2019

Linking Microbial Community Structure and Ecosystem Functions in Acidic Soil from Pennsylvania, USA

Madison T. James

Sophia T. Farrisi

Shreya Shah

Vishal Shah

Linking Microbial Community Structure and Ecosystem Functions in Acidic Soil from Pennsylvania, USA

Madison T. James, Sophia T. Farrisi, Shreya Shah, Vishal Shah*

West Chester University, West Chester, PA, USA

RUNNING TITLE: Soil microbial community structure and function

*Corresponding Author

Phone: 610-436-3522

Email: vshah@wcupa.edu

Abstract: Microorganisms play a critical role in the structure and functioning of soil ecosystems. Within acidic soil across the northeastern United States and Canada, we have little understanding of the microbial diversity present and its relationship to the biochemical cycles. The current study is aimed at understanding the taxonomical and functional diversities in the acidic soil obtained from near various types of trees, how the diversities change as a function of depth, and the linkage between taxonomical and functional diversities. From eight sampling locations, soil samples were collected from three horizons (depths). The three depths were 0-10 cm (A), 11-25 cm (B), and 26-40 cm (C). Results indicate that across all the samples analyzed, *Bradyrhizobium* and *Candidatus Solibacter* are the most abundant bacteria in the soil microbiome. The differences in the soil microbiome across the samples were attributed to the abundance of individual organism's present in the soil and not to the presence or absence of individual organisms. Subsystem level analysis of the soil microbiome sequences indicate that there is higher level of abundance of genes attributed to regulation and cell signaling. A low level of sequences were detected for sulfur metabolism, potassium metabolism, iron acquisition and metabolism, and phosphorous metabolism. Structure-functional analysis indicate that *Bradyrhizobium*, *Rhodopseudomonas*, and *Burkholderia* are the major organisms involved in the nutritional ecosystem functioning within acidic soil. Based on the results, we propose utilizing a consortium of these organisms as an environmentally friendly alternative to the use of chemicals to maintain soil fertility and ecosystem functioning.

Keywords: Acidic soil, microbial ecology, shotgun metagenome, microbiome, Bradyrhizobium, Rhodopseudomonas, Burkholderia.

1.0. INTRODUCTION

Soil microorganisms are the largest biodiversity pool on earth, with more than 10^{30} microbial cells, $10^4 – 10^6$ species, and nearly 1000 Gbp of microbial genome per gram of soil (Vogel et al. 2009; Mendes and Tsai, 2018). They are the primary factors that affect the soil ecosystem functioning and play key roles in forming and maintaining a multitude of soil characteristics including integrity, fertility, ecology, and overall soil function (Shah et al. 2011). Soil microorganisms are also vital for decomposition, pollutant removal, recycling of essential elements, suppressing plant diseases found in soil, and promoting growth for vegetation (Garbeva, 2004). Much is known of the microbial taxa present in soils from across the planet and the impact of perturbation of soil conditions. A Google Scholar search for the term 'soil microbial diversity' reveals over 1.6 million hits. Nevertheless, our understanding of how microbial diversity and ecosystem functions are linked, and how each of the microbial taxa present in the soil are linked to the individual ecological functions remain limited.

Increased use of 16s rDNA metagenomic methodology using pyrosequencing and Illumina Miseq and Hiseq techniques, has increased our understanding of the taxonomy of soil microorganisms by orders of magnitude. However, the 16S rDNA sequencing method has numerous limitations including differentiating closely related species (Hasan et al. 2014), non-uniform distribution of sequence dissimilarity among taxa, presence of multiple copies of the 16S rRNA gene (Garrity et al. 2009), failure of target amplification of polymerase chain reaction (PCR) primers (Venter et al. 2004), and generation of chimeric sequences (Quince et al. 2009). Further, in majority of these research, the role of individual microorganisms in the soil remains at the level of hypothesis based on prior literature (examples include our own prior research: Kumar et al. 2011; Collins et al. 2012). Methods such as Biolog, Fungilog, and soil enzyme activity are many

68 times used in studies as indicators of the ecosystem functioning and correlation to the taxonomic
69 data (Rutgers et al. 2016; Sobek et al. 2003; Nannipieri et al. 2002). While a step forward, these
70 methods are primarily predictive of soil microbial functional dynamics (Bell et al. 2009).

71 Whole genome shotgun metagenomics provide a better approach for obtaining the
72 taxonomic and functional aspects of the entire soil microbial genome. This method yields millions
73 to billions of short reads, providing necessary sequencing depth as needed. It also offers an
74 opportunity to identify organisms present in the microbiome and the biochemical pathway
75 information present at the genomic level in each of the identified organisms. In this study, we
76 employ shotgun metagenomic approach to identify and quantitate bacterial species present in the
77 acidic soil and elucidate the major ecological functions of major organisms.

78 Acidic soil typically has a pH range of 4.0-4.5, is high in iron and aluminum, and is often
79 considered nutrient-poor. Across the eastern United States and southeastern Canada, soil is
80 primarily acidic (Bruulsema 2006). The acidic conditions in the soil of the region is primarily
81 attributed to the parent materials of the soil and increased precipitation that leaches cations from
82 the soil. The soil is optimal for the growth of trees like Apple, Beech, Dogwood, Oak, and
83 Magnolia, and Pears. Literature search indicates that no reports are available studying the structure
84 – function relationship in the natural acidic soil from the region.

85 The current study focused on understanding the taxonomical and functional diversities in
86 the acidic soil obtained from near various types of trees, how the diversities change as a function
87 of depth, and the linkage between taxonomical and functional diversities. We address three
88 questions in the study: (1) what are the major microorganisms that are common to all soil types

and depth? (2) what are the biochemical pathways that can be generalized across all the soil types and depth? (3) What role do each of these organisms play in the ecosystem functioning in the soil?

2.0. METHODOLOGY

2.1. Sample Collection: Samples were collected initially from eight sampling locations across the West Chester University Campus in West Chester, PA. No permit was required to obtain samples. Table 1 describes the analyzed eight locations, the types of vegetation present at each sampling location, as well as sampling coordinates. The locations were selected based on the vegetation present and initial sampling of 25 different locations. The final locations were selected based on the similarity of the vegetation between sampling locations, pH levels, and the quantity/quality of the DNA isolated. The protocols for the safety of data collection were strictly followed as recommended by the U.S. Fish and Wildlife Services and the Foundation for Ecological Research in the Northeast (Batcher, 2005). At each location, soil samples were collected from three horizons: 0–5 cm (Horizon A); 6-15 cm (Horizon B) and 16–30 cm (Horizon C).

2.2. pH Measurement: 5 g soil samples were mixed with 10 mL d/w and vortexed for 10 minutes. The solution was allowed to sit for 1 hour and pH measured of the settled solution. All measurements were done in triplicates.

2.3. DNA Extraction and Shotgun Metagenomics: DNA extraction from each soil sample was carried out using the Qiagen DNeasy PowerSoil DNA Isolation Kit, according to the manufacturer's protocol. DNA concentration in all samples was determined using the Qubit 3 Fluorometer (Invitrogen Technologies). All the samples were diluted to 100ng/µl and used for the library preparation, using the Nextera Dna Flex Library Preparation kit (Illumina, San Diego, CA), according to the manufacturer's protocol. Cluster generation and sequencing were performed with

the MiSeq Reagent Kit v2 500-cycles Paired-End in a MiSeq instrument at West Chester University. Samples were sequenced in a batch of 24 samples on a single flow cell. DNA sequences were annotated with Metagenomics Rapid Annotation (MG-RAST) pipeline version 4.0 for downstream analyses. Taxonomic and functional profiles were generated using the normalized abundance of sequence matches to the Refseq and Subsystems databases, respectively. All settings were set at default values prior to analysis. The sequences have been deposited and are available through the NCBI BioProject Database ID: PRJNA 719140.

