### 

**Table of Contents**

[**Methodology:**](#_kmkw911vv4pg) **4**

[**Results:**](#_a1jokrr3z9xz) **5**

[Table 1: Sample Overview](#_7ulq5cjw2gt9) 5

[Table 2: Summarized Quality Statistics for Raw Data and Preprocessed data using MultiQC](#_g48019xyh0ti) 7

[Figure 01(a): Taxonomic bar plot across the samples at the phylum level](#_9x2gb9riysz) 8

[Figure 01(b): Taxonomic bar plot across the samples at the class level](#_ob2m8mnpqpbm) 9

[Figure 01(c): Taxonomic bar plot across the samples at the order level](#_th8t6xjzo4o8) 10

[Figure 01(d): Taxonomic bar plot across the samples at the family level](#_lwomhtnc2bif) 11

[Figure 01(e): Taxonomic bar plot across the samples at the genus level](#_emtwwaiylcn5) 12

[Figure 01(f): Taxonomic bar plot across the samples at the species level](#_gpfbvvo2uj3s) 13

[Figure 02: Alpha Rarefaction Curves](#_vrr88dmpt6l5) 14

[Figure 03: Beta Diversity PCoA Weighted Unifrac](#_thj8ercksozy) 14

[Figure 04: Alpha Diversity Plot sample wise](#_juvr331nxldb) 16

[Figure 05(a): Bar plot at phylum level using phyloseq](#_kcfym72mgxoh) 17

[Figure 05(b): Bar plot at class level using phyloseq](#_kglfnl2a06xf) 18

[Figure 05(c): Bar plot at order level using phyloseq](#_y5lhpwoka102) 19

[Figure 05(d): Bar plot at family level using phyloseq](#_ty4cooiqhxah) 20

[Figure 05(e): Bar plot at genus level using phyloseq](#_tqhgd7207lvj) 21

[Figure 05(f): Bar plot at species level using phyloseq](#_p6xpoektdknb) 22

[Figure 06(a): Phylum Level Heatmap using Ampvis](#_cy47o2nux9zn) 23

[Figure 06(b): Family Level Heatmap using Ampvis](#_zdxyb6t9tfqi) 24

[Figure 06(c): Genus Level Heatmap using Ampvis](#_wameu2dge5uz) 25

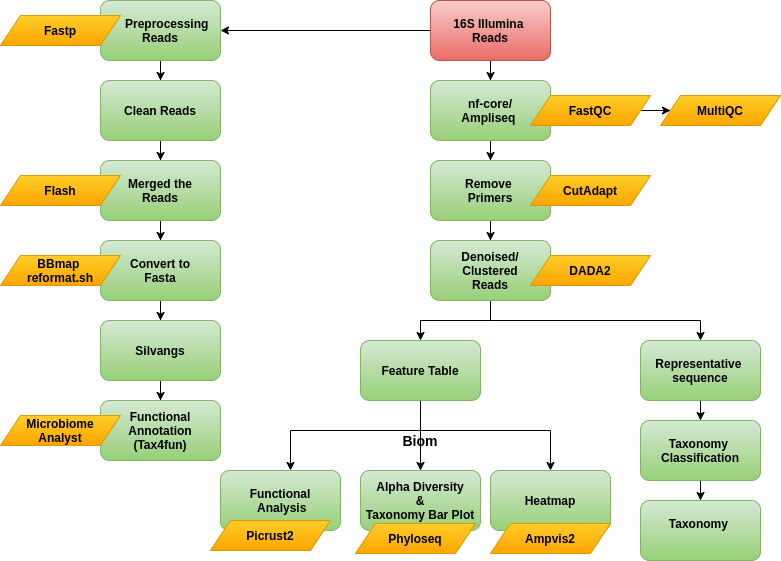
[Figure 06(c): Species Level Heatmap using Ampvis](#_lqym3s163ap8) 26

[Reference:](#_3obrf69ld2ph) 27

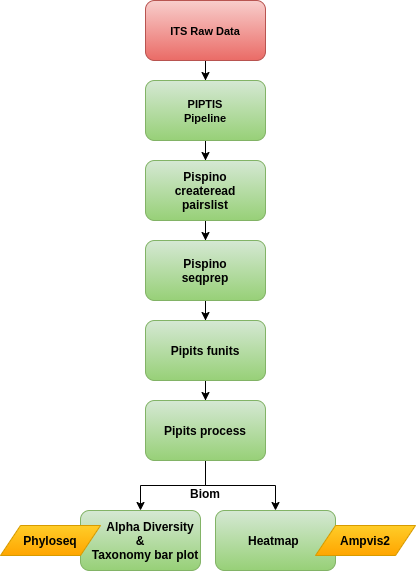
### 

### Analysis Workflow:

**16S workflow**



**ITS workflow**

****

### Methodology:

**16S metagenomic analysis**

Illumina paired-end reads were processed through the amplicon sequencing analysis workflow (ampliseq v1.1.2) (https://github.com/nf-core/ampliseq ) using DADA2 [1] and QIIME2 [2] implemented in the nf-core framework [3], which support data analysis that follows FAIR principles [4]. Several bio tools were used in sequence to analyse the raw data.  firstly, we used FASTQC  to check the quality of the raw data, second MultiQC was applied to summarise the same.  Cut adapt plugin removes the primer fragments of the sequencing reads.

First, raw data quality control analysis was done using FASTQC [5] and summarized using MultiQC [6]. Primer fragments of matched sequences were removed with the cutadapt [7] plugin. Sequences were then quality filtered, denoised, merged and the chimera removed, using the DADA2 plugin. SILVA database (Release 138) [8] was downloaded and sequences flanking the forward [GCCTACGGGNGGCWGCAG] and reverse [ACTACHVGGGTATCTAATCC] primers were extracted followed by training the QIIME2 Naïve Bayes feature classifier [9].The representative sequences were classified by taxon using the fitted classifier. nf-core ampliseq pipeline was executed with standard parameters, with DADA2 quality settings with --trunclenf and --trunclenr parameters of 277 and 215 respectively. ASVs annotated as mitochondria or chloroplasts were removed by nf-core/ampliseq because these sequences are typically considered unwanted. Relative abundance was calculated based on the taxonomic classification. An iterative abundance of the taxa were plotted using barplot with respect to their samples. For each specimen, alpha diversity was estimated using Chao1 and Abundance-based Coverage Estimator indices, which measure the species richness, and using Shannon and Simpson indices, which measure the richness and distribution of taxa. Beta diversity indices distance matrices for the Jaccard, BrayCurtis, unweighted UniFrac, and weighted UniFrac measures, as well as visualizations of the principal co-ordinates analysis. The cleaned reads were analysed using Silvangs (<https://ngs.arb-silva.de/silvangs/>) (default parameters) and for prediction of functional annotation tax4fun was done with the aid of the marker data profiling module of “MicrobiomeAnalyst” (<https://www.microbiomeanalyst.ca/>) [10, 11]. Further downstream analysis was performed by filtered the biom and phylogenetic tree using Shiny-Phyloseq [12] (<https://github.com/joey711/shiny-phyloseq>) (Parameters:- Subset Taxa:Taxonomic Ranks=Rank1, Select Taxa=d\_Bacteria, Subset Samples: Sample Variables=NULL, Variable Classes= NULL, Total Sums Filtering: Sample Min= 1000, Taxa min= 100, kOverA OTU Filtering: A=2e-05, k=10 ) for Alpha diversity plot (parameters:- Aesthetic Mapping: x = samples, color = samples, alpha Measures= chao1, Shannon, Simpson, ACE, Observed, and set all other parameters as default), Bar-plot (parameters:- Aesthetic Mapping: X-Axis=sample, color=Rank1, Rank2, Rank3, Rank4, Rank5, Rank6, Rank7 and Data= Counts, with adjusted color palette, size and format). Heatmap was plotted using Ampvis2 R package [13].

Predictive functional analysis of bacterial communities was done using Phylogenetic Investigation of Communities by Reconstruction of Unobserved States 2 (PICRUSt2) [14] To predict functions, the representative sequences were placed into a reference tree using the place\_seqs.py utility.

**Fungal ITx metagenomic analysis**

Analysis of fungal internal transcribed spacer (ITS) sequences from the Illumina sequencing platform was done using PIPITS v2.7 [15], which was reported to perform better than QIIME2 and Galaxy [16]. The illumina fastq sequences were assembled with VSEARCH [17] and quality-filtering was done with fastx through the PIPITS command pispino\_createreadpairslist. The ITSx [18] was executed through the PIPITS command pipits\_funits. Chimera filtering and clustering was done using VSEARCH through the PIPITS command pipits\_process. The classic tabular OTU table was converted into a BIOM format and taxonomy was assigned with UNITE [RDP Classifier].

Number of reads used to generate OTU table: 667566  
Number of OTUs: 90  
Number of phylotypes: 11  
Number of samples: 4

**Commands**

1. **ampliseq pipeline**

nextflow run nf-core/ampliseq -profile singularity --input ./Illumina\_PE\_16s --FW\_primer GCCTACGGGNGGCWGCAG --RV\_primer ACTACHVGGGTATCTAATCC --trunclenf 277 --trunclenr 215 --metadata Metadata.tsv --reference\_database ./Silva\_138\_release.zip

1. **PIPTIS pipeline**pispino\_createreadpairslist -i rawdata -o readpairslist.txt  
   pispino\_seqprep -i rawdata -o out\_seqprep -l readpairslist.txt  
   pipits\_funits -i out\_seqprep/prepped.fasta -o out\_funits\_ITS1 -x ITS1 -v -r pipits\_process -i out\_funits/ITS.fasta -o out\_process -t 50 -v -r
2. **PICRUST2**

picrust2\_pipeline.py -s sequences.fasta -i otu\_table.biom -o picrust2\_out -p 50 --in\_traits COG,EC,KO,PFAM,TIGRFAM

### 

### 16S Results:

#### Table 1: Sample Overview

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Type** | **Library type** | **Number of samples** | **Approach used** |
| Illumina 16S | Paired-end | 4 | 16S Metagenomics Analysis |

#### 

#### Table 2: Summarized Quality Statistics for Raw Data and Preprocessed data using MultiQC

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample Name** | **Read Orientations** | **Average GC (%)** | **Length**  **(bp)** | **Total Sequences (Millions)** | **% BP Trimmed** | **Average GC (%)** | **Length**  **(bp)** | **Total Sequences (Millions)** |
| cpapaya | R1 | 55 | 266 | 0.1 |  | 55 | 266 | 0.1 |
| cpapaya | R2 | 54 | 235 | 0.1 | 8.80 | 54 | 235 | 0.1 |
| vcauliflora | R1 | 55 | 264 | 0.1 |  | 55 | 264 | 0.1 |
| vcauliflora | R2 | 54 | 240 | 0.1 | 8.60 | 54 | 240 | 0.1 |
| vcundinamarcensis | R1 | 55 | 264 | 0.1 |  | 55 | 264 | 0.1 |
| vcundinamarcensis | R2 | 54 | 236 | 0.1 | 8.70 | 54 | 236 | 0.1 |
| vgoudotiana | R1 | 55 | 266 | 0.2 |  | 55 | 266 | 0.2 |
| vgoudotiana | R2 | 54 | 237 | 0.2 | 8.60 | 54 | 237 | 0.2 |

