16S rRNA Sequencing Report

Project ID: M01022021001

Sample received date: 01-02-2021

Reported date: 15-02-2021

Sample type: Fecal samples from swine

Sample size: 15 samples

16S rRNA Metagenomic Sequencing Report

☑ Quality profiles
☑ Alpha diversity
☑ Beta diversity
☑ Taxonomic profiles and Krona plots
☑ Linear discriminant analysis effect size (LEfSe) and Cladogram

Optional Bioinformatics

☑ Phylogenetic tree
☑ Heatmap of relative abundance
☑ Venn diagram
☑ Metabolic pathway prediction
☑ NCBI Genbank submission

16S rRNA Sequencing
(ตัวอย่างผลการวิเคราะห์เบื้องต้น)

Results

1. Sample information

The sample information table shows your sample ID, sample groups, and the names of fastq files obtained from 16S rRNA Metagenomic Sequencing. In this report, only 10 samples are shown in the following table. If there are more than 10 samples in a project, the complete sample information table can be accessed from the link below the table.

No.	Customer label	Sample ID	Group	Raw seq R1 file (fastq)	Raw seq R2 file (fastq)
1	SW1_1	SW_1_01	week 1	SW1_1_R1_001.fastq	SW1_1_R2_001.fastq
2	SW1_2	SW_1_02	week 1	SW1_2_R1_001.fastq	SW1_2_R2_001.fastq
3	SW1_3	SW_1_03	week 1	SW1_3_R1_001.fastq	SW1_3_R2_001.fastq
4	SW1_4	SW_1_04	week 1	SW1_4_R1_001.fastq	SW1_3_R2_001.fastq
5	SW1_5	SW_1_05	week 1	SW1_5_R1_001.fastq	SW1_3_R2_001.fastq
6	SW2_1	SW_2_01	week 2	SW2_1_R1_001.fastq	SW2_1_R2_001.fastq
7	SW2_2	SW_2_02	week 2	SW2_2_R1_001.fastq	SW2_2_R2_001.fastq
8	SW2_3	SW_2_03	week 2	SW2_3_R1_001.fastq	SW2_3_R2_001.fastq
9	SW2_4	SW_2_04	week 2	SW2_4_R1_001.fastq	SW2_3_R2_001.fastq
10	SW2_5	SW_2_05	week 2	SW2_5_R1_001.fastq	SW2_3_R2_001.fastq

Table:
Summary_samples_information.txt
Raw seq files:
All FASTQ files

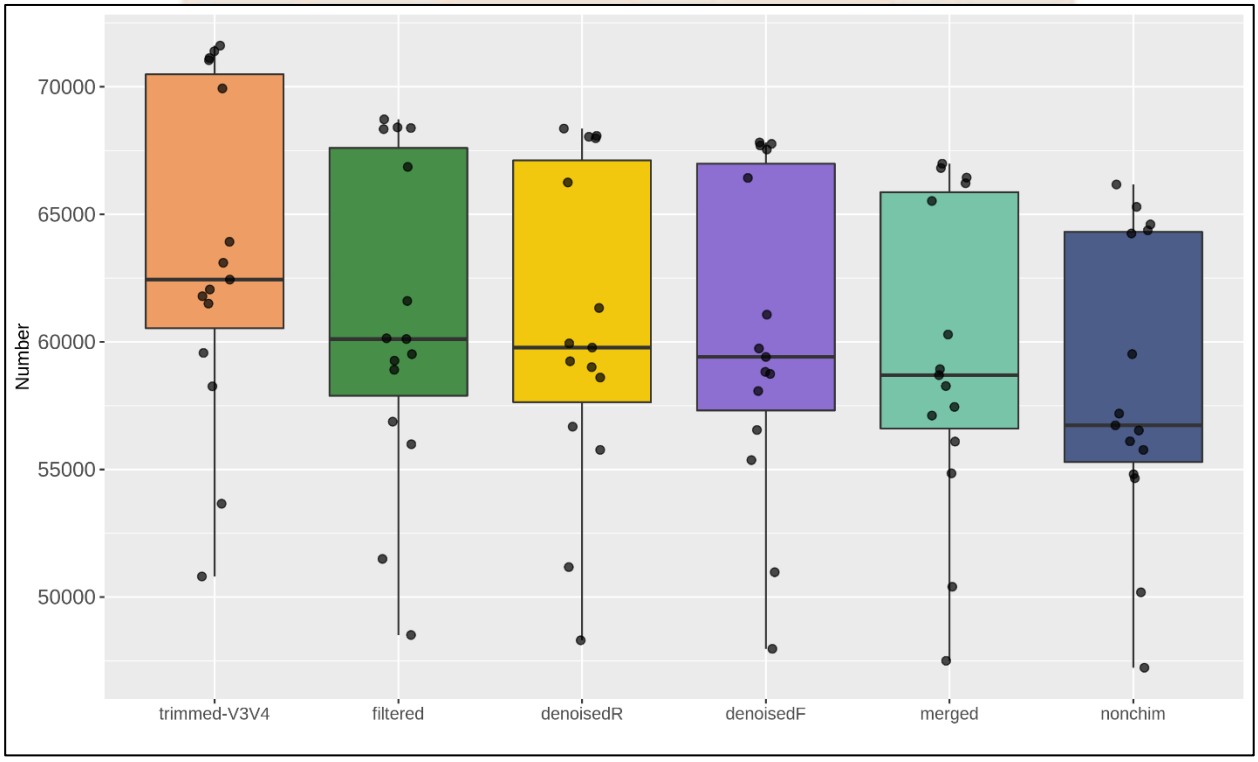
2. Quality profiles

Track Read Changes

The numbers of sequence read from all samples in each step of the pipeline analysis are shown to estimate the performance of the run and how do reads change through the pipeline. More than 50,000 reads are recommended from raw sequencing reads.

- Trimming is used for removing the primer and adaptor sequences from V3V4 amplicon reads.
- Filtering is used to filter low-quality sequences of both forward and reverse reads, which are unexpected reads. Normally, the quality of reverse read drops off at the end of read more than in the forward read.
- Denoising refers to a process that removes sequence errors from amplicon reads.
- Merging of paired-end reads generates one consensus sequence by assembly between the forward and reverse overlapping reads.
- Nonchim (refers to non-chimeric reads) is a process to remove chimeric reads. Chimeras occur during PCR reaction step when two sequencing reads incorrectly joined together. They are indicative of 16S structural variation.
- Percentage calculates the percent of remaining sequencing reads after non-chimeric process from total reads at the first step.

In this report, only 10 samples are provided in the following table. If there are more than 10 samples in your project, the complete sample information table can be accessed from the link below the table.



No.	Sample ID	Trimmed-V3V4	Filtered	DenoisedF	DenoisedR	Merged	Nonchim	%
1	SW_1_01	121,127	117,752	116,979	117,387	115,686	113,752	89.63
2	SW_1_02	113,924	111,328	110,795	111,050	109,994	109,229	92.66
3	SW_1_03	112,443	109,782	109,377	109,537	108,532	106,816	90.99
4	SW_1_04	121,044	118,046	117,329	117,732	116,070	114,335	90.56
5	SW_1_05	111,791	109,224	108,449	108,859	107,045	105,485	89.79
6	SW_2_01	100,805	98,095	97,556	97,838	97,028	96,781	92.08
7	SW_2_02	103,658	101,034	100,500	100,790	99,927	99,721	92.66
8	SW_2_03	111,501	108,504	107,681	108,149	106,663	105,715	90.59
9	SW_2_04	121,392	118,331	117,427	117,988	116,446	115,843	92.23
10	SW_2_05	108,258	105,563	104,940	105,347	104,394	104,202	93.04

PDF:
Box_plot_summary_filter.pdf
Table:
Summary_filter_reads.txt

Feature prevalence

Feature prevalence provides the number of samples in which a phylum appears at least once. The prevalence of phylum Actinobacteriota, Bacteroidota, Campilobacterota, Deferribacterota, Desulfobacterota, Fibrobacterota, Firmicutes, Patescibacteria, Planctomycetota, Proteobacteria, Spirochaetota, Synergistota, and Verrucomicrobiota are shown. The figure of total abundance from each phylum can be downloaded from the same filename below the figure.

