CHAPTER 7: METAGENOMIC ANALYSIS

7.1. General

The study of soil microbes is crucial for understanding the role of microorganisms in maintaining ecosystem health and functioning. Soil microorganisms play a vital role in nutrient cycling, decomposition, and soil structure maintenance, and they also influence plant growth, disease resistance, and stress tolerance. A significant portion of microorganisms present on the Earth remains unidentified due to challenges in isolating and cultivating them on appropriate media. Despite ongoing improvements in cultivation techniques, the lab-based culture of these microbes is still impractical, particularly in complex environments like soils (Lewis et al., 2021; Dedysh, 2011). Traditional culturing methods have limitations, as they only allow for the isolation and study of a small fraction of the microbial community. It is well established that only approximately 1% of bacteria on Earth can be readily cultivated in vitro– the so-called 'great plate count anomaly', based on the observation that microscopic counts are considerably larger than the equivalent total viable counts (Staley & Konopka, 1985; Amann et al., 1995; Hugenholtz et al., 1998). This limitation hampers our understanding of microbial community composition, a crucial factor in comprehending soil ecosystem functioning.

In contrast, metagenomics enables the simultaneous study of all organisms present in a sample, providing a more comprehensive picture of the microbial diversity and community structure (Zhou et al., 2015). The metagenomic approach, integrating genomics, bioinformatics, and systems biology, enables the examination of microbial diversity, taxonomic structure, and richness in a sample. Metagenomics libraries can be explored for both functional and genetic diversity, identifying new catabolic genes for xenobiotics degradation. This approach holds promise in predicting *in-situ* microbial responses, activities, and dynamics, including bioremediation of pollutants. Additionally, *in-situ* analysis of rRNA and mRNA provides insights into active taxa and functional gene expression, offering a comprehensive understanding of soil microbial communities.

Metagenomics is a powerful and effective approach for studying the biodegradation of insecticides in contaminated environments. One of the significant advantages of metagenomics is that it enables the analysis of the entire microbial community present in the environment, which can provide a more comprehensive understanding of the biodegradation process. Metagenomics allows for the identification of the specific genes and enzymes that are involved in the biodegradation of insecticides. This knowledge can help researchers develop better bioremediation strategies to

minimize the harmful effects of these compounds on the environment. Furthermore, cutting-edge molecular techniques offer the potential to understand the roles played by different microbial groups in the biodegradation process. This information can be used to identify the dominant taxonomic groups in specific soils and monitor changes in the microbial community over time.

Metagenomic approach allows to look at the diversity of microorganisms present in a sample and to characterize their taxonomic structure and richness. The metagenomic approach involves the extraction of DNA from an environmental sample, followed by the construction of a metagenomic library. This library can then be sequenced, allowing for the identification and analysis of the genetic material of the microorganisms present in the sample. Metagenomics libraries can be assessed for both functional and genetic diversity. New catabolic genes for the degradation of xenobiotics are identified through functional metagenomics. Metagenomics propose considerable assurance for prediction of *in-situ* microbial responses, activities, and dynamics along with bioremediation of pollutants (Prabha et al., 2017). Furthermore, in-situ analysis of rRNA permits to identify the active taxa and mRNA allow detecting an expression of functional genes in soil which gives a complete image of the structure and function of the soil microbial communities (Urich et al., 2008). Metagenomics studies are considered the most efficient way to see the complete microbiological profile of the bulk soil as well as rhizosphere (Soni et al., 2017).

Metagenomic libraries can be evaluated for both genetic and functional diversity. By analyzing the genetic makeup of the microorganisms, researchers can gain insights into their evolutionary history and the potential functions they perform. Functional metagenomics involves screening the library to identify genes that are involved in specific metabolic pathways or other cellular processes. This approach can be used to identify new catabolic genes that help to degrade xenobiotics or other compounds that are difficult to break down.

Metagenomics also offers significant potential for predicting the behavior and activity of microbial communities in natural environments. By analyzing the genetic material of the microorganisms, researchers can gain insights into their metabolic capabilities and the environmental conditions that are necessary for their growth and survival (Dubey et al., 2019). This information can be used to predict how microbial communities will respond to changes in the environment, or to design strategies for the bioremediation of pollutants. With advances in sequencing technology and data

analysis, this approach will continue to play an important role in the study of microbiology and environmental science.

Soil metagenomics is a rapidly growing field that employs advanced molecular techniques to analyze the genetic material extracted from soil samples. This approach allows researchers to study the entire microbial community present in the soil, including bacteria, fungi, viruses, and other microorganisms that cannot be cultured using traditional methods (Lahlali et al., 2021). Metagenomics has proven particularly useful in studying the adaptations and interactions of microorganism communities in response to environmental changes, such as climate change, pollution, and land use changes (Kumar et al., 2021). This approach allows researchers to identify changes in microbial community composition and function, providing insights into the resilience and adaptability of soil microbial communities. Soil metagenomics is an essential tool for understanding the complex interactions between soil microorganisms and their environment. By providing a comprehensive view of the microbial community, metagenomics has the potential to revolutionize our understanding of soil ecology and inform efforts to promote soil health and sustainability (Nwachukwu and Babalola, 2022).