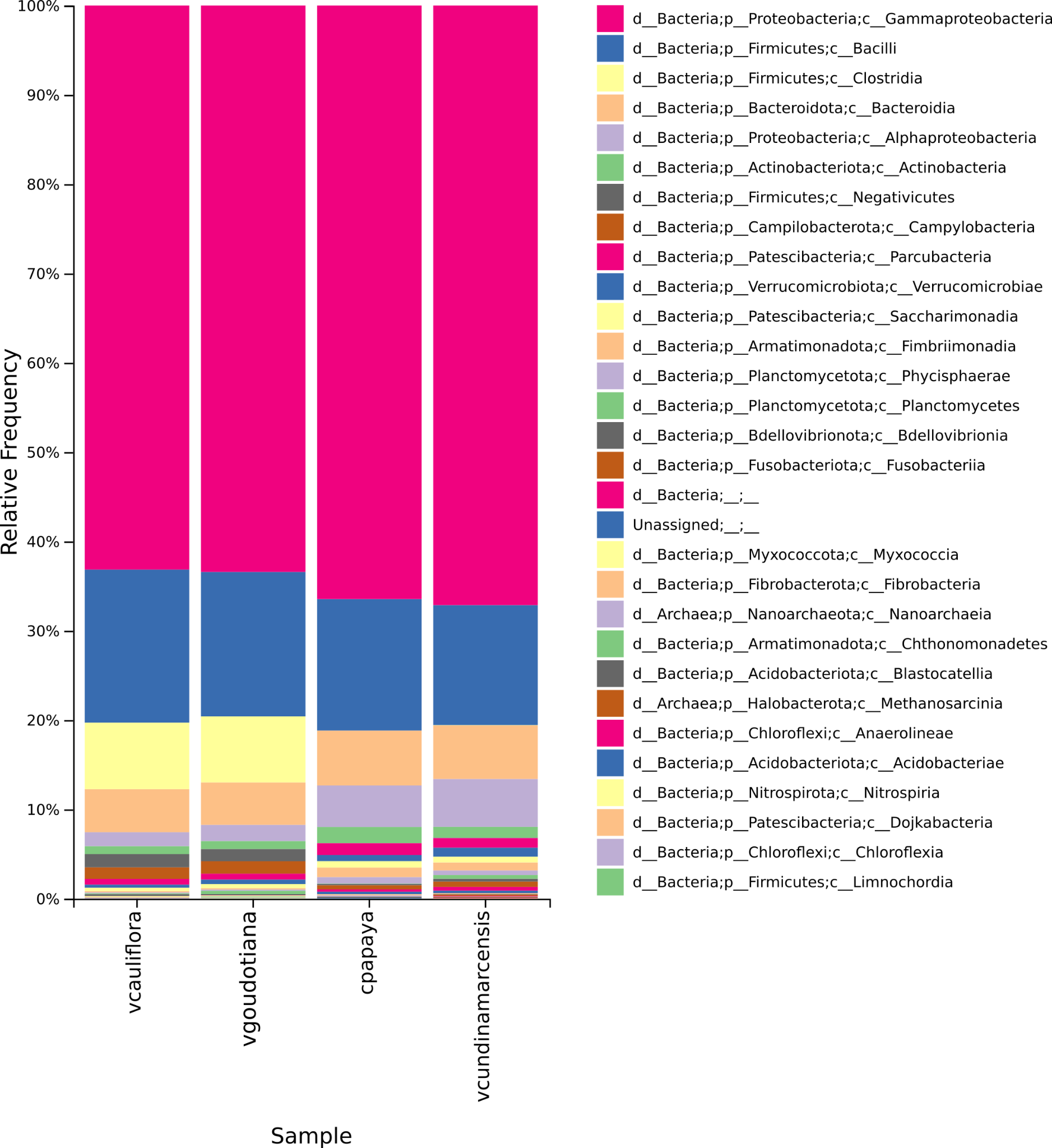
#### 

#### 

#### Figure 01(a): Taxonomic bar plot across the samples at the phylum level

Deliverables/16S\_ampliseq/01\_16S\_ampliseq/barplot/index.html

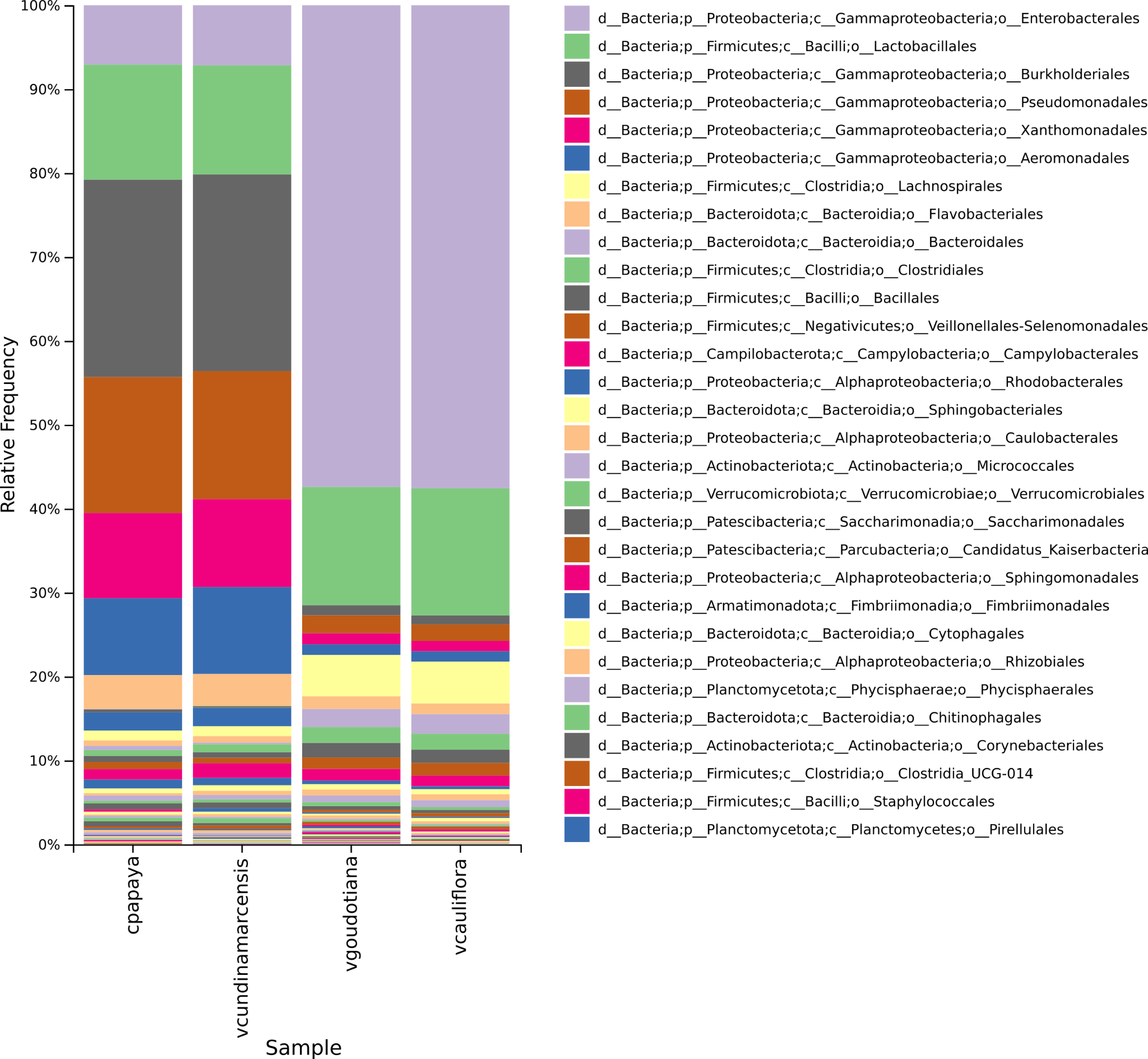
Deliverables/16S\_ampliseq/04\_Figures/QIIME2/01\_a\_QIIME2\_barplot\_Phylum.png



#### Figure 01(b): Taxonomic bar plot across the samples at the class level

Deliverables/16S\_ampliseq/01\_16S\_ampliseq/barplot/index.html

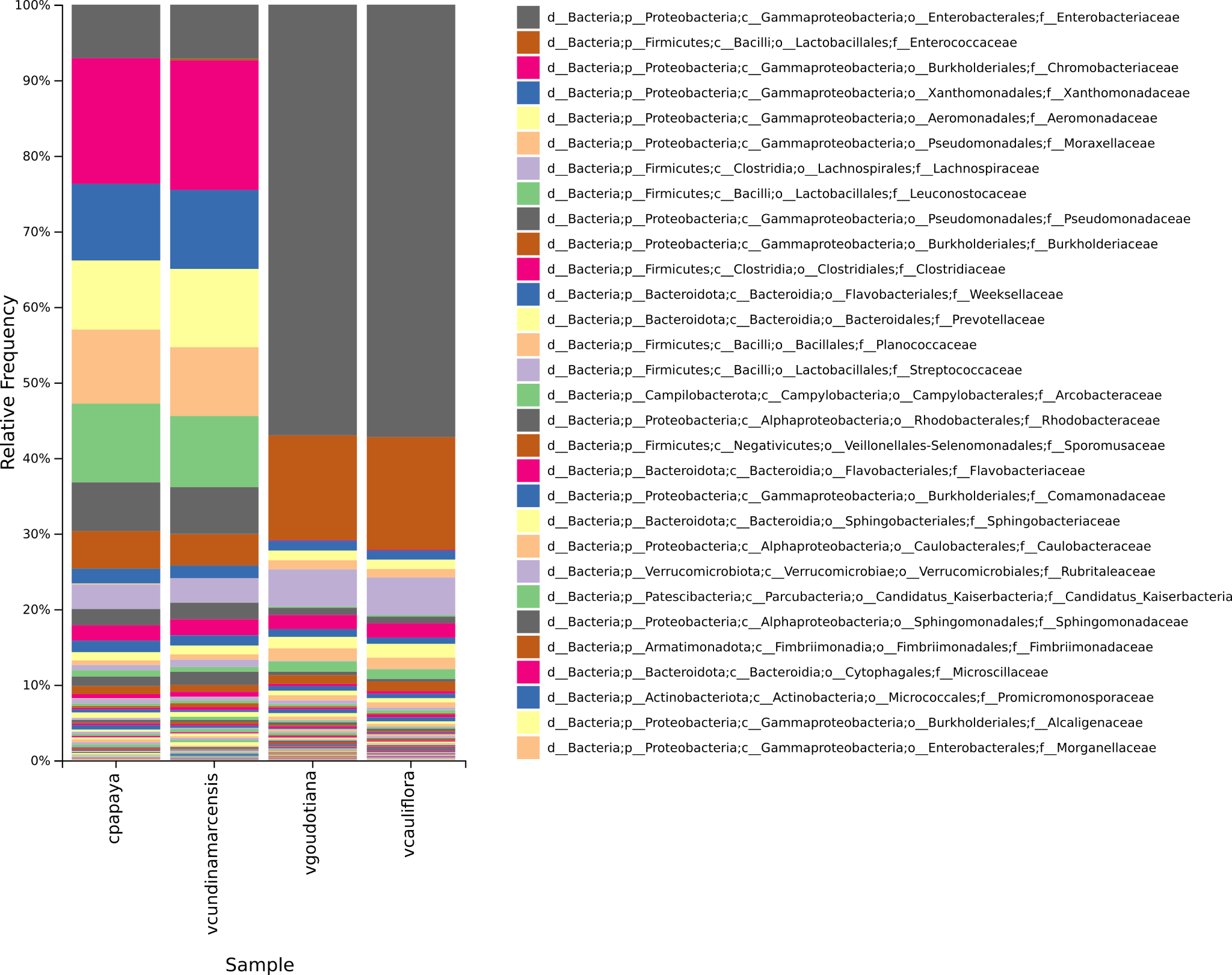
Deliverables/16S\_ampliseq/04\_Figures/QIIME2/01\_b\_QIIME2\_barplot\_Class.png



#### Figure 01(c): Taxonomic bar plot across the samples at the order level

Deliverables/16S\_ampliseq/01\_16S\_ampliseq/barplot/index.html

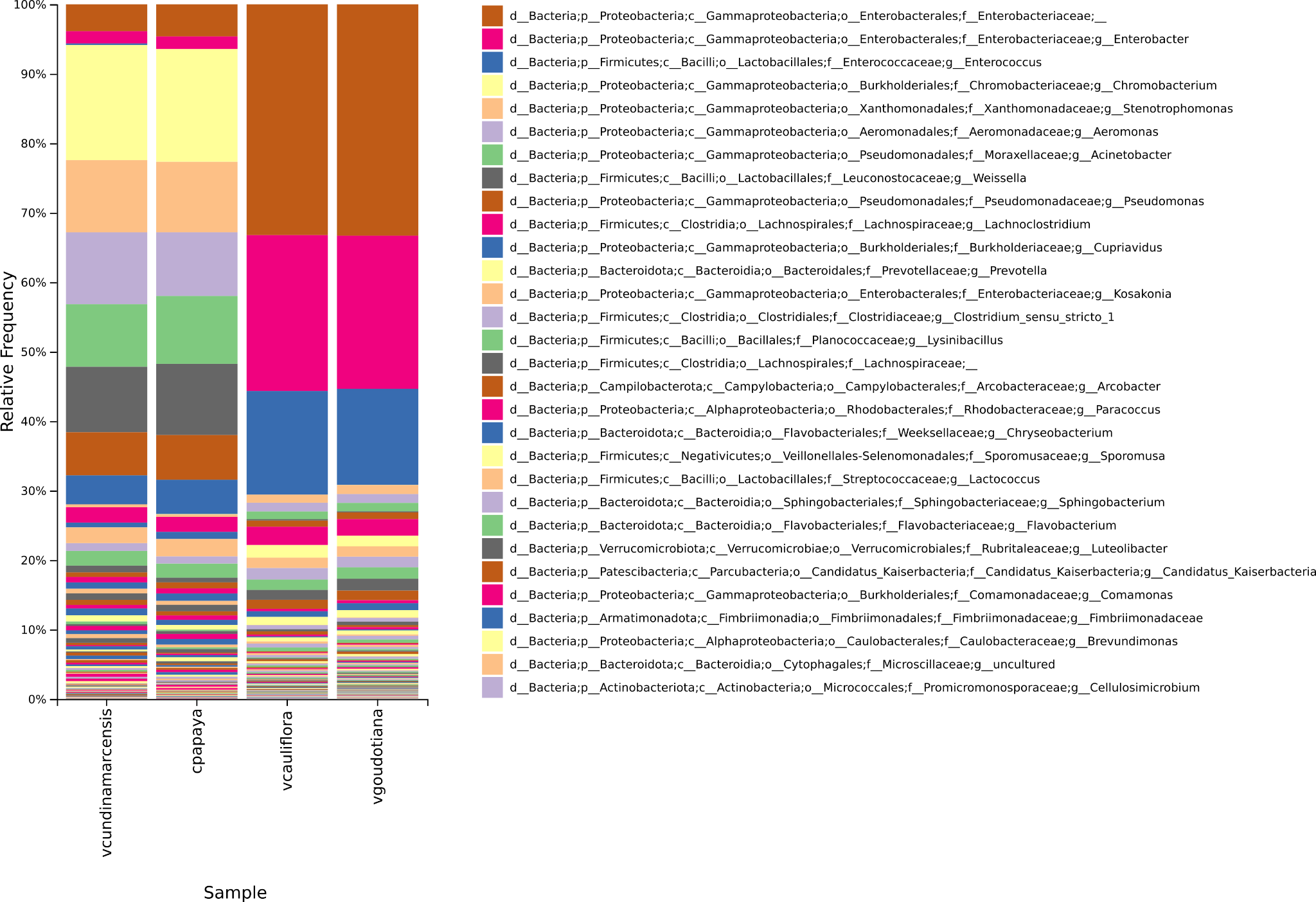
Deliverables/16S\_ampliseq/04\_Figures/QIIME2/01\_c\_QIIME2\_barplot\_Order.png



