SEL *Soil Ecology Letters*

RESEARCH ARTICLE

Global-scale analysis reveals distinct patterns of non-ribosomal peptide and polyketide synthase encoding genes in root and soil bacterial communities

Barak Dror[1,2], Edouard Jurkevitch[2], Eddie Cytryn[1,*]

1 Institute of Soil, Water and Environmental Sciences, Agricultural Research Organization, Rishon-Lezion, Israel
2 Department of Plant Pathology and Microbiology, The Robert H. Smith Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, Rehovot, Israel

HIGHLIGHTS

• The overall abundance of secondary metabolites-encoding genes in soil and root microbiomes is similar.
• Certain biosynthetic gene clusters (BGCs) are ubiquitous and more abundant in roots compared with soil.
• The majority of identified BGCs are potentially novel.

GRAPHICAL ABSTRACT

A survey of 20 soil and root-associated shotgun metagenomics shows a distinct profile of secondary metabolites encoding genes

ARTICLE INFO
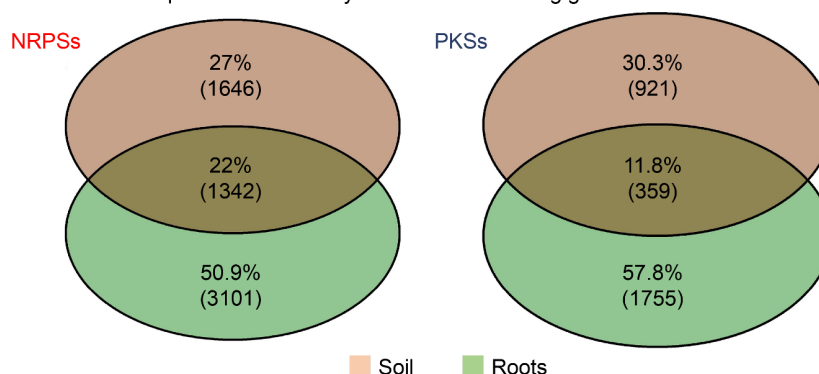
ABSTRACT

Secondary metabolites (SMs) produced by soil bacteria, for instance antimicrobials and siderophores, play a vital role in bacterial adaptation to soil and root ecosystems and can contribute to plant health. Many SMs are non-ribosomal peptides and polyketides, assembled by non-ribosomal peptides synthetase (NRPS) and polyketide synthase (PKS) and encoded by biosynthetic gene clusters (BGCs). Despite their ecological importance, little is known about the occurrence and diversity of NRPs and PKs in soil. We extracted NRPS- and PKS-encoding BGCs from 20 publicly available soil and root-associated metagenomes and annotated them using antiSMASH-DB. We found that the overall abundance of NRPSs and PKSs is similar in both environments, however NRPSs and PKSs were significantly clustered between soil and root samples. Moreover, the majority of identified sequences were unique to either soil- or root-associated datasets and had low identity to known BGCs, suggesting their novelty. Overall, this study illuminates the huge untapped diversity of predicted SMs in soil and root microbiomes, and indicates presence of specific SMs, which may play a role in inter- and intra-bacterial interactions in root ecosystems.

© Higher Education Press 2022

* Corresponding author
E-mail address: eddie@volcani.agri.gov.il (E. Cytryn)

1 Introduction

Plant roots closely interact with the microbial community surrounding them, creating a close and inter-dependent cross-talk (Schlaeppi and Bulgarelli, 2015; Cordovez et al., 2019). Plants secrete different exudates through their root system to 'attract' specific assemblages of microbial communities (Sasse et al., 2018). This 'selection' enriches specific bacterial populations from the surrounding soil, and different plants share a core root microbiome that is distinct from the bulk soil, despite belonging to different botanical families (Levy et al., 2018). Specific root microbial community members play key roles in plant development, growth and health through secretion of enzymes and phytohormones, nutrient provision (e.g., nitrogen and phosphorus) and production of metabolites that help defend plants against micro- and macro-pathogens, through direct antagonism or induced resistance (Tyc et al., 2017; Sharrar et al., 2020).

Polyketides and non-ribosomal peptides are specific secondary metabolite (SM) families synthesized by large enzyme complexes termed non-ribosomal peptide synthases (NRPSs) and polyketide synthetases (PKSs). These enzymes are typically organized on operon-like structures termed biosynthetic gene clusters (BGCs), which employ a modular arrangement composed of repeating domains that assemble the compound in a step-wise manner (Jenke-Kodama et al., 2005; McErlean et al., 2019). Polyketides and non-ribosomal peptides are known for their diverse structure and broad array of functions, which include antimicrobials, pigments and siderophores (Wang et al., 2014). While many of these compounds have been exploited in medicine, industry and agriculture, we now realize that soil microbiomes contain a diverse array of unknown BGCs, and we have only touched the tip of the iceberg regarding their potential ecological function and applicative potential (Dror et al., 2020b; Waschulin et al., 2021).