7.2. Metagenomic analysis

The high-throughput metagenomic approach was used to study the potential functional microbiome and composition of the taxonomic community in Sample 1-16S, Sample 2-16S and Sample 3-16S. The amount of guanine-cytosine based on the V3-V4 amplicon region was estimated in all the soil samples. The GC content, or the percentage of guanine (G) and cytosine (C) nucleotides in DNA, can vary widely among different organisms and environments. It is a significant molecular characteristic with various implications for the structure and function of DNA molecules. In soil samples, the GC content can be influenced by the types of microorganisms present. The amount of guanine-cytosine in all the samples has been presented in Figure 7.1.

It can be observed that the GC content remains almost constant across all the samples. Table 7.1 gives the details for the V3-V4 amplicon region of all the samples. It can be observed that the number of reads is maximum in Sample 1-16S

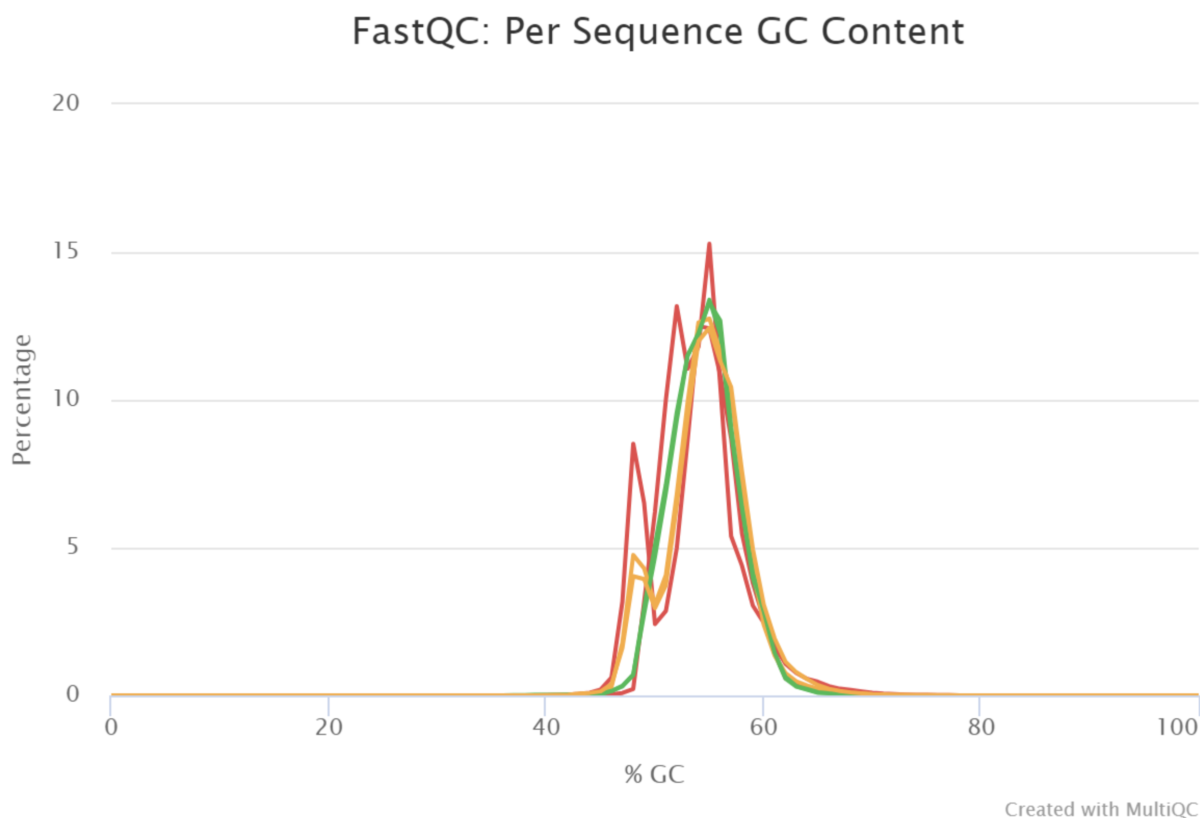

Figure 7.1: GC content of the sequences (Yellow color represents Sample 1-16S, Green color represents Sample 2-16S, and red color represents Sample 3-16S)

Table 7.1: Sample details for V3-V4 amplicon region

S. No.	Sample name	No of reads (in million)	GC content (%)
1	Sample 1-16S	0.6M	53%
2	Sample 2-16S	0.2M	54%
3	Sample 3-16S	0.2M	54%

In the analysis, features with identical values (i.e. zeros) across all samples and features that appear in only one sample (considering artifacts) have been excluded. Data filtering aims to remove low-quality or uninformative features to improve downstream statistical analysis.

• Low count filter - features with very small counts in very few samples are likely due to sequencing errors or low-level contaminations. The minimum count needs to be specified (default 4). Data can also be filtered based on mean and median values.

• Low variance filter - features that are close to constant throughout the experiment conditions are unlikely to be associated with the conditions under study. Their variances can be measured using inter-quantile range (IQR), standard deviation or coefficient of variation (CV).

This project followed the default low count filter at a 20% prevalence with a minimum count of 0. The low variance was filtered based on standard deviation with 10% to remove.