#### Figure 01(d): Taxonomic bar plot across the samples at the family level

Deliverables/16S\_ampliseq/01\_16S\_ampliseq/barplot/index.html

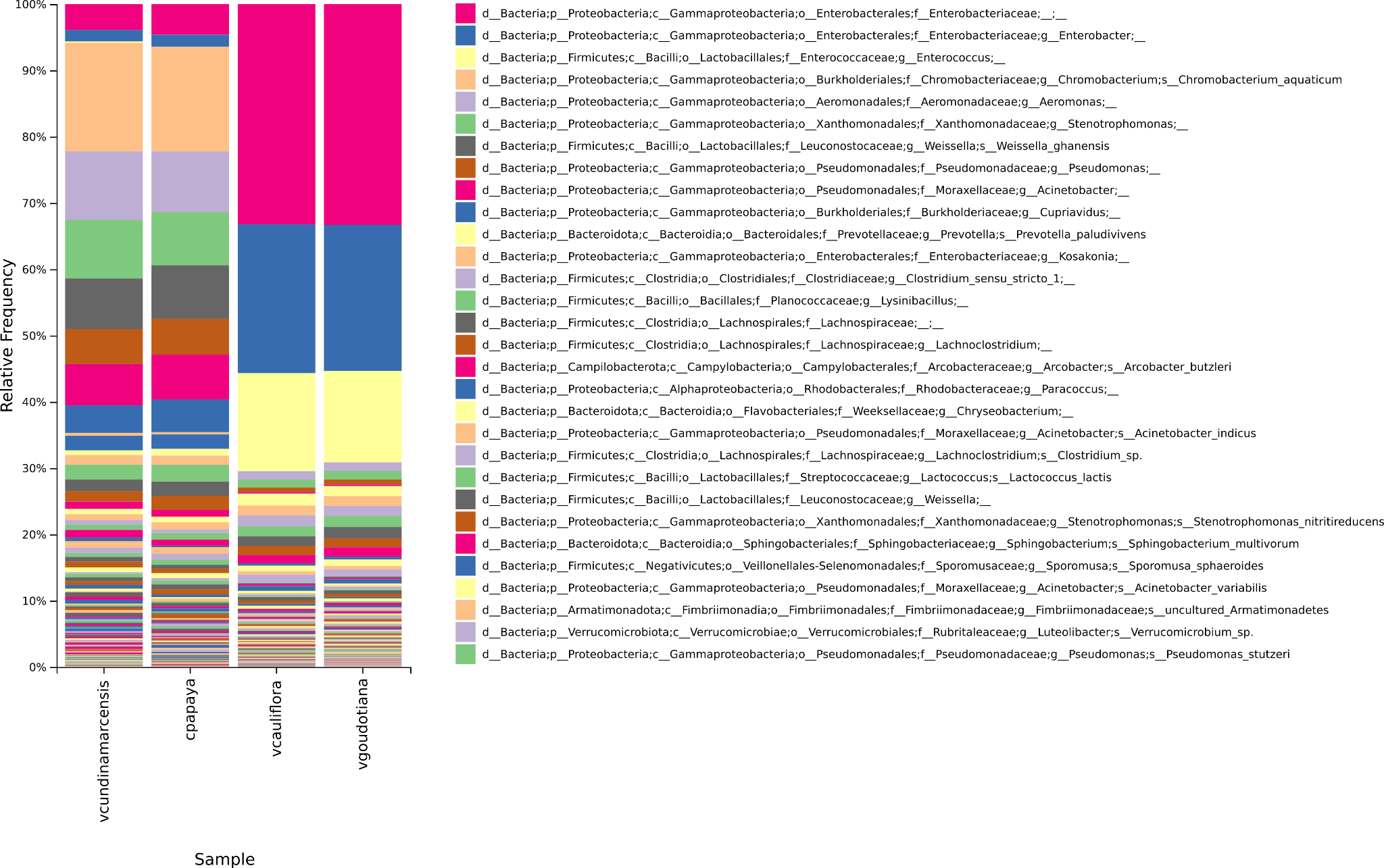
Deliverables/16S\_ampliseq/04\_Figures/QIIME2/01\_d\_QIIME2\_barplot\_Family.png



#### Figure 01(e): Taxonomic bar plot across the samples at the genus level

Deliverables/16S\_ampliseq/01\_16S\_ampliseq/barplot/index.html

Deliverables/16S\_ampliseq/04\_Figures/QIIME2/01\_e\_QIIME2\_barplot\_Genus.png

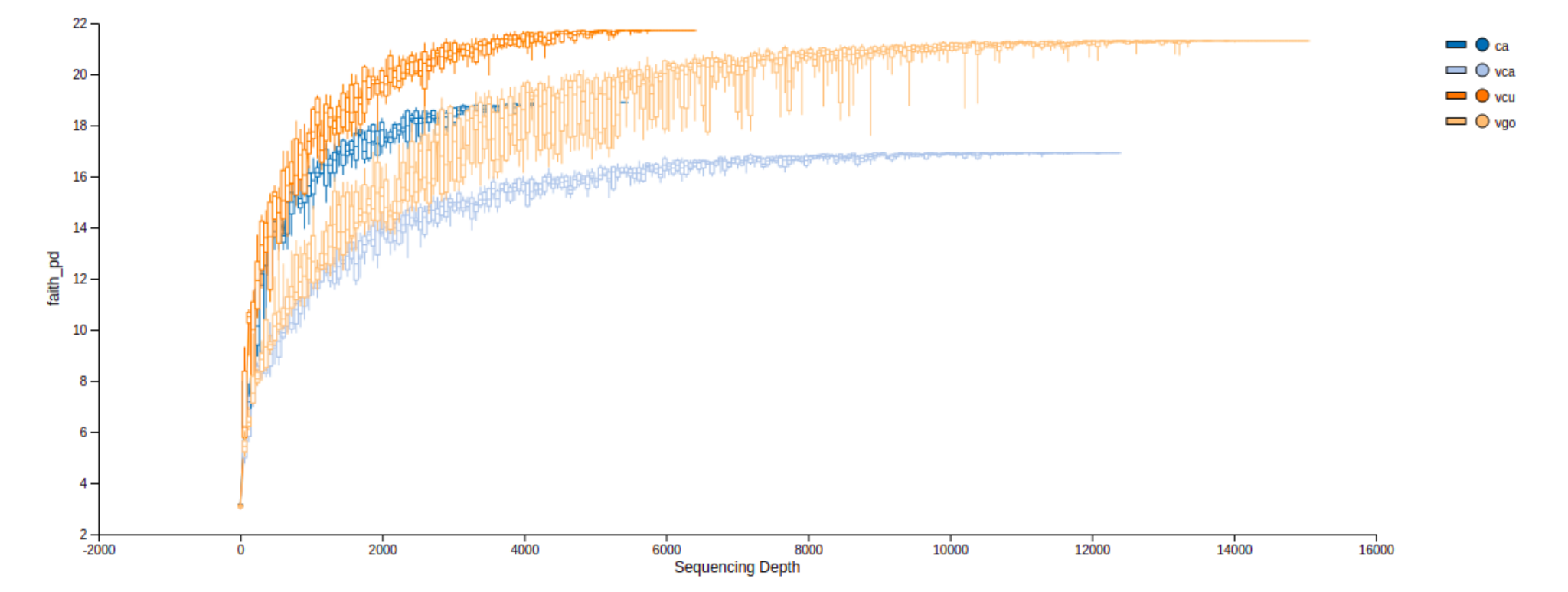


#### 

#### Figure 01(f): Taxonomic bar plot across the samples at the species level

Deliverables/16S\_ampliseq/01\_16S\_ampliseq/barplot/index.html

Deliverables/16S\_ampliseq/04\_Figures/QIIME2/01\_f\_QIIME2\_barplot\_Species.png



#### Figure 02: Alpha Rarefaction Curves

Deliverables/16S\_ampliseq/01\_16S\_ampliseq/alpha-rarefaction/index.html

Deliverables/16S\_ampliseq/04\_Figures/QIIME2/02\_QIIME2\_Alpha\_rarefaction.png

#### Figure 03: Beta Diversity PCoA Weighted Unifrac

Deliverables/16S\_ampliseq/04\_Figures/QIIME2/03\_QIIME2\_PCoA\_weighted\_unifrac.png

Deliverables/16S\_ampliseq/01\_16S\_ampliseq/beta-diversity/weighted\_unifrac\_pcoa\_results-PCoA/index.html

Path: Deliverables\_nextflow\_16s/MicrobiomAnalyst/functional\_KO\_Description.csv

**Path to the picrust functional analysis results:**

Deliverables/16S\_ampliseq/03\_PICRUST2/picrust2\_out/COG\_metagenome\_out/COG\_annotated.tsv

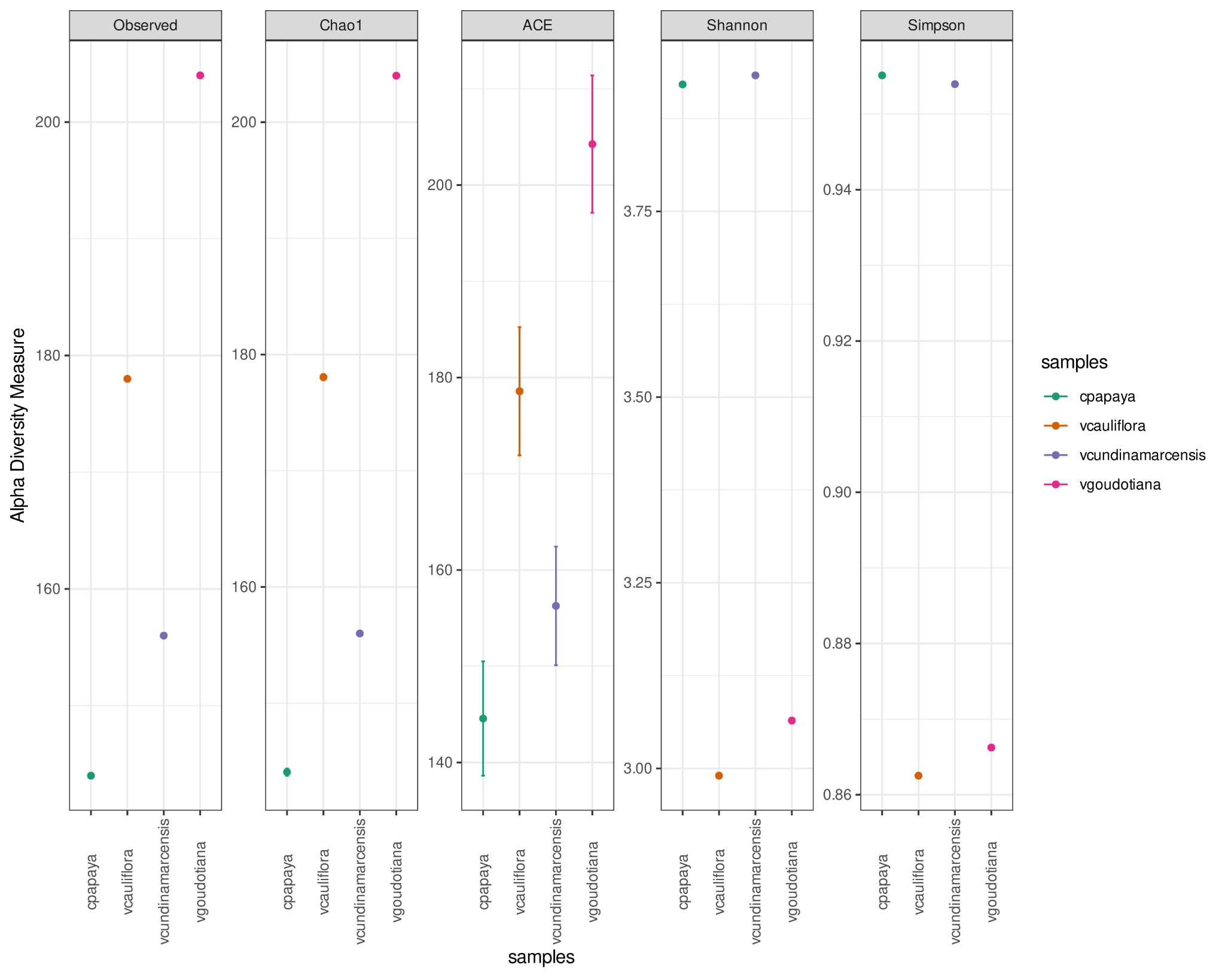
Deliverables/16S\_ampliseq/03\_PICRUST2/picrust2\_out/EC\_metagenome\_out/EC\_Annotated.tsv

Deliverables/16S\_ampliseq/03\_PICRUST2/picrust2\_out/KO\_metagenome\_out/KO\_Annotated.tsv

Deliverables/16S\_ampliseq/03\_PICRUST2/picrust2\_out/pathways\_out/Pathway\_Annotated.tsv

Deliverables/16S\_ampliseq/03\_PICRUST2/picrust2\_out/PFAM\_metagenome\_out/PFAM\_Annotated.tsv

Deliverables/16S\_ampliseq/03\_PICRUST2/picrust2\_out/TIGRFAM\_metagenome\_out/TIGRFAM\_Annotated.tsv