2.4. Clustering analysis:

Clustering analysis were performed using Statistica (release 14.0) software. The tree cluster analysis was performed using Ward's method as the amalgamation rule and the distance measured as the Euclidean distances. Prior to clustering analysis, data obtained from MG-RAST were log2 transformed and DSeq normalized.

RESULTS

3.1. Soil pH: pH for all the soil samples analyzed in this study were in the acidic range of 4.1 to 6.3 (Table 1). Results show that the type of tree clearly influences the soil pH, with soil around Douglas Fir being the most acidic soil amongst all the types studied. No significant difference in pH was observed across the depths, except for soil obtained around the Oak tree (Table 1), where a stark drop in pH was observed as we go from depth A (pH, 5.7) to depth C (pH, 4.4).

3.2. Sequencing analyses and microbial community diversity: A total of 22,745,412 raw sequence reads were generated for the 24 samples using the Illumina Miseq sequencing platform. 96.2% of the sequences passed the Illumina Chasity Filter for a total of 21,880,208 PF reads. MG-RAST analysis of the submitted reads yielded 8,219,706 total sequences (Table S1). Over 99% of

the sequences were annotated, with almost equal distribution of known proteins (4,073,051) and proteins of unknown function (4,073,868) (Table S1).

The rarefaction curves indicate high genetic diversity, with no complete saturation observed even after almost 8 million sequences (Figure S1). For all the samples, the curve has slowly begun to flatten, indicating a reasonable number of species have been sampled. The mean alpha diversity observed was 479, with the range from 417 to 547 species (Table S1).

Taxonomic characterization of soil microbiome: Taxonomically, all soil samples had bacterial populations from 50-57 phylum. Bacteria belonging to Proteobacteria and Actinobacteria were the most predominant bacteria, comprising over 60% of the total microbial community in each of the samples analyzed (Figure 1).

Hierarchical structure analysis was performed on the normalized genus level abundance data using Ward's linkage method to investigate the link between soil microbiota and plant type/soil pH and depth. Results indicate that the soil samples analyzed can be divided into six major clusters, after which the linkage distance separating the sub-clusters is small (Figure 2). Table 2 describes the members of each cluster and K-means clustering confirms the results. While overall the samples from individual locations from each of the horizon are clustered together or are in close clusters, samples from horizon B of location 10 (Pine tree vegetation) and horizon A of location 12 (Tulip tree) have unique microbiota to form its own cluster. ANOVA analysis indicates that the mean abundance for all the genera within a cluster are statistically different between clusters ($p < 0.05$), except for the abundance of seven genera (Table S2). The seven genera whose abundance are not statistically different between clusters ($p > 0.05$) are *Nitrosopumilus*, *Carboxydothermus*, unclassified genera derived from Deltaproteobacteria, *Pelotomaculum*, *Oceanicola*, *Thermotoga*, and *Bdellovibrio* (Table S2).

157 Table 3 shows the abundance of the top 30 microbial genera in the representative samples
158 from each of the clusters and the average abundance of the organisms across all the 24 samples
159 analyzed. Results show that *Bradyrhizobium* and *Candidatus Solibacter*, both Gram-negative
160 bacteria, are the most abundant microorganism in the soil samples analyzed. *Streptomyces* and
161 *Mycobacterium* are the two most abundant Gram-positive bacteria found in the soil samples.
162 Figure 3 provides an overview of the total microbial community present in the soil sample obtained
163 from the top horizon near Oak tree (sample 1A). As described earlier, the primary difference in
164 microbial community in all the samples is the % abundance of individual organisms.

165 **Functional characterization of soil microbiome**

166 A heat map illustrating the functional annotation of sequence reads containing predicated proteins
167 of known functions across all the 24 soil samples is shown in Figure 4. Variation was observed
168 between samples primarily related to proteins involved in virulence, disease and defense; cell wall
169 and capsule; membrane transport; DNA metabolism; and respiration. Among the functional
170 categories identified by MG-RAST, the five most dominant categories based on the relative
171 abundance of assigned reads were carbohydrates (13.3±0.4%), the clustering-based subsystems
172 (functional coupling evidence but unknown function; 12.9±0.2%), amino acids and derivatives
173 (9.6±0.3%), miscellaneous (6.8±0.2%), and protein metabolism (7.7±0.3%).

174 Relative abundance of the predicated proteins annotated at subsystem level 2 for each of
175 the soil samples is presented in the supplemental table (Table S3). Hierarchical structure analysis
176 was performed on the normalized values, similar to that performed for taxonomic data. Results
177 indicate that the soil samples can be divided into 5 clusters, after which the linkage distance
178 separating the sub-clusters is small (Figure 5). Table 4 describes the members of each cluster and
179 K-means clustering confirms the results. Similar to taxonomic clustering, samples from horizon

B of location 10 (Pine tree vegetation) and horizon A of location 12 (Tulip tree) have unique composition of functional proteins to form its own cluster.

Table 5 shows the abundance of top 30 predicated proteins in the representative samples from each of the clusters. Results show that unidentified proteins involved in regulation and cell signaling comprise nearly 1 in 5 proteins predicated from the sequences. Nearly 6% of the predicated proteins are from the miscellaneous SEED category comprising a diverse set of genes identified during investigation of plant-prokaryote interactions by a project at the Department of Energy (DOE), USA (Thureborn et al. 2016). Protein biosynthesis, central carbohydrate metabolism, and resistance to antibiotics and toxic compounds were the other top predicated functions of the proteins. Figure 6 provides an overview of the functional hierarchical profiles of the microbiome from the top horizon soil obtained near Oak tree (sample 1A) analyzed at subsystem level 3.

Linking diversity to function

To identify the key microorganisms playing significant role in the biochemistry of soil, Refseq and Subsystems analysis were performed together on MGRAST platform. The Subsystem analysis was performed at level 3 wherever possible. Top 5 genera having the largest quantity of annotated reads within each of the metabolic class were identified (Table S4). Data indicates that *Bradyrhizobium*, *Rhodopseudomonas*, and *Burkholderia* are the key bacteria within the soil microbiota. Both, *Bradyrhizobium* and *Rhodopseudomonas* are top contributors in 24 of the 44 metabolic classes analyzed (Figure 7). *Burkholderia* is a top organism in 16 of the metabolic classes (Figure 7).

From an agricultural perspective, Nitrogen, Phosophorus, Sulphur and Iron metabolic pathways are significant. In the Nitrogen, Iron and Sulphur pathways, beyond the three genera identified, *Mycobacteria* also plays a significant role. Organisms from *Anaeromyxobacter* and *Aromatoleum* genera are key contributors in the nitrosative stress and dissimilatory nitrile reductase pathways respectively (Table S4). Organisms from *Sorangium* genera have the most genes coding for Sulphate reduction associated complexes. Similarly, organisms from *Cupriavidus* and *Pseudomonas* genera are other top bacteria involved in Phosphate pathways (Table S4). In Iron pathways, *Bacillus*, *Frankia*, and *Pseudomonas* were the top genera involved (Table S4). The catabolic genes related to the degradation of xenobiotics were also annotated and linked to the microbial genera. Beyond *Bradyrhizobium*, *Rhodopseudomonas,* and *Burkholderia,* bacteria from *Pseudomonas* and *Cupriavidus* play a key role in degradation of xenobiotic compounds (Table S4). Results indicate that for each of the biochemical functions, there is redundancy within the soil microbiome.

DISCUSSION:

We investigated the microbial structural and functional diversity within the top acidic soil associated with a wide variety of plants. Results indicate that irrespective of the level of acidity in the soil, most of the microorganisms associated with the soil generally remains the same. The differences observed between soil samples, could be attributed to the abundance of individual organism's present in the soil based on the soil chemistry and the vegetation present. The change in microbial abundance results in change in the abundance of functional genes within the soil microbiome. Literature is replete with scientific studies showing soil microbiome changes with the structure of the soil (eg. Fierer and Jackson, 2006; Fierer et al. 2012; Mendes and Tsai, 2018; Shah et al. 2021). Based on our results, we suggest that one needs to consider whether the type of

organisms present in the soil are different or if the abundance of individual organisms is different before reaching the conclusions related to microbiome difference amongst different soil samples. Further, current methods of calculating alpha and beta diversity may not capture the true similarities in the microbiome from different soil types. As further advances are made in the next-generation sequencing techniques, we believe similarities in the microbiome across soil type could become more evident.