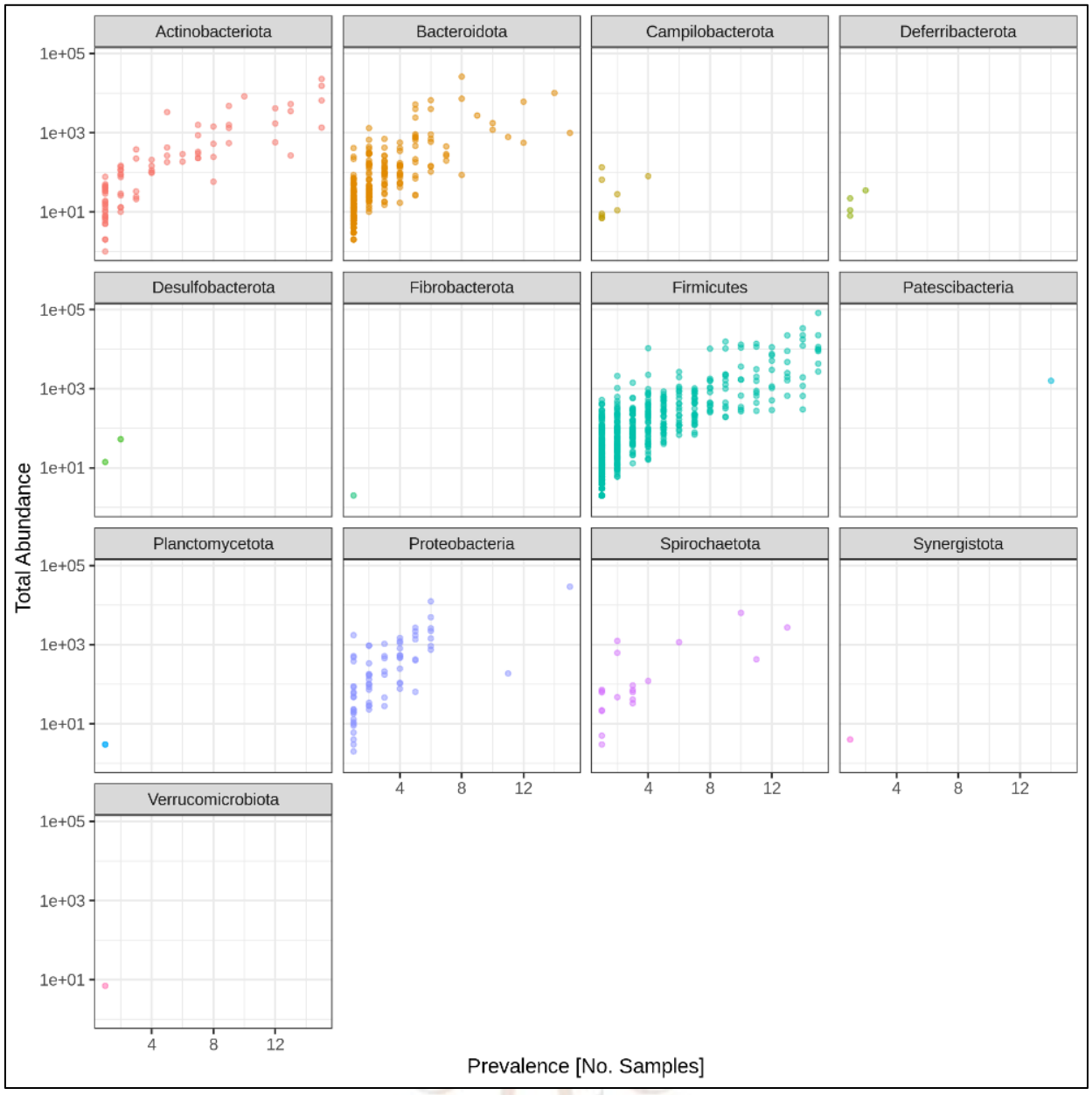


Figure:

Prevalence Phylum.png

Rarefaction analysis

Rarefaction curve represents the species richness (the number of different species) within and between sequencing reads. It can be used to estimate how many amplicon sequence variants (ASVs) or taxa would have been found in the same size of reads (1). Rarefaction curve rapidly increases at first where every read in the samples are identified (like the exponential phase), then slowly reaches the plateau stage when the rare species remain to be sampled (like the stationary phase). The plateau curve in rarefaction analysis determines whether sufficient reads have been detected to get a good representation of the microbial compositions in an environment. More samples can increase the number of reads with good representatives of all taxa. The figure of rarefaction curve can be downloaded with the same filename below the figure.

Suggestion by Porcinotec:

The approximate saturation of microbial richness of all samples was 42,000 sequencing depths, as estimated by the rarefaction curves. This finding sufficiently estimated the true bacterial compositions of gut microbiome in swine among the sample groups. Fecal samples obtained from week2 showed the highest number of observed ASVs, suggesting that the abundance of gut microbiota in week2 was relatively high compared to other groups.

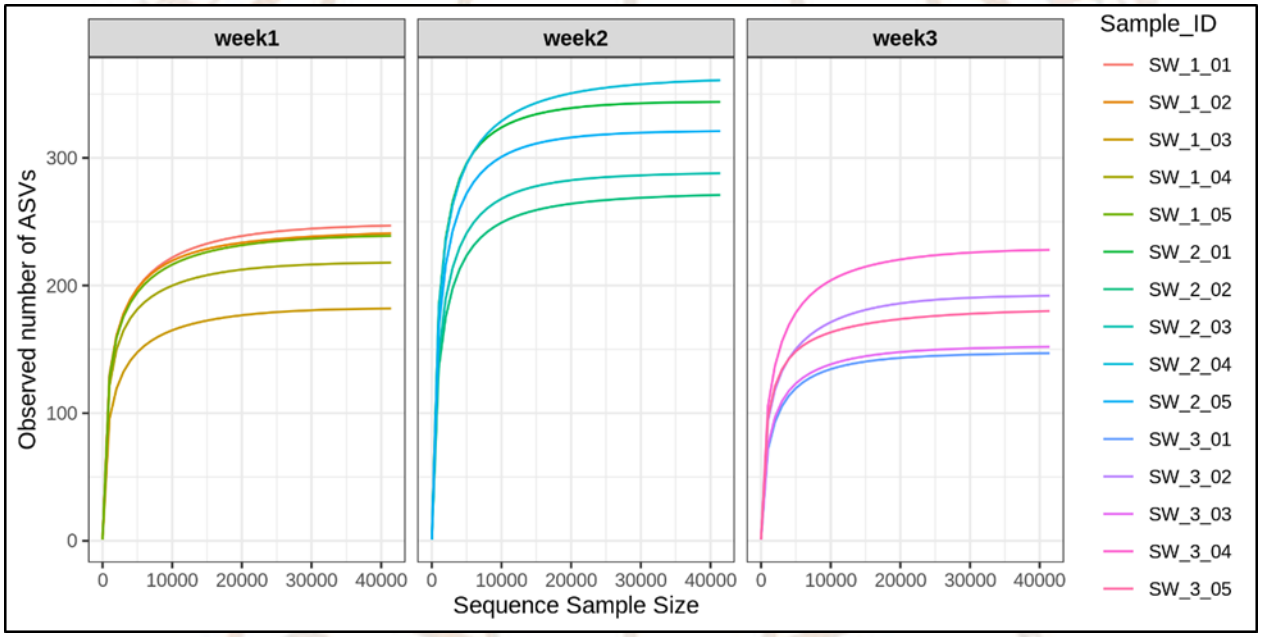


Figure:

Rarefaction_curves.png

3. Alpha diversity

Alpha diversity analyses refer to bacterial diversity within each community (1). Box plots of alpha diversity (observed species, Chao1, Shannon, and phylogenetic diversity (PD) whole tree) in each sample group are shown. The black dots represent individual samples in each group. Alpha diversity values of each sample and quartiles of the distribution (minimum, first quartile, median, third quartile, and maximum of boxes) can be found. The alpha-diversity values are also provided as the table from the link below figure.

Suggestion by Porcinotec:

High quality reads of 16S rRNA after processing were 1,500,000 reads. The percentage of Good's coverage for bacteria ranged from 98.34% to 99.96%, indicating that this data covered most of the bacterial diversity (>98%) in each environment. The observed abundance of ASVs in week2 was significantly higher than that in week1 and week3 ($p=0.0079$ and 0.0079, respectively). This result was consistent with the rarefaction curves in the previous data. Chao1 richness index showed that there were statistically significant differences in week2 when compared to other groups ($p=0.0079$). However, there was no significant difference in richness between week1 and week3. This data indicated that week2 group had the highest bacterial abundance. In contrast, both Shannon and PD whole tree significantly decreased in week3 compared with the other two groups ($p<0.05$ and 0.01, respectively). These findings illustrated that the highest bacterial abundance (Chao1) and diversity (Shannon) were found in week2, but the lowest microbiota was observed in week3 group. The decrease in microbial diversity might be affected by the ages of swine.

No.	Sample ID	Read	Observed ASVs	Chao1	Shannon	PD whole tree
1	SW_1_01	113,752	247	250.50	4.06	0.96
2	SW_1_02	109,229	244	244.33	4.01	0.97
3	SW_1_03	106,816	182	183.91	3.51	0.94
4	SW_1_04	114,335	222	224.33	3.94	0.96
5	SW_1_05	105,485	240	242.80	3.97	0.96
6	SW_2_01	96,781	347	348.50	4.44	0.97
7	SW_2_02	99,721	271	272.88	3.94	0.96
8	SW_2_03	105,715	289	291.14	4.07	0.96
9	SW_2_04	115,843	365	369.23	4.44	0.97
10	SW_2_05	104,202	325	325.00	4.21	0.96

