

D4006- Gut Microbiota Analysis

UC Davis MMPC - Microbiome & Host Response Core

Contents

1 Methods:	1
1.1 Sequencing	1
1.2 Data processing	1
2 Summary of Findings:	2
2.1 Sequencing analysis	2
2.2 Microbial diversity	2
2.2.1 Alpha Diversity	2
2.2.2 Beta Diversity	10
2.3 Data analysis using taxa abundance data	13
2.3.1 Stacked bar graphs of Taxa abundances at each level	14
A Appendix 1 (Taxa Abundance Tables)	30
B Appendix 2 (Taxa removed from filtered datasets at each level)	37
References:	43

Core Contacts:

Helen E. Raybould, Ph.D., Core Leader (heraybould@ucdavis.edu)
 Trina A. Knotts, Ph.D., Core Co-Leader (taknotts@ucdavis.edu)
 Michael L. Goodson, Ph.D., Core Scientist (mlgoodson@ucdavis.edu)

Client(s): Kent Lloyd, DVM PhD ;MMRRC; UC Davis

Project #: MBP-3995

MMRRC strain ID: MMRRC_044024

Animal Information: The strain was donated to the MMRRC by John Rubenstein at University of California, San Francisco. Fecal samples were obtained from animals housed under the care of John Rubenstein at University of California, San Francisco consuming LabDiet PicoLab Mouse Diet 20(5058).

1 Methods:

Brief Project Description: MMRRC strains are often contributed to the MMRRC to fulfill the resource sharing aspects of NIH grants. Since transporting mice to another facility often causes a microbiota shift, having a record of the original fecal microbiota from the donor institution where the original phenotyping or testing was performed may prove helpful if a phenotype is lost after transfer. Several MMRRC mouse lines were selected for fecal microbiota profiling of the microbiota.

Table 1: Animal-Strain Information

X.SampleID	TreatmentGroup	Animal_ID	Genotype	Line	Sex
MMRRC.44024_M418	MMRRC.44024_Tg_M418	M418	Tg	MMRRC.44024	M
MMRRC.44024_M419	MMRRC.44024_Tg_M419	M419	Tg	MMRRC.44024	M

1.1 Sequencing

Frozen fecal or regional gut samples were shipped on dry ice to UC Davis MMPC and Host Microbe Systems Biology Core. Total DNA was extracted using Mo-Bio (now Qiagen) PowerFecal kit. Sample libraries were prepared and analyzed by barcoded amplicon sequencing. In brief, the purified DNA was amplified on the V4 region of the 16S rRNA genes via PCR using the following primers: F319 (5'-ACTCCTACGGGAGGCAGCAGT-3') and R806 (5'-GGACTACNVGGGTWTCTAAT-3'). High-throughput sequencing was performed with Illumina MiSeq paired end 250-bp run.

1.2 Data processing

The data derived from sequencing was processed using QIIME2 for 16S based microbiota analyses (QIIME 2 Development Team (2017)). Demultiplexed paired end sequences that already had barcodes and adapters removed were analyzed using Qiime 2 version 2020.8.0. For quality filtering and feature (OTU) prediction, we used DADA2 (Callahan et al. (2016)). Upon reviewing the sequence quality data, we trimmed 0 nucleotides (nts) from the 5' end of the forward and 0 nts from the reverse reads. Forward reads were truncated to 270 nts and reverse reads to 200 nts. Representative sequences were aligned using MAFFT (Kato and Standley (2013)). A phylogenetic tree of the aligned sequences was made using FastTree 2 (Price, Dehal, and Arkin (2010)). OTUs/features were taxonomically classified using a pre-trained Naive Bayes taxonomy classifier. The classifier was trained using the Silva 128 97% OTUs (Quast et al. (2013)) for the 319F-806R region. Tables of taxonomic counts and percentage (relative frequency) were generated. Diversity analyses were run on the resulting OTU/feature .biom tables to provide both phylogenetic and non-phylogenetic metrics of alpha and beta diversity (Lozupone et al. (2011)). Additional data analysis (PLS-DA) and statistics were performed with R . Visualization files (.qzv) can be viewed through <http://view.qiime2.org> . The most relevant files for browsing in this manner are the feature tables (table.qzv (seq count data by taxa) and relative_freq_table.qzv (% abundance by taxa)), and taxa bar plots (taxa-bar-plots.qzv).

2 Summary of Findings:

2.1 Sequencing analysis

We obtained a mean of 48600 ± 2854 (SD) individual sequencing reads per sample (Min= 46582; Max= 50618). After data processing the average number of sequences for each sample passing through to OTU classification was $NA \pm NA$ (SD). The average number of OTUs per sample was 145.