Taxonomically, prior research has shown that Gram-negative organisms are predominant organisms present in the soil (Shah and Subramaniam, 2018). Results obtained in the current study supports the prior observation. When one considers similar observations in microbiome studies conducted in marine environments, and even in human, fish and animals, a theme starts to emerge - in the microbial communities across the matrices, Gram-negative bacteria are the predominant organisms.

Functionally, high levels of genes attributed to regulation and cell signaling (level 1) appear to be an identifying indicator for acidic soils. cAMP is a major gene annotated to this category. Delmont et al. (2012) reported abundance of cAMP related annotation within the soil metagenome. Considering acidic soil is poor in nutrition and the northeast region of the United States has varying weather patterns, soil bacteria might be required to deal with constantly fluctuating substrates and environmental conditions. cAMP is a universal cell energy and metabolism regulator. Higher level of this and other genes involved in regulation and cell signaling can be attributed to the requirement of bacteria to adapt to the changing soil chemistry. Surprisingly, we noticed low levels of the abundance of genes related to nutrient cycling (sulfur metabolism, potassium metabolism, iron acquisition and metabolism, and phosphorous metabolism). Genes annotated to virulence disease and defense were significantly prevalent in the soil samples analyzed. The cluster-based

248 subsystems contain diverse functions, such as resistance to antibiotics and toxic compounds, and
249 pathogenicity islands.
250
251 Results of our study indicate that in the acidic soil, *Bradyrhizobium*, *Rhodopseudomonas*,
252 and *Burkholderia* are the major organisms in the soil involved in the nutritional ecosystem
253 functioning. *Bradyrhizobium* and *Candidatus solibacter* are taxonomically the most abundant
254 organisms in the soil samples analyzed. Collectively, it is evident from taxonomic and functional
255 analysis of the soil microbiome, bacteria from *Bradyrhizobium* are highly critical to maintaining
256 soil fertility, irrespective of soil type. Analyzing the microbiota present in 52 soil samples from
257 different countries, Shah and Subramaniam (2018) found that bacteria from *Bradyrhizobium*
258 genera were the most abundant organisms in the microbiota. The structure-function linkage results
259 indicate that the organism is not only responsible for nitrogen fixation and other pathways in N
260 cycle, but also plays a key role in S and Fe cycles, and degradation pathways of xenobiotic
261 compounds. *Bradyrhizobium* bacteria are present as symbiotic and non-symbiotic organisms in the
262 soil, and literature is replete with the importance of the organism in the Nitrogen cycle (Ormeño-
263 Orrillo and Martínez-Romero, 2019). Many strains of *Bradyrhizobium* are used commercially to
264 improve crop production (Environmental Protection Agency, n.d.). We suggest that the beneficial
265 impact of the organism in improving soil fertility could also be attributed to its role in other
266 biochemical pathways.

267 Acidic soils provide a unique environment for soil microorganisms due to iron, manganese
268 and aluminum toxicity, low nitrogen, phosphorus, and molybdenum levels, toxic levels of phenolic
269 acids, and hydrogen ion toxicity (Kidd and Proctor, 2001; Shah et al. 2011). Often to overcome
270 this issue, nitrogen fertilizers and other chemicals are used to improve soil fertility, but these
271 methods can cause other environmental issues including increase in nitrous oxide emissions (Xu,

2014). As a substitute to the use of chemicals for improving soil fertility and crop production, we suggest to the scientific community to study the possibility of using consortia of organisms including *Bradyrhizobium*, *Rhodopseudomonas*, and *Burkholderia.* Considering the importance and ubiquity of these organisms in the soil, the consortia could be used by farmers across the globe, irrespective of soil chemistry and geographical location.

Next-generation sequencing methods are increasingly used to study how the soil microbiome responds to changes in environmental conditions or to addition of contaminants in the soil. We suggest that in addition to analyzing general community-based diversity changes, scientists should specifically look for changes in the *Bradyrhizobium*, *Rhodopseudomonas*, and *Burkholderia* population to understand the impact. Our results suggest that changes in abundance of these organisms may greatly impact the soil fertility.

Considering that the soil samples analyzed were from the West Chester, PA region only, further studies are warranted using acidic soil samples from across the globe to validate the observations. Nevertheless, the metagenomic data reported here furthers our knowledge on the acidic soil microbial communities at structural and functional level. There is a large degree of similarity in the soil microbiome associated with different vegetation and soil pH. Increasing our attention to similarities in soil microbiome may allow us to further the biotechnological potential of microbial based products to improve soil fertility in the future.

Acknowledgement: This research is based upon work supported by the National Science Foundation under Grant No. CBET-1748439 and West Chester University Sustainability Grant.

Conflicts of Interest: Authors have no conflicts of interest to declare.

REFERENCES

Batcher, M. S. Monitoring Protocols for Central Pine Barrens Field Plots, v. 1.01. Prepared for Upton Ecological Research Reserve, Brookhaven National Laboratory, June 12, 2005

Bell, C. W., Acosta-Martinez, V., McIntyre, N. E., Cox, S., Tissue, D. T., & Zak, J. C. (2009). Linking microbial community structure and function to seasonal differences in soil moisture and temperature in a Chihuahuan desert grassland. *Microbial ecology*, *58*(4), 827-842.

Bruulsema, T. W. (2006). Soil Fertility in the Northeast Region. *Better Crops*, *90*(1), 8.

Collins, D., Luxton, T., Kumar, N., Shah, S., Walker, V. K., & Shah, V. (2012). Assessing the impact of copper and zinc oxide nanoparticles on soil: a field study. *PLoS One*, *7*(8), e42663.

Delmont, T. O., Prestat, E., Keegan, K. P., Faubladier, M., Robe, P., Clark, I. M., ... & Vogel, T. M. (2012). Structure, fluctuation and magnitude of a natural grassland soil metagenome. *The ISME journal*, *6*(9), 1677-1687.

Environmental Protection Agency (n.d.). TSCA Experimental Release Application Approved for Bradyrhizobium japonicum Strains (fact sheet). Web citation, https://www.epa.gov/regulation-biotechnology-under-tsca-and-fifra/tsca-experimental-release-application-approved. Accessed on June 10, 2021.

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences*, *103*(3), 626-631.

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., ... & Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, *109*(52), 21390-21395.

315 Garbeva, P. V., Van Veen, J. A., & Van Elsas, J. D. (2004). Microbial diversity in soil: selection
316 of microbial populations by plant and soil type and implications for disease suppressiveness.
317 *Annu. Rev. Phytopathol.*, *42*, 243-270.
318 Garrity, G. M., Thompson, L. M., Ussery, D. W., Paskin, N., Baker, D., Desmeth, P., ... & Ong,
319 P. S. (2009). Studies on monitoring and tracking genetic resources: an executive summary.
320 *Standards in genomic sciences*, *1*(1), 78-86.
321 Hasan, N. A., Young, B. A., Minard-Smith, A. T., Saeed, K., Li, H., Heizer, E. M., ... & Colwell,
322 R. R. (2014). Microbial community profiling of human saliva using shotgun metagenomic
323 sequencing. *PLoS One*, *9*(5), e97699.
324 Kidd, P. S., & Proctor, J. (2001). Why plants grow poorly on very acid soils: are ecologists missing
325 the obvious?. *Journal of Experimental Botany*, *52*(357), 791-799.
326 Kumar, N., Shah, V., & Walker, V. K. (2011). Perturbation of an arctic soil microbial community
327 by metal nanoparticles. *Journal of Hazardous Materials*, *190*(1-3), 816-822.
328 Mendes, L. W., & Tsai, S. M. (2018). Distinct taxonomic and functional composition of soil
329 microbiomes along the gradient forest-restinga-mangrove in southeastern Brazil. *Antonie Van*
330 *Leeuwenhoek*, *111*(1), 101-114.
331 Nannipieri, P., Kandeler, E., & Ruggiero, P. (2002). Enzyme activities and microbiological and
332 biochemical processes in soil. *Enzymes in the Environment. Marcel Dekker, New York*, 1-33.
333 Ormeño-Orrillo, E., & Martínez-Romero, E. (2019). A genomotaxonomy view of the
334 Bradyrhizobium genus. *Frontiers in microbiology*, *10*, 1334.
335 Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., ... & Sloan, W. T.
336 (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature*
337 *methods*, *6*(9), 639-641.