#### Figure 04: Alpha Diversity Plot sample wise for 16S Reads

Deliverables/16S\_ampliseq/04\_Figures/Phyloseq/4\_Alpha\_diversity.svg

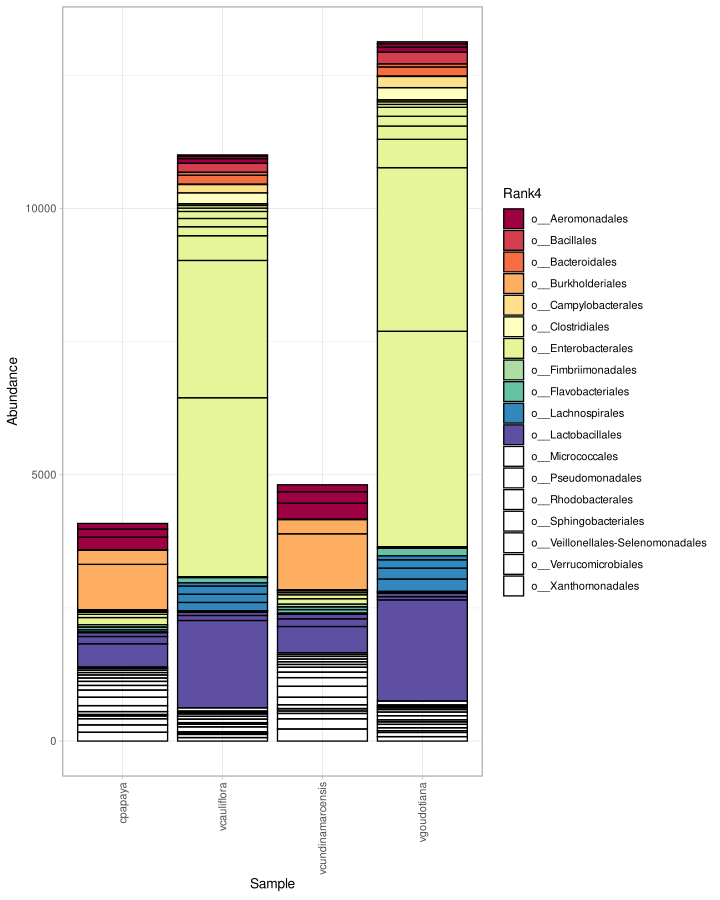
#### Figure 05(a): Bar plot at phylum level using phyloseq

Deliverables/16S\_ampliseq/04\_Figures/Phyloseq/png/5\_a\_Phylum.png



#### Figure 05(b): Bar plot at class level using phyloseq

Deliverables/16S\_ampliseq/04\_Figures/Phyloseq/png/5\_b\_Class.png

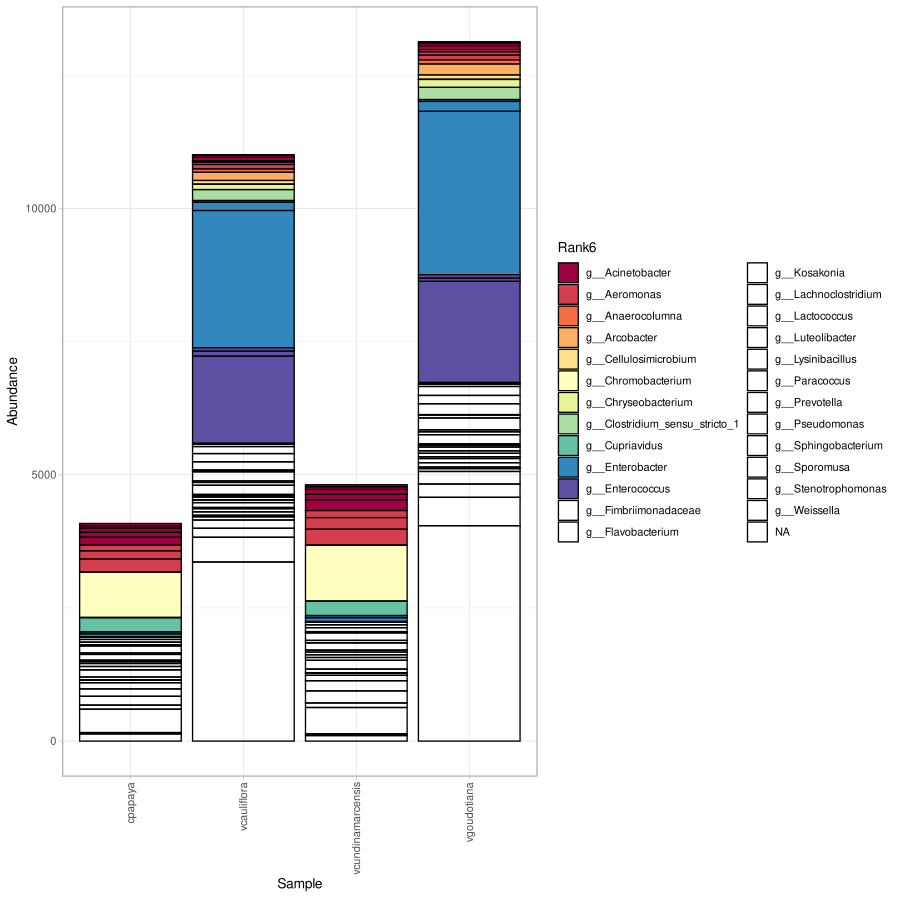


#### Figure 05(c): Bar plot at order level using phyloseq

Deliverables/16S\_ampliseq/04\_Figures/Phyloseq/png/5\_c\_Order.png

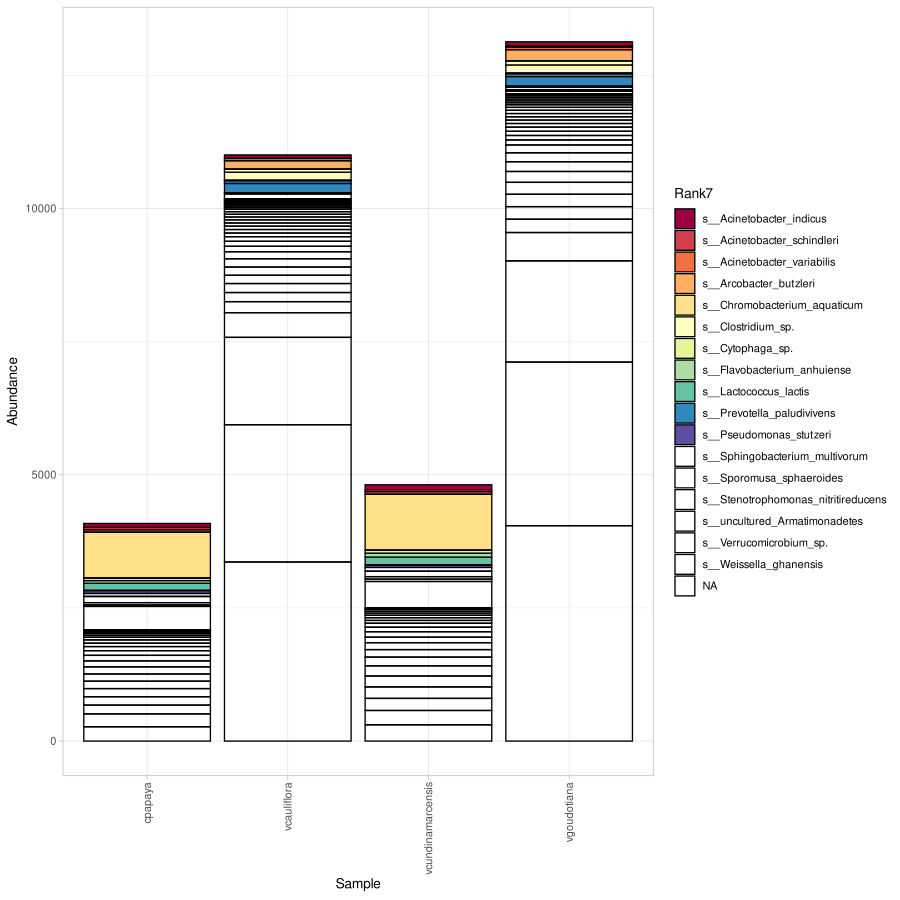
#### Figure 05(d): Bar plot at family level using phyloseq