Data Normalization

Normalization aims to address the variability in sampling depth and the sparsity of the data to enable more biologically meaningful comparisons. The data was rarefied to minimum library size. The data was scaled by the Total sum scaling (TSS) factor to bring all the samples to the same scale.

Core Microbiome

The core microbiome refers to the set of taxa that are detected in a high fraction of the population above a given abundance threshold. The count data is transformed to compositional (relative) abundance in order to perform such analysis. The data was visualized with a sample prevalence of 20% and a relative abundance of 0.1 %.

In metagenomics, OTU is used to classify bacteria based on sequence similarity of the 16S marker gene. An OTU consists of a group of bacteria whose 16S marker gene shows a sequence identity of 97 percent and above. It is used to classify bacteria at the genus level.

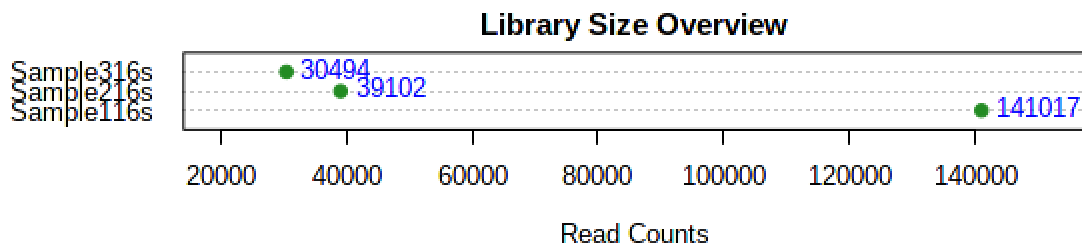

Figure 7.2: Library size overview of all the samples

A total of 141017 reads in Sample 1-16S, 39102 in Sample 2-16S and 30494 in Sample 3-16S were obtained. The library size overview is represented in Figure 7.2. The study concludes that the number of reads in Sample 1-16S is much higher as compared to Sample 2-16S and Sample 3-16S, suggesting a decrease in the number of reads in the presence of imidacloprid.

Similarity among the samples was also determined. It can be inferred from Figure 7.3 that Sample 2-16S and Sample 3-16S are closely related as compared to Sample 1-16S. Imidacloprid was added in sample 2, which was subjected to bacterial degradation for 30 days. The soil sample obtained after 30 days was sample 3. However, there was no exposure of imidacloprid in sample 1, which could be the possible reason for the larger difference between Sample-16S and Sample 2-16S. The Sample 2-16S and Sample 3-16S are closely related because of the addition of imidacloprid in Sample 2-16S and incomplete degradation of imidacloprid in Sample 3-16S.

