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.agri.gov.il

Abstract

Polyketides (PKs) and nonribosomal peptides (NRPs) 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 SMs are largely unexplored due to the limited cultivability of bacteria. Here, we analyzed the diversity, composition 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 (NRPS) and polyketide synthases (PKS) 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 NRPS and PKS BGCs of unknown function were retrieved. Three clusters were taxonomically affiliated with Actinobacteria, while the remaining were associated with bacteria of unknown phylogenetic origin. One Streptomyces-derived BGC was predicted to encode for a polyene with potential antifungal activity, while the others were too novel to predict chemical structure. Screening against a suite of metagenomic datasets revealed a higher abundance 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 yet undiscovered SMs.

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 may encode for novel metabolites and 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 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 "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 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 (SMs) 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 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 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 NRPSs

and PKSs 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 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 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 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 synthases (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 identity and activity, of the synthesized metabolites (20, 21). To date, a handful of studies have explored the diversity and composition of bacterial SM-encoding BGCs in soil, demonstrating the vast genetic diversity and novelty of NRPSs and PKSs genes (22, 23). However, little is known regarding the distribution and of these gene families in the root microbiome, and their functional role in this complex community remains an enigma (24).

This study proposes a novel 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 NRPSs and PKSs encoding genes in the root environment relative to adjacent bulk soil; (ii) NRPSs and PKSs 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.

Results

1. Composition and diversity of NRPSs and PKSs genes in roots vs. bulk soil

To determine the composition and diversity of NRPSs and PKSs genes in tomato and lettuce root samples relative to bulk soil (previous studies targeting this controlled lysimeter system showed that bulk soil 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 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 (table S1). Further filtering steps using QIIME2 and DADA2 denoising methods, resulted in 2,980 and 3,269 non-redundant KS and AD domain 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, a principal coordinate analysis (PCoA) and analysis of similarity (ANOSIM) using the Bray-Curtis similarity index were performed (Fig. **1A** and **1B**). 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 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 producing 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 >50% amino acid sequence identity cut-off), and then grouped according to their identity to the MIBiG reference genes (**Table 1**). Almost 40% of the KS and AD domain sequences in the root environment showed less than 50% amino acid identity with genes found in the MIBiG database (characterized as 'unassigned') whereas less than 6.5% of the sequences shared over 85% similarity to the reference MIBiG genes. These results demonstrate the profusion of potentially novel SMs in both root and soil environments.

2. Pinpointing predicted root-enriched NRPs and PKs

As SMs are known to play critical roles in bacteria-bacteria and bacteria-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 tomato soil. To do so, the MIBiG-aligned amplicons were annotated to corresponding BGC-associated metabolites, using a cut-off e-Value of $< 10^{-40}$. Sequences that did not meet these criteria were defined as 'unknown'. 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 BGC family as the reference sequence, and thus may be inferred to encode a similar function (26-29).

DESeq2 analysis of the AD and KS amplicons revealed that 55% (11/20) and 70% (14/20) of the highly enriched AD and KS domain amplicons in both of the plant root samples (tomato and lettuce respectively, adjusted p-value < 0.1, log2FoldChange>5), were not associated with known BGCs, thus their associated metabolites cannot be inferred (**Fig. 2**). Nonetheless, several root-enriched AD and KS domains sequences (9 in tomato and 6 in lettuce) were above these threshold values, thus associated with known metabolites. These included the non-ribosomal peptides stenothricin (30) and griselimycin (31) whose BGC families were predicted to be highly abundant in the tomato and lettuce roots, respectively, relative to the bulk soil. While we cannot determine the actual role of these enriched metabolites in the root environment, both of them are known for their antimicrobial activity.

Next, we calculated the relative abundance of amplicons that were associated with known metabolites (based on the criteria above) in the different root and soil samples (**Fig. 3**). For NRPs, we observed several associated metabolites that were highly abundant in both roots and the soil (*e.g.*, the co-regulated Pseudomonas-derived NRPs sessilin and orfamide). In contrast, stenothricin was significantly more abundant in both of the root samples relative to soil (**Fig. 3A**, $p<0.05$ Kruskal-Wallis test). For PKs, stambomycin and concanamycin A were significantly more abundant in lettuce roots, whereas PM100117/8 was more abundant in the tomato roots (**Fig. 3B**, $p<0.05$). In addition, 26 associated metabolites were present in at least one of the root-associated samples, and totally absent in the soil samples, *e.g*. diaphorin (in lettuce) and basiliskamides (present in both tomato and lettuce).

Due to the potential biases associated with the above PCR-based approach, we analyzed previously published (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 fifty most abundant NRPSs and PKSs in each of the root datasets (representing the normalized abundance within-samples by plotting the coefficient of variance (CV) for each gene, **Fig. 4A** and **B**). Highly abundant sequences in tomato roots (**Fig. 4A**) are predicted to encode for metabolites that have previously been shown to possess antimicrobial, (e.g., streptazone E (I)), antifungal (heat-stable antifungal factor- HSAF, II and ECO-02301 (III)), and insecticidal (melingmycin, V) activity (33-35). Other genes were associated with metabolites such as nemadectin (nematocidal activity (36)) and the immunosuppressant meridamycin (IV and VI, respectively, (37)). In lettuce (**Fig. 4B**), we also identified several NRPSs and PKSs associated with the insecticide orfamide A (I and VI, (38)), the antimicrobial sessilin (III- V, (39)) and HSAF as well (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 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 NRPSs and PKSs with enhanced expression in lettuce 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. 4**). These included sequences associated with HSAF in tomato and orfamide and sessilin in lettuce. Not only were these highly abundant, but they also displayed rather low CV values, indicating that they were equally present in all of the replicates tested. We also analyzed NRPs and PKSs 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**). Using DESEq2, we found one sequence in tomato roots and nine in lettuce roots that were significantly expressed relative to their presence.