Deliverables/16S\_ampliseq/04\_Figures/Phyloseq/png/5\_d\_Family.png



#### Figure 05(e): Bar plot at genus level using phyloseq

Deliverables/16S\_ampliseq/04\_Figures/Phyloseq/png/5\_e\_Genus.png



#### Figure 05(f): Bar plot at species level using phyloseq

Deliverables/16S\_ampliseq/04\_Figures/Phyloseq/png/5\_f\_Species.png

#### Figure 06(a): Phylum Level Heatmap using Ampvis

Deliverables/16S\_ampliseq/04\_Figures/Ampvis2/6\_a\_phylum.png

#### 

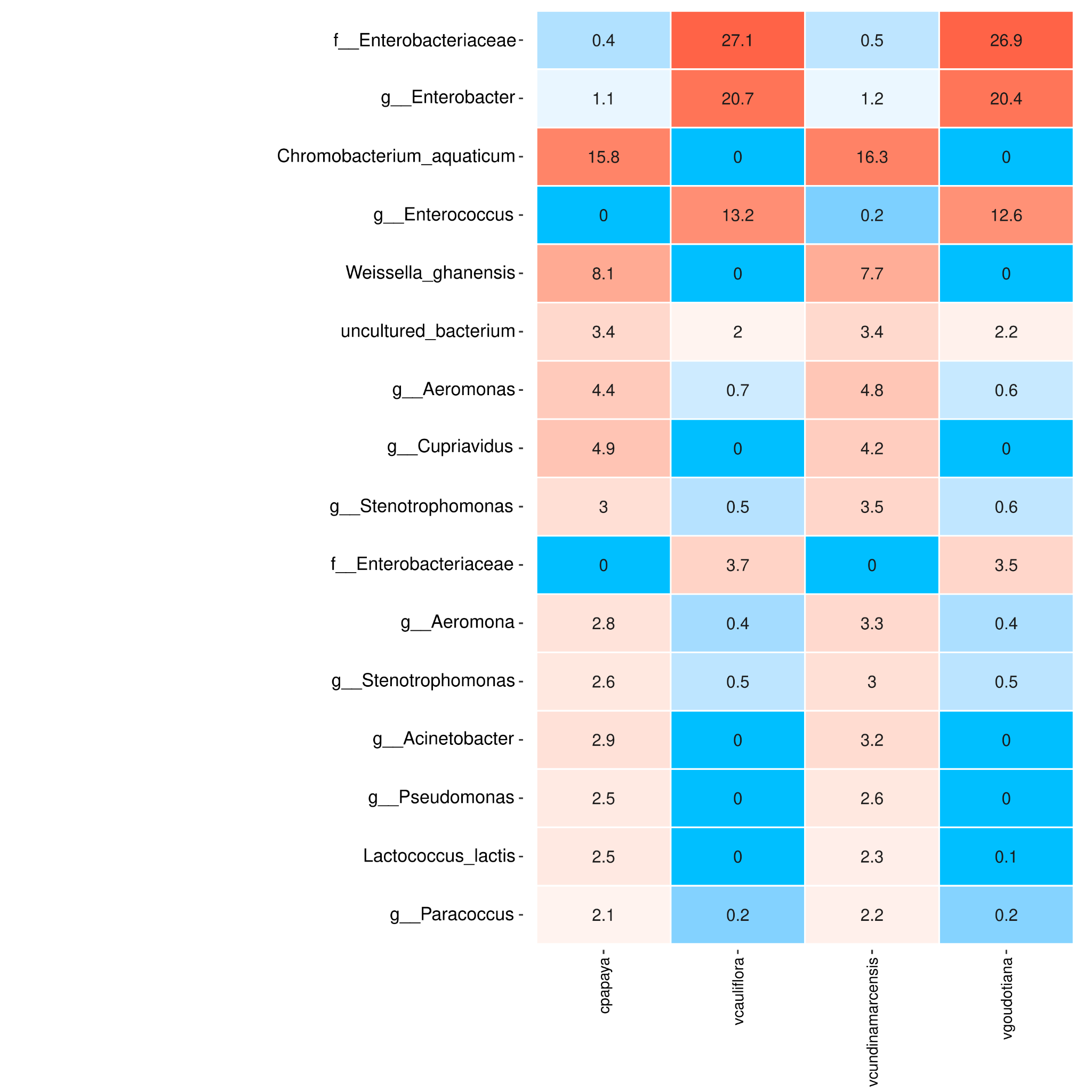


#### Figure 06(b): Family Level Heatmap using Ampvis

Deliverables/16S\_ampliseq/04\_Figures/Ampvis/6\_b\_family.png

#### Figure 06(c): Genus Level Heatmap using Ampvis

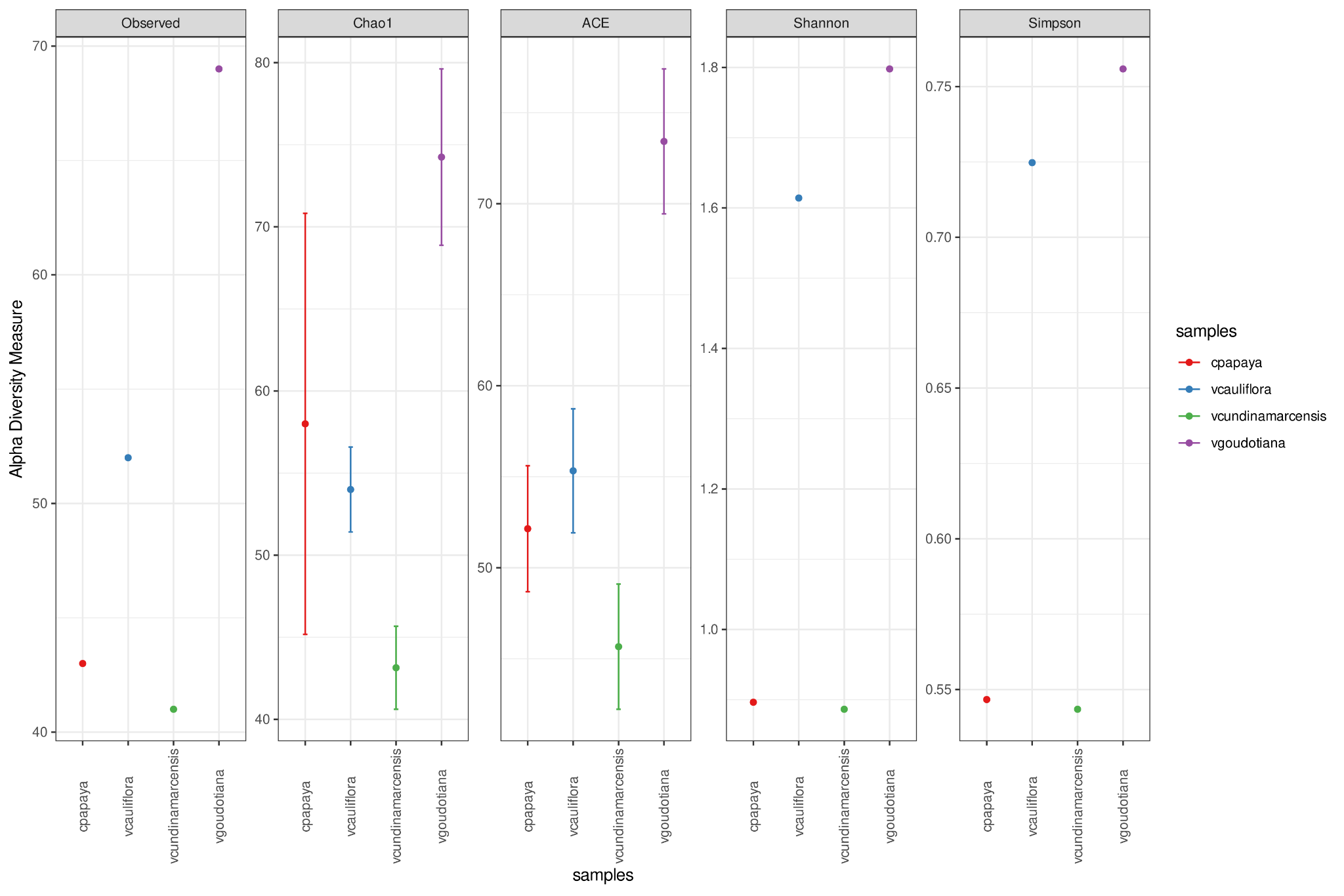
Deliverables/16S\_ampliseq/04\_Figures/Ampvis/6\_c\_genus.png



#### Figure 06(c): Species Level Heatmap using Ampvis

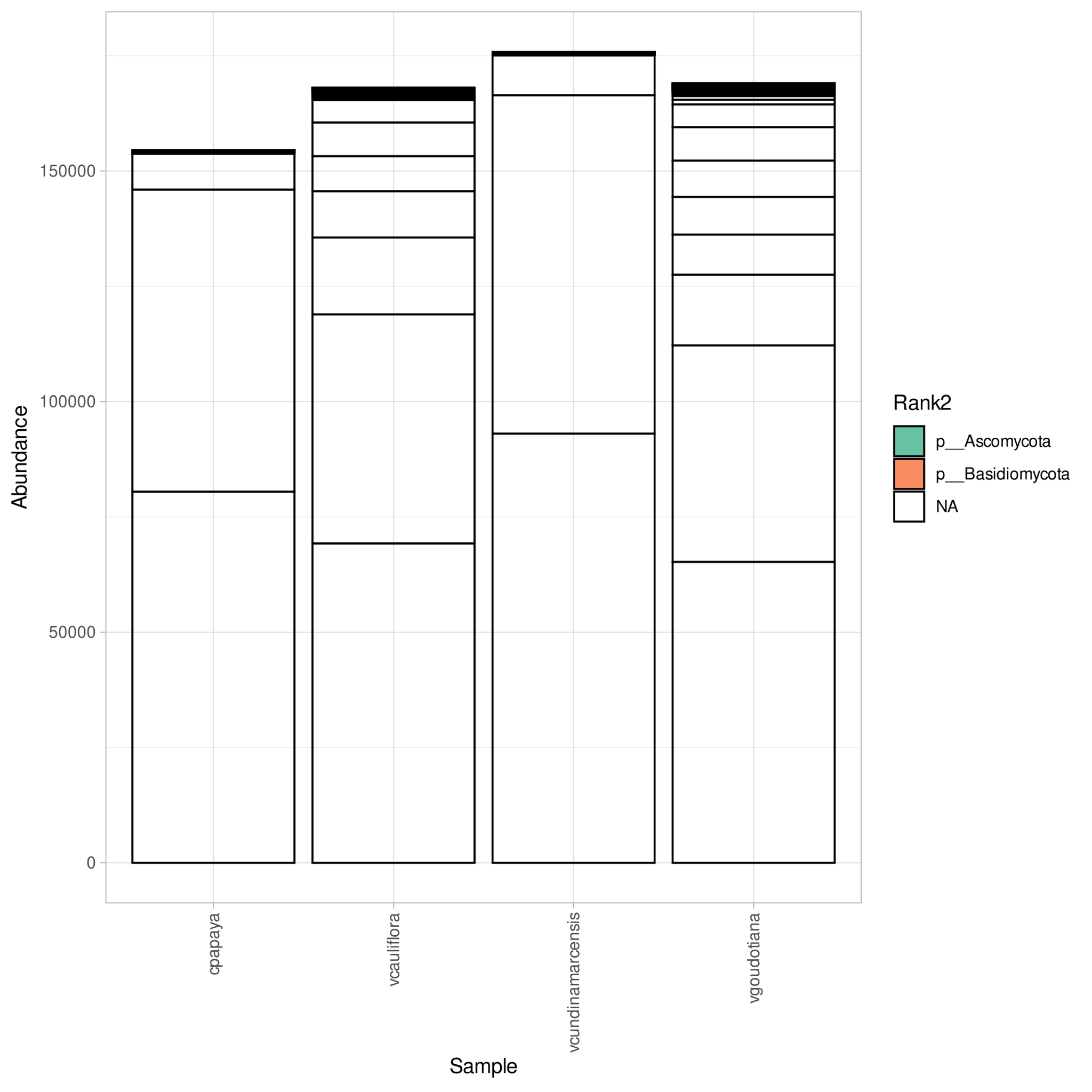
Deliverables/16S\_ampliseq/04\_Figures/Ampvis2/6\_d\_Species.png

### ITS Results:



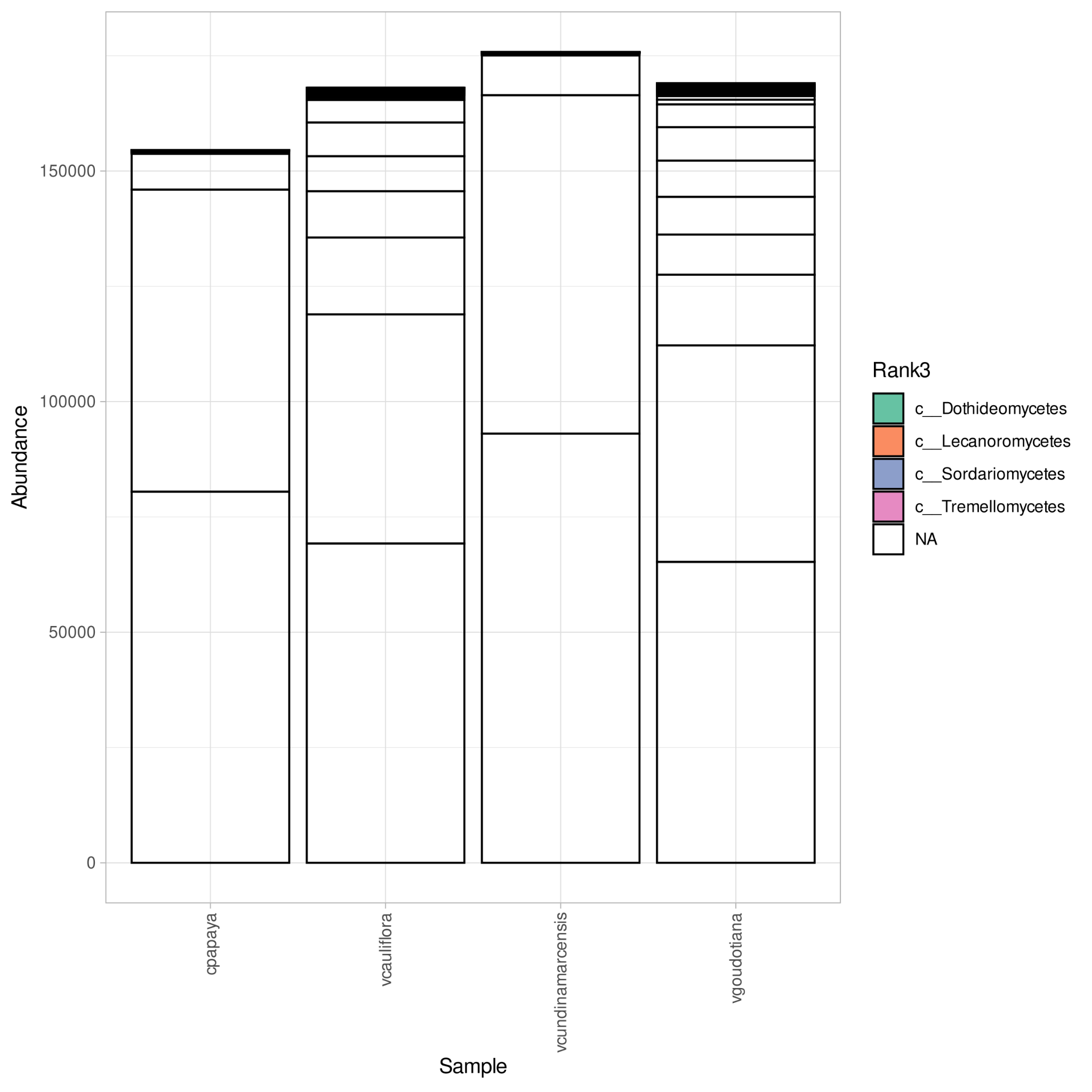
#### Figure 07: Alpha Diversity Plot sample wise for ITS Reads

Deliverables/ITS\_PIPTIS/Figures/ITS\_phyloseq/7\_Alpha\_Diversity.png



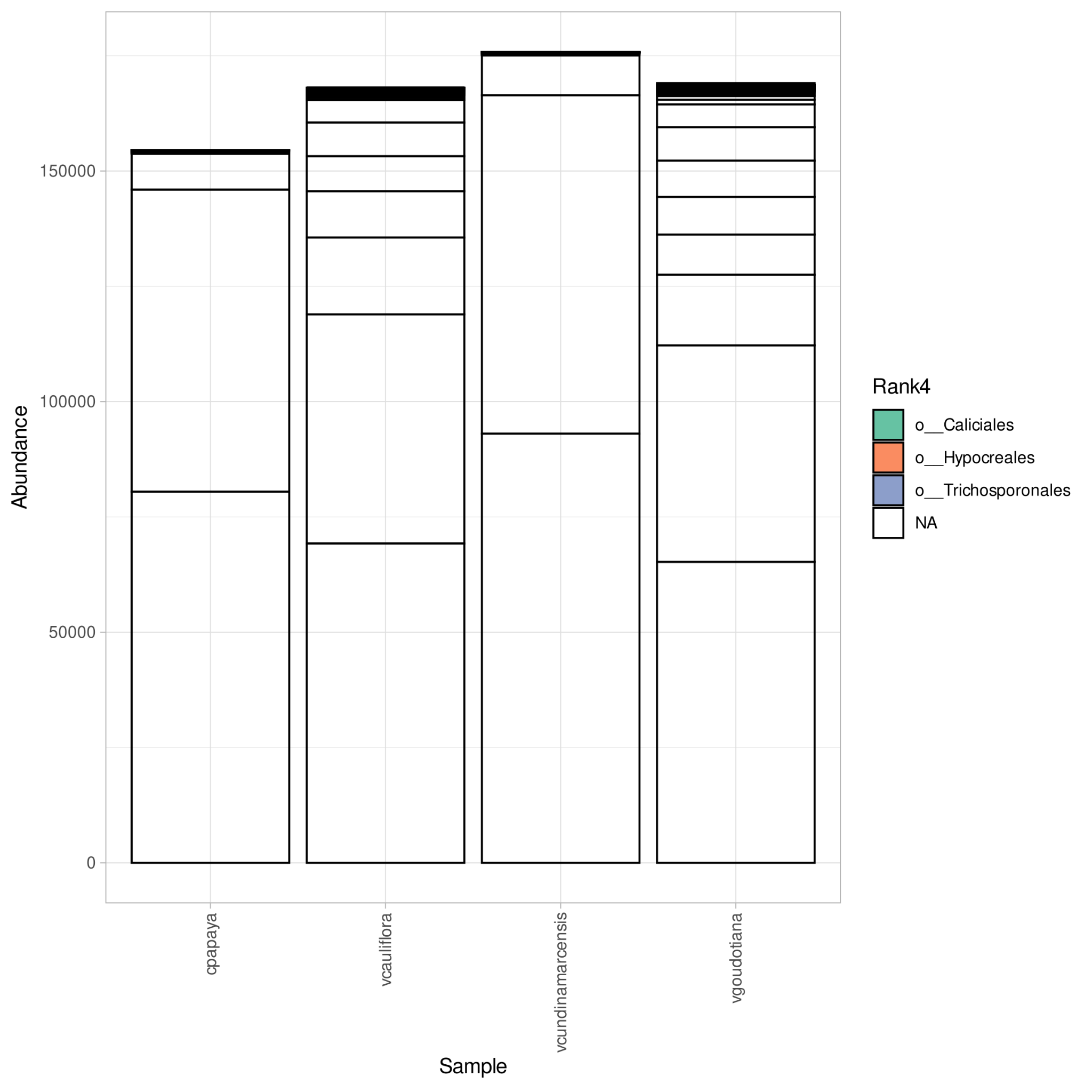
#### Figure 08(a): Bar plot at phylum level using phyloseq