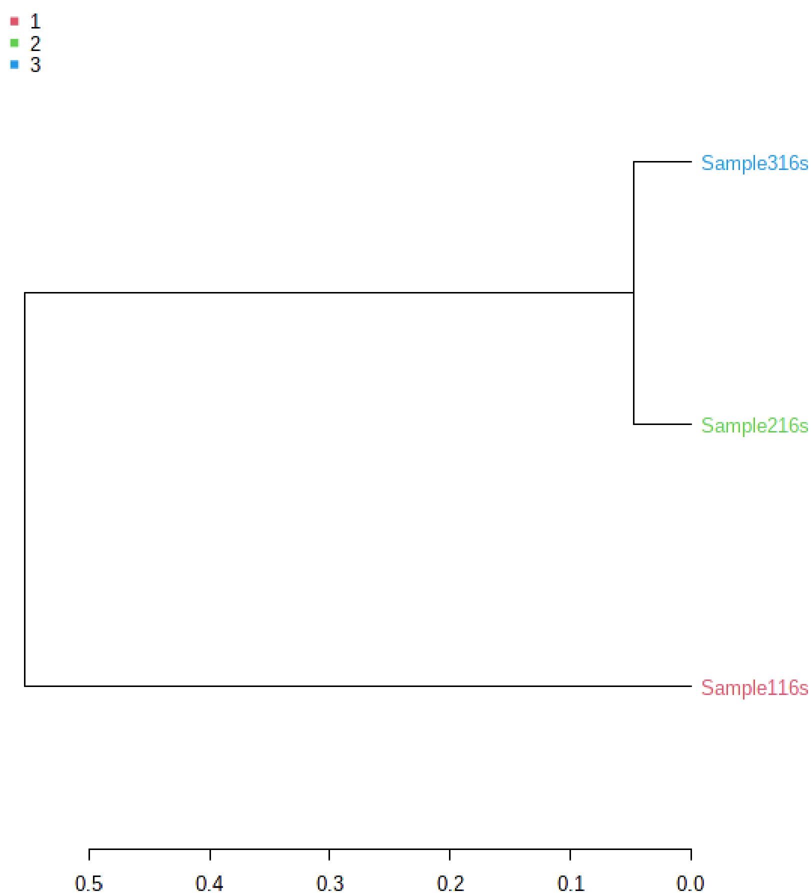

Figure 7.3: Dendrogram representing similarity among the samples

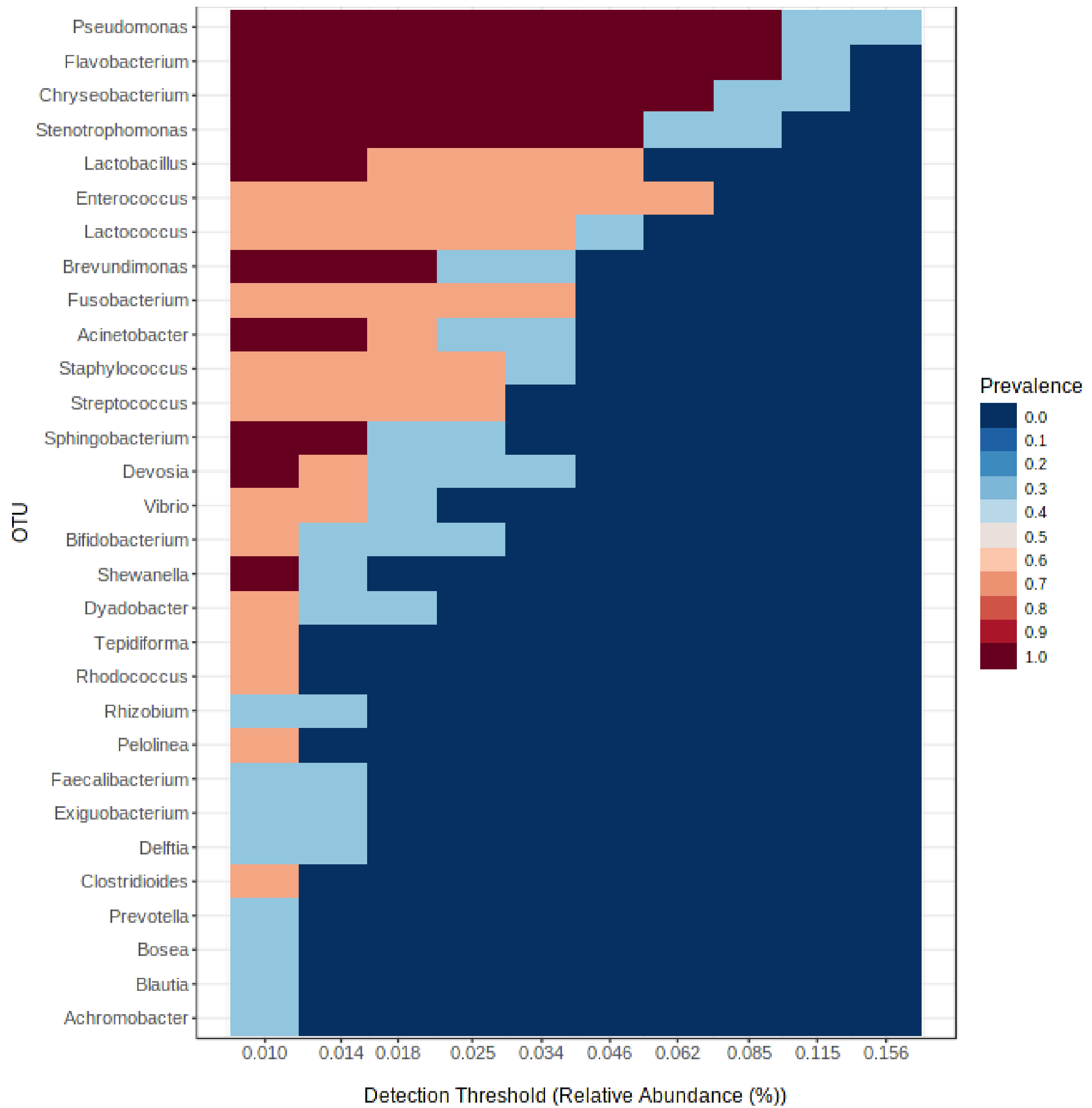

Figure 7.4: Heatmap depicting the Core microbiome at the order level, where the Y-axis represents the prevalence level of core features across the detection threshold (Relative abundance) range on the X-axis

It can be inferred from Figure 7.4 that *Pseudomonas* are the most prevalent operational taxonomic units present in the samples and *Achromobacter* are the least prevalent.

7.3. Microbial quantity and community analysis in soil samples

3 samples were sequenced and compared completely in this study. These results indicate that the imidacloprid changed the soil microbial community structures. Similarity analyses among the tested soil samples were conducted based on Illumina MiSeq next-generation sequencing technology. According to the clustering results, the overall community structures were affected by the imidacloprid treatment.

A community of abundantly present top 10 genera in both the soil samples were analyzed. *Pseudomonas* (21%) and *Flavobacterium* (21%) constitute major genera in Sample 1, followed by *Chryseobacterium* (17%), *Stenotrophomonas* (13%), *Brevundimonas* (6%), *Devolsa* (5%), *Acinetobacter* (5%), *Sphingobacterium* (5%), *Bifidobacterium* (4%) and *Dyadobacter* (3%). Figure 7.5 shows the pie chart representing the 10 most abundant genera in Sample 1-16S.

Similarly, the most abundant genera in Sample 2-16S and Sample 3-16S were also analyzed. Graphical representations of both samples are given in Figures 7.6 and 7.7, respectively.