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 datasets (lettuce and tomato NRPSs and PKSs amplicons, tomato and lettuce metagenomes, **Fig. S3**). In general, approximately 35% and 40% of the MBiG-characterized genes were found in 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 NRPSs and PKSs 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. <u>Extraction and environmental distribution of whole SM-encoding gene clusters</u>

The identification of NRPSs and PKSs that were either enriched in roots relative to adjacent bulk soil or were 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 previously identified NRPSs and PKSs candidate sequences against a large set of previously collated soil and rhizosphere cosmid libraries using the bioinformatic platform eSNaPD (29)(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 three NRPS and two hybrid NRPS/PKS gene clusters (**Fig. 5**). 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 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 we could glean from the recovered BGC was for 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 revealed that the gene cluster likely encodes for 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 polyene substructures like this are seen in a number of natural products (40, 41), 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 may encode for an 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**)(35)

The five recovered BGCs were initially 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 assessed the abundance 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 demonstrated that the recovered BGCs are highly ubiquitous in most of the queried root samples and in some of the soil samples (**Fig. 7A**). Only B893 showed a significantly higher abundance in the root samples when compared with the soil samples ($P<0.05$, Wilcoxon test, **Fig. S4**). B893 was found in 16/20 of the root-associated datasets we examined (compared with 8/20 soil datasets, **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 (42, 43). Several studies have identified and characterized BGCs and/or associated metabolites in prominent plant growth promoting and biocontrol agents originally isolated from plant roots (44, 45). 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 SMs diversity and distribution in bulk soil and roots (17, 26, 46, 47), 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 expression of these sequences in root ecosystems.

Our results demonstrate a distinct composition 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 (48-50). It is well established that plants interact with soil bacterial communities through secretion of root exudates (51, 52), 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 (53-55). Thus, while at this point we cannot infer the actual function of the highly abundant and expressed NRPs and PKs encoding genes in the root environment, we can infer that these NRPSs and PKSs likely play a role in various processes, *e.g.* competition and root colonization (56, 57).

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. Sessilin and orfamide, for example, were both highly abundant in the 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 (38, 58), and were found to play a role in the biocontrol of the soilborne fungal pathogen Rhizoctonia, the causative agent of root rot in beans (59). In tomato roots we found sequences that were inferred to encode for additional predicted

antifungals, including the recently identified polyketide selvamicin (60) and HSAF (34), the latter was also found to be highly abundant in lettuce. The NRP/PK hybrid HSAF is produced by the bacterium *Lysobacter enzymogenes* strain C3, a known biocontrol of a wide range of fungal species (34, 61). The fact that its homologue was ubiquitous, highly abundant and highly expressed in both lettuce and tomato roots, suggests that this metabolite may play a pivotal role in the ecology of the root microbiome (6). In this regard, 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).

A large fraction of the highly abundant and highly expressed NRPSs and PKSs identified in this study were not assigned to, or had low identity (50-70%), when compared 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. (62), highlighting the vast amount of root-associated SMs with unidentified functional roles, that undoubtedly play a pivotal role in bacterial-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, 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 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 metabolite's actual activity, they were

associated with bacterial groups well-known for their antimicrobial capacity. *Saccharohrix saharansis (*Actinobacteria*)*, for instance, which contains a BGC closely related to clone B893, is a soil-dwelling bacteria known to produce an array of antimicrobials (64). This BGC likely encodes for a polyene substructure, often seen in antifungal agents, as it capable of directly disrupting 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 this BGC 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 SM capacity of microbial communities. The pipeline adopted in this study, where potentially important AD and KS sequences are first identified, followed by extraction of BGCs from cosmid libraries in order to identify potentially novel BGCs, has the advantages of being resource-efficient, while spawning deeper knowledge regarding potential 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* activity against various phytopathogens. Our results coincide with studies conducted in other plants, showcasing the yet unexplored diversity of NRPSs and PKSs (47, 67).

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, 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

Fig. S5 present a conceptual description of the pipeline applied in this study.

Amplicon sequencing, shotgun metagenomic and meta-transcriptomics analyses

Tomato soil and root samples and lettuce root 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. Each sample type (soil, tomato roots and lettuce roots) was analyzed in triplicates samples from three different lysimeters (thus n=3 for each sample type). Each triplicate 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. As the same soil was used throughout the experiment, and given soil sample collection was taken for sufficient distance from growing plants, bulk soil from the tomato lysimeters served as a reference point for both tomato and lettuce soils, as previous study showed that they harbored almost identical microbial communities (Zolti et al. 2019). 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 and briefly rinsing. The roots were then lightly dried and immediately frozen in liquid nitrogen on site and kept at −80 °C until processed. in this study, extracted DNA was used as template for NRPSs and PKSs PCR amplification using degenerated primers (A3F/A7R for AD-NRPS and degKS2F/degKS2R for

KS-PKS domains, as previously reported (23). 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 and KS domains) 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 are available via the MG-RAST data depository, under mgm4862150.3. In addition, shotgun metagenome and metatranscriptome datasets of lettuce and tomato roots (n=3 for each dataset type, hence 6 for tomato and 6 for lettuce), previously generated and analyzed from same samples were also used for NRPSs and PKSs identification as described below (32). Shotgun sequences data is available via the NCBI Sequence Read Archive (SRA) data depository, with the project number PRJNA602301.

Identification and annotation of NRPSs and PKSs

For chemical diversity analysis of NRPS and PKS gene clusters, 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 NRPSs and PKSs genes (AD and KS domains) were included in the analysis. Alignment was performed using diamond blastx command line, with >50% amino acid sequence identity. To associate each NRPS or PKS hit with its potentially synthesized metabolite, e-Value $<10^{-40}$ was used as a cut-off. Hits that did not pass this threshold were regarded as 'unknown'.

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 NRPSs and PKSs, 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% 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 in the degenerate primers (27). 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 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 eValue threshold (10^{-40}) and identity (>85%). For each hit, counts were normalized using gyrB and rpoB housekeeping genes counts (obtained via the IMG/MER platform). A relative abundance heatmap created using pheatmap R package (74). For gene clusters with more than one AD/KS domain, results are shown only for datasets that contained all cluster-belonging domains. Gene clusters presence/absence plot created using ggplot2 R package. Further information regarding selected metagenome datasets is shown in table S2.

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 an adjusted P-value <0.1 (Wald test p-values corrected for multiple testing by Benjamini-Hochberg method) 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.