Most of the research conducted to date on bacterial SM-encoding activity has focused on individual cultured isolates, using customized bioinformatic tools (Eastman et al., 2014; de Bruijn et al., 2015; Nguyen et al., 2018). However, as most soil bacteria cannot be cultivated by existing methods, a huge fraction of bacterial SM remains untapped (Bodor et al., 2020; Dror et al., 2020a). To assist in revealing this 'uncultured fraction', culture-independent methods and pipelines have been developed and applied to mine BGC from environmental microbiomes. These include functional gene amplicon sequencing and extracting SM-encoding genes from shotgun meta–omics data (Amos et al., 2015; Charlop-Powers et al., 2016). By applying these techniques, we recently found that the composition and diversity of NRPS and PKS encoding BGCs are distinct in lettuce vs. tomato roots and that both are significantly different when compared to bulk soil. We also identified specific BGCs ubiquitous to plant roots that may play a role in plant growth and health (Dror et al., 2020b).

This study exploited *in-silico* pipelines previously developed in our lab to investigate NRPS and PKS composition, diversity, taxonomy and potential function in soil and root-associated microbial communities from a broad range of geographic regions and plant types. We pinpointed novel SM families unique to each of these environments, which may play important ecological functions and can have future translational potential.

2 Methods

2.1 Shotgun metagenome datasets

Assembled soil and root-associated shotgun metagenomes were searched and downloaded from the MG-RAST and JGI-IMG repositories using the keywords: 'shotgun metagenome', 'terrestrial biome', 'root' and 'rhizosphere'. The source of each dataset was validated by referring to publications linked to the data. Only well characterized datasets with sufficient metadata and >200 000 contigs were selected for further analyses (Table 1). Soil texture and pH for the selected samples were obtained from the website www.SoilGrid.com, using coordinates provided in the repositories or associated manuscripts.

2.2 NRPSs/PKSs mining and annotation

Raw metagenome FASTA files were filtered to include only contigs longer than 300 bp. These were used as input for ORF prediction with Prodigal, and filtered again for ORFs > 80 bp. To avoid redundancy, CD-HIT was applied (using a >97% identity cutoff), and subsequently, ORFs were randomly selected from each dataset using the seqtk sample command to avoid dataset size bias. NRPS/PKS annotation was performed by aligning sampled ORFs to AD or KS domains (for NRPSs and PKSs, respectively) downloaded from the antiSMASH-DB API platform (Blin et al., 2019) using DIAMOND blastx (>50% identity, >50% query cover, max E value 10^{-10}). Only alignments longer than 70 bp were used for downstream analyses.

2.3 Taxonomy annotation

Annotated AD and KS genes were taxonomically assigned by running Diamond against the non-redundant (nr) Blast NCBI database (08/2016 version). The lowest common ancestor classification was determined using MEGAN V6.21.12 (Huson et al., 2016) by taking the top 20 percent of the hits and filtering for a minimum score of 50 and maximum expect value of 0.0. Genes identifiers were then converted to full taxonomy using the mapping feature provided by MEGAN (11.2018 version). A custom Python code was used to aggregate all genes by environment (soil

Table 1 Datasets included in the study.

Number	Dataset	Environment	Acession No.	Latitude	Longitude	Location	Soil type	pH
1	Tomato	Roots	PRJNA602301	31.606	34.791	Israel	Loamy sand	7.9
2	Lettuce		PRJNA602301	31.606	34.791	Israel	Loamy sand	7.9
3	Wheat		PRJNA664890	31.990	34.820	Israel	Clay	7.8
4	*Arabidopsis*		1119352 (JGI)	35.908	−79.051	North Carolina, USA	Loam	5.5
5	Corn		mgm4504797.3	41.770	−92.961	Iowa, USA	Silt loam	6.3
6	Poplar		3300009147	35.844	−83.961	Tenesee, USA	Silt loam	5.5
7	Switchgrass		mgp6787	42.395	−85.374	Michigan, USA	Sandy loam	6
8	Miscanthus		mgp6787	42.395	−85.374	Michigan, USA	Sandy loam	6
9	Potato		mgm4599960	4.745	−75.440	Colombia	Loam	5.4
10	Paramo		mgm4599954	4.732	−75.734	Colombia	Clay loam	5.5
11	Cucumber		SAMN02202255	31.410	35.167	Israel	Sandy loam	7.79
12	Peat	Soil	3300018086	4.063	−73.195	Colombia	Clay loam	5.3
13	Desert		3300034008	34.300	−117.610	California, USA	Silt loam	7.3
14	Tundra		3300025774	71.299	−156.610	Alaska, USA	Loam	5.5
15	Forest soil		3300024284	40.445	−87.030	Indiana, USA	Loam	6.2
16	Florida		3300033158	26.500	−80.300	Florida, USA	Sandy loam	6.3
17	Prairie		mgm4504798.3	41.771	−92.961	Iowa, USA	Silt loam	6.4
18	Amatitan		3300030511	21.053	−103.902	Mexico	Clay loam	6.8
19	San Felipe		3300014488	21.766	−100.163	Mexico	Sandy loam	6.25
20	Colorado		3300044737	40.368	−106.174	Colorado, USA	Sandy loam	7.5