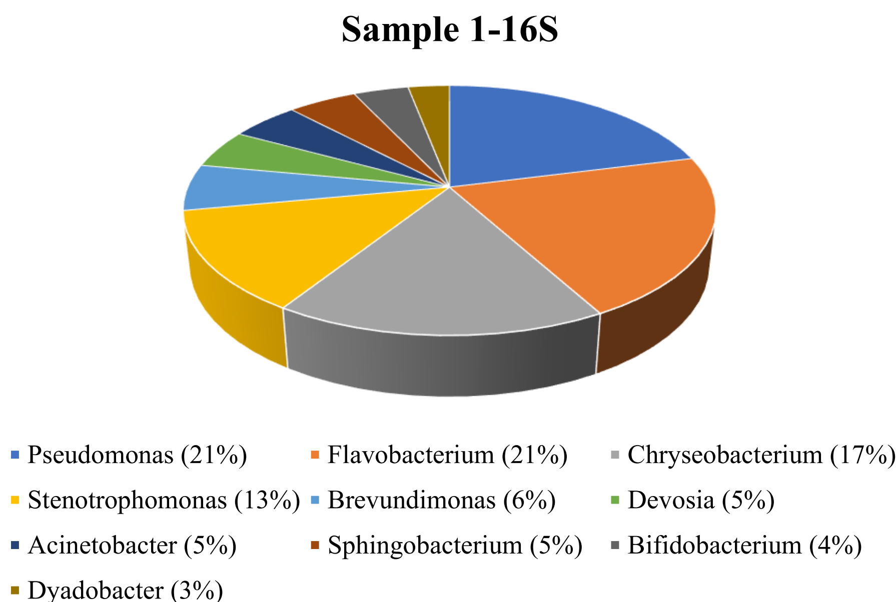

Figure 7.5: 10 most abundant genus present in Sample 1-16S

Sample 2-16S

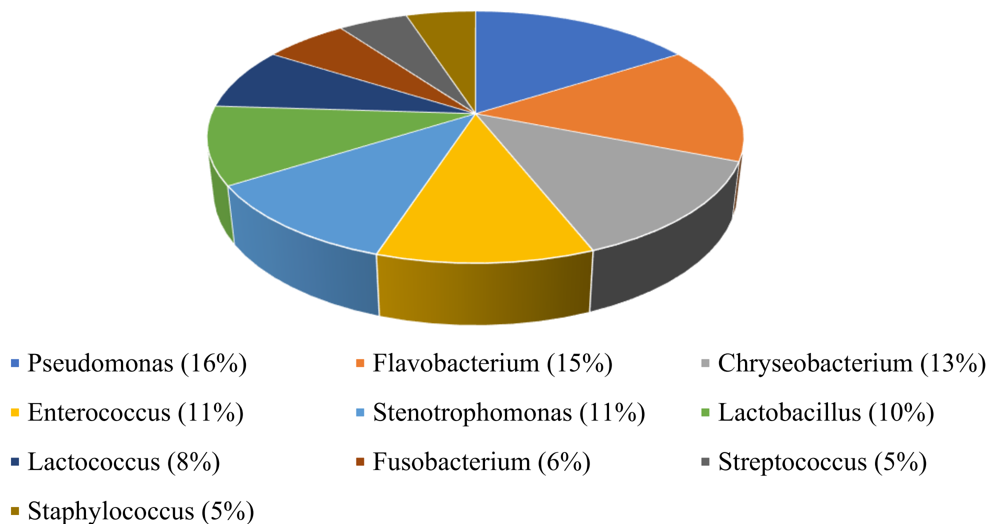

Figure 7.6: 10 most abundant genus present in Sample 2-16S

Sample 3-16S

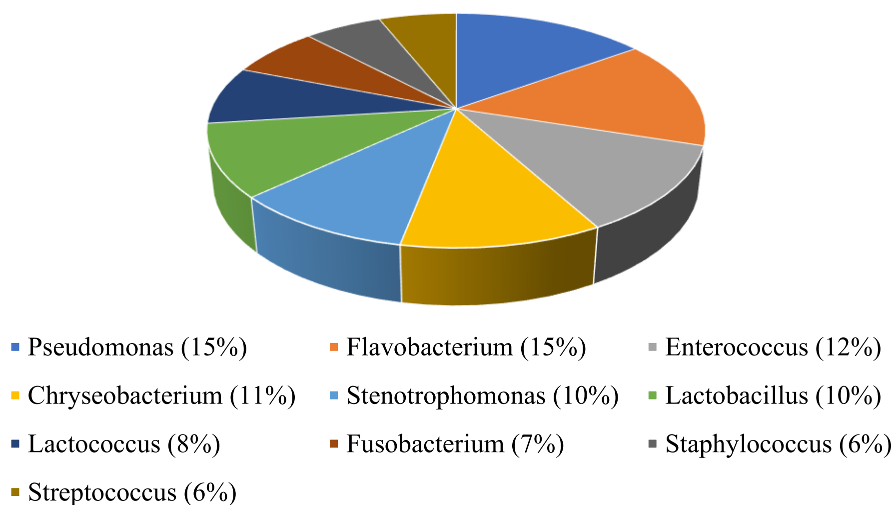

Figure 7.7: 10 most abundant genus present in Sample 3-16S

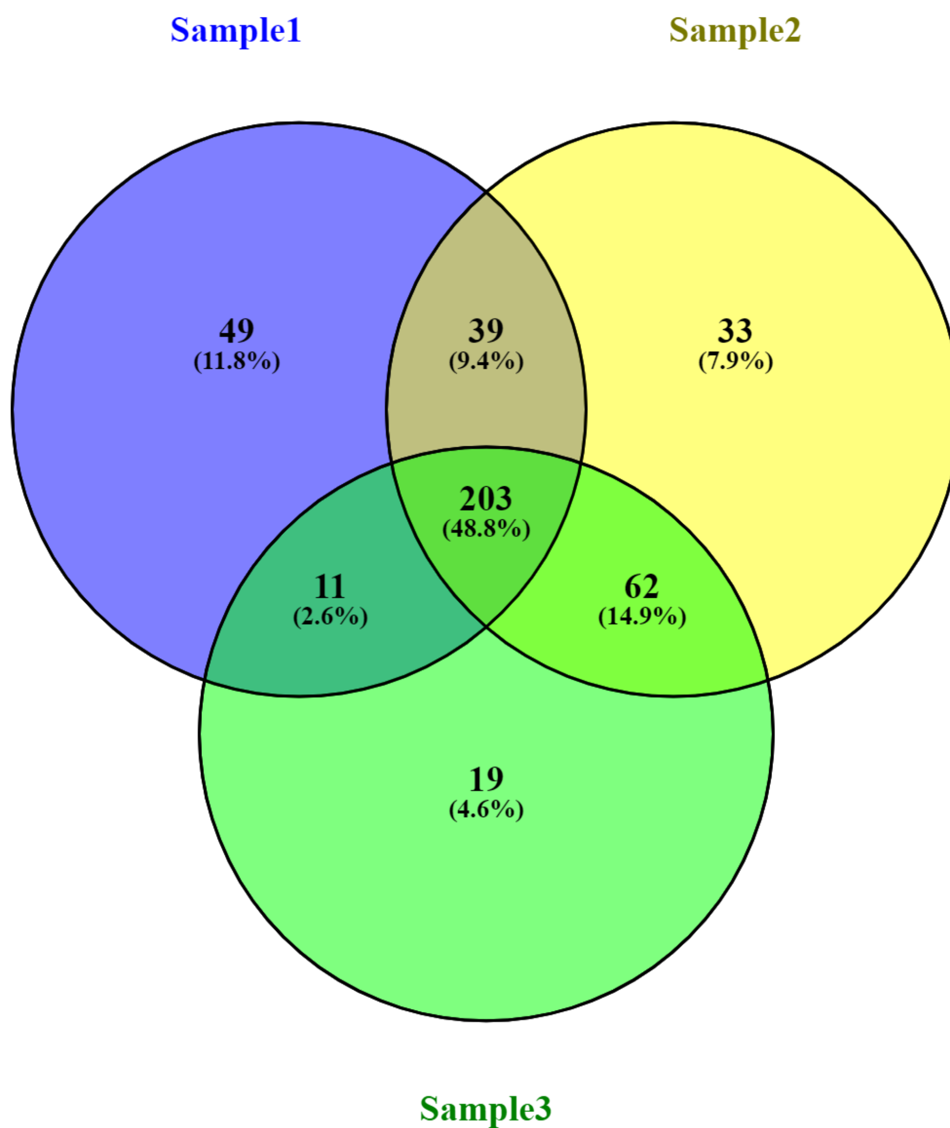

Figure 7.8: Venn diagram representing a relation between the genus present across all the samples

In the analysis of three different samples, a total of 203 common elements were identified. These common elements were found to be present in all three samples. Furthermore, there were specific elements that were found exclusively in each sample. Sample 1-16S had 49 elements that were not present in the other two samples, indicating a unique composition or source for Sample 1-16S. Similarly, Sample 2-16S had 33 exclusive elements, while Sample 3-16S had 19 unique elements.

Additionally, there were elements that were found to be common between pairs of samples. For instance, Sample 1-16S and Sample 2-16S had 39 common elements. On the other hand, Sample 1-16S and Sample 3-16S had 11 common elements, indicating a potential connection between these two samples, but to a lesser extent compared to Sample 1-16S and Sample 2-16S. Finally, Sample 2-16S and Sample 3-16S had the highest number of common elements, with 62 overlapping elements.