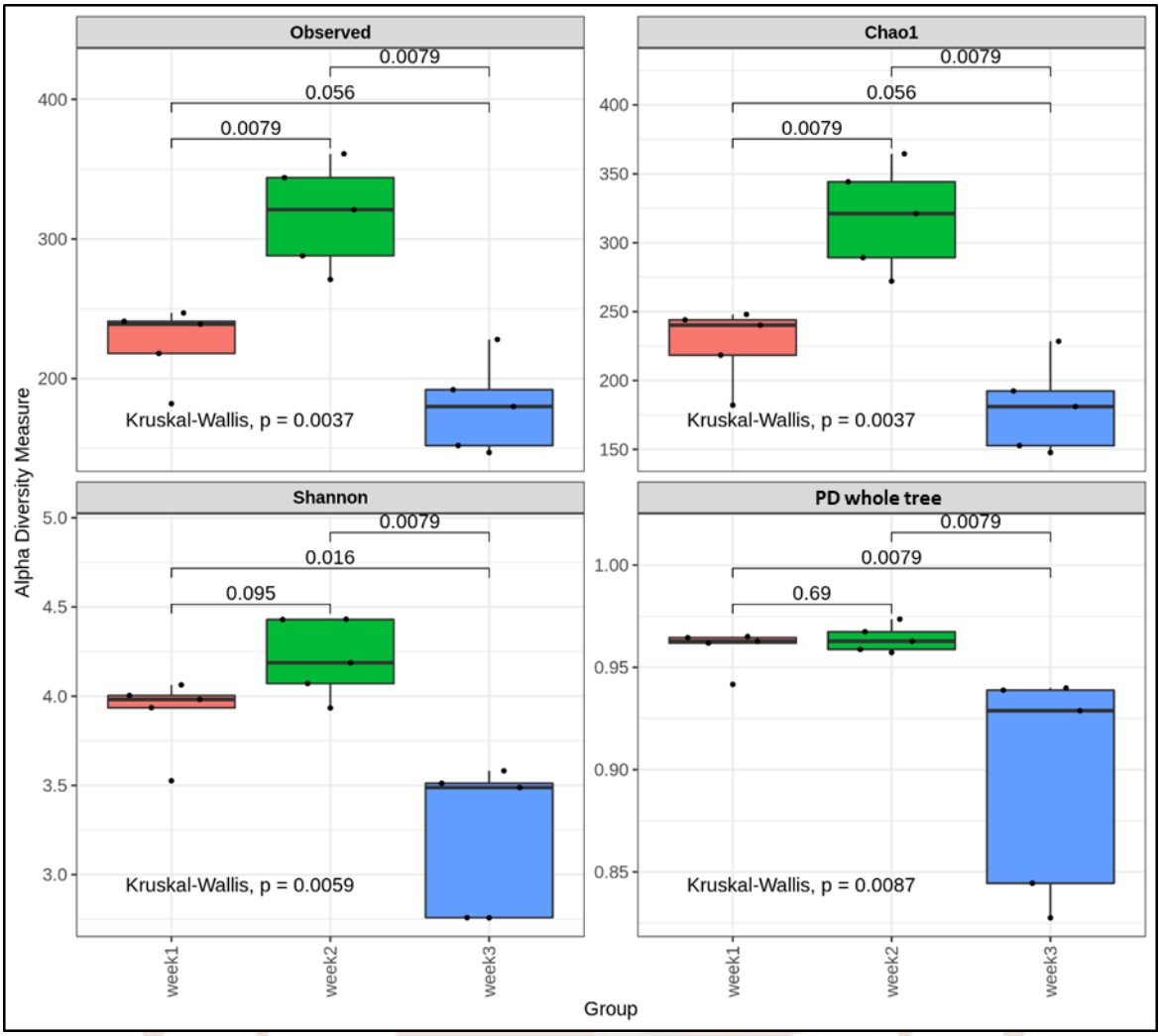


Table:
Summary_ richness.txt
Figure:
plot_richness_boxplot_Group.png

4. Beta diversity

Beta diversity analyses represent the difference in microbial community between samples, or a simpler definition would be, how similar or different it is between your samples. To estimate the beta diversity, principal coordinate analysis (PCoA) on weight/unweight Unifrac distances, generalized UniFrac (GUniFrac) distances, and non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity are usually performed (2).

PCoA is a multidimensional scaling method to visualize the dissimilarity of data (3). It uses the actual distances between samples as input for the dissimilarity matric. UniFrac distances measure the phylogenetic distance between a pair of samples (2). Weighted distance takes into account the relative abundance of taxa shared between samples combined with phylogenetic distance, but unweighted UniFrac distance considers a qualitative diversity metric (only presence/absence of taxa in a sample). However, these are limited to both rare phylogenies for unweighted UniFrac distance and most abundant lineages for weighted UniFrac distances. Therefore, GUniFrac distance has been developed to overcome the limitations on two UniFrac distances (4). Thus, GUniFrac distance can be used to detect a much wider range of changed microbiota composition.

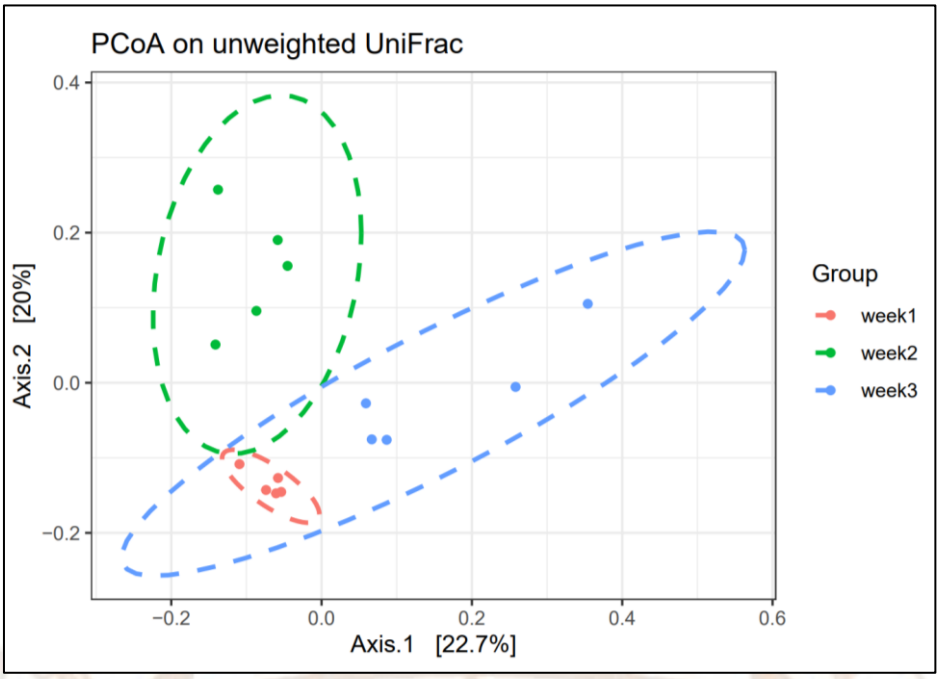
NMDS is non-parametric approach (3). The actual distances between samples are converted into rank orders for creating the dissimilarity matric. The Bray-Curtis dissimilarity is calculated based on non-phylogenetic measurement with microbial abundance between a pair of samples (similar to weighted Unifrac). Please note that NMDS is calculated based on non-metric distances, while PCoA is a matrix of dissimilarities between samples.

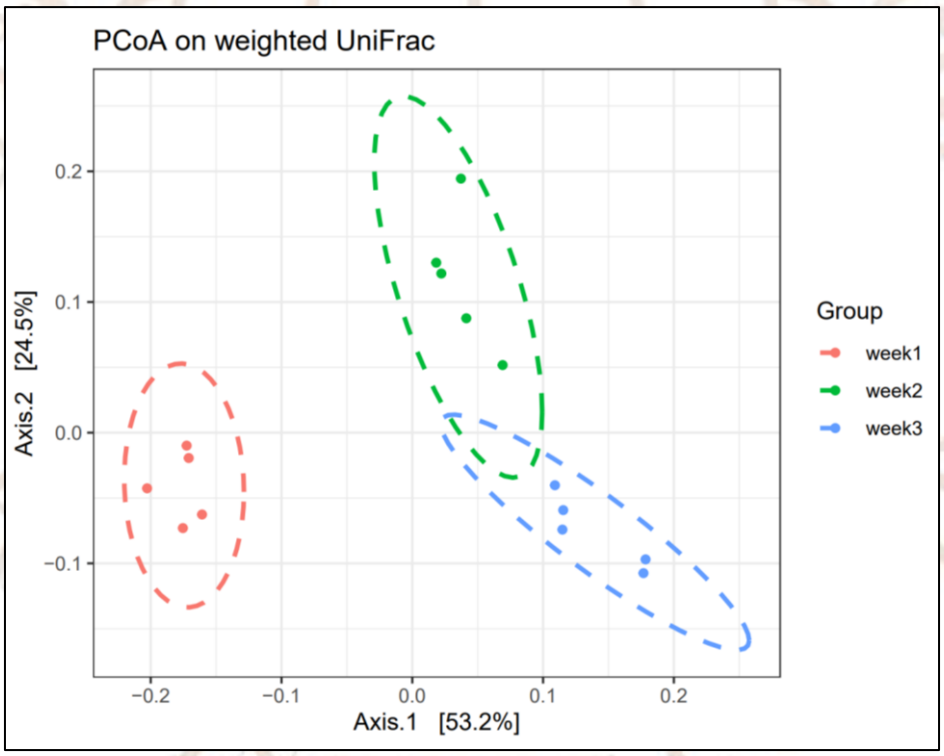
	Parametric distances	**Non-parametric distances**	**Phylogenetic**	**Relative abundance**
PCoA Weighted UniFrac	✓	-	✓	✓
PCoA Unweighted UniFrac	✓	-	✓	-
NMDS Bray-Curtis	-	✓	-	✓

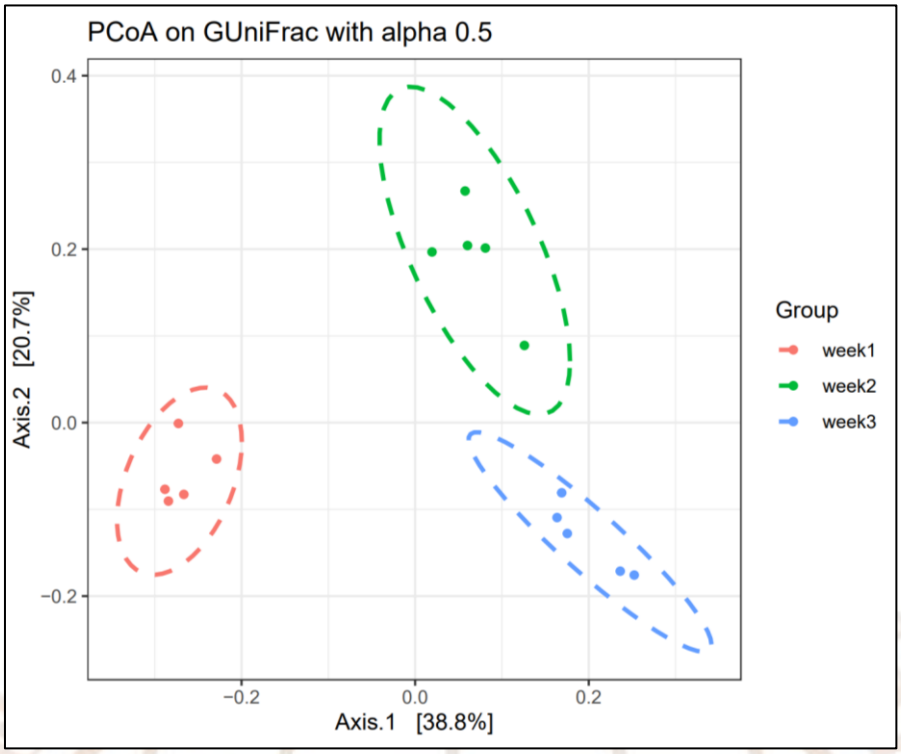
For beta-diversity plots, the 2-dimensional PCoA on weighted/unweighted Unifrac distances, GUniFrac distances, and NMDS based on Bray-Curtis dissimilarity are generated as a pdf document. In addition, PERMANOVA testing is also provided across all sample groups and pairwise between samples. The customer can propose the parameters of interest to be analyzed.