and roots) and according to different taxonomic levels. These outputs were visualized using R.

2.4 Statistics

Statistical analyses were performed using R. Statistical differences in the presence of AD and KS sequences between studied environments (soil vs. roots) were determined using the Student *t*-test (following Shapiro-Wilk test for normality). Alpha and beta diversities of samples were calculated using the R package vegan (Oksanen et al., 2013), with Aitchison transformation to account for compositional data. To assess differences in diversity between environments, principal coordinate analysis (PCoA) and analysis of similarity (ANOSIM) were performed. Differences were considered statistically significant when P values were lower than 0.05. Python and R scripts can be found on github using the link https://github.com/barakdror/microbial_ecology_R.

3 Results

3.1 Soil and root shotgun metagenome datasets

Over 100 metagenome datasets were initially screened for our analysis, however, following a robust filtering process to prevent methodological biases, only 11 root-associated and 9 soil metagenomes (primarily from North and South America) were selected for *in-silico* analyses (Table 1). The root-associated datasets included a variety of host vegetables (tomato, lettuce, cucumber, and potato), cereals (wheat, corn, switchgrass and miscanthus) and trees (poplar).

3.2 NRPSs and PKSs abundance and presence in rhizospheres and soils

We aligned the predicted ORFs within each dataset against a reference database composed of antiSMASH-generated adenylation (AD) and ketosynthase (KS) domain sequences (markers for NRPSs and PKSs, respectively), and subsequently calculated the percentage of AD and KS genes in each dataset (i.e., number of unique AD/KS genes found divided by the total number of ORFs predicted in the dataset). Generally, we saw a significant positive correlation between the relative abundance of AD and KS domains (Fig. 1A, R=0.85, p-value< 0.01, Pearson correlation). This trend was not related to sample environment (roots vs. soil), as among the top five AD/KS domains-rich datasets, three were root-associated (tomato, wheat and lettuce) and two were soil-associated (Colorado and San Felipe). Moreover, we saw no significant difference in the presence of AD or KS domains in root vs. soil samples when pooled together (Fig. 1B and C, respectively). We also did not find any correlation between AD or KS domain abundance and soil pH or latitude (Figs. S1 and S2).
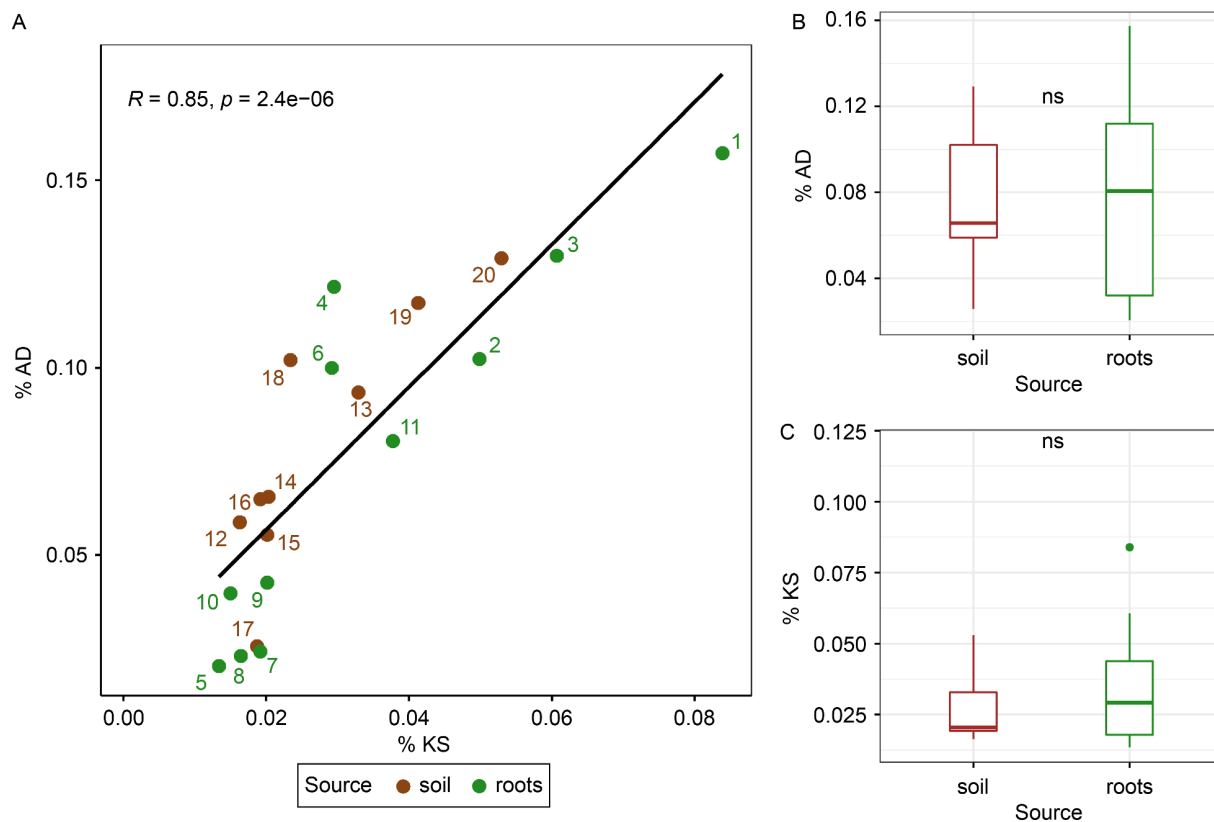
Fig. 1 Pearson correlation of %AD vs. %KS in studied samples (A). Percentage was calculated as number of AD/KS genes/total ORFs in each dataset. Boxplots figures of %AD (B), and %KS (C) of soil and root-associated datasets (NS=non-significant, Student *t*-test).