475 7. Philippot L, Raaijmakers JM, Lemanceau P, Van Der Putten WH. 2013. Going back to the roots:
476 the microbial ecology of the rhizosphere. Nature Reviews Microbiology 11:789.
477 8. Reddy BVB, Kallifidas D, Kim JH, Charlop-Powers Z, Feng Z, Brady SF. 2012. Natural product
478 biosynthetic gene diversity in geographically distinct soil microbiomes. Applied and
479 environmental microbiology:AEM. 00102-12.
480 9. Payne JA, Schoppet M, Hansen MH, Cryle MJ. 2017. Diversity of nature's assembly lines–recent
481 discoveries in non-ribosomal peptide synthesis. Molecular BioSystems 13:9-22.
482 10. Hertweck C. 2009. The biosynthetic logic of polyketide diversity. Angewandte Chemie
483 International Edition 48:4688-4716.
484 11. Newman DJ, Cragg GM. 2016. Natural products as sources of new drugs from 1981 to 2014.
485 Journal of natural products 79:629-661.
486 12. Chowdhury SP, Hartmann A, Gao X, Borriss R. 2015. Biocontrol mechanism by root-associated
487 Bacillus amyloliquefaciens FZB42–a review. Frontiers in microbiology 6:780.
488 13. Pham VH, Kim J. 2012. Cultivation of unculturable soil bacteria. Trends in biotechnology 30:475-
489 484.
490 14. Alain K, Querellou J. 2009. Cultivating the uncultured: limits, advances and future challenges.
491 Extremophiles 13:583-594.
492 15. Abrudan MI, Smakman F, Grimbergen AJ, Westhoff S, Miller EL, van Wezel GP, Rozen DE. 2015.
493 Socially mediated induction and suppression of antibiosis during bacterial coexistence.
494 Proceedings of the National Academy of Sciences 112:11054-11059.
495 16. Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. 2018. Novel soil bacteria
496 possess diverse genes for secondary metabolite biosynthesis. Nature 558:440.
497 17. Hover BM, Kim S-H, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, Maniko J, Estrela AB,
498 Molina H, Park S. 2018. Culture-independent discovery of the malacidins as calcium-dependent
499 antibiotics with activity against multidrug-resistant Gram-positive pathogens. Nature
500 microbiology 3:415.
501 18. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos EL,
502 Kim HU, Nave M. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster
503 boundary identification. Nucleic acids research 45:W36-W41.
504 19. Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. 2017. PRISM 3: expanded prediction of
505 natural product chemical structures from microbial genomes. Nucleic acids research 45:W49-
506 W54.
507 20. Lemetre C, Maniko J, Charlop-Powers Z, Sparrow B, Lowe AJ, Brady SF. 2017. Bacterial natural
508 product biosynthetic domain composition in soil correlates with changes in latitude on a
509 continent-wide scale. Proceedings of the National Academy of Sciences:201710262.
510 21. Chevrette MG, Gutiérrez-García K, Selem-Mojica N, Aguilar-Martínez C, Yañez-Olvera A, Ramos-
511 Aboites HE, Hoskisson PA, Barona-Gómez F. 2020. Evolutionary dynamics of natural product
512 biosynthesis in bacteria. Natural Product Reports 37:566-599.
513 22. Amos GC, Borsetto C, Laskaris P, Krsek M, Berry AE, Newsham KK, Calvo-Bado L, Pearce DA,
514 Vallin C, Wellington EM. 2015. Designing and implementing an assay for the detection of rare
515 and divergent NRPS and PKS clones in European, Antarctic and Cuban soils. PloS one
516 10:e0138327.
517 23. Charlop-Powers Z, Pregitzer CC, Lemetre C, Ternei MA, Maniko J, Hover BM, Calle PY, McGuire
518 KL, Garbarino J, Forgione HM. 2016. Urban park soil microbiomes are a rich reservoir of natural
519 product biosynthetic diversity. Proceedings of the national academy of sciences 113:14811-
520 14816.

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 Biology and Biochemistry:107838.
25. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJ, Van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic acids research 48:D454-D458.
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. Journal of the American Chemical Society 141:3910-3919.
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. Proceedings of the National Academy of Sciences 112:4221-4226.
28. Peek J, Lilic M, Montiel D, Milshteyn A, Woodworth I, Biggins JB, Ternei MA, Calle PY, Danziger M, Warrier T. 2018. Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. Nature communications 9:1-15.
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. Chemistry & biology 21:1023-1033.
30. Liu W-T, Lamsa A, Wong WR, Boudreau PD, Kersten R, Peng Y, Moree WJ, Duggan BM, Moore BS, Gerwick WH. 2014. MS/MS-based networking and peptidogenomics guided genome mining revealed the stenothricin gene cluster in Streptomyces roseosporus. The Journal of antibiotics 67:99-104.
31. Terlain B, Thomas J. 1971. Structure of griselimycin, polypeptide antibiotic extracted Streptomyces cultures. I. Identification of the products liberated by hydrolysis. Bulletin de la Societe chimique de France 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.
33. Ohno S, Katsuyama Y, Tajima Y, Izumikawa M, Takagi M, Fujie M, Satoh N, Shin‐ya K, Ohnishi Y. 2015. Identification and characterization of the streptazone E biosynthetic gene cluster in Streptomyces sp. MSC090213JE08. ChemBioChem 16:2385-2391.
34. Yu F, Zaleta-Rivera K, Zhu X, Huffman J, Millet JC, Harris SD, Yuen G, Li X-C, Du L. 2007. Structure and biosynthesis of heat-stable antifungal factor (HSAF), a broad-spectrum antimycotic with a novel mode of action. Antimicrobial agents and chemotherapy 51:64-72.
35. McAlpine JB, Bachmann BO, Piraee M, Tremblay S, Alarco A-M, Zazopoulos E, Farnet CM. 2005. Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. Journal of natural products 68:493-496.
36. Doscher M, Wood I, Pankavich J, Ricks C. 1989. Efficacy of nemadectin, a new broad-spectrum endectocide, against natural infections of canine gastrointestinal helminths. Veterinary parasitology 34:255-259.
37. Salituro GM, Zink DL, Dahl A, Nielsen J, Wu E, Huang L, Kastner C, Dumont FJ. 1995. Meridamycin: a novel nonimmunosuppressive FKBP12 ligand from Streptomyces hygroscopicus. Tetrahedron letters 36:997-1000.
38. Jang JY, Yang SY, Kim YC, Lee CW, Park MS, Kim JC, Kim IS. 2013. Identification of orfamide A as an insecticidal metabolite produced by Pseudomonas protegens F6. Journal of agricultural and food chemistry 61:6786-6791.