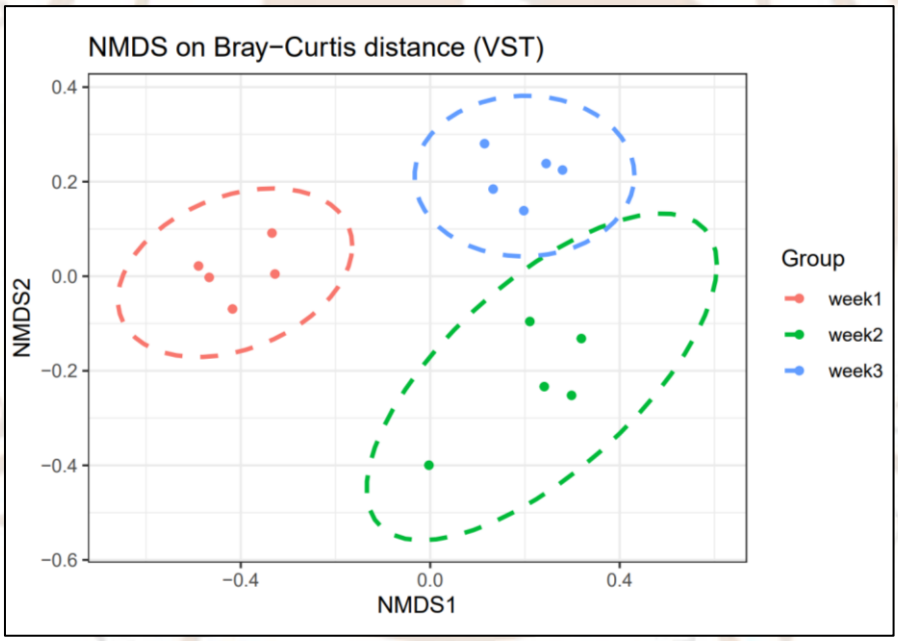
Suggestion by Porcinotec:

The weighted and unweighted UniFrac PCoA and NMDS based on Bray-Curtis distance showed that microbiota communities were clearly distinct in all groups (PERMANOVA test; $p<0.001$). Moreover, pairwise testing between each sample group was also statically significant different ($p<0.01$). This finding indicated that the age of animals affected on microbial community.


PCoA on unweighted UniFrac
0.4
0.2
0.0
−0.2
Axis.2 [20%]
−0.2
0.0
0.2
0.4
0.6
Axis.1 [22.7%]
Group
week1
week2
week3


PCoA on weighted UniFrac
0.2
0.1
0.0
−0.1
Axis.2 [24.5%]
−0.2
−0.1
0.0
0.1
0.2
Axis.1 [53.2%]
Group
week1
week2
week3





PDF:

plot_ordination_Group.pdf

PERMANOVA Table:

statistic_PERMANOVA_Unweighted_UniFrac_beta_diversity.txt

statistic_PERMANOVA_Unweighted_UniFrac_ Pairwise_beta_diversity.txt

statistic_PERMANOVA_Weighted_UniFrac_beta_diversity.txt

statistic_PERMANOVA_Weighted_UniFrac_Pairwise_beta_diversity.txt

statistic_PERMANOVA_GUniFrac_with_alpha0.5_beta_diversity.txt

statistic_PERMANOVA_GUniFrac_with_alpha0.5_beta_Pairwise_diversity.txt

statistic_PERMANOVA_Bray-Curtis_beta_diversity.txt

statistic_PERMANOVA_Bray-Curtis_ Pairwise_beta_diversity.txt

5. Taxonomic profiles

Bar chart

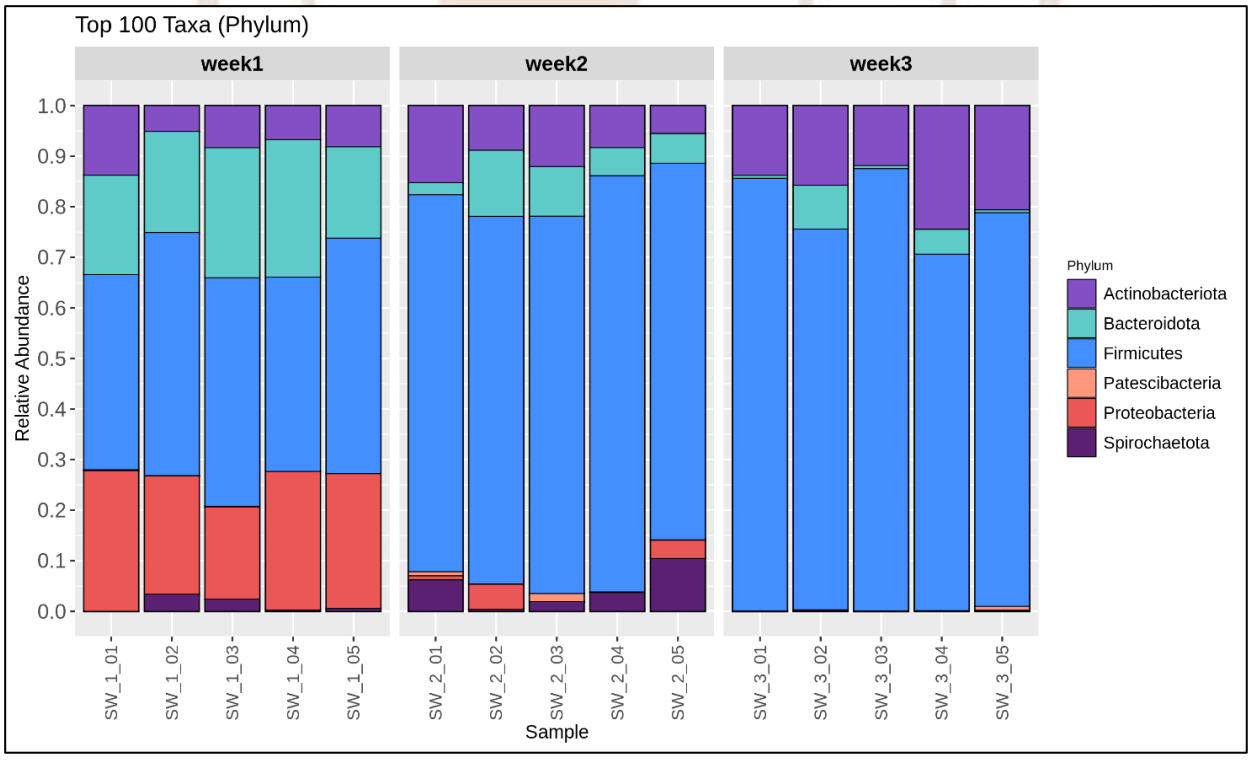
Microbiota compositions of different taxa profiles (phylum, class, order, family, and genus) are frequently visualized by bar charts. Different color represents different taxa compositions. Microbial composition graph of each taxonomy level and abundance tables can be accessed from the link below the figure. You can visualize the abundance of each bacterium on the HTML files.

Suggestion by Porcinotec:

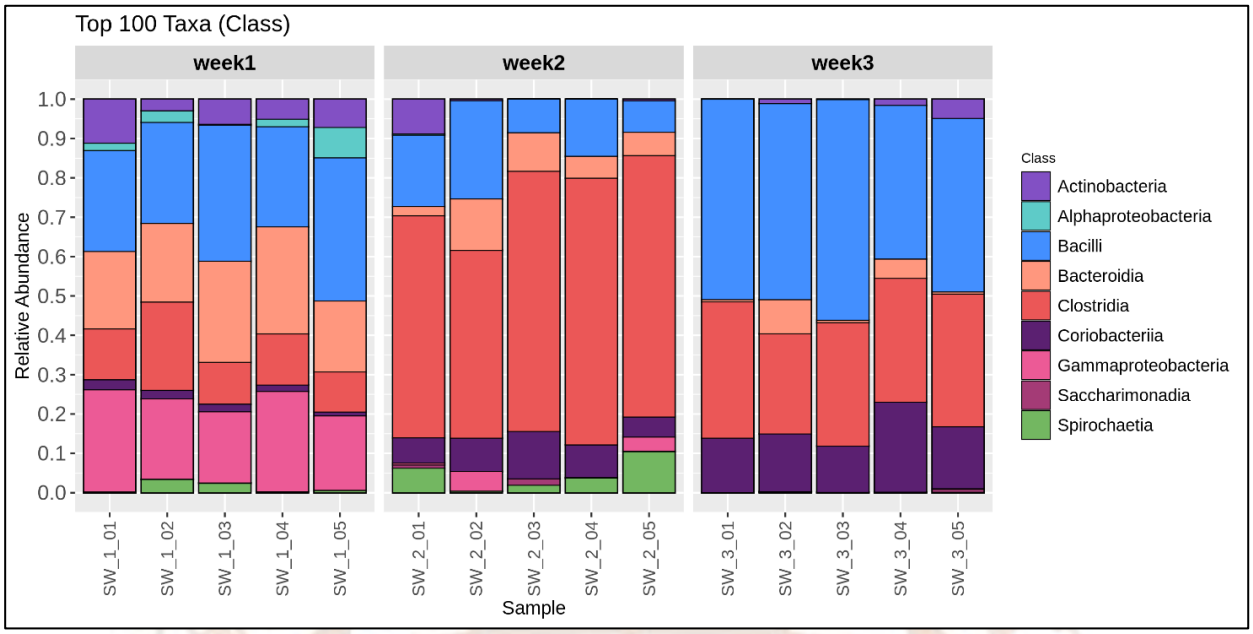
A total of 45 different bacterial phyla were identified. Only six enriched phyla were shown for the top 100 taxonomic classification. The bacteria of phylum Firmicutes were highly prevalent (avg. 63.42±12.16%), followed by Proteobacteria, Bacteriodota, and Actinobacteriota, respectively. Proteobacteria, Firmicutes, and Bacteroidota dominated in week1. In week3, the abundance of Proteobacteria and Bacteriodota were replaced by Firmicutes phyla compared to week1 group ($p<0.05$). This finding indicated that age was associated with gut microbiome of swine. The variation in microbial diversity at class level decreased from week1 to week3. Bacteria in the class of Clostridia significantly increased in week2, while Bacilli and Coriobacteria were relatively high in week3.

Overall, 355 genera were detected among samples. The relative abundance of bacteria in members of *Myroides*, *Kurthia*, *Escherichia*/*Shigella*, *Enterococcus*, and *Acinetobacter* were significantly increased in week1 when compared to other groups ($p<0.05$). More variation in relative abundance of microbial taxa was observed in week2. Interestingly, *Catenibacterium* genera which belongs to the Firmicutes was significantly predominant in week3, followed by *Subdoligranulum*, and *Lactobacillus*. This finding suggested that the gut microbiome of swine varied according to the ages.