Rutgers, M., Wouterse, M., Drost, S. M., Breure, A. M., Mulder, C., Stone, D., ... & Bloem, J. (2016). Monitoring soil bacteria with community-level physiological profiles using Biolog™ ECO-plates in the Netherlands and Europe. *Applied Soil Ecology*, *97*, 23-35.

Shah, A., Shah, S., & Shah, V. (2021). Impact of flooding on the soil microbiota. *Environmental Challenges*, 100134.

Shah, V., & Subramaniam, S. (2018). Bradyrhizobium japonicum USDA110: a representative model organism for studying the impact of pollutants on soil microbiota. *Science of the total environment*, *624*, 963-967.

Shah, V., Shah, S., Kambhampati, M. S., Ambrose, J., Smith, N., Dowd, S. E., ... & Green, T. (2011). Bacterial and archaea community present in the Pine Barrens Forest of Long Island, NY: unusually high percentage of ammonia oxidizing bacteria. *PLoS One*, *6*(10), e26263.

Sobek, E. A., & Zak, J. C. (2003). The Soil FungiLog procedure: method and analytical approaches toward understanding fungal functional diversity. *Mycologia*, *95*(4), 590-602.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., ... & Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *science*, *304*(5667), 66-74.

Vogel, T. M., Simonet, P., Jansson, J. K., Hirsch, P. R., Tiedje, J. M., Van Elsas, J. D., ... & Philippot, L. (2009). TerraGenome: a consortium for the sequencing of a soil metagenome. *Nature Reviews Microbiology*, *7*(4), 252-252.

Xu, H. J., Wang, X. H., Li, H., Yao, H. Y., Su, J. Q., & Zhu, Y. G. (2014). Biochar impacts soil microbial community composition and nitrogen cycling in an acidic soil planted with rape. *Environmental science & technology*, *48*(16), 9391-9399.

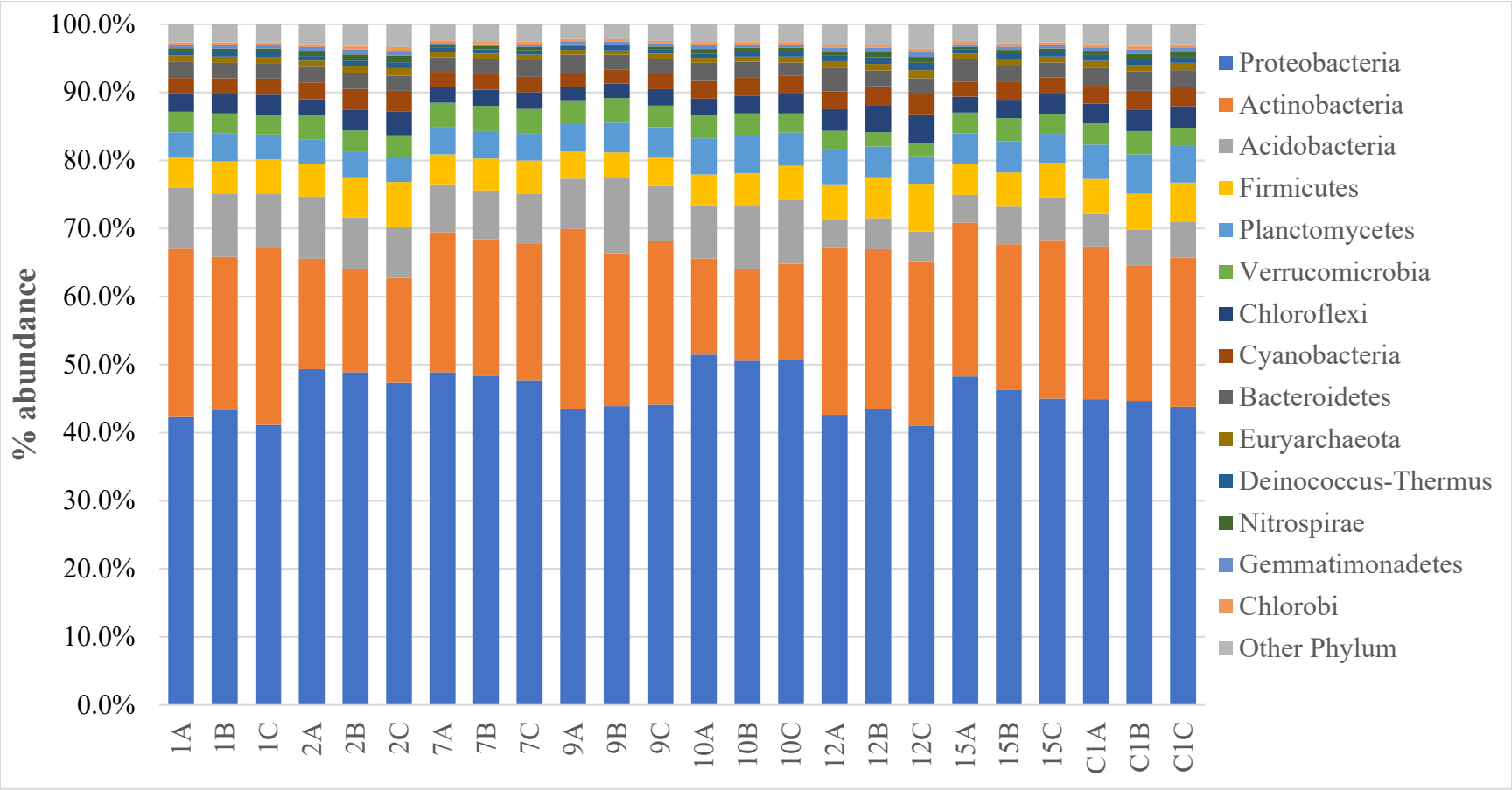

Figure 1. Phyla comparison of soil microbiomes collected from eight sampling locations at three different horizons (A, 0-10 cm; B, 10-20 cm; C, 20-30 cm).