2.2 Microbial diversity

Alpha and beta diversity measures were calculated using the QIIME pipeline (QIIME 2). Rarefaction was used to sample the same number of random reads from each sample for the diversity analyses. The sampling depth was set at 17000 sequences per sample.

2.2.1 Alpha Diversity

Alpha diversity is a measure of a specific sample's richness or diversity. Six different measures of alpha diversity were analyzed (Chao1, Good's coverage, Observed species, Faith's phylogenetic diversity (PD)), Shannon index, and Simpson's index. To account for differences in the number of reads in a sample, we performed rarefaction to collect a random set of sequences on a set number of sequences to rule out sample depth bias.

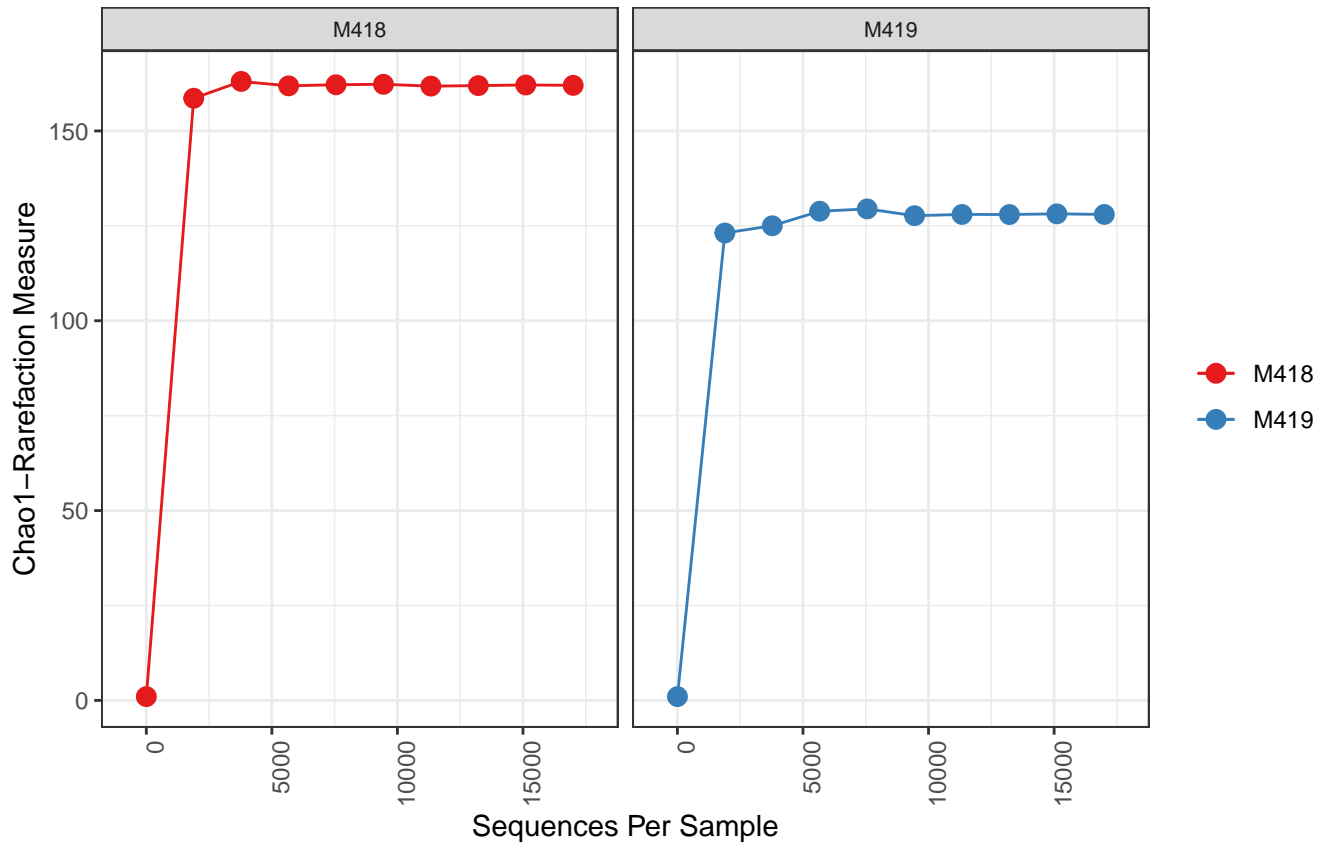


Figure 1: Chao1 Measure of Alpha Diversity- Rarefaction curve-Animal_ID.