39. D'aes J, Kieu NP, Léclère V, Tokarski C, Olorunleke FE, De Maeyer K, Jacques P, Höfte M, Ongena M. 2014. To settle or to move? The interplay between two classes of cyclic lipopeptides in the biocontrol strain P seudomonas CMR 12a. Environmental microbiology 16:2282-2300.
40. Hamilton-Miller J. 1973. Chemistry and biology of the polyene macrolide antibiotics. Bacteriological reviews 37:166.
41. Chevrette MG, Carlson CM, Ortega HE, Thomas C, Ananiev GE, Barns KJ, Book AJ, Cagnazzo J, Carlos C, Flanigan W. 2019. The antimicrobial potential of Streptomyces from insect microbiomes. Nature communications 10:1-11.
42. 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. Proceedings of the National Academy of Sciences 111:9259-9264.
43. Vannier N, Agler M, Hacquard S. 2019. Microbiota-mediated disease resistance in plants. PLoS pathogens 15.
44. Loper JE, Gross H. 2007. Genomic analysis of antifungal metabolite production by Pseudomonas fluorescens Pf-5, p 265-278, New Perspectives and Approaches in Plant Growth-Promoting Rhizobacteria Research. Springer.
45. Tran H, Ficke A, Asiimwe 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 Phytologist 175:731-742.
46. 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.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.
52. 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.
53. 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.
54. 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.

613 55. Mendes R, Kruijt M, De Bruijn I, Dekkers E, van der Voort M, Schneider JH, Piceno YM, DeSantis
614 TZ, Andersen GL, Bakker PA. 2011. Deciphering the rhizosphere microbiome for disease-
615 suppressive bacteria. Science 332:1097-1100.
616 56. Schlatter D, Kinkel L, Thomashow L, Weller D, Paulitz T. 2017. Disease suppressive soils: new
617 insights from the soil microbiome. Phytopathology 107:1284-1297.
618 57. Bach E, dos Santos Seger GD, de Carvalho Fernandes G, Lisboa BB, Passaglia LMP. 2016.
619 Evaluation of biological control and rhizosphere competence of plant growth promoting
620 bacteria. Applied soil ecology 99:141-149.
621 58. Olorunleke FE, Hua GKH, Kieu NP, Ma Z, Höfte M. 2015. Interplay between orfamides, sessilins
622 and phenazines in the control of R hizoctonia diseases by P seudomonas sp. CMR 12a.
623 Environmental microbiology reports 7:774-781.
624 59. Hua GKH, Höfte M. 2015. The involvement of phenazines and cyclic lipopeptide sessilin in
625 biocontrol of Rhizoctonia root rot on bean (Phaseolus vulgaris) by Pseudomonas sp. CMR12a is
626 influenced by substrate composition. Plant and soil 388:243-253.
627 60. Van Arnam EB, Ruzzini AC, Sit CS, Horn H, Pinto-Tomás AA, Currie CR, Clardy J. 2016. Selvamicin,
628 an atypical antifungal polyene from two alternative genomic contexts. Proceedings of the
629 National Academy of Sciences 113:12940-12945.
630 61. Zhao Y, Cheng C, Jiang T, Xu H, Chen Y, Ma Z, Qian G, Liu F. 2019. Control of wheat fusarium
631 head blight by heat-stable antifungal factor (HSAF) from Lysobacter enzymogenes. Plant disease
632 103:1286-1292.
633 62. Carrión VJ, Perez-Jaramillo J, Cordovez V, Tracanna V, De Hollander M, Ruiz-Buck D, Mendes LW,
634 van Ijcken WF, Gomez-Exposito R, Elsayed SS. 2019. Pathogen-induced activation of disease-
635 suppressive functions in the endophytic root microbiome. Science 366:606-612.
636 63. Owen JG, Reddy BVB, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH, Brady SF. 2013. Mapping
637 gene clusters within arrayed metagenomic libraries to expand the structural diversity of
638 biomedically relevant natural products. Proceedings of the National Academy of Sciences
639 110:11797-11802.
640 64. Lamari L, Zitouni A, Boudjella H, Badji B, Sabaou N, Lebrihi A, Lefebvre G, Seguin E, Tillequin F.
641 2002. New Dithiolopyrrolone Antibiotics from Saccharothrix sp. SA 233. The Journal of
642 antibiotics 55:696-701.
643 65. te Welscher YM, van Leeuwen MR, de Kruijff B, Dijksterhuis J, Breukink E. 2012. Polyene
644 antibiotic that inhibits membrane transport proteins. Proceedings of the National Academy of
645 Sciences 109:11156-11159.
646 66. Wang H, Li C, Zhang B, He H, Jin P, Wang J, Zhang J, Wang X, Xiang W. 2015. Complete genome
647 sequence of Streptomyces cyaneogriseus ssp. noncyanogenus, the thermotolerant producer of
648 commercial antibiotics nemadectin. Journal of biotechnology 204:1-2.
649 67. Zhao B, Gao Z, Shao Y, Yan J, Hu Y, Yu J, Liu Q, Chen F. 2012. Diversity analysis of type I
650 ketosynthase in rhizosphere soil of cucumber. Journal of basic microbiology 52:224-231.
651 68. Zolti A, Green SJ, Mordechay EB, Hadar Y, Minz D. 2019. Root microbiome response to treated
652 wastewater irrigation. Science of The Total Environment 655:899-907.
653 69. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, Alexander H, Alm EJ,
654 Arumugam M, Asnicar F. 2018. QIIME 2: Reproducible, interactive, scalable, and extensible
655 microbiome data science. PeerJ Preprints,
656 70. Brady SF. 2007. Construction of soil environmental DNA cosmid libraries and screening for
657 clones that produce biologically active small molecules. Nature protocols 2:1297.
658 71. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics
659 26:2460-2461.