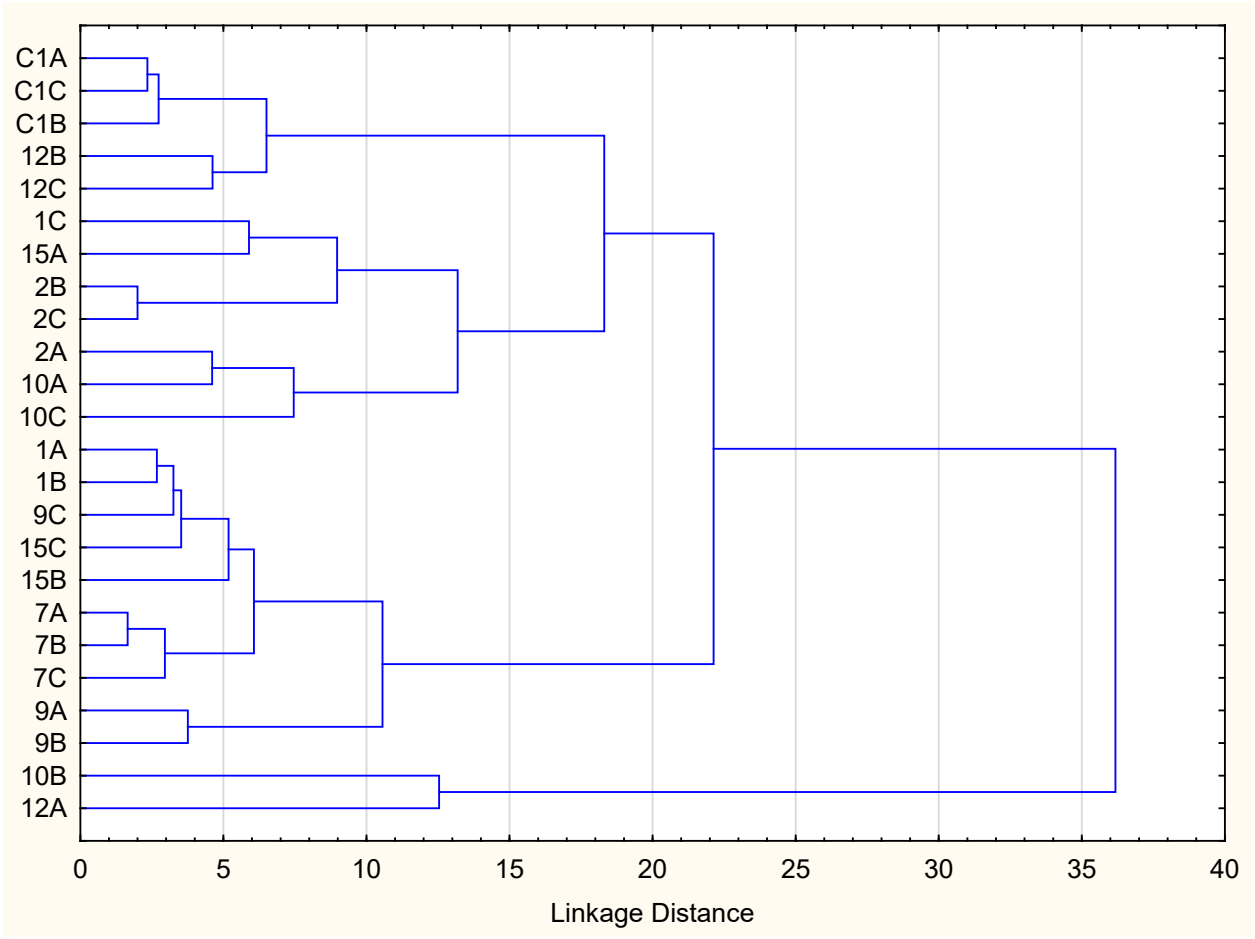

Figure 2. Hierarchical structure analysis used to identify the number of major clusters and the cluster members based on the normalized genus level abundance data.

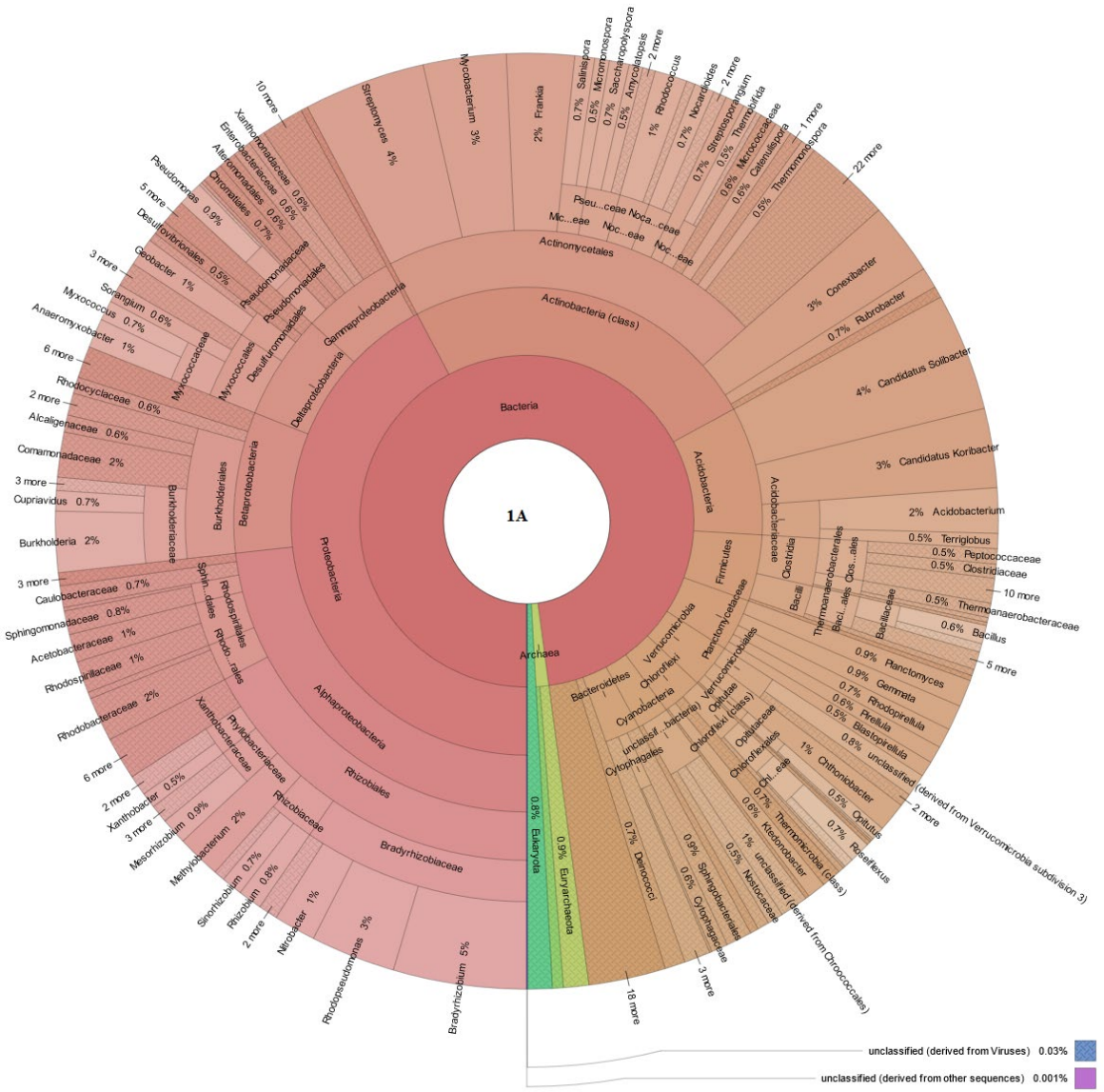

Figure 3. Krona plot of the microorganisms identified in the soil sample obtained from the top horizon near an Oak tree (sample 1A).