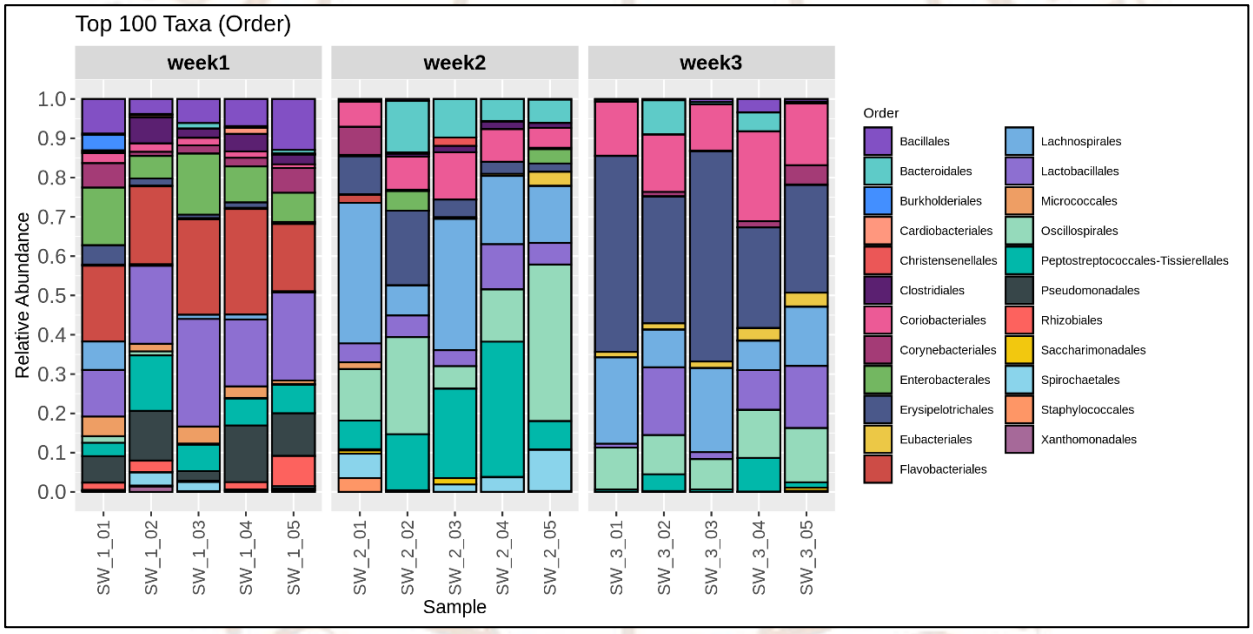
Phylum



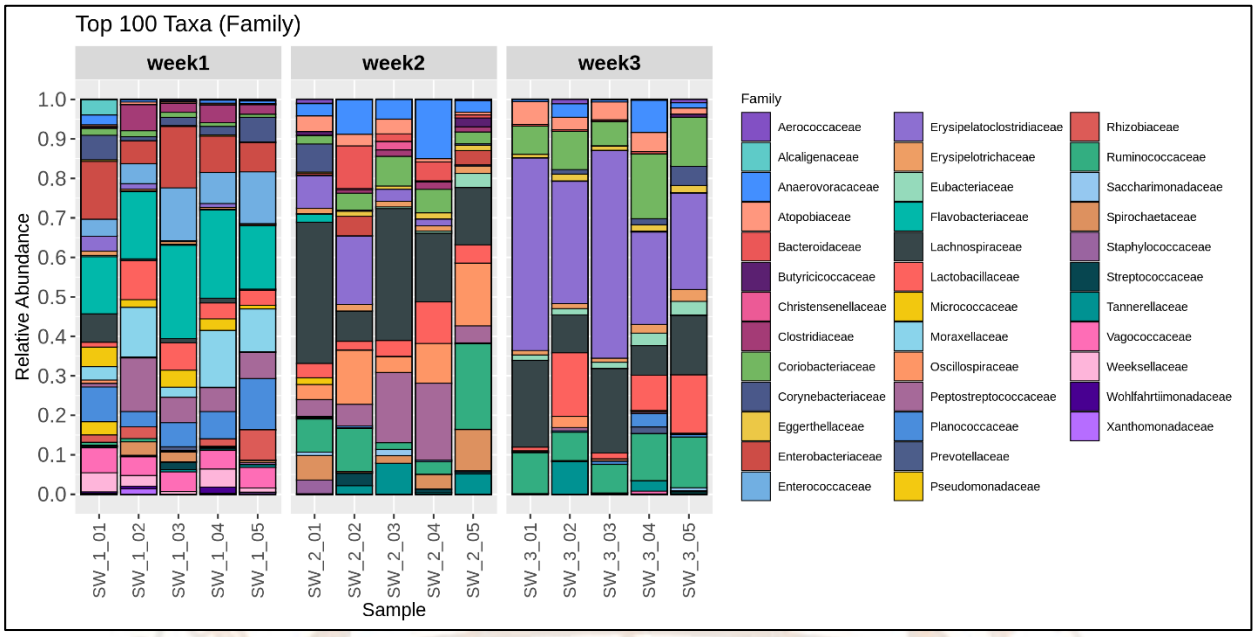
Class


Top 100 Taxa (Class)
week1
week2
week3
Relative Abundance
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
SW_1_01
SW_1_02
SW_1_03
SW_1_04
SW_1_05
SW_2_01
SW_2_02
SW_2_03
SW_2_04
SW_2_05
SW_3_01
SW_3_02
SW_3_03
SW_3_04
SW_3_05
Sample
Class
Actinobacteria
Alphaproteobacteria
Bacilli
Bacteroidia
Clostridia
Coriobacteriia
Gammaproteobacteria
Saccharimonadia
Spirochaetia

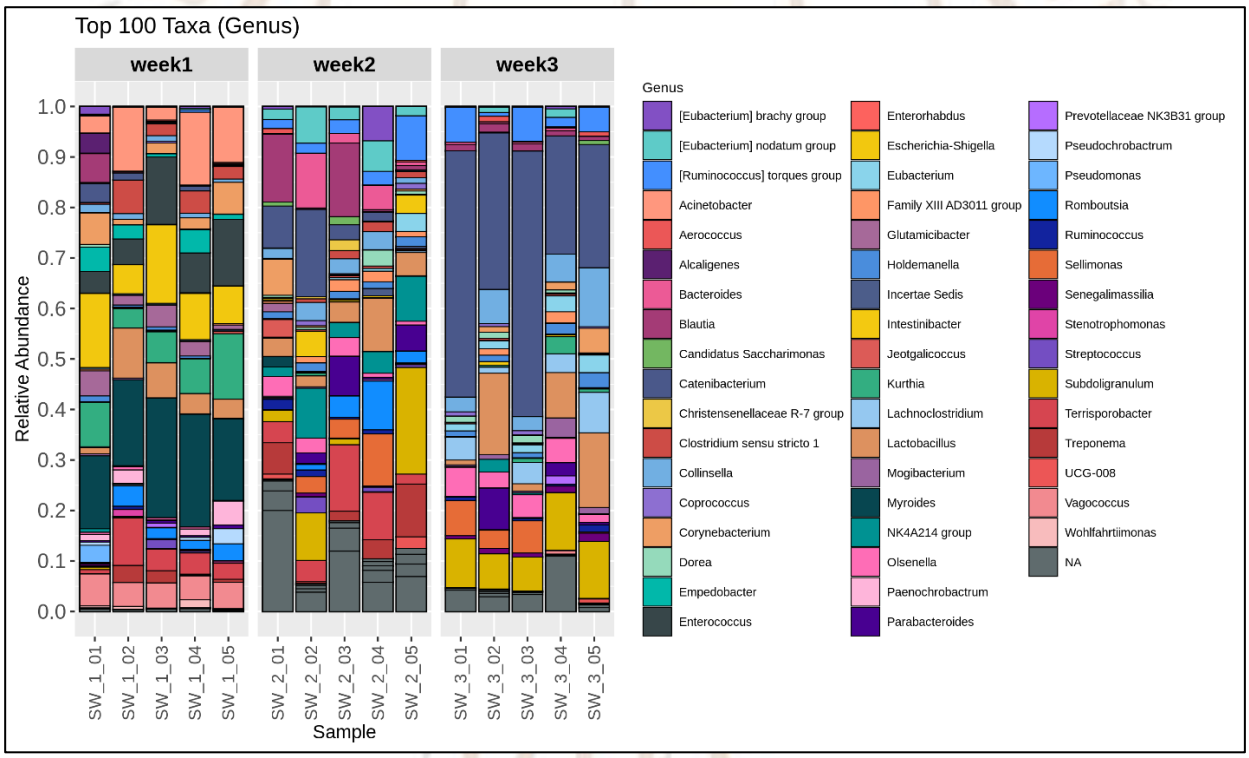
Order


Top 100 Taxa (Order)
week1
week2
week3
Relative Abundance
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
SW_1_01
SW_1_02
SW_1_03
SW_1_04
SW_1_05
SW_2_01
SW_2_02
SW_2_03
SW_2_04
SW_2_05
SW_3_01
SW_3_02
SW_3_03
SW_3_04
SW_3_05
Sample
Order
Bacillales
Bacteroidales
Burkholderiales
Cardiobacteriales
Christensenellales
Clostridiales
Coriobacteriales
Corynebacteriales
Enterobacterales
Erysipelotrichales
Eubacteriales
Flavobacteriales
Lachnospirales
Lactobacillales
Micrococcales
Oscillospirales
Peptostreptococcales-Tissierellales
Pseudomonadales
Rhizobiales
Saccharimonadales
Spirochaetales
Staphylococcales
Xanthomonadales

Family



Genus



Microbial all taxa (Figure):

Phylum; plot_bar_phylum_Group.png

Class; plot_bar_class_Group.png

Order; plot_bar_order_Group.png

Family; plot_bar_family_Group.png

Microbial top 100 taxa (Figure):

Phylum; plot_bar_phylum_top100taxa_Group.png

Class; plot_bar_class_top100taxa_Group.png

Order; plot_bar_order_top100taxa_Group.png

Family; plot_bar_family_top100taxa_Group.png
Genus; plot_bar_genus_top100taxa_Group.png

Microbial all taxa (HTML):

Phylum; plot_bar_phylum_Group.html
Class; plot_bar_class_Group.html
Order; plot_bar_order_Group.html
Family; plot_bar_family_Group.html

Microbial top 100 taxa (HTML):

Phylum; plot_bar_phylum_top100taxa_Group.html
Class; plot_bar_class_top100taxa_Group.html
Order; plot_bar_order_top100taxa_Group.html
Family; plot_bar_family_top100taxa_Group.html
Genus; plot_bar_genus_top100taxa_Group.html

Abundance tables:

expr.abundance.all.txt
expr.relative_abundance.all.txt
expr.asv.fasta

Krona plot

Krona plot is multilevel pie chart used for visualizing taxonomic classification (5). The output file is an HTML file that can be accessed in the link below the figure. Figures with specific taxonomy can be captured on the HTML for publication. If your samples do not have any replicate, we can provide individual plot of Krona. In contrast, if your samples are defined into a group, the Krona plot for grouped data can be created to compare the taxa among groups.