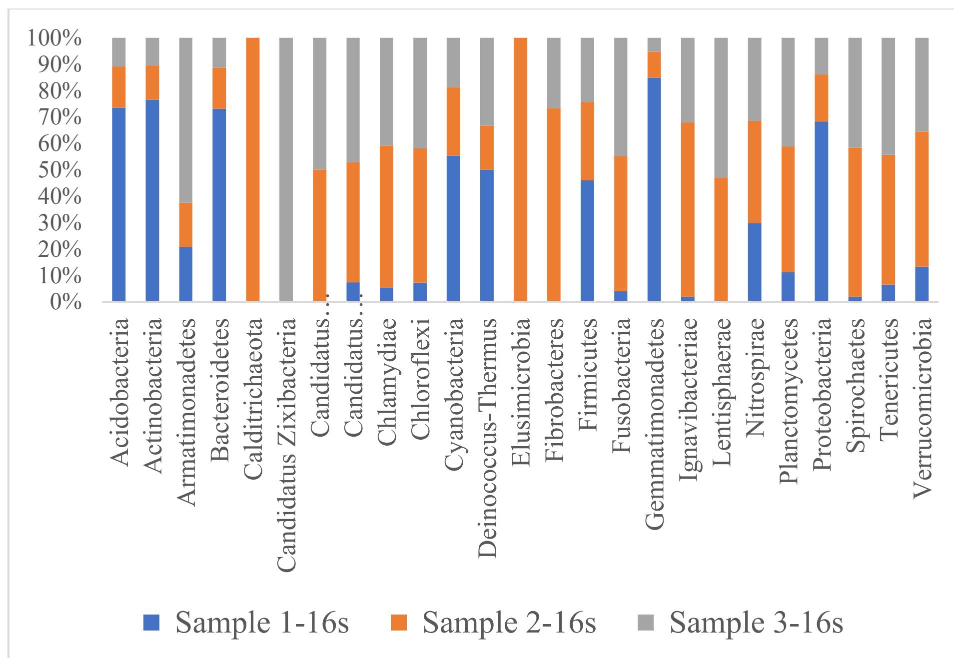

Figure 7.9: Variation in Sample 1-16S, Sample 2-16S and Sample 3-16S at phylum level

A significant difference in the bacterial community between imidacloprid-contaminated and non-contaminated soil was observed in this study. The analysis revealed distinct variations in microbial diversity, highlighting the impact of pesticide contamination on soil ecosystems. Specifically, when comparing the contaminated samples, Sample 2-16S and Sample 3-16S exhibited less variation at the phylum level. This indicates a more homogeneous bacterial community structure in the presence of imidacloprid. In contrast, the control Sample 1-16S, which was not exposed to the insecticide, showed a more diverse occurrence of different phyla. This observation is supported by Figure 7.9, which illustrates the differences in microbial composition between the samples at phylum level. These findings align with the research conducted by Fang et al. (2018), who found that pesticide contamination not only reduces microbial diversity and richness but also tends to enrich specific bacterial communities that can tolerate the chemical stress. Similarly, a study by

Regar et al. (2019) reported comparable results, reinforcing the idea that pesticides like imidacloprid significantly alter soil microbial communities by diminishing overall diversity.

7.4. Metabolic pathway prediction

The metabolic pathways within the samples were predicted using KEGG (Kyoto Encyclopedia of Genes and Genomes). The KEGG signaling pathway is represented as a network of gene products that are primarily proteins but also include functional RNA. The KEGG database contains a list of metabolic pathways, and the program attempts to map each read to a gene with a unique valid KO identifier, and thus to one or more pathways. Nicotinate and nicotinamide metabolic genes were identified according to the utilization of neonicotinoid insecticide, imidacloprid.

The different microbial community structures and abundances might lead to different selective degradation pathways of imidacloprid in different soil samples. A total of 47 pathways, based on the KEGG database were predicted in Sample 1-16S. However, one additional pathway was predicted in Sample 2-16S and Sample 3-16S. Nicotinate and nicotinamide metabolism was the additional pathway observed in the samples contaminated with imidacloprid. Therefore, the pathway may be correlated with the degradation of imidacloprid since imidacloprid belongs to the neonicotinoid class of insecticides. The different pathways predicted have been tabulated in Table 7.2.

Table 7.2: Pathways predicted by KEGG

S. No.	Pathways (Sample 1-16S)	Pathways (Sample 2-16S)	Pathways (Sample 3-16S)
1.	Biosynthesis of amino acids	Biosynthesis of amino acids	Biosynthesis of amino acids
2.	Carbon metabolism	Carbon metabolism	Carbon metabolism
3.	Pyruvate metabolism	Pyruvate metabolism	Citrate cycle (TCA cycle)
4.	Citrate cycle (TCA cycle)	Citrate cycle (TCA cycle)	Pyruvate metabolism
5.	Methane metabolism	Methane metabolism	Cysteine and methionine metabolism

6.	Glycolysis / Gluconeogenesis	Cysteine and methionine metabolism	Methane metabolism
7.	Cysteine and methionine metabolism	Glycolysis / Gluconeogenesis	Lipopolysaccharide biosynthesis
8.	Pentose phosphate pathway	Lipopolysaccharide biosynthesis	Glycolysis / Gluconeogenesis
9.	Lipopolysaccharide biosynthesis	Pentose phosphate pathway	Butanoate metabolism
10.	Butanoate metabolism	Butanoate metabolism	Terpenoid backbone biosynthesis
11.	Pantothenate and CoA biosynthesis	Pantothenate and CoA biosynthesis	Nitrogen metabolism
12.	Carbon fixation in photosynthetic organisms	Carbon fixation in photosynthetic organisms	Pentose phosphate pathway
13.	Terpenoid backbone biosynthesis	Terpenoid backbone biosynthesis	Nitrotoluene degradation
14.	Nitrotoluene degradation	Nitrotoluene degradation	Glyoxylate and dicarboxylate metabolism
15.	Pyrimidine metabolism	Porphyrin and chlorophyll metabolism	Porphyrin and chlorophyll metabolism
16.	Porphyrin and chlorophyll metabolism	Carbon fixation pathways in prokaryotes	Carbon fixation pathways in prokaryotes
17.	Carbon fixation pathways in prokaryotes	Pyrimidine metabolism	Pantothenate and CoA biosynthesis
18.	One carbon pool by folate	Valine, leucine and isoleucine degradation	One carbon pool by folate
19.	Nitrogen metabolism	One carbon pool by folate	Seleno-compound metabolism
20.	Drug metabolism - other enzymes	Drug metabolism - other enzymes	Thiamine metabolism