Deliverables/ITS\_PIPTIS/Figures/ITS\_phyloseq/8\_a\_Phylum.png



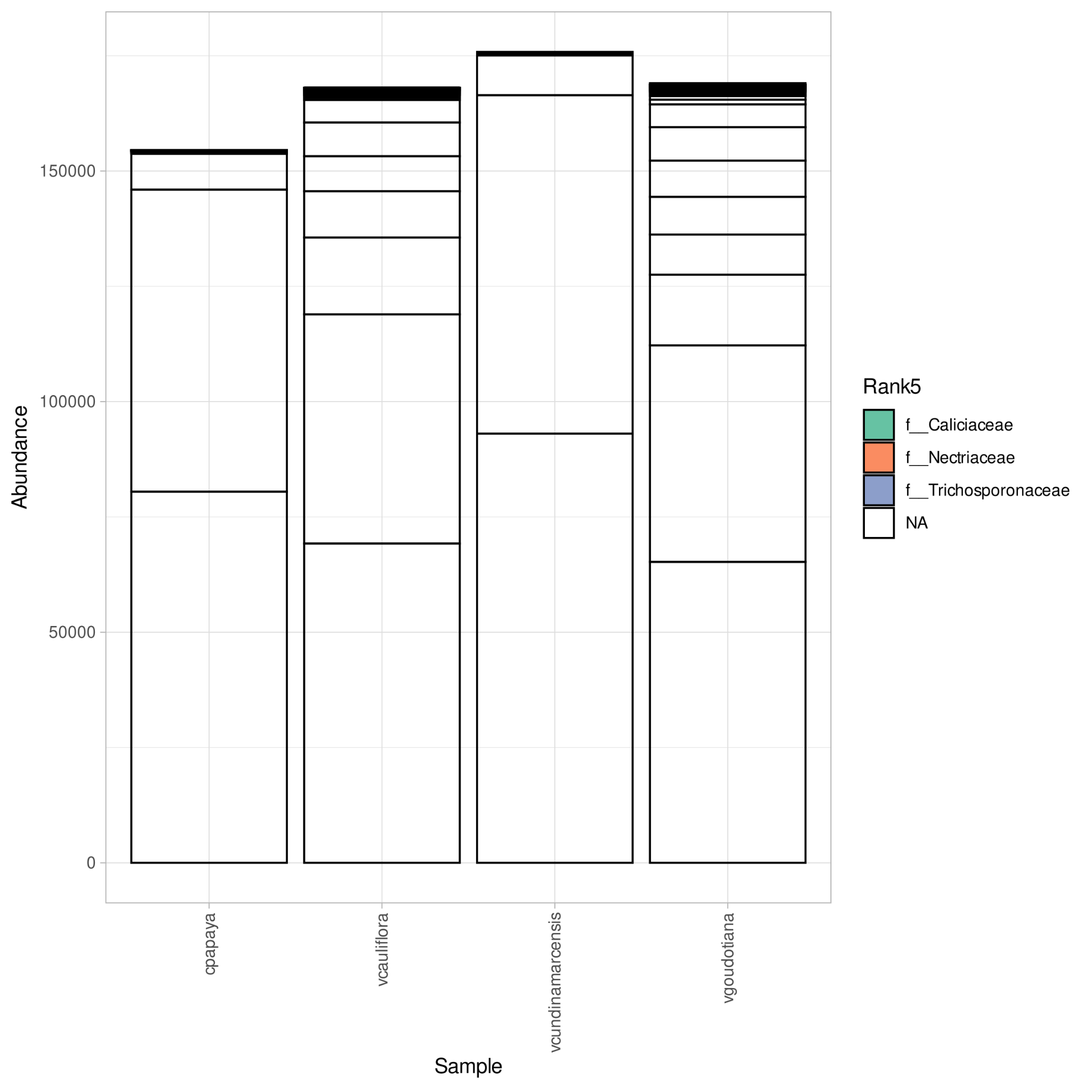
#### Figure 08(b): Bar plot at class level using phyloseq

Deliverables/ITS\_PIPTIS/Figures/ITS\_phyloseq/8\_b\_Class.png



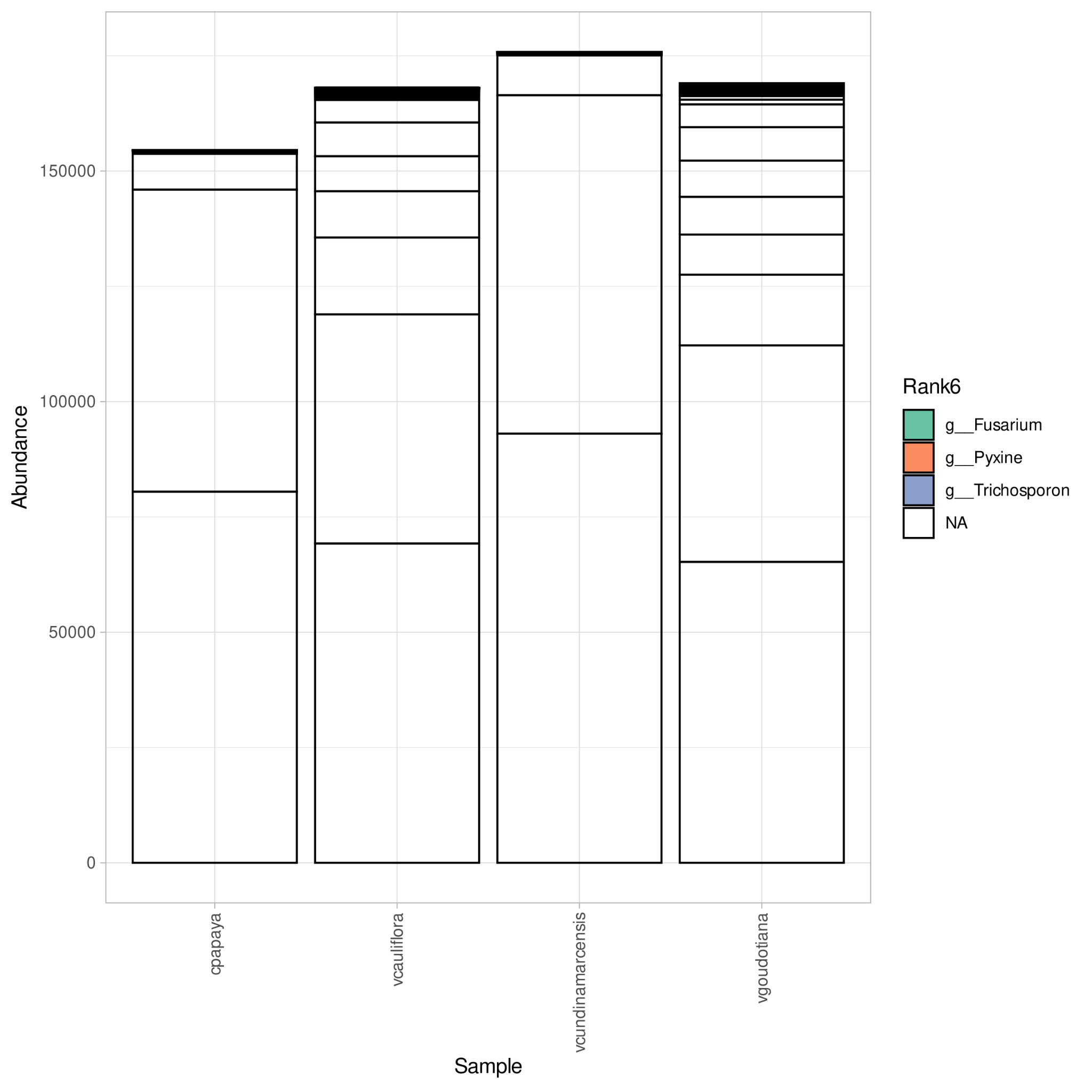
#### Figure 08(c): Bar plot at order level using phyloseq

Deliverables/ITS\_PIPTIS/Figures/ITS\_phyloseq/8\_c\_Order.png



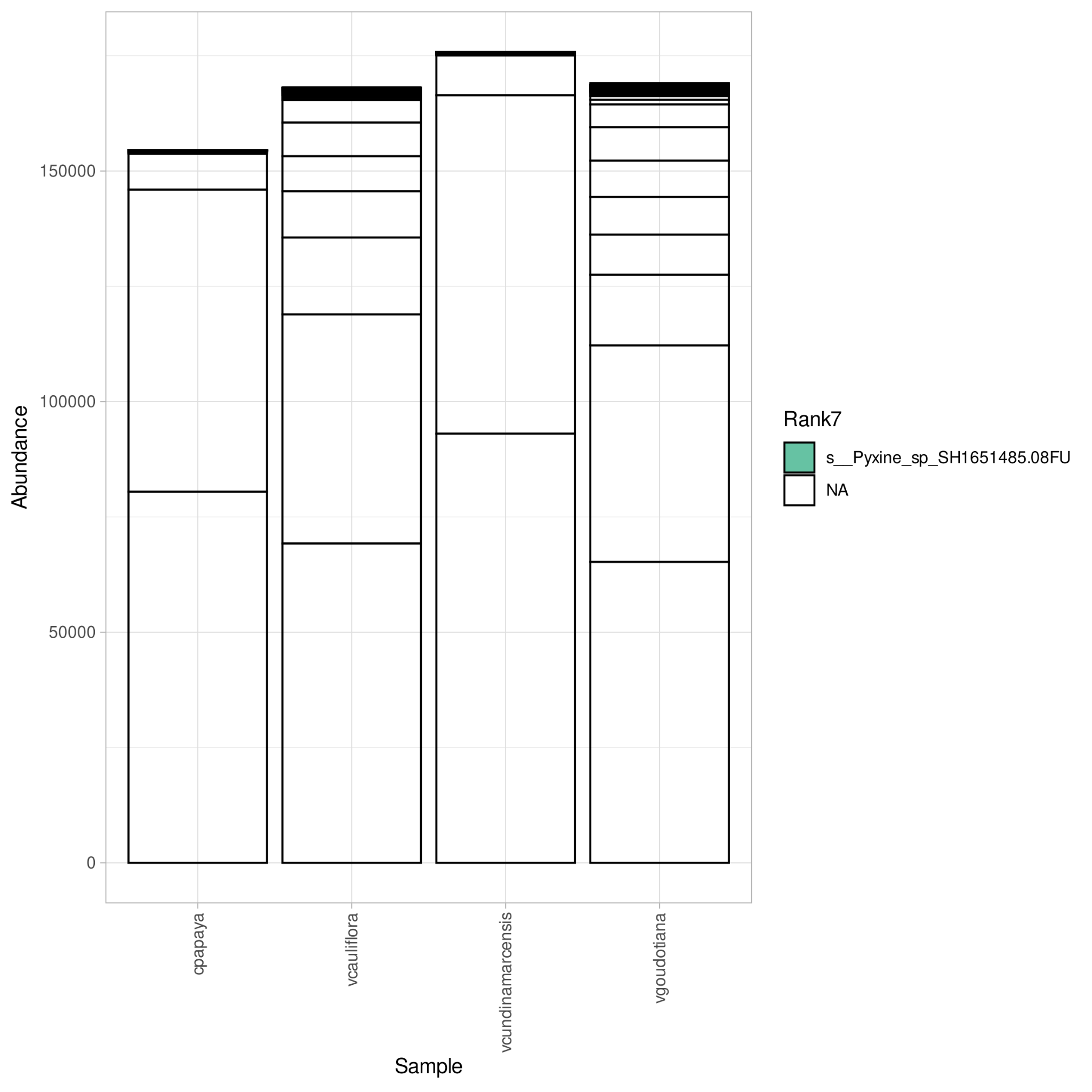
#### Figure 08(d): Bar plot at family level using phyloseq