HTML:

SampleID-krona.html

Group-krona.html

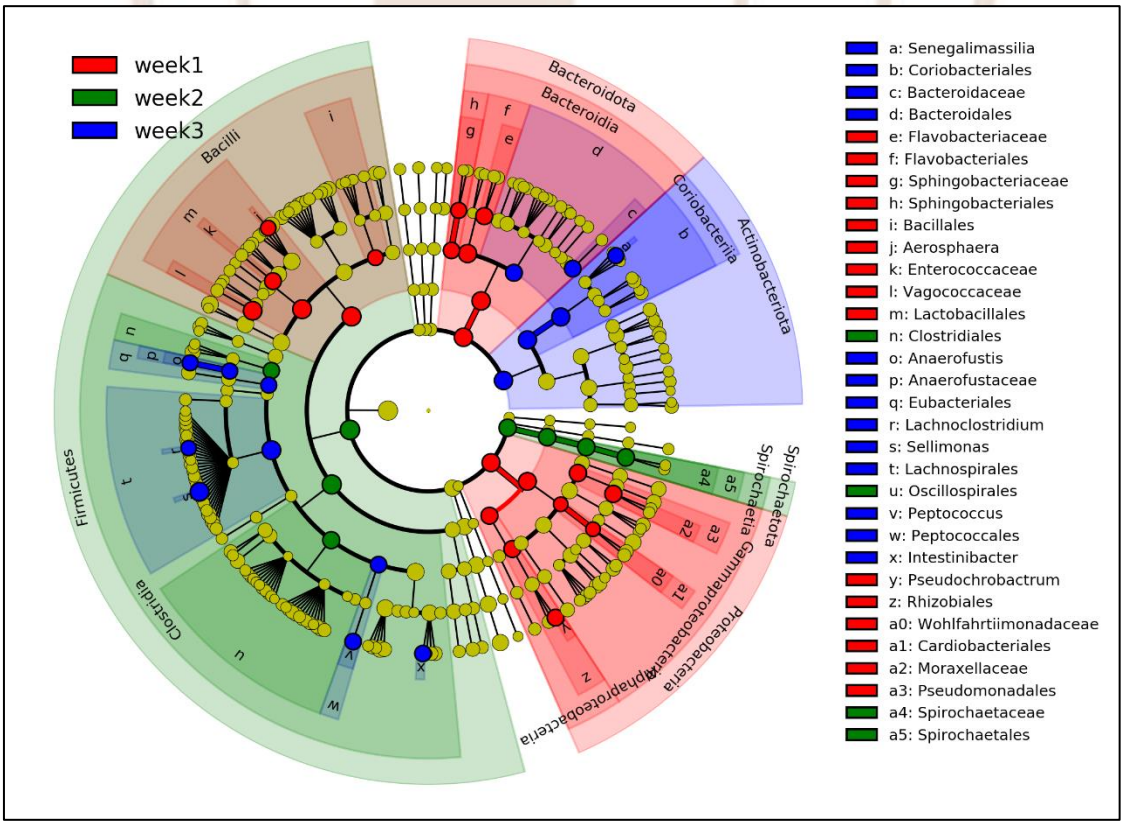
6. Linear discriminant analysis effect size (LEfSe) and Cladogram

LEfSe is used to determine the significantly higher taxonomy, genes, or functions, which can explain the difference in taxa between groups (6). It is usually used to identify biomarkers between 2 or more sample groups based on bacterial relative abundances. The bar plot represents the effect size (LDA) for a significant taxon in a certain group. The length of the bar represents a log10 transformed LDA score. The colors represent which group that taxa are highly presented compared to the other group. We also provide the taxonomic table of LEfSe input as the following table.

Cladogram explains the differentially abundant taxonomic clades according to LEfSe analysis. The dot color and shading represent significantly higher abundance of taxon in a certain group. The significant phyla are presented as dots in the centre, while the significant genera are shown in the outer circle. The name of significant phyla is given in the outermost circle for colored shading. The results from both LEfSe and cladogram are similar. Please notice that the result of LEfSe shows significant difference of single taxa level, while cladogram shows the different taxonomic clades (from phylum to genus) among groups.

Suggestion by Porcinotec:

Bacterial taxa with LDA scores greater than 4 was shown in this report. Bacteria in phylum Proteobacteria and Bacteroidota, class Gammaproteobacteria and Bacteroidia, *Bacilli*, and *Lactobacillales* were the core gut microbiota in swine at week1 ($p<0.05$). The Firmicutes, Clostridia, order Oscillospirales, and Spirochaetales contributed to dominant biomarkers in week2 ($p<0.05$). The Actinobacteriota, Coriobacteriia, *Lachnospirales*, *Lachnoclostridium*, and *peptococcus* (belonging to phylum Firmicutes) were highly prevalent in week3 ($p<0.05$). These results indicated that this core microbiota contributed to different phylotypes between gut microbiota of swine at different ages.



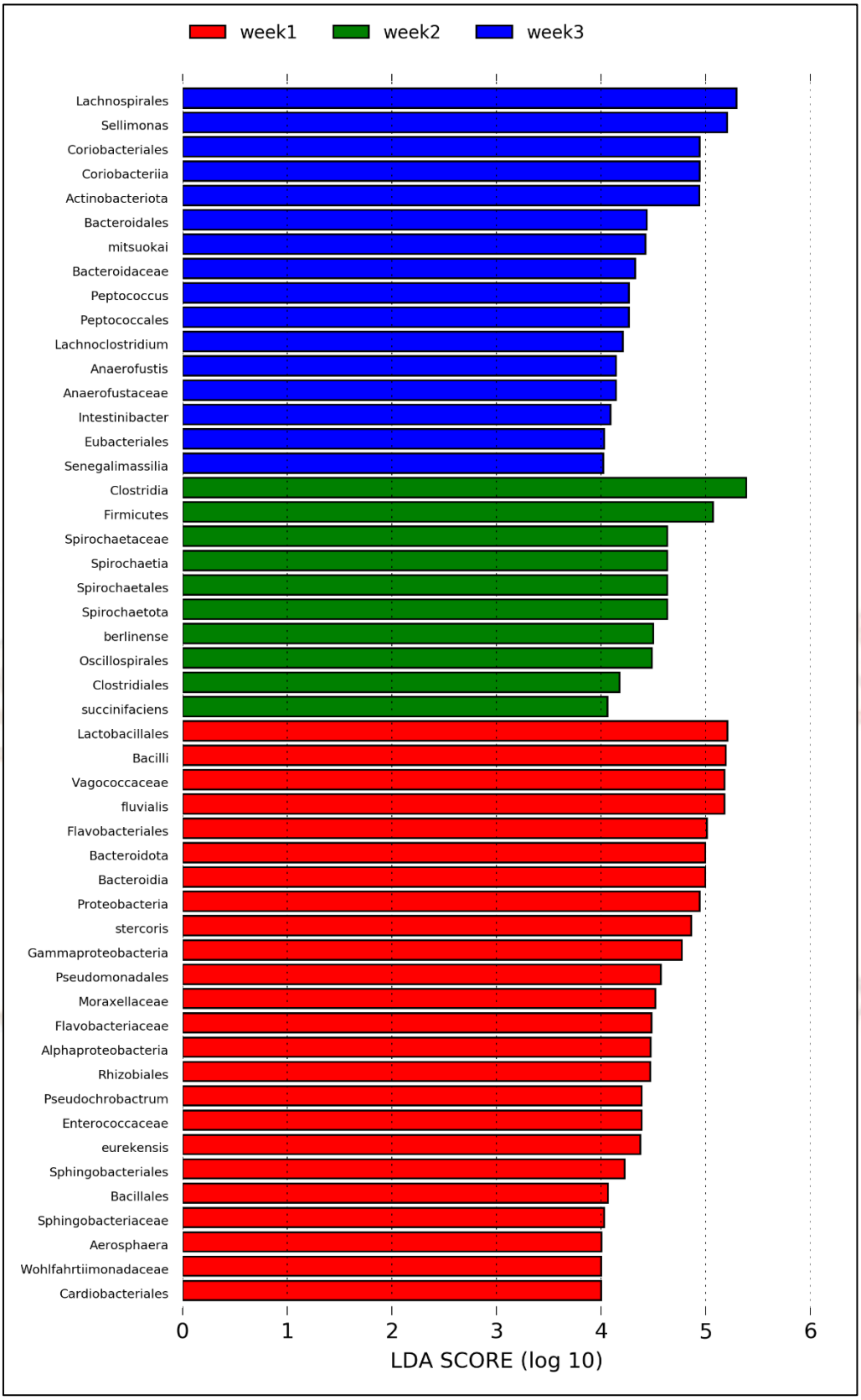


Figure:

Plot_LDA_score.png

Plot_Cladogram.png

Table:

expr1.lefse_table.txt

7. Optional requirements

7.1 Phylogenetic tree

Phylogenetic tree shows the evolutionary relationships among a set of taxa. The phylogenetic tree is important to study the bacterial evolution based on the difference or similarity in the genetic (7). The branching in the phylogenetic tree indicates how each bacterium evolved from a series of common ancestors. In this report, the phylogenetic tree is given in couple with the proportions of bacterial abundance. The colors and sizes represent different groups and taxa abundance, respectively. The figure of phylogenetic tree can be downloaded with the same filename below the figure.