3.3 Diversity and taxonomy of NRPSs and PKSs in studied samples

No statistical differences were detected in alpha diversity of soil- and root-associated samples for either NRPS or PKS associated genes (Shannon index, Student *t*-test, *p*-value>0.05, Fig. S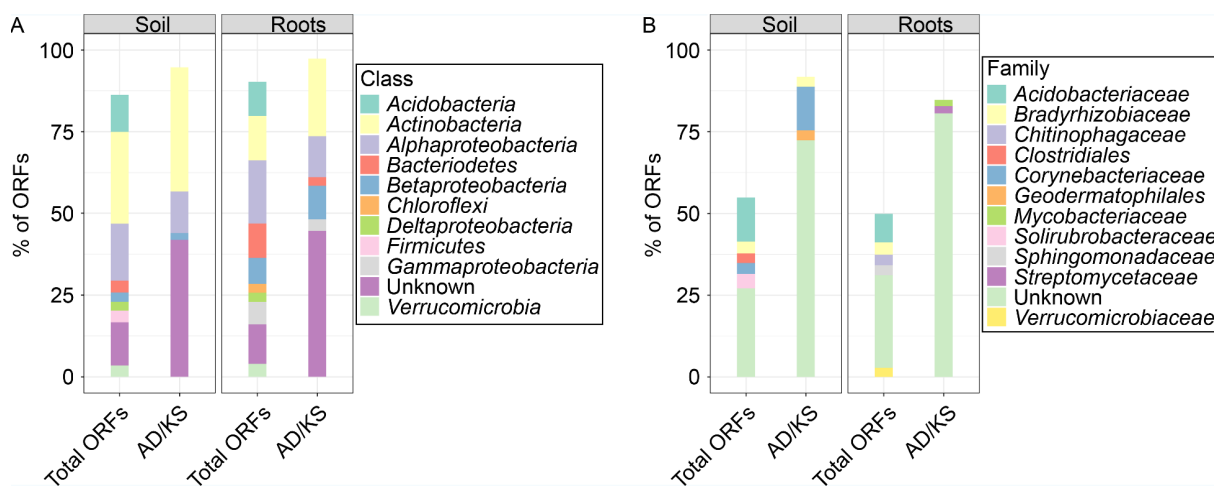3). However, principal coordinate analysis revealed significant differences in the composition of root-associated NRPS and PKS relative to those from bulk soil samples, based on analysis of AD (ANOSIM *R*=0.16, *p*-value<0.05, Fig. 2A) and KS domains (ANOSIM *R*=0.21, *p*-value<0.05, Fig. 2B), respectively.