660 72. Zhang T, Luo Y, Chen Y, Li X, Yu J. 2012. BIGrat: a repeat resolver for pyrosequencing-based re-
661 sequencing with Newbler. BMC research notes 5:567.
662 73. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH
663 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic acids research
664 47:W81-W87.
665 74. Kolde R, Kolde MR. 2015. Package 'pheatmap'. R Package 1.
666 75. Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M. 2007. The
667 vegan package. Community ecology package 10:631-637.
668 76. Love M, Anders S, Huber M. 2014. Differential gene expression analysis based on the negative
669 binomial distribution. Genome Biol 15:550.

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

Tables

Table 1. Abundance of NRPS (AD) and PKS (KS) amplicons based on levels of similarity to MIBiG NRPSs and PKSs reference genes, in the tomato and cucumber root and tomato soil samples. Identity groups were determined as unassigned (< 50% amino acid sequence identity), 50-70%, 70-85%, 85-95% 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	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 2. Predicted taxonomy and source dataset of recovered cosmids. UA- unassigned, only affiliated organisms with >80% identity are shown.

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

Figures legend

Figure 1. Principal coordinate analysis (PCoA) of lettuce roots (green circles), tomato roots (red squares) and soil (from tomato lysimeters, brown triangles) 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. Root-enriched NRPS and PKS amplicons in tomato (A) and lettuce (B), relative to bulk tomato soil. Only the top 20 sequences with an adjusted p-value <0.1 are shown. Associated metabolites (based on the MIBiG repository) are presented only for sequences with e-Value<10-40 and >50% AA identity. Identity bars represent amino acid identity to MIBiG reference sequence (0-not identified).

Figure 3. Relative abundance of NRPS (A) and PKS (B) associated metabolites based on MIBiG annotations. Only AD and KS amplicon hits to reference MIBiG sequences over >50% amino acid identity and e-Value<10-40 were included. Counts were normalized and log10-transformed. Only associated metabolites that were present in >2 different samples are shown.

Figure 4. The fifty most abundant NRPSs and PKSs in tomato (A) and lettuce (B) root microbiomes extrapolated from shotgun metagenomic data. 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 indicating that these are also highly expressed. Roman numbers represent predicted associated metabolites for hits with e-Value > 10-40. (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. 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 6. A) Comparison of B481 cluster and the related BGC in *S. cyaneogriseus.* B) Structure prediction for PKS modules present on clone B893. Examples of known polyene antifungal agents are shown below.

Figure 7. A) Relative abundance of clones-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 IMG BlastN feature with eValue<10-5 and identity >85%. Counts were normalized with rpoB and gyrB for each dataset. B) Presence/absence of each recovered gene cluster in the different environment. Only gene clusters where all AD/KS domains were found are considered present.