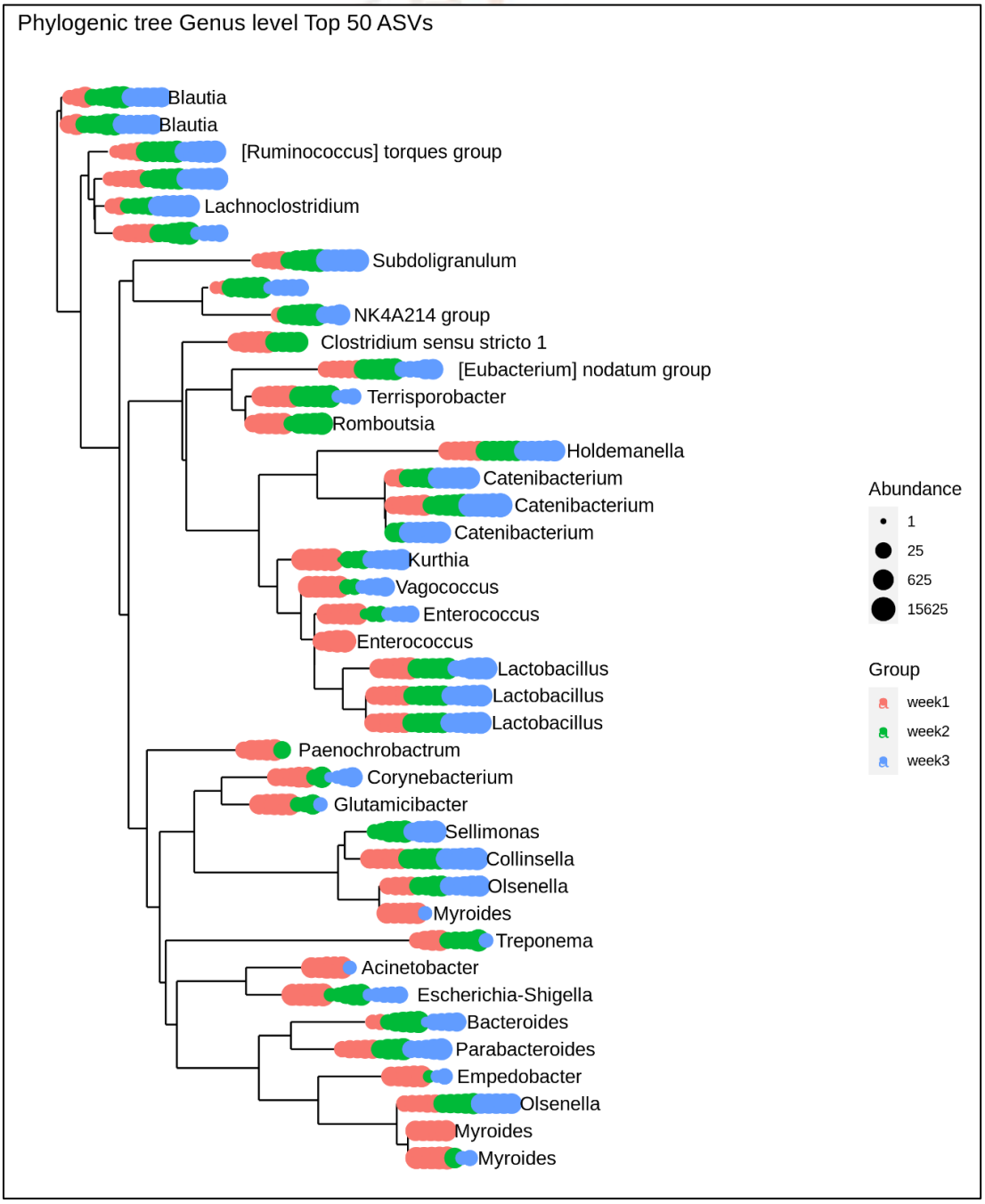


Figure:
Phylogenic_tree_Genus_level_top50ASVs_Group.png

7.2 Heatmap of relative abundance

Heatmap is a graphical representation of data where different matric values are represented as different shading of colors. It is necessary for visualizing the concentration of values between two dimensions of a matrix. In this case, heatmap is used for visualizing the bacterial abundance between samples. The heatmap figure can be downloaded with the same filename below the figure.

Suggestion by Porcinotec:

Twenty-eight ASVs were more abundant in week1, 24 were more abundant in week2, and 29 were more abundant in week3. The Abundance of bacteria in genus *Enterococcus*, *Paenochrobactrium*, *Acenitobacter*, *Myroides*, *Kurthia*, and *Escherichia-Shigella* were highly enriched in week1. However, most of them were not found in gut microbiome of week2 and week3 groups. On the other hand, unclassified species in genus *Catenibacterium*, *Eubacterium*, *Sellimonas*, *NK4A214*, and family *Oscillospiraceae* were rarely found in week1 ($p<0.05$). Interestingly, the abundance of members for *Terrisporobacter*, *Streptococcus*, *Treponema*, *Clostridium sensu stricto 1*, and *Romboutsia* significantly decreased in only week2 when compared to week1 and week3 ($p<0.05$). This result suggested that the distribution and abundance of gut microbiome in swine could vary according to age resulting from solid and liquid feedings.

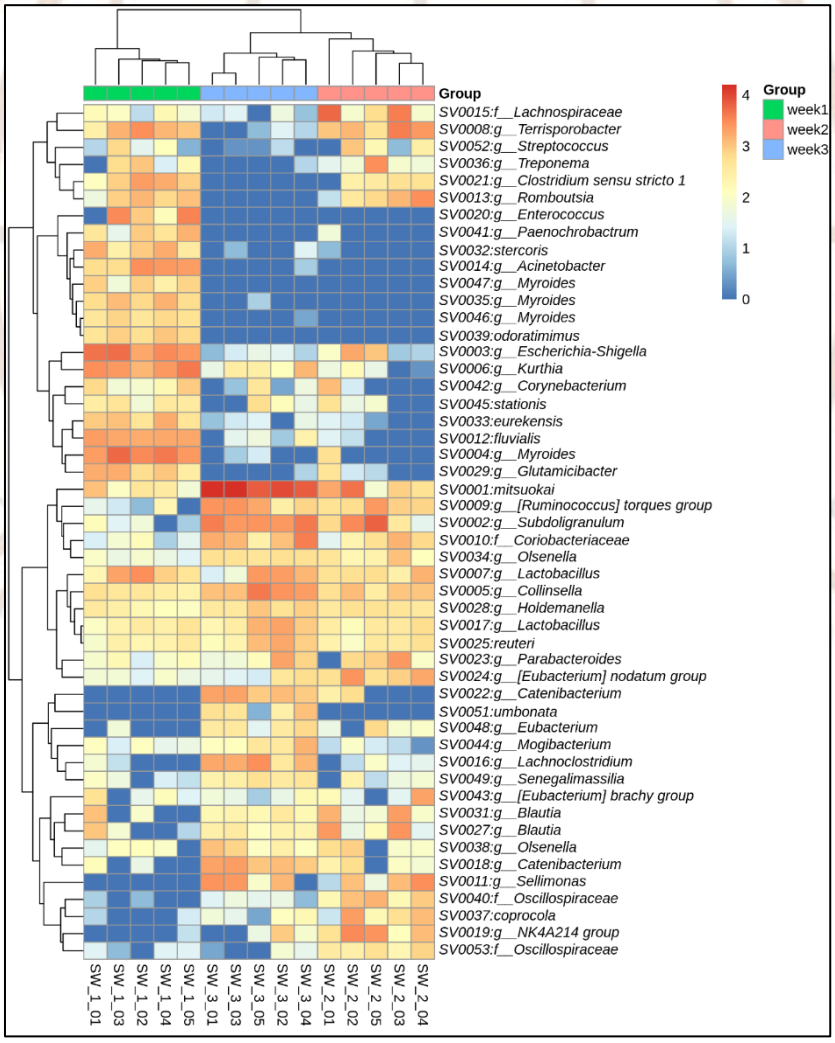


Figure:

Heat_map_label_ASV_number_top50ASVs.png

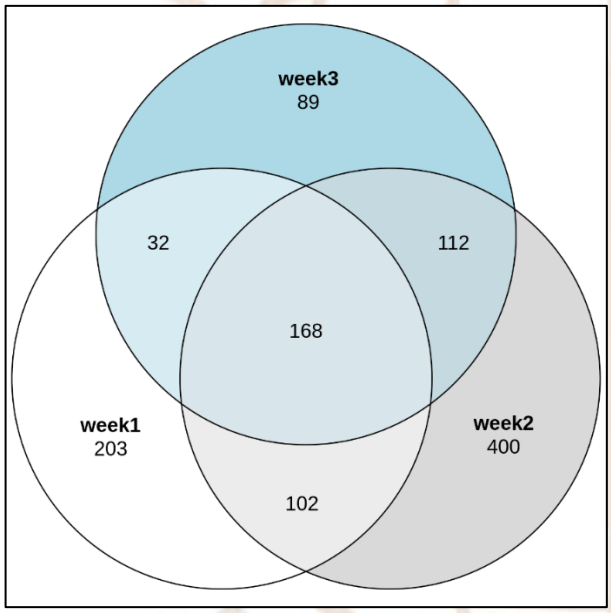
7.3 Venn diagram

Venn diagram shows the core, shared, and individual microbiomes among groups. A core microbiota in the Venn diagram is based on shared ASV occurrences across groups (membership), which can explain the distribution of ASV in your samples (8). Understanding which ASV members are the core microbiota can predict community responses to perturbation. This core microbiota can identify what microorganisms are uniquely associated with your conditions. In this report, we provide the numbers of ASV occurrences and the percentages of ASV across groups. The figure of Venn diagram can be downloaded with the same filename below the figure.

Suggestion by Porcinotec:

A Venn diagram showed that 168 genera (avg. 15%) were common in all groups and that 203, 400, and 89 ASVs were unique in week 1, week2, and week3, respectively. Week1 and week3 groups shared the lowest number of ASVs (200 ASVs or 18%). Week2 shared 270 ASVs with week1 and 280 ASVs with week3. Overall, 168 ASVs were considered as the core microbiota in gut microbiome in different ages of swine.