We assessed whether the observed difference in NRPSs and PKSs composition between soil and roots derives from
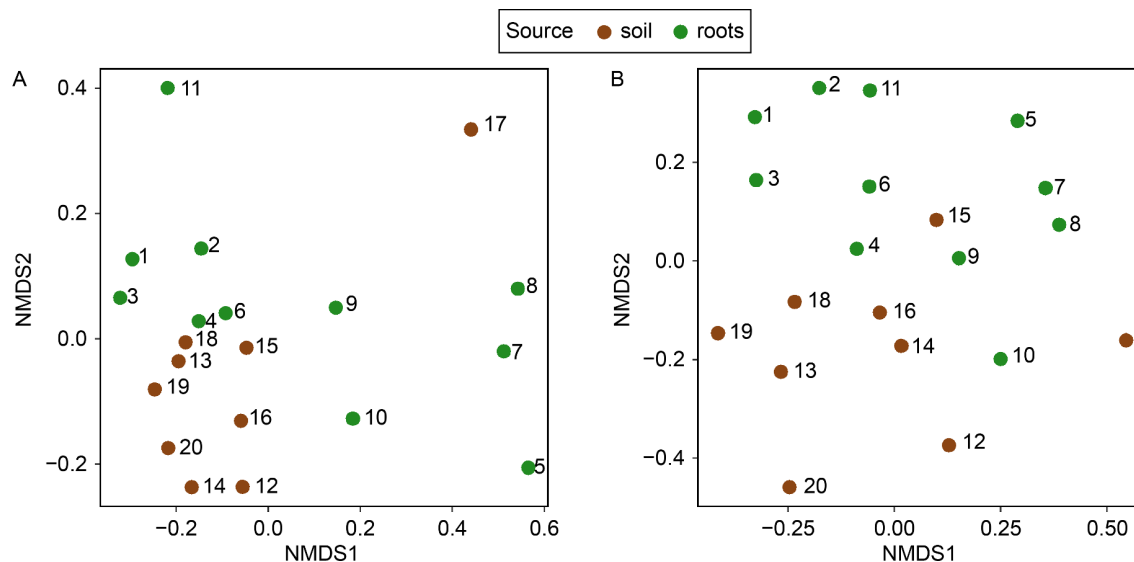

Fig. 2 Nonmetric multidimensional scaling using Aitchison dissimilarity of NRPS (A) and PKS (B) genes found in studied samples. ANOSIM statistics: *R*=0.16, *p*-value <0.05 for NRPS, *R*=0.21, *p*-value<0.001 for PKS, calculated with >1000 permutations.

an inherent difference in the taxonomy of the bacteria harboring these genes. We taxonomically annotated both the total ORFs sampled within each dataset, and the mined NRPS and PKS within these datasets using the NCBI nr database. Resulting annotations were taxonomically assigned using MEGAN and grouped at different taxonomic levels. At the class level the overall bacterial communities were phylogenetically diverse and soil samples were substantially different than roots. However, the taxonomic affiliation of NRPS and PKS was much narrower, and these communities were primarily annotated to *Actinobacteria, Alphaproteobacteria* and *Betaproteobacteria* (Fig. 3A). In addition, ~35%−40% of the NRPS and PKS in both soil and roots datasets could not be assigned to specific bacterial taxa (compared with ~10%−15% of the total soil and root ORFs). A similar trend was observed at the family level (Fig. 3B). Most of the NRPSs and PKSs in soil and root samples could not be taxonomically characterized (~70%−75%), though some root-associated NRPS and PKS were associated with *Streptomycetaceae* and *Mycobacteriaceae*, whereas certain soil NRPS and PKS were associated with *Corynebacteriaceae* and *Bradyrhizobiaceae*.

3.4 Commonality and novelty of NRPSs and PKSs in soil vs. root-associated environments

Analysis of the distribution of extracted NRPSs and PKSs between soil- and root-associated datasets, revealed a relatively low fraction of common SM-encoding genes between environments (Fig. 4). To identify potentially novel NRPSs and PKSs, we aggregated the soil and root associated AD and KS domains, and divided them into three groups (70%−85%, 85%−97% and >97%), based on identity of amino-acids to antiSMASH-DB reference sequences. Overall, most of the AD and KS hits in both soils and roots were classified as potentially novel (70%−85% identity, Fig. 5), whereas a very small fraction (less than 4%) were classified as known (i.e., identified in previous studies, >97% identity to reference sequences).

4 Discussion

The rhizosphere microbiome is a subset of the surrounding soil microbial community, however, the organic oasis created by root exudates and mucilage facilitate proliferation