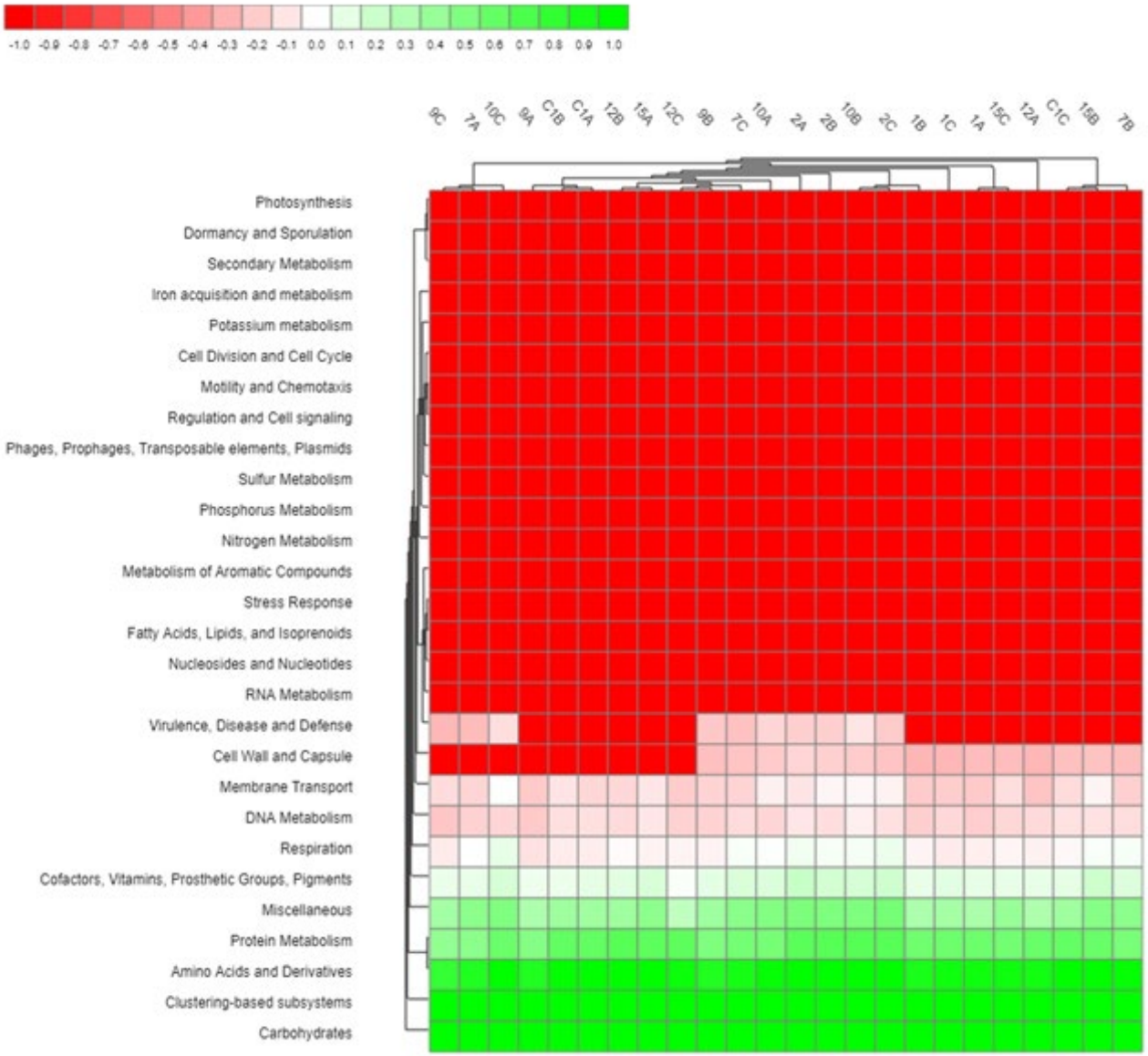

Figure 4. Heat map showing the differential abundance of functional categories (subsystem Level 1) between different soil samples.

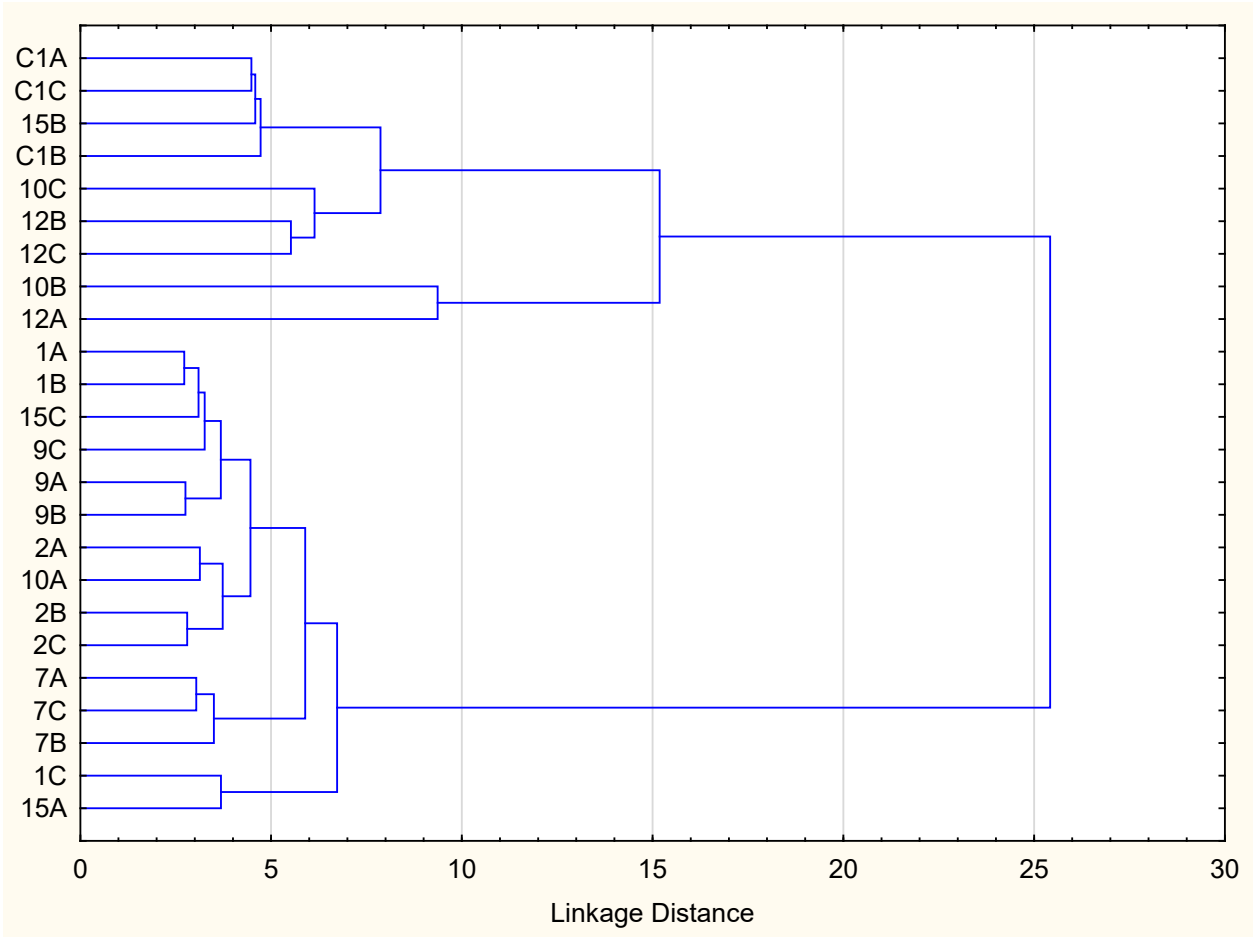

Figure 5. Hierarchical structure analysis used to identify the number of major clusters and the cluster members based on the normalized predicated proteins annotated at subsystem level 2.