ASV count



Percentage

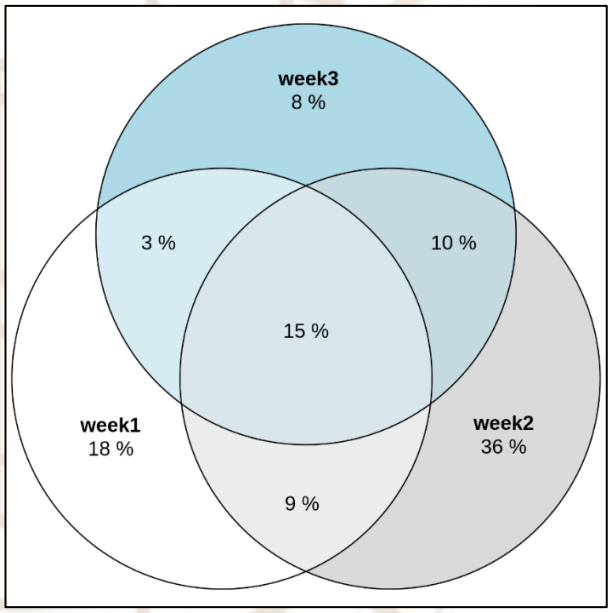


Figure:

Venn_diagram_counts.png

Venn_diagram_percent.png

7.4 Metabolic pathway prediction

The metabolic profiles cannot be directly identified by 16S rRNA gene sequences. However, these metabolic functions can be predicted on database with a reference genome. The Kyoto Encyclopedia of Genes and Genome (KEGG) Orthology database is used for prediction of the metabolic potentials into the categories of genes (COGs). Phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) is an analytical tool used for predicting the abundance of metabolic potential of a microbial community (9). The prediction of metabolic potentials is very useful to understand the change of metabolic pathways from microbiota in your conditions. The figure of metabolic potentials can be downloaded with the same filename below the figure. If you need to create other levels of metabolic functions, you can download spf file below this figure and visualize on Statistical Analysis of Metagenomic Profiles (STAMP) analysis (10).

Suggestion by Porcinotec:

The metabolic potentials from gut microbiota between week1 and week2 were compared based on their 16S rRNA profiles. Twenty-six KEGG pathways showed statistically significant difference between week1 and week2 (in range of p<1e-15 to p=0.024). Several metabolic functions were significantly higher in week1 than those in week2 (p<0.001), including carbohydrate metabolism, amino acid metabolism, cellular processes and signaling, metabolism, nucleotide metabolism, membrane transport, and energy metabolism. On the other hand, the xenobiotic biodegradation and metabolism, lipid metabolism, signal transduction, cell motility, metabolism of other amino acids, biosynthesis of other secondary metabolites, and infectious diseases were highly prevalence in week2 than week1 (p<0.001). The different metabolic potentials resulted from variation of gut microbiota.

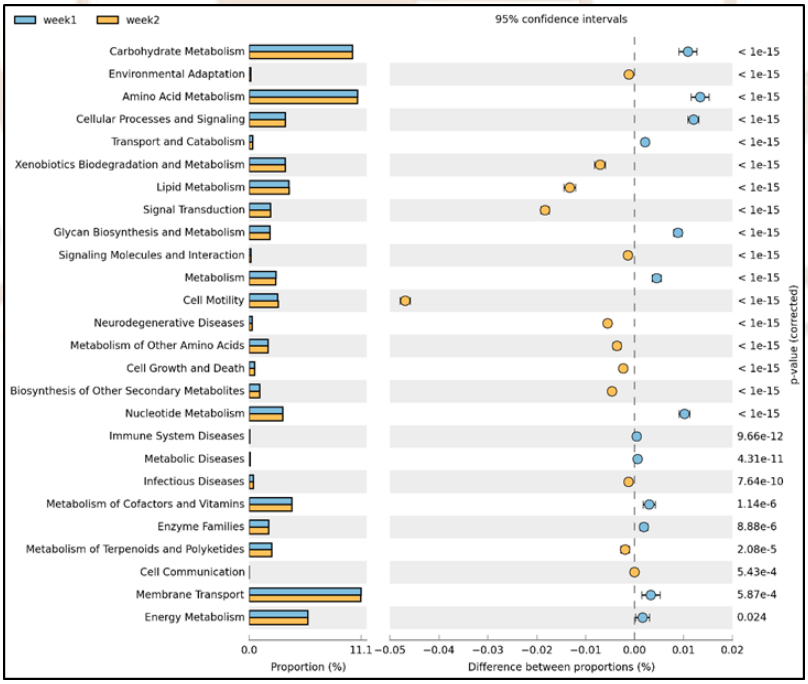


Figure:

Metabolic_pathway.png

File:

Metabolic_pathway.spf

8. Pipeline analysis

	Pipeline	**Detail**
Taxonomic reference data	Silva version 138	https://www.arb-silva.de/documentation/release-138/
Bioinformatics' pipelines	DADA2 v1.16.0	https://benjjneb.github.io/dada2/
R	R version 4.0.4 (2021-02-15)	https://www.r-project.org/
R packages	venneuler_1.1-0 rJava_0.9-13 dendextend_1.14.0 RColorBrewer_1.1-2 viridis_0.5.1 viridisLite_0.3.0 microbiomeutilities_1.00.15 microbiome_1.10.0 ggpubr_0.4.0 ranacapa_0.1.0 plotly_4.9.3 vegan_2.5-7 lattice_0.20-41 permute_0.9-5 ggrepel_0.9.1 ggbeeswarm_0.6.0 DESeq2_1.30.1 SummarizedExperiment_1.20.0 Biobase_2.50.0 MatrixGenerics_1.2.1 matrixStats_0.58.0 GenomicRanges_1.42.0 GenomeInfoDb_1.26.4 rexmap_1.1	forcats_0.5.1 stringr_1.4.0 dplyr_1.0.5 purrr_0.3.4 readr_1.4.0 tidyr_1.1.3 tibble_3.1.0 tidyverse_1.3.0 DECIPHER_2.16.1 RSQLite_2.2.5 Biostrings_2.58.0 XVector_0.30.0 IRanges_2.24.1 S4Vectors_0.28.1 BiocGenerics_0.36.0 ggplot2_3.3.3 phangorn_2.6.2 ape_5.4-1 phyloseq_1.34.0 data.table_1.14.0 dada2_1.16.0 Rcpp_1.0.6 microbiomeMarker_0.0.1.9000 pairwiseAdonis_0.0.1

9. Material and methods

Metagenomic DNA extraction

Metagenomic DNA for prokaryotes was isolated using QIAamp PowerFecal Pro DNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocols. Briefly, 0.25 g of fecal samples were extracted. The quality of the extracted DNA was determined via DeNovix QFX Fluorometer.

16S rRNA library sequencing

The prokaryotic 16S rRNA gene at V3V4 region was performed using the Qiagen QIAseq 16S/ITS Region panel (Qiagen, Hilden, Germany). 16S rRNA amplicons were labeled with different sequencing adaptors using QIAseq 16S/ITS Region Panel Sample Index PCR Reaction (Qiagen, Hilden, Germany). The quality and quantity of DNA libraries were evaluated using DeNovix QFX Fluorometer and QIAxcel Advanced (Qiagen, Hilden, Germany), respectively. 16S rRNA libraries were sequenced using an illumina Miseq600 platform (Illumina, San Diego, CA, USA).

Bioinformatics analyses

The raw sequences were categorized into groups based on the 5′ barcode sequences. The sequences were processed following DADA2 v1.16.0 pipeline (https://benjjneb.github.io/dada2/). The DADA2 pipeline describes microbial diversity and community structures using unique amplicon sequence variants (ASVs) (11). Microbial taxa were classified from Silva version 138 as a reference database (12). Alpha diversity index (Chao1 richness, Shannon, and PD whole tree) was computed using DADA2 software. For Bata diversity, non-metric multidimensional scaling (NMDS) based on Bray-Curtis dissimilarity and principal coordinate analysis (PCoA) were plotted from Phyloseq data. Linear discriminant analysis effect size (LEfSe) and cladogram plot were performed to identify the bacterial biomarkers. Venn diagram was also used to identify the numbers of core microbiome. To study the bacterial correlated evolution, the phylogenetic tree was computed. Moreover, the metabolic potentials were conducted using PICRUSt software via the KEGG database (9, 10).

Data analysis

Pairwise comparison of alpha diversity (Observed ASVs, Chao1, Shannon, and PD whole tree) was calculated using Kruskal-Wallis test ($p<0.05$). Permutational multivariate analysis of variance (PERMANOVA) was performed to evaluate the significant differences for beta diversity among groups at $p<0.05$. Moreover, the Kruskal-Wallis sum-rank test was also used in LEfSe analysis to identify bacterial biomarkers that differed significantly in abundant taxon between sample groups.

Availability of Supporting Data

Nucleic acid sequences in this study were deposited in an open access Sequence Read Archive database of NCBI, accession number SRPxxxxxx.

10. References

1. Willis AD. Rarefaction, Alpha Diversity, and Statistics. Front Microbiol. 2019;10:2407.
2. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, et al. Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. Frontiers in Microbiology. 2016;7.
3. Ramette A. Multivariate analyses in microbial ecology. Fems Microbiol Ecol. 2007;62(2):142-60.
4. Xia Y, Sun J. Hypothesis Testing and Statistical Analysis of Microbiome. Genes Dis. 2017;4(3):138-48.
5. Ondov BD, Bergman NH, Phillippy AM. Krona: Interactive Metagenomic Visualization in a Web Browser. In: Nelson KE, editor. Encyclopedia of Metagenomics. New York, NY: Springer New York; 2013. p. 1-8.
6. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011;12(6):R60.
7. Washburne AD, Morton JT, Sanders J, McDonald D, Zhu Q, Oliverio AM, et al. Methods for phylogenetic analysis of microbiome data. Nat Microbiol. 2018;3(6):652-61.
8. Shade A, Handelsman J. Beyond the Venn diagram: the hunt for a core microbiome. Environ Microbiol. 2012;14(1):4-12.
9. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. Nat Biotechnol. 2020;38(6):685-8.
10. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. Bioinformatics. 2014;30(21):3123-4.
11. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.
12. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(Database issue):D590-6.