Fig. 3 Class-level (A) and family-level (B) taxonomy of total ORFs vs. NRPSs and PKSs found in soil- and root-associated datasets, based on shotgun metagenome predicted ORFs. Only classes with >2% and families with >1.5% relative abundance are shown.
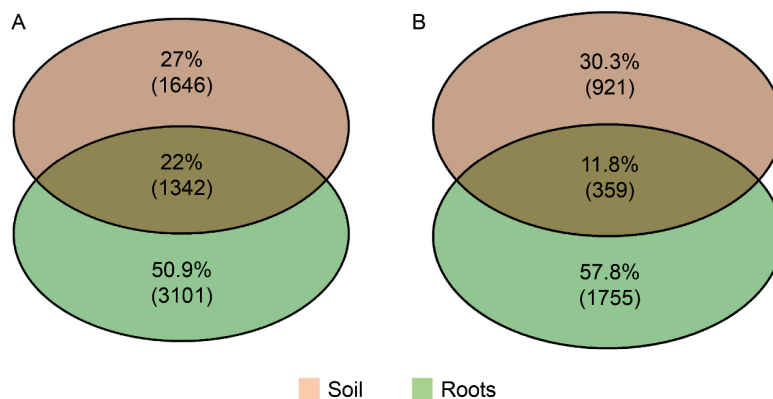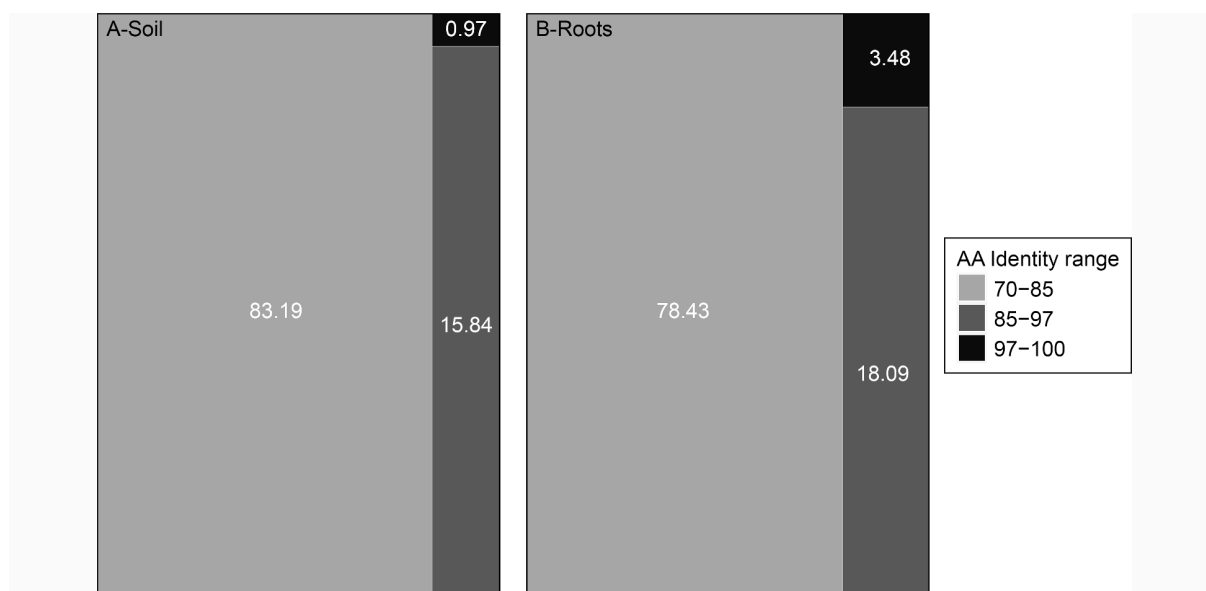

Fig. 4 A Venn diagram of extracted NRPSs (A) and PKSs (B) sequences found in soil- and root-associated datasets (following a random rarefaction of total ORFs within each dataset to avoid datasets uneven sampling bias).

Fig. 5 Amino acid identity of NRPSs and PKSs ORFs annotated using the antiSMASH-DB database for soil-associate (A) and root datasets (B) clustered into three groups based on %similarity. NRPSs and PKSs were pooled together by environment.

of specific bulk soil populations, whose composition are dictated by specific plant species (Peiffer et al., 2013; Kawasaki et al., 2016; Qiao et al., 2017). Secondary metabolites (SMs) are believed to play a critical role in both soil and root-associated microbiomes (Tyc et al., 2017), but little is known about the abundance and potential function of SMs in these environments and their distribution in soil vs. root ecosystems. In this study, we explored and characterized soil- and root-associated SM-encoding genes specifically focusing on root-associated SMs with potential impact on host plants.

The composition of bacterial NRPS and PKS differed between soil and roots, implying a distinct SM-encoding capacity within each environment. While the observed pattern may stem from issues related to the complexity of the soil versus the root microbiome and consequently from sampling depth as the rhizosphere microbiome is a derivative of the larger soil microbiome pool, our findings point toward the selective power the host plant may have on the surrounding bacterial community. These results coincide with previous studies that found the rhizosphere microbiome to be enriched in genes related to ecosystem-specific functional processes such as transport, primary metabolism (Yan et al., 2017; Lee et al., 2019), nitrogen cycling (Schmidt et al., 2019), motility, chemotaxis and plant cell-wall degradation (Ofek-Lalzar et al., 2014).