Figure 1

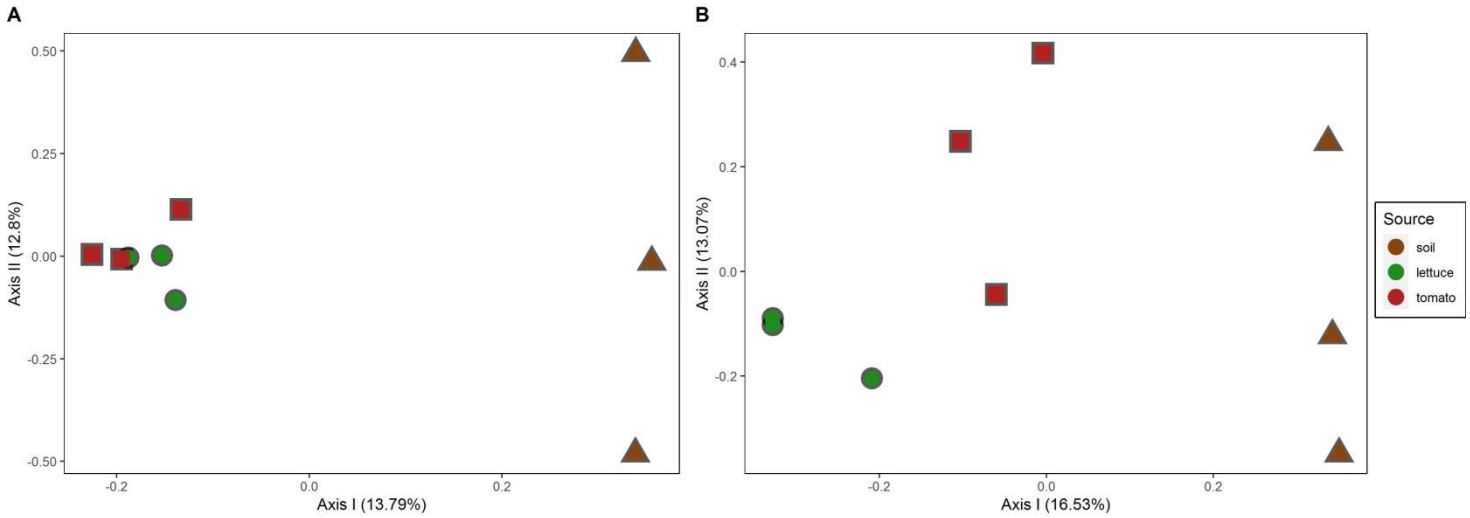


Figure 2

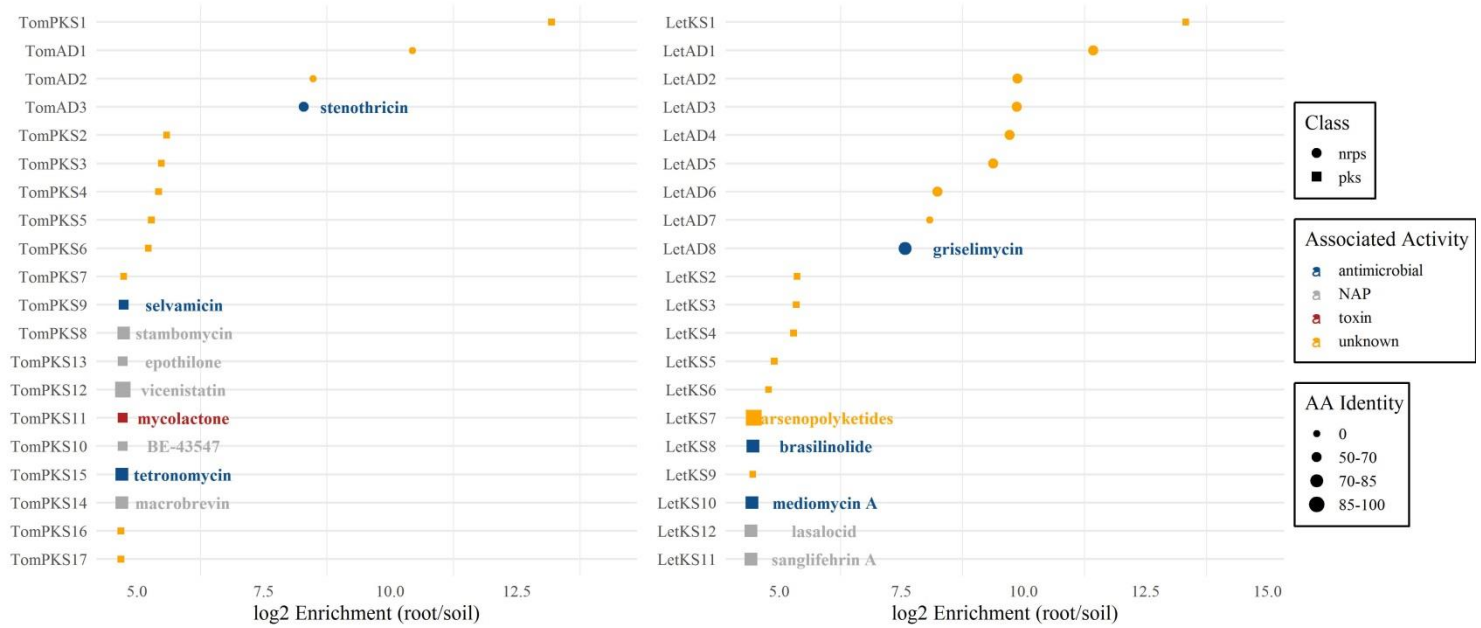
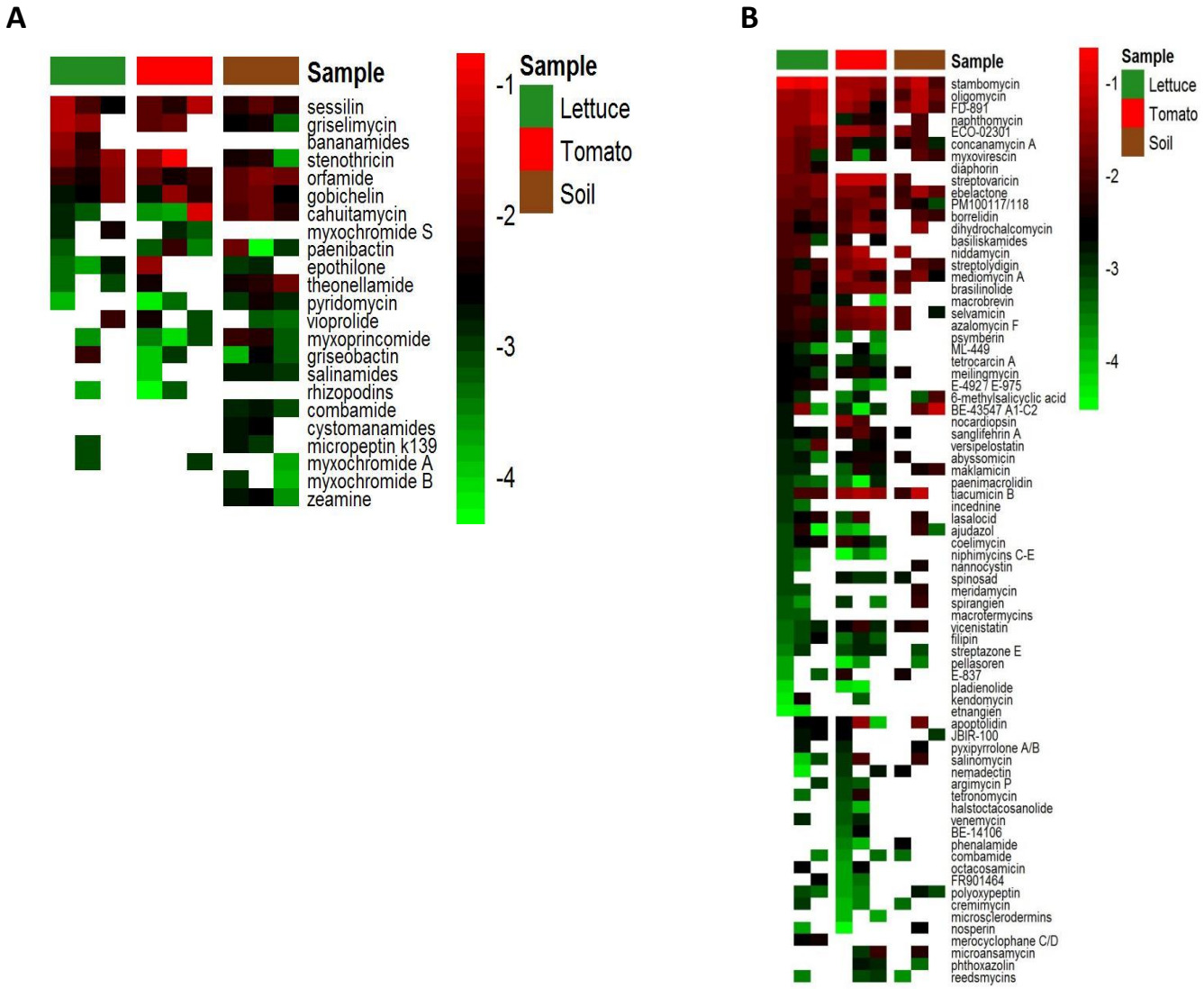
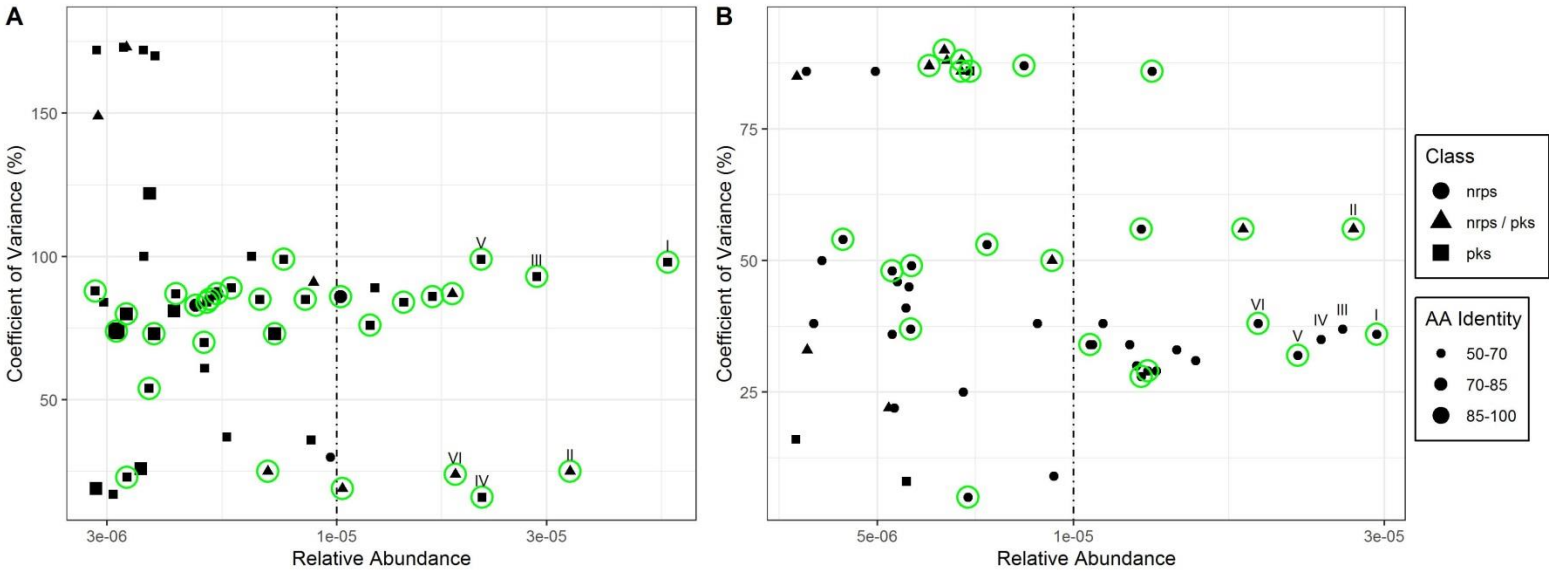


