BGC in *S. cyaneogriseus.* (b) Structure prediction for PKS modules present on clone B893. Examples of know polyene antifungal agents.

Gene	Identity	Tomato	Lettuce	Average (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 (1114)
	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 (1138)	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)

Table 1. Abundance of NRPS (AD) and PKS (KS) amplicons grouped based on levels of similarity to MIBIG NRPS/PKS reference genes, in the tomato and cucumber root and tomato soil samples. Groups were characterized as unassigned (< 50% sequence identity), 50-70% sequence identity, 70-85% sequence identity, 85-95% sequence identity and 95-100% sequence identity. Numbers outside parentheses indicate % of sequences associated with each group, numbers inside parentheses indicate total number of hits for each group.

Gene Cluster	Domain	Coverage	Identity	Accession ID	Source Dataset	Affiliated Organism	Taxa
1711	B1711-1 (NRPS)	84%	58.92%	*WP_136726459.1*	*Lettuce*	*Streptomyces sp. NEAU-C40*	*Actinobacteria*
	B1711-2 (NRPS)	97%	69.36%	*WP_136726460.1*		*Streptomyces sp. NEAU-C40*	*Actinobacteria*
481	B481-1 (NRPS)	99%	92.97%	*WP_052808563.1*	*Tomato*	*Streptomyces cyaneogriseus*	*Actinobacteria*
	B481-2 (NRPS)	99%	96.60%	*WP_052808562.1*		*Streptomyces cyaneogriseus*	*Actinobacteria*
	B481-3 (NRPS)	99%	93.30%	*WP_044379030.1*		*Streptomyces cyaneogriseus*	*Actinobacteria*
385	B385 (NRPS)	UA	UA		*Lettuce*	UA	
326	B326-1 (NRPS)	UA	UA		*Lettuce*	UA	
	B326-2 (PKS)	UA	UA			UA	
893	B893-1 (PKS)	98%	*90.90%*	*WP_141974662.1*	*Tomato*	*Saccharothrix Saharansis*	*Actinobacteria*
	B893-2 (PKS)	99%	*89.02%*	*WP_141974663.1*		*Saccharothrix Saharansis*	*Actinobacteria*
	B893-3 (PKS)	98%	*92.31%*	*TQM77832.1*		*Saccharothrix Saharansis*	*Actinobacteria*
	B893-4 (PKS)	100%	*87.62%*	*WP_141974664.1*		*Saccharothrix Saharansis*	*Actinobacteria*
	B893-5 (PKS)	97%	*89.78%*	*WP_141974665.1*		*Saccharothrix Saharansis*	*Actinobacteria*

Table 2. Predicted taxonomy information and source dataset of recovered cosmids. UA-unassigned, only affiliated organisms with >50% identity are shown.