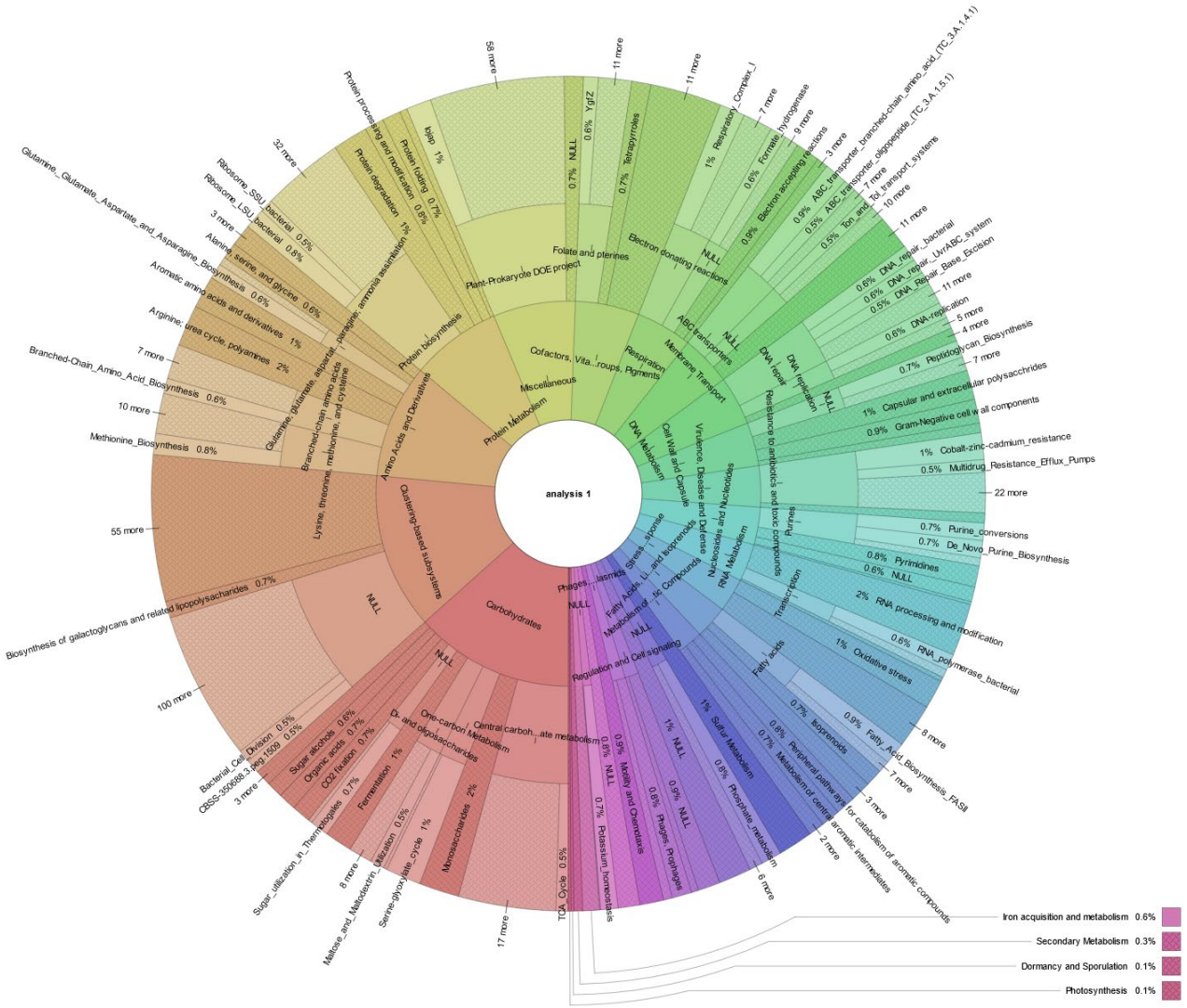

Figure 6. Krona plot of the functional sequences identified in the soil sample obtained from the top horizon near an Oak tree (sample 1A).

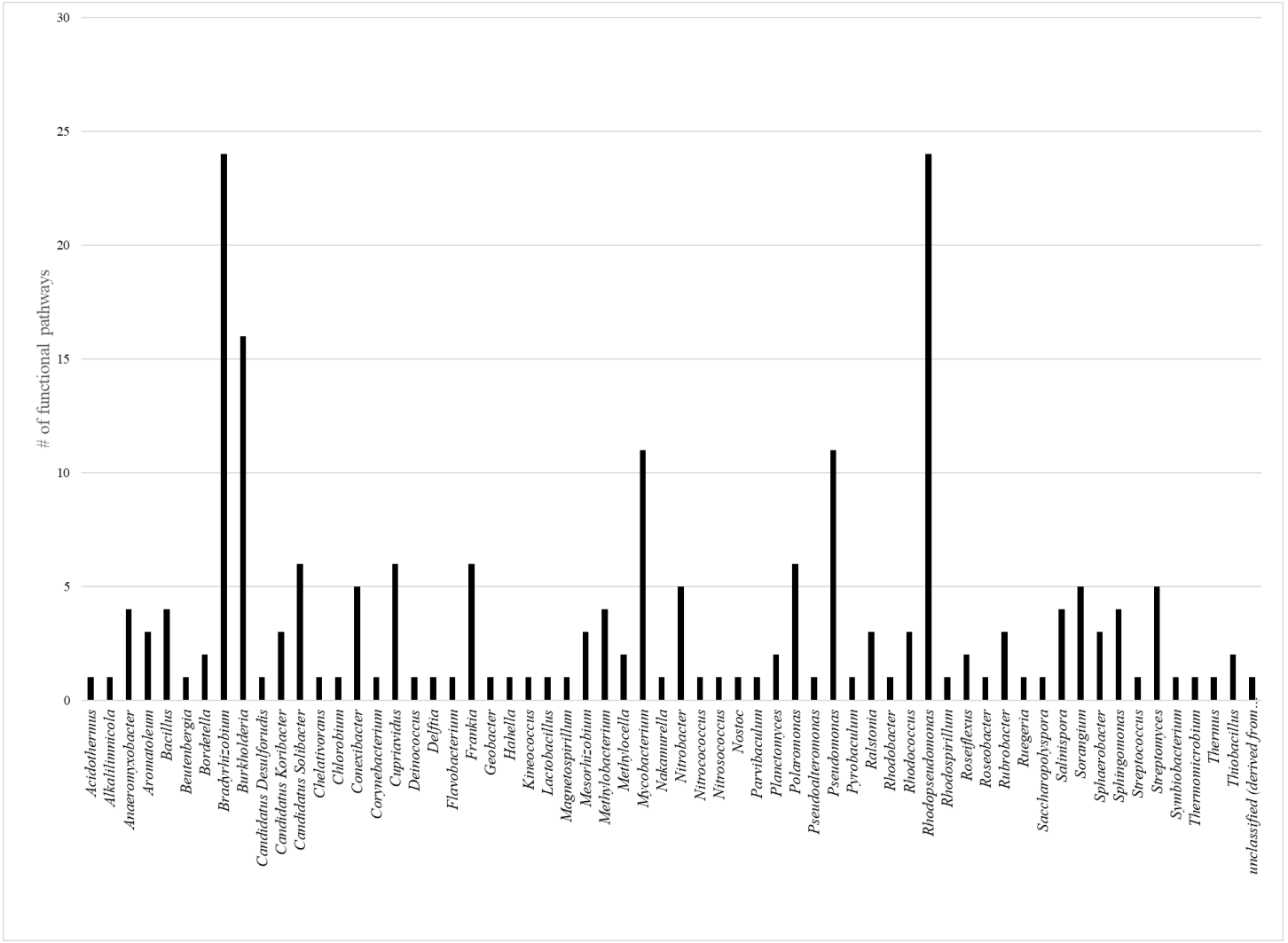

Figure 7. Total number of functional pathways top microbial genera are involved in within the acidic soil samples analyzed.

Site No.	Plant	Coordinates	Top (A)	Middle (B)	Bottom (C)
1	Oak	39°57'10.4"N 75°35'54.9"W	5.7 ± 0.3	5 ± 0.2	4.4 ± 0.2
2	Douglas Fir	39°56'59.1"N 75°35'38.9"W	4.1 ± 0.06	4.3 ± 0.06	4.4 ± 0.1
7	Pine - 1	39°55'57.1"N 75°36'09.8"W	5.4 ± 0.1	5.4 ± 0.03	5.4 ± 0.04
9	Pine - 2	39°56'55.9"N 75°36'06.6"W	6.2 ± 0.2	5.9 ± 0.1	5.7 ± 0.1
10	Pine - 3	39°56'22.7"N 75°35'36.5"W	5.8 ± 0.04	5.9 ± 0.1	5.9 ± 0.09
12	Tulip Tree	39°57'09.0"N 75°35'58.2"W	5.8 ± 0.1	6.1 ± 0.1	6 ± 0.6
15	Willow Oak	39°57'00.4"N 75°36'01.3"W	6.2 ± 0.1	6.3 ± 0.05	6.2 ± 0.1
C1	Grass	39°57'04.8"N 75°35'58.7"W	5.7 ± 0.2	6.1 ± 0.08	6.2 ± 0.03

Table 1. Vegetation type, coordinates, and pH for the locations from where the soil samples were obtained along with the pH values at each of the three sampling depths for each location. (0–5 cm, Horizon A; 6-15 cm, Horizon B; 16–30 cm, Horizon C)

Cluster 1	10B
Cluster 2	C1A, C1B, C1C, 12B, 12C
Cluster 3	12A
Cluster 4	2A, 10A, 10C
Cluster 5	1A, 1B, 7A, 7B, 7C, 9A, 9B, 9C, 15B, 15C
Cluster 6	1C, 2B, 2C, 15A

Table 2. Clustering of the soil samples based on the hierarchical structure analysis of taxonomic data at genera level. Soil samples were collected from eight sampling locations at three different horizons (A, 0-10 cm; B, 10-20 cm; C, 20-30 cm).