Figure 3

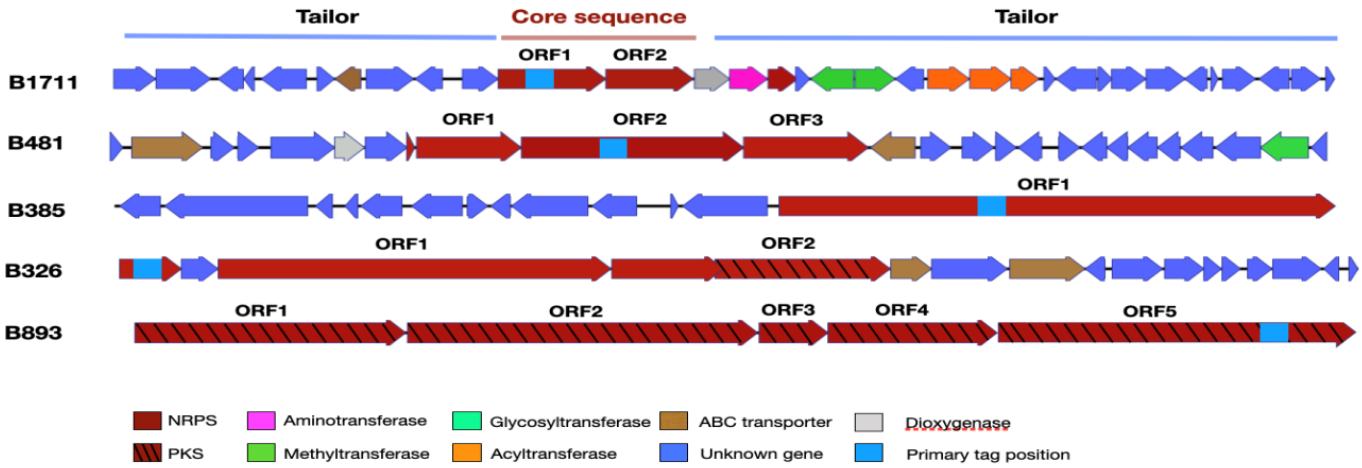

A
Sample
sessilin
griselimycin
bananamides
stenothricin
orfamide
gobichelin
cahuitamycin
myxochromide S
paenibactin
epothilone
theonellamide
pyridomycin
vioprolide
myxoprincomide
griseobactin
salinamides
rhizopodins
combamide
cystomanamides
micropeptin k139
myxochromide A
myxochromide B
zeamine
-1
-2
-3
-4
Sample
Lettuce
Tomato
Soil
B
Sample
stambomycin
oligomycin
FD-891
naphthomycin
ECO-02301
concanamycin A
myxovirescin
diaphorin
streptovaricin
ebelactone
PM100117/118
borrelidin
dihydrochalcomycin
basiliskamides
niddamycin
streptolydigin
mediomycin A
brasilinolide
macrobrevin
selvamicin
azalomycin F
psymberin
ML-449
tetrocarcin A
meilingmycin
E-492 7 E-975
6-methylsalicyclic acid
BE-43547 A1-C2
nocardiopsin
sanglifehrin A
versipelostatin
abyssomicin
maklamicin
paenimacrolidin
tiacumicin B
incednine
lasalocid
ajudazol
coelimycin
niphimycins C-E
nannocystin
spinosad
meridamycin
spirangien
macrotermycins
vicenistatin
filipin
streptazone E
pellasoren
E-837
pladienolide
kendomycin
etnangien
apoptolidin
JBIR-100
pyxipyrrolone A/B
salinomycin
nemadectin
argimycin P
tetronomycin
halstoctacosanolide
venemycin
BE-14106
phenalamide
combamide
octacosamicin
FR901464
polyoxypeptin
cremimycin
microsclerodermins
nosperin
merocyclophane C/D
microansamycin
phthoxazolin
reedsmycins
-1
-2
-3
-4
Sample
Lettuce
Tomato
Soil