21.	Glyoxylate and dicarboxylate metabolism	Nitrogen metabolism	Histidine metabolism
22.	Thiamine metabolism	Seleno-compound metabolism	Carbon fixation in photosynthetic organisms
23.	Seleno-compound metabolism	Thiamine metabolism	Pyrimidine metabolism
24.	Histidine metabolism	Glyoxylate and dicarboxylate metabolism	Alanine, aspartate and glutamate metabolism
25.	Valine, leucine and isoleucine degradation	Histidine metabolism	Arginine and proline metabolism
26.	Arginine and proline metabolism	Alanine, aspartate and glutamate metabolism	Valine, leucine and isoleucine degradation
27.	Geraniol degradation	Geraniol degradation	Geraniol degradation
28.	Alanine, aspartate and glutamate metabolism	Arginine and proline metabolism	Valine, leucine and isoleucine biosynthesis
29.	Biotin metabolism	Biotin metabolism	Biotin metabolism
30.	Valine, leucine and isoleucine biosynthesis	Valine, leucine and isoleucine biosynthesis	Drug metabolism - other enzymes
31.	Glycine, serine and threonine metabolism	Glycine, serine and threonine metabolism	Glycine, serine and threonine metabolism
32.	Folate biosynthesis	Folate biosynthesis	Folate biosynthesis
33.	Purine metabolism	Starch and sucrose metabolism	2-Oxocarboxylic acid metabolism
34.	2-Oxocarboxylic acid metabolism	Propanoate metabolism	Phenylalanine, tyrosine and tryptophan biosynthesis
35.	Fatty acid biosynthesis	Purine metabolism	Starch and sucrose metabolism
36.	Phenylalanine, tyrosine and tryptophan biosynthesis	2-Oxocarboxylic acid metabolism	Fatty acid biosynthesis

37.	Starch and sucrose metabolism	Fatty acid biosynthesis	Lysine biosynthesis
38.	Propanoate metabolism	Phenylalanine, tyrosine and tryptophan biosynthesis	Propanoate metabolism
39.	Nicotinate and nicotinamide metabolism	Nicotinate and nicotinamide metabolism	Purine metabolism
40.	D-Glutamine and D-glutamate metabolism	D-Glutamine and D-glutamate metabolism	D-Glutamine and D-glutamate metabolism
41.	Synthesis and degradation of ketone bodies	Synthesis and degradation of ketone bodies	Synthesis and degradation of ketone bodies
42.	Lysine biosynthesis	Lysine biosynthesis	Nicotinate and nicotinamide metabolism
43.	Peptidoglycan biosynthesis	Peptidoglycan biosynthesis	Peptidoglycan biosynthesis
44.	Vitamin B6 metabolism	Vitamin B6 metabolism	Vitamin B6 metabolism
45.	Streptomycin biosynthesis	Streptomycin biosynthesis	Streptomycin biosynthesis
46.	Limonene and pinene degradation	Limonene and pinene degradation	Caprolactam degradation
47.	Glycosaminoglycan degradation	Glycosaminoglycan degradation	Biosynthesis of siderophore group non-ribosomal peptides
48.	-	Fatty acid degradation	Limonene and pinene degradation

7.5. Summary

In the present study, metagenomic analysis of three soil samples was conducted to better understand the microbial community structure and their functions. This information is crucial for

environmental management, bioremediation strategies, and advancing our knowledge of microbial ecology in various environments.

The results reveal a substantial shift in the bacterial microflora of the soil samples. *Pseudomonas* was the most prevalent genus across all samples. However, the abundance of *Pseudomonas* decreased with the addition of imidacloprid. This study has highlighted the complex diversity of bacteria present in imidacloprid-contaminated samples.

The findings suggest that imidacloprid significantly impacts soil microflora. In the imidacloprid-supplemented sample, the number of reads was the lowest, while the non-contaminated sample had the highest number of reads. A total of 203 bacterial genera were identified, each playing various roles in different metabolic pathways.

The detailed insights gained from this study underscore the importance of maintaining microbial diversity in soil, which is essential for nutrient cycling, soil structure maintenance, and overall ecosystem resilience. The observed reduction in microbial diversity and the shift in community composition in response to imidacloprid highlight the need for careful management of pesticide use to prevent long-term detrimental effects on soil health and ecosystem stability. The impact of imidacloprid on microbial communities calls for further research to develop sustainable agricultural practices that minimize negative environmental impacts while ensuring crop protection.