We also found soil and root microbial communities are rather similar in their potential to encode for NRPS and PKS, regardless of soil pH. However, there was a clear and significant positive correlation between NRPS and PKS ratios in individual samples. This can reflect the presence of SM-rich bacterial taxonomic families within some samples, and their absence in others. This is supported by previous studies that found soil bacterial structure and functional traits to differ between biomes (Fierer et al., 2009; Noronha

et al., 2017). We found that in root-associated samples, NRPSs and PKSs were commonly associated with *Actinobacteria*, *Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria* and *Bacteroidetes,* whereas in soil they were exclusively linked to a higher share of *Actinobacteria* and *Alphaproteobacteria*. The fact most of the mined NRPSs and PKSs had low identity (70%−85%) when compared to the antiSMASH-DB database, emphasize the huge untapped diversity of these SM families.

Comparative analysis of environmental shotgun metagenomic data such as the one presented here, have the potential to link functional genes and processes to specific ecosystems, and therefore, future research should complement this study and expand the number of datasets used, to quantitatively determine the abundance of selected NRPSs and PKSs in soil vs. root environments. However, as previously discussed (Eckert et al., 2020; Vangay et al., 2021), a large fraction of environmental shotgun metagenome data is not deposited into public repositories, and those that are frequently lack details and metadata that is crucial for enabling robust comparative analyses. There is a vital need for repositories to draft submission standards or adopt previously established guidelines (e.g., MIxS) (Yilmaz et al., 2011) that will ensure a minimal set of background data for submission of environmental shotgun metagenome.

5 Conclusions

This study found soil and root microbial communities to differ in their SM-encoding genes at a global scale, regardless of plant host and soil properties. We found this distinction is likely derived from specific taxonomies of the bacterial strains residing in each environment. Given the immense amount of potentially novel sequences that likely

encode metabolites with unknown function, future research will aim to pinpoint, characterize, and express root-specific BGCs in order to elucidate their *in-vitro* and *in-planta* function. This process also needs to be conducted for BGC that are profuse in soil and root environments that encode for metabolites that are currently only known for clinical activity.

Electronic supplementary material

References

Amos, G.C.A., Borsetto, C., Laskaris, P., Krsek, M., Berry, A.E., Newsham, K.K., Calvo-Bado, L., Pearce, D.A., Vallin, C., Wellington, E.M.H., 2015. Designing and implementing an assay for the detection of rare and divergent NRPS and PKS clones in European, Antarctic and Cuban soils. PLoS ONE 10, 1–15.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. AntiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. Nucleic Acids Research 47, W81–W87.

Bodor, A., Bounedjoum, N., Vincze, G.E., Kis, Á.E., Laczi, K., Bende, G., Szilágyi, Á., Kovács, T., Perei, K., Rákhely, G., 2020. Challenges of unculturable bacteria: environmental perspectives. Reviews in Environmental Science and Bio/Technology 19, 1–22.

de Bruijn, I., Cheng, X., de Jager, V., Expósito, R.G., Watrous, J., Patel, N., Postma, J., Dorrestein, P.C., Kobayashi, D., Raaijmakers, J.M., 2015. Comparative genomics and metabolic profiling of the genus *Lysobacter*. BMC Genomics 16, 1–16.

Charlop-Powers, Z., Pregitzer, C.C., Lemetre, C., Ternei, M.A., Maniko, J., Hover, B.M., Calle, P.Y., McGuire, K.L., Garbarino, J., Forgione, H.M., Charlop-Powers, S., Brady, S.F., 2016. Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. Proceedings of the National Academy of Sciences of the United States of America 113, 14811–14816.

Cordovez, V., Dini-Andreote, F., Carrión, V.J., Raaijmakers, J.M., 2019. Ecology and evolution of plant microbiomes. Annual review of Microbiology 73, 69–88.

Dror, B., Jurkevitch, E., Cytryn, E., 2020a. State-of-the-art methodologies to identify antimicrobial secondary metabolites in soil bacterial communities−A review. Soil Biology and Biochemistry 147, 107838.

Dror, B, Wang, Z., Brady, S.F., Jurkevitch, E., Cytryn, E., 2020b. Elucidating the diversity and potential function of nonribosomal peptide and polyketide biosynthetic gene clusters in the root microbiome. mSystems 5, e00866–20.

Eastman, A.W., Heinrichs, D.E., Yuan, Z.C., 2014. Comparative and genetic analysis of the four sequenced Paenibacillus polymyxa genomes reveals a diverse metabolism and conservation of genes relevant to plant-growth promotion and competitiveness. BMC Genomics 15, 851.