Genus	C1A	1A	2A	10B	12A	15A	Average
Bradyrhizobium	2.7	4.6	5.0	5.1	2.7	4.7	4.4
Candidatus Solibacter	2.9	4.1	4.4	4.5	2.3	2.4	3.6
Streptomyces	3.6	4.2	2.4	2.2	4.1	3.6	3.4
Mycobacterium	3.0	2.8	3.6	3.0	2.7	3.1	2.8
Rhodopseudomonas	1.9	2.9	3.3	3.2	1.9	3.1	2.8
Burkholderia	1.9	2.1	2.2	2.6	1.7	2.1	2.2
Candidatus Koribacter	1.1	2.7	2.7	2.7	1.0	0.9	1.9
Conexibacter	2.0	2.5	1.1	0.7	2.3	2.0	1.7
Frankia	1.7	2.3	1.3	1.2	1.9	1.8	1.8
Methylobacterium	1.4	1.5	2.0	1.8	1.3	1.7	1.7
Nitrobacter	1.0	1.4	1.7	1.7	0.9	1.5	1.4
Anaeromyxobacter	1.5	1.0	1.1	1.3	1.4	1.1	1.3
Geobacter	1.3	1.0	1.2	1.3	1.2	1.2	1.2
Planctomyces	1.2	0.9	1.0	1.4	1.3	1.1	1.1
Mesorhizobium	1.1	0.9	1.0	1.0	1.2	1.2	1.0
Acidobacterium	0.6	1.6	1.5	1.6	0.6	0.6	1.1
Gemmata	1.0	0.9	0.8	1.4	1.1	1.0	1.1
Rhodococcus	1.1	1.1	0.8	0.7	1.1	1.0	1.0
Pseudomonas	1.0	0.9	0.9	1.1	0.9	1.0	0.9
Chthoniobacter	0.9	1.0	1.2	1.0	0.7	1.0	1.0
Rhizobium	0.9	0.8	1.0	1.0	0.9	1.1	1.0

Sinorhizobium	1.0	0.7	1.0	0.9	0.9	0.9	0.9
Unclassified (derived from Verrucomicrobia subdivision 3)	0.9	0.8	1.1	1.1	0.8	0.8	0.9
Sorangium	1.1	0.6	0.7	0.9	1.0	1.0	0.8
Cupriavidus	0.9	0.7	0.9	1.0	0.8	0.9	0.8
Rhodopirellula	1.0	0.7	0.7	0.9	1.0	0.9	0.8
Pirellula	1.0	0.6	0.6	0.9	1.1	0.8	0.8
Myxococcus	1.0	0.7	0.7	0.8	0.9	0.8	0.8
Bacillus	0.7	0.6	0.9	0.8	0.9	0.8	0.8
Roseiflexus	0.9	0.7	0.6	0.7	0.9	0.7	0.8

Table 3. Relative Abundance of the top 30 bacterial genera in the representative soil samples from each cluster and average values across all the soil samples analyzed.

Cluster 1	12A
Cluster 2	C1A, C1B, 10C, 12B, 12C, 15B
Cluster 3	10B
Cluster 4	1C, 7A, 7B, 7C, 15A, 15C
Cluster 5	C1C, 1A, 1B, 2A, 2B, 2C, 9A, 9B, 9C, 10A

Table 4. Clustering of the soil samples based on the functional annotation of the sequence reads by MG-RAST at subsystem level 2. Soil samples were collected from eight sampling locations at three different horizons (A, 0-10 cm; B, 10-20 cm; C, 20-30 cm).

Level 1	Level 2	C1A	1A	2A	10B	12A	15A
Regulation and Cell signaling	NULL	21.4	21.6	21.7	21.6	21.9	21.8
Miscellaneous	Plant-Prokaryote DOE project	6.0	5.9	6.2	6.2	5.9	6.1
Protein Metabolism	Protein biosynthesis	4.7	4.6	4.5	4.3	4.8	4.6
Carbohydrates	Central carbohydrate metabolism	4.2	4.2	4.1	4.1	4.0	4.1
Virulence, Disease and Defense	Resistance to antibiotics and toxic compounds	2.8	3.0	3.3	3.7	2.9	2.8
DNA Metabolism	DNA repair	2.7	2.4	2.5	2.7	2.5	2.6
Amino Acids and Derivatives	Lysine, threonine, methionine, and cysteine	2.5	2.5	2.5	2.5	2.6	2.6
Respiration	Electron donating reactions	2.3	2.2	2.2	2.1	2.4	2.1
RNA Metabolism	RNA processing and modification	2.2	2.1	2.1	2.1	2.3	2.2
Cofactors, Vitamins, Prosthetic Groups, Pigments	Folate and pterines	2.0	2.0	1.9	1.8	2.0	2.0
Amino Acids and Derivatives	Branched-chain amino acids	1.9	2.0	1.9	1.8	1.9	1.9
Membrane Transport	ABC transporters	1.6	1.6	1.6	1.8	1.5	1.7
Amino Acids and Derivatives	Arginine; urea cycle, polyamines	1.8	1.5	1.5	1.6	1.6	1.6
Nucleosides and Nucleotides	Purines	1.7	1.6	1.6	1.6	1.6	1.6
Carbohydrates	Monosaccharides	1.6	1.6	1.4	1.4	1.8	1.6
Carbohydrates	Di- and oligosaccharides	1.4	1.7	1.4	1.3	1.5	1.4
Carbohydrates	One-carbon Metabolism	1.5	1.4	1.4	1.4	1.5	1.3
Protein Metabolism	Protein degradation	1.4	1.4	1.3	1.5	1.4	1.3

Fatty Acids, Lipids, and Isoprenoids	Fatty acids	1.2	1.3	1.4	1.3	1.3	1.3
Amino Acids and Derivatives	Aromatic amino acids and derivatives	1.3	1.3	1.3	1.3	1.3	1.3
Cell Wall and Capsule	Capsular and extracellular polysacchrides	1.1	1.2	1.2	1.1	1.3	1.1
Stress Response	Oxidative stress	1.1	1.0	1.1	1.2	1.1	1.2
DNA Metabolism	DNA replication	1.1	1.1	1.1	1.1	1.1	1.2
Carbohydrates	Fermentation	0.9	1.0	1.1	1.0	1.1	1.1
RNA Metabolism	Transcription	1.0	1.0	0.9	0.8	1.0	0.9
Amino Acids and Derivatives	Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	1.0	0.9	0.9	1.0	0.9	1.0
Respiration	Electron accepting reactions	0.8	1.0	0.9	0.9	0.9	0.9
Cell Wall and Capsule	Gram-Negative cell wall components	0.8	1.0	1.0	1.0	0.8	0.8
Phages, Prophages, Transposable elements, Plasmids	Phages, Prophages	0.8	0.8	0.9	0.8	0.7	0.8
Nucleosides and Nucleotides	Pyrimidines	0.9	0.9	0.8	0.7	0.8	0.8

Table 5. Relative Abundance of the top 30 functional genes (level 2) bacterial genera in the representative soil samples from each cluster.