Deliverables/ITS\_PIPTIS/Figures/ITS\_phyloseq/8\_d\_Family.png



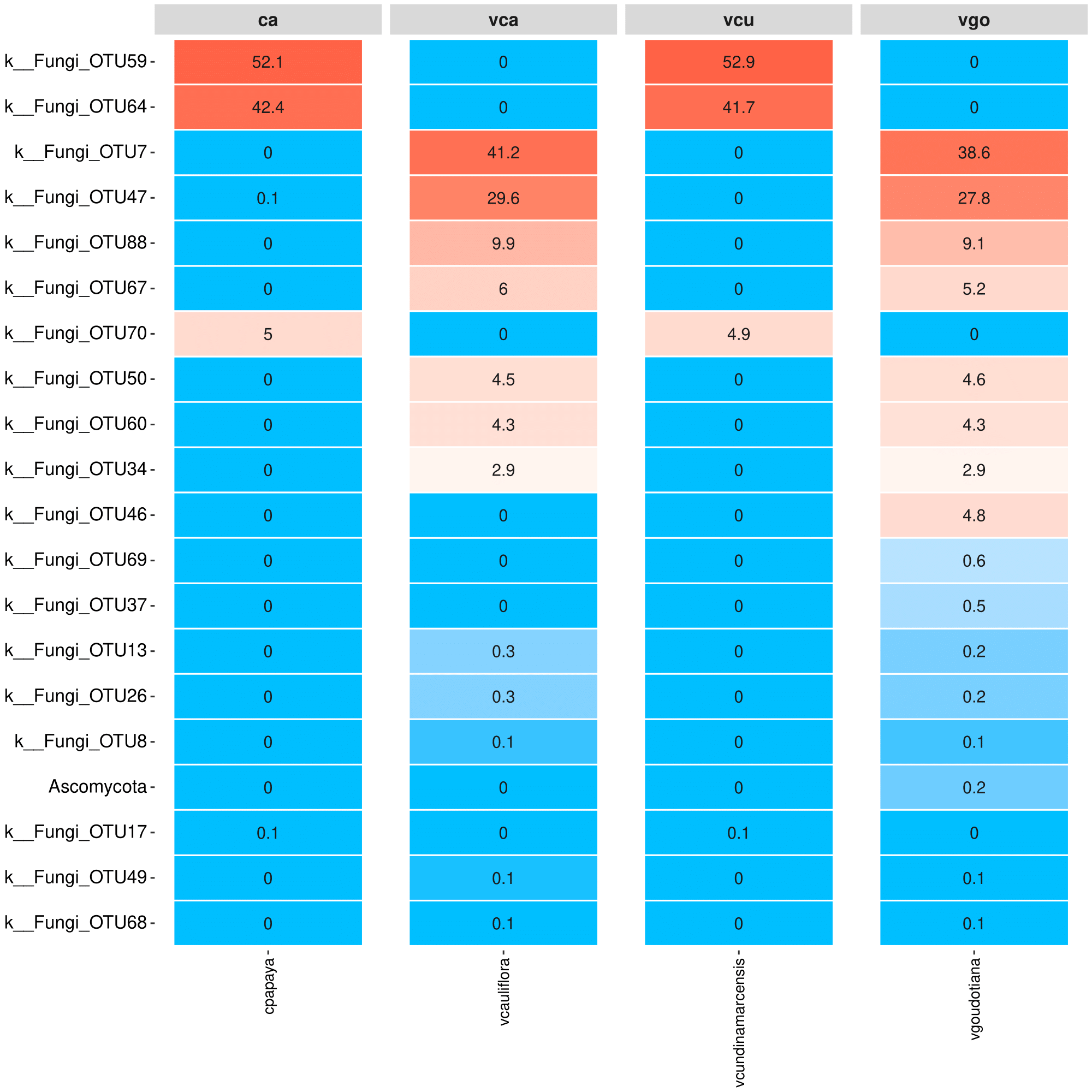
#### Figure 08(e): Bar plot at genus level using phyloseq

Deliverables/ITS\_PIPTIS/Figures/ITS\_phyloseq/8\_e\_Genus.png



#### Figure 08(c): Bar plot at species level using phyloseq

Deliverables/ITS\_PIPTIS/Figures/ITS\_phyloseq/8\_f\_Species.png



#### Figure 09(a): Phylum Level Heatmap using Ampvis

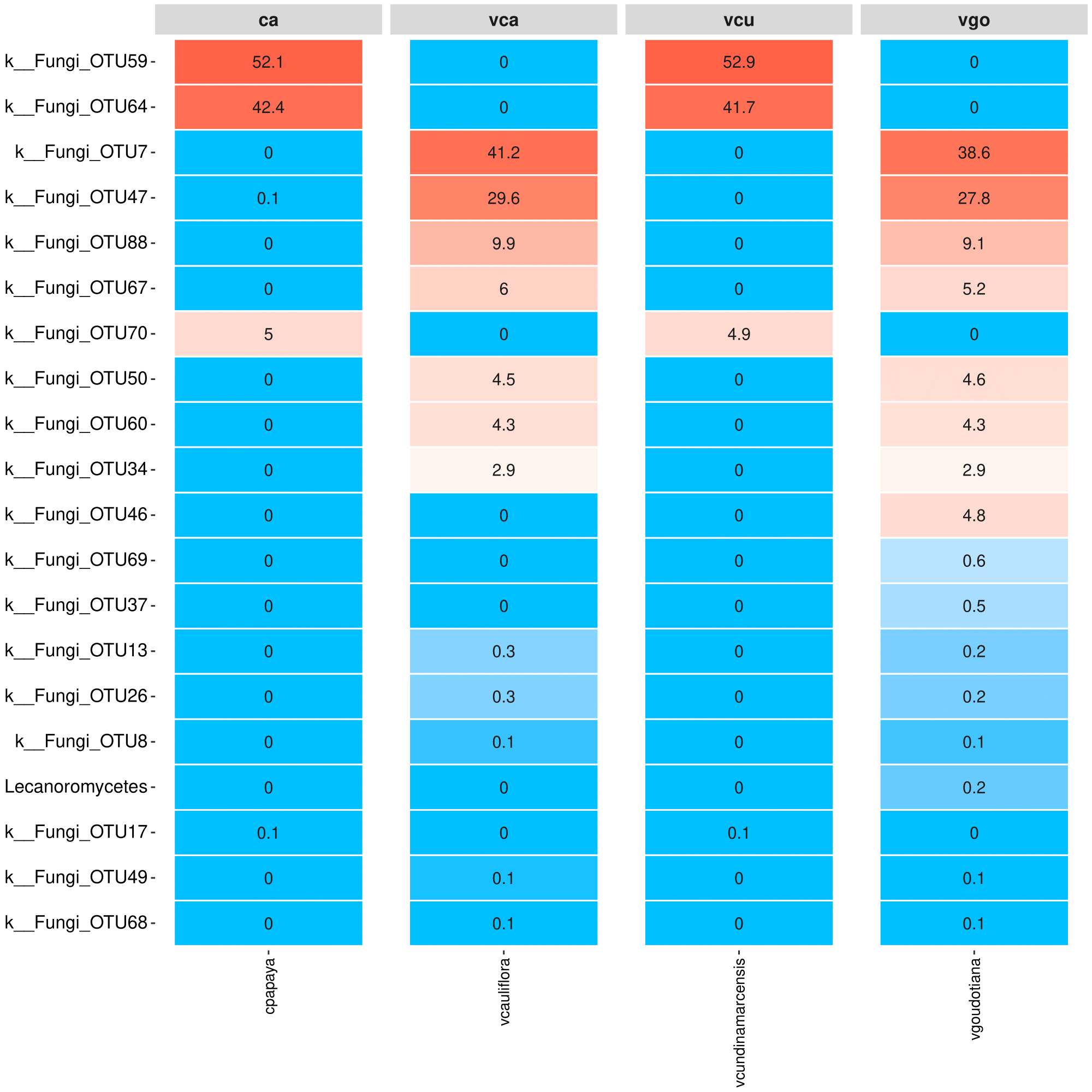
Deliverables/ITS\_PIPTIS/Figures/ITS\_ampvis2\_heatmap/png/9\_a\_Phylum\_Heatmap.png

ca: cpapaya

vca: vcauliflora

vcu: vcundinamarcensis

vgo: vgoudotiana



#### Figure 09(b):Class Level Heatmap using Ampvis

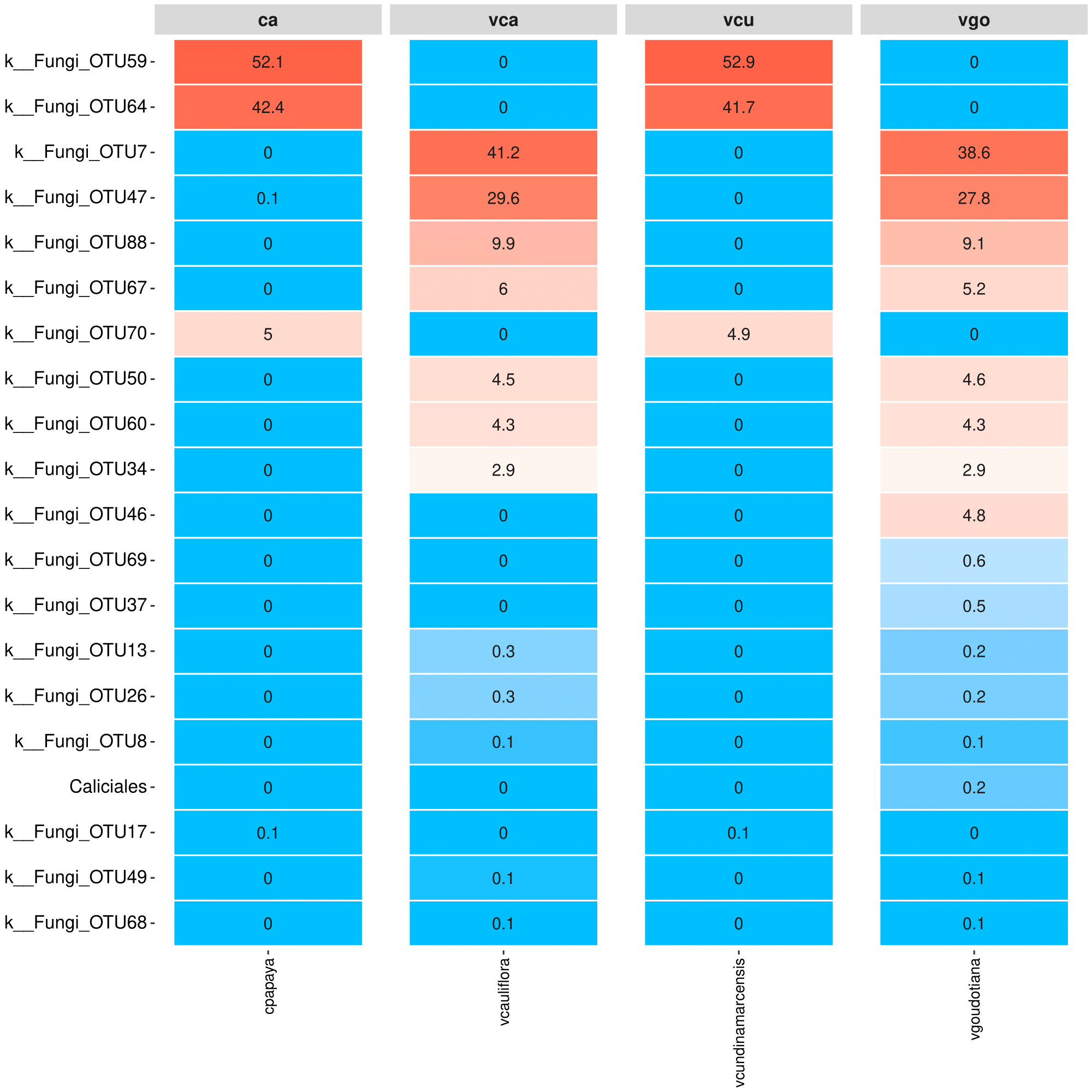
Deliverables/ITS\_PIPTIS/Figures/ITS\_ampvis2\_heatmap/png/9\_b\_Class\_Heatmap.png