Eckert, E.M., Di Cesare, A., Fontaneto, D., Berendonk, T.U., Bürgmann, H., Cytryn, E., Fatta-Kassinos, D., Franzetti, A., Larsson, D.G.J., Manaia, C.M., 2020. Every fifth published metagenome is not available to science. PLoS Biology 18, e3000698.

Fierer, N., Stricklandd, S., Liptzind, D., Bradford, A., Cleveland, C., 2009. Global patterns in belowground communities. Ecology Letters 12, 1–12.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J., Tappu, R., 2016. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Computational Biology 12, e1004957.

Jenke-Kodama, H., Sandmann, A., Müller, R., Dittmann, E., 2005. Evolutionary implications of bacterial polyketide synthases. Molecular Biology and Evolution 22, 2027–2039.

Kawasaki, A., Donn, S., Ryan, P.R., Mathesius, U., Devilla, R., Jones, A., Watt, M., 2016. Microbiome and exudates of the root and rhizosphere of Brachypodium distachyon, a model for wheat. PloS One 11, e0164533.

Lee, S.A., Kim, Y., Kim, J.M., Chu, B., Joa, J.H., Sang, M.K., Song, J., Weon, H.Y., 2019. A preliminary examination of bacterial, archaeal, and fungal communities inhabiting different rhizocompartments of tomato plants under real-world environments. Scientific Reports 9, 1–15.

Levy, A., Conway, J.M., Dangl, J.L., Woyke, T., 2018. Elucidating bacterial gene functions in the plant microbiome. Cell Host & Microbe 24, 475–485.

McErlean, M., Overbay, J., Van Lanen, S., 2019. Refining and expanding nonribosomal peptide synthetase function and mechanism. Journal of Industrial Microbiology and Biotechnology 46, 493–513.

Nguyen, T.M., Seo, C., Ji, M., Paik, M.J., Myung, S.W., Kim, J., 2018. Effective soil extraction method for cultivating previously uncultured soil bacteria. Applied and Environmental Microbiology 84, e01145–18.

Noronha, M.F., Lacerda Júnior, G.V., Gilbert, J.A., de Oliveira, V.M., 2017. Taxonomic and functional patterns across soil microbial communities of global biomes. Science of The Total Environment 609, 1064–1074.

Ofek-Lalzar, M., Sela, N., Goldman-Voronov, M., Green, S.J., Hadar, Y., Minz, D., 2014. Niche and host-associated functional signatures of the root surface microbiome. Nature Communications 5, 1–9.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H., 2013. Package 'vegan'. Community ecology package, version 2, 1–295.

Peiffer, J.A., Spor, A., Koren, O., Jin, Z., Tringe, S.G., Dangl, J.L., Buckler, E.S., Ley, R.E., 2013. Diversity and heritability of the maize rhizosphere microbiome under field conditions. Proceedings of the National Academy of Sciences 110, 6548–6553.

Qiao, Q., Wang, F., Zhang, J., Chen, Y., Zhang, C., Liu, G., Zhang, H., Ma, C., Zhang, J., 2017. The variation in the rhizosphere

microbiome of cotton with soil type, genotype and developmental stage. Scientific Reports 7, 1–10.

Sasse, J., Martinoia, E., Northen, T., 2018. Feed your friends: do plant exudates shape the root microbiome? Trends in Plant Science 23, 25–41.

Schlaeppi, K., Bulgarelli, D., 2015. The plant microbiome at work. Molecular Plant-microbe Interactions 28, 212–217.

Schmidt, J.E., Kent, A.D., Brisson, V.L., Gaudin, A.C.M., 2019. Agricultural management and plant selection interactively affect rhizosphere microbial community structure and nitrogen cycling. Microbiome 7: 1–18.

Sharrar, A.M., Crits-Christoph, A., Méheust, R., Diamond, S., Starr, E.P., Banfield, J.F., 2020. Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. Mbio 11, e00416–20.

Tyc, O., Song, C., Dickschat, J.S., Vos, M., Garbeva, P., 2017. The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. Trends in Microbiology 25, 280–292.

Vangay, P., Burgin, J., Johnston, A., Beck, K.L., Berrios, D.C., Blumberg, K., Canon, S., Chain, P., Chandonia, J.M., Christianson, D., 2021. Microbiome metadata standards: report of the National Microbiome Data Collaborative's Workshop and follow-on activities. mSystems 6, e01194–20.

Wang, H., Fewer, D.P., 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.

Waschulin, V., Borsetto, C., James, R., Newsham, K.K., Donadio, S., Corre, C., Wellington, E., 2021. Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. The ISME Journal 16, 1–11.

Yan, Y., Kuramae, E.E., De Hollander, M., Klinkhamer, P.G., Van Veen, J.A., 2017. Functional traits dominate the diversity-related selection of bacterial communities in the rhizosphere. The ISME Journal 11, 56–66.