752 **Figure 4**



753

754 **Figure 5**



755

756

757

Figure 6

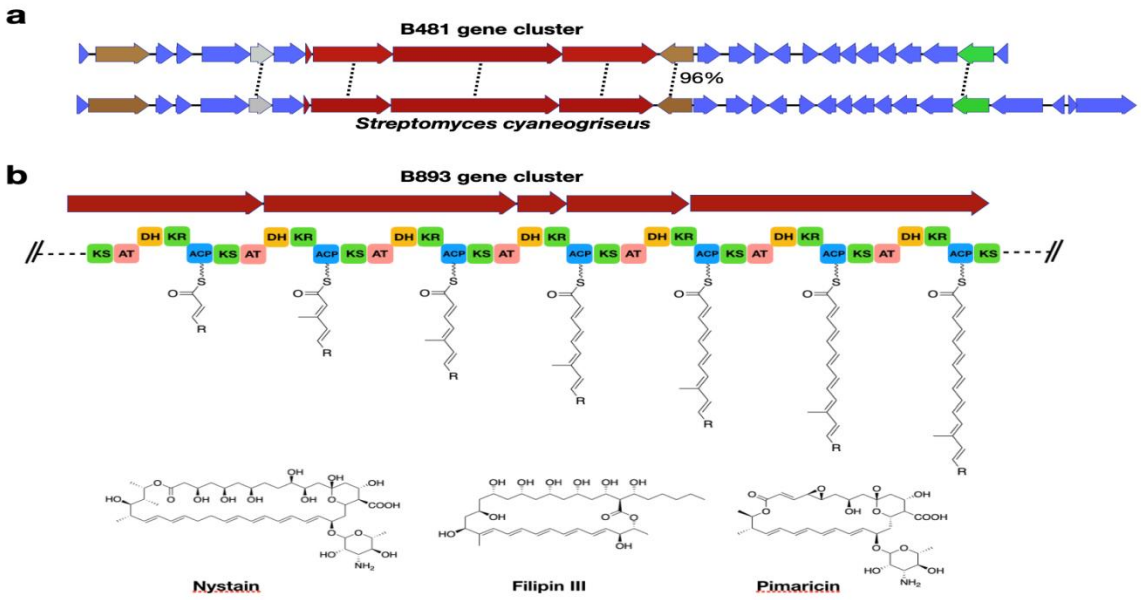
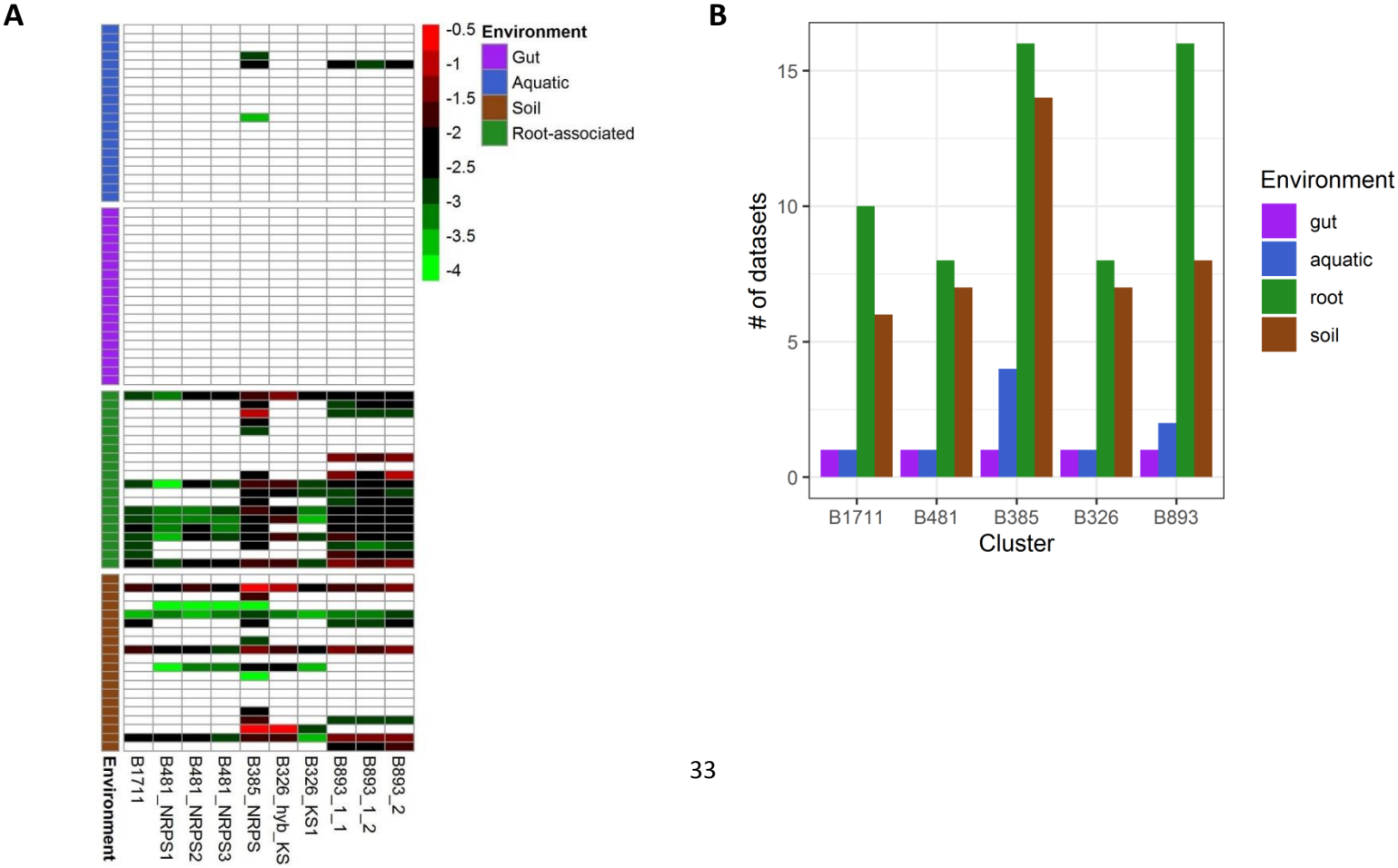



Figure 7