ca: cpapaya

vca: vcauliflora

vcu: vcundinamarcensis

vgo: vgoudotiana



#### Figure 09(c): Order Level Heatmap using Ampvis

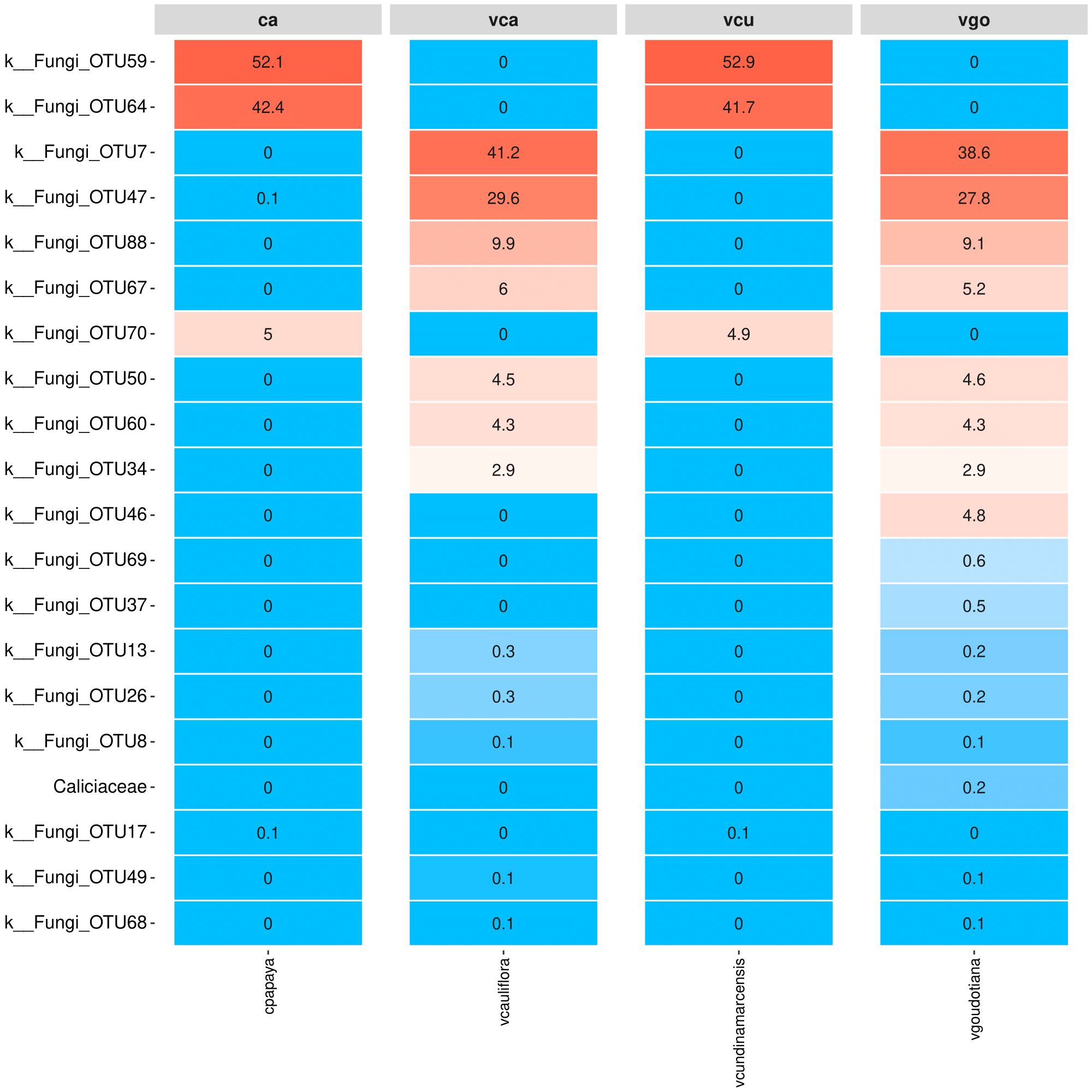
Deliverables/ITS\_PIPTIS/Figures/ITS\_ampvis2\_heatmap/png/9\_c\_Order\_Heatmap.png

ca: cpapaya

vca: vcauliflora

vcu: vcundinamarcensis

vgo: vgoudotiana



#### Figure 09(d): Family Level Heatmap using Ampvis

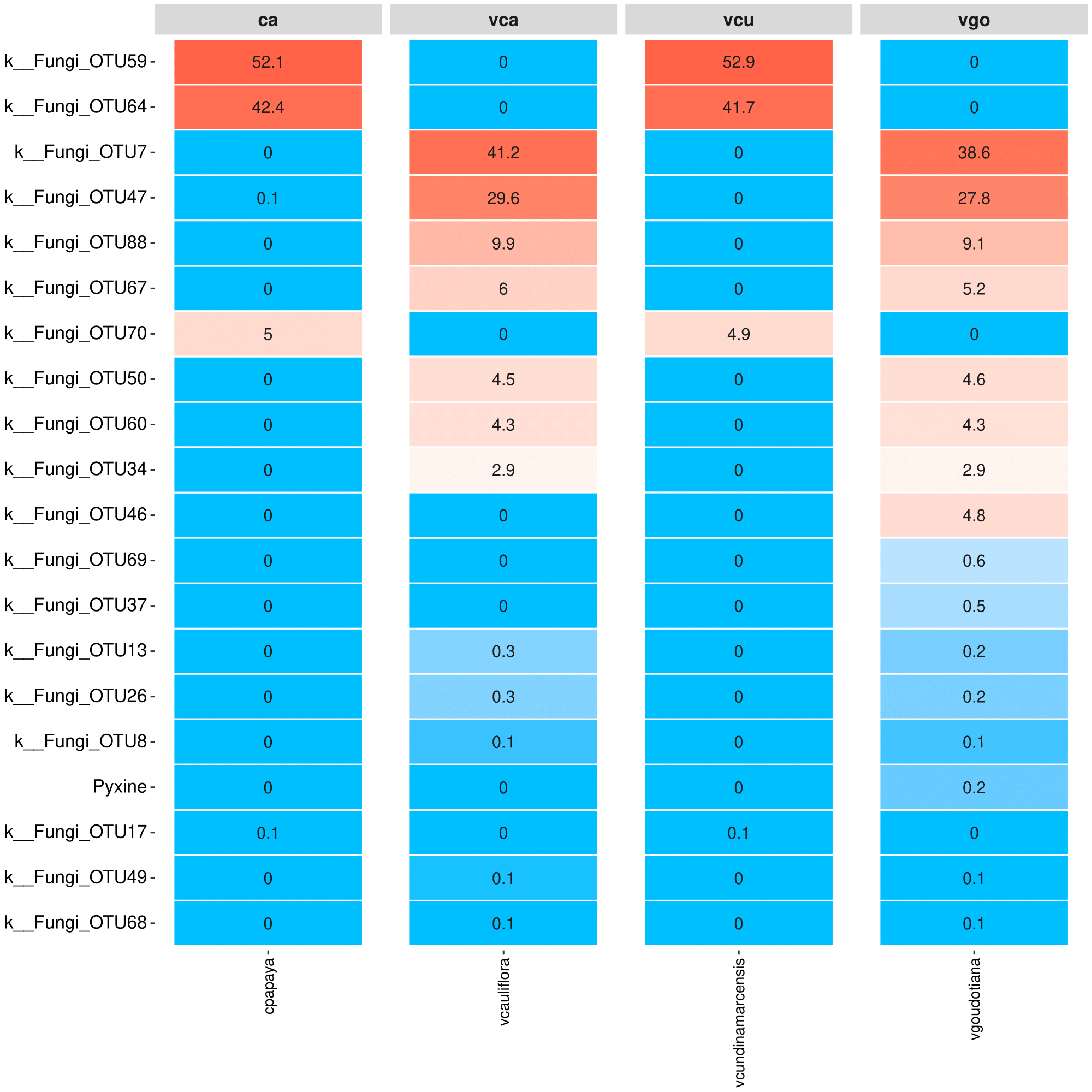
Deliverables/ITS\_PIPTIS/Figures/ITS\_ampvis2\_heatmap/png/9\_d\_Family\_Heatmap.png

ca: cpapaya

vca: vcauliflora

vcu: vcundinamarcensis

vgo: vgoudotiana



#### Figure 09(e): Genus Level Heatmap using Ampvis

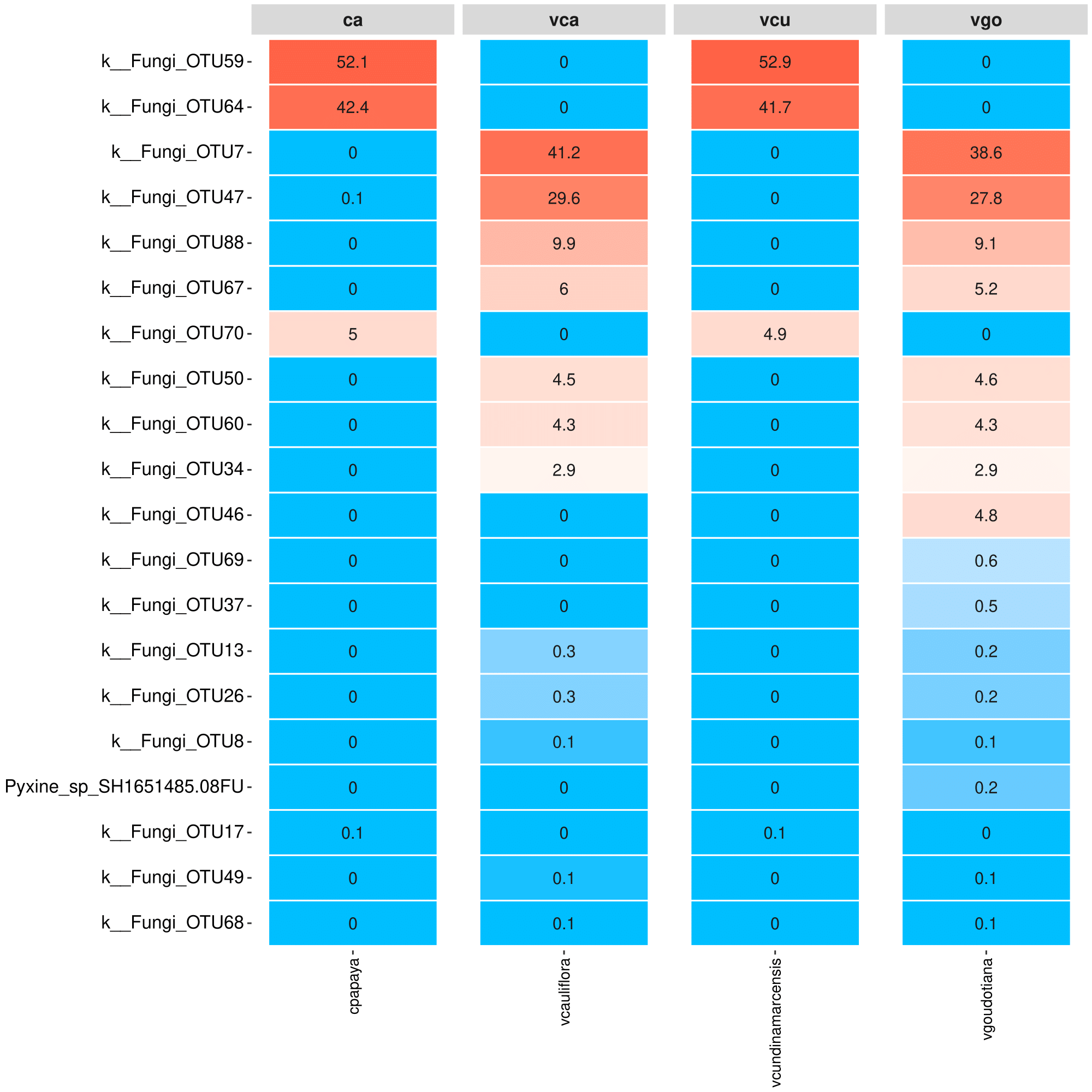
Deliverables/ITS\_PIPTIS/Figures/ITS\_ampvis2\_heatmap/png/9\_e\_Genus\_Heatmap.png

ca: cpapaya

vca: vcauliflora

vcu: vcundinamarcensis

vgo: vgoudotiana



#### Figure 09(f): Species Level Heatmap using Ampvis

Deliverables/ITS\_PIPTIS/Figures/ITS\_ampvis2\_heatmap/png/9\_f\_Species\_Heatmap.png

ca: cpapaya

vca: vcauliflora

vcu: vcundinamarcensis

vgo: vgoudotiana

#### 

#### Reference:

1. DADA2, "Callahan, Benjamin J., et al. "DADA2: high-resolution sample inference from Illumina amplicon data." Nature methods 13.7 (2016): 581."
2. QIIME2, "Bolyen, Evan, et al. "Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2." Nature Biotechnology 37 (2019): 852â€“857."
3. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, Garcia MU, Di Tommaso P, Nahnsen S. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020 Mar;38(3):276-278.
4. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3:160018. doi: 10.1038/sdata.2016.18
5. FastQC, "Andrews, Simon. "FastQC: a quality control tool for high throughput sequence data." (2010)."
6. MultiQC, "Ewels, Philip, et al. "MultiQC: summarize analysis results for multiple tools and samples in a single report." Bioinformatics 32.19 (2016): 3047-3048."
7. Cutadapt "Martin, Marcel. "Cutadapt removes adapter sequences from high-throughput sequencing reads." EMBnet. journal 17.1 (2011): pp-10."
8. SILVA database, "Quast, Pruesse, et al. 2013. 'The SILVA ribosomal RNA gene database project: improved data processing and web-based tools', Nucleic Acids Research, 41: D590-D96."
9. Feature-classifier, "Bokulich, Kaehler, et al. "Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin." Microbiome 6 (2018): 90.
10. Dhariwal, A., Chong, J., Habib, S., King, I. L., Agellon, L. B., & Xia, J. (2017). MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic acids research*, *45*(W1), W180-W188.
11. Aßhauer, Kathrin P., Bernd Wemheuer, Rolf Daniel, and Peter Meinicke. "Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data." *Bioinformatics* 31, no. 17 (2015): 2882-2884.
12. McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*, *8*(4), e61217.
13. Andersen, K. S., Kirkegaard, R. H., Karst, S. M., & Albertsen, M. (2018). ampvis2: an R package to analyse and visualise 16S rRNA amplicon data. BioRxiv, 299537.
14. Douglas, G. M., Maffei, V. J., Zaneveld, J., Yurgel, S. N., Brown, J. R., Taylor, C. M., ... & Langille, M. G. (2020). PICRUSt2: An improved and customizable approach for metagenome inference. *BioRxiv*, 672295.
15. Gweon, H. S., Oliver, A., Taylor, J., Booth, T., Gibbs, M., Read, D. S., ... & Schonrogge, K. (2015). PIPITS: an automated pipeline for analyses of fungal internal transcribed spacer sequences from the I llumina sequencing platform. *Methods in ecology and evolution*, *6*(8), 973-980.
16. Anslan, S., Nilsson, R. H., Wurzbacher, C., Baldrian, P., Tedersoo, L., & Bahram, M. (2018). Great differences in performance and outcome of high-throughput sequencing data analysis platforms for fungal metabarcoding. *MycoKeys*, (39), 29.
17. Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584.
18. Bengtsson‐Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., ... & Nilsson, R. H. (2013). Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in ecology and evolution*, *4*(10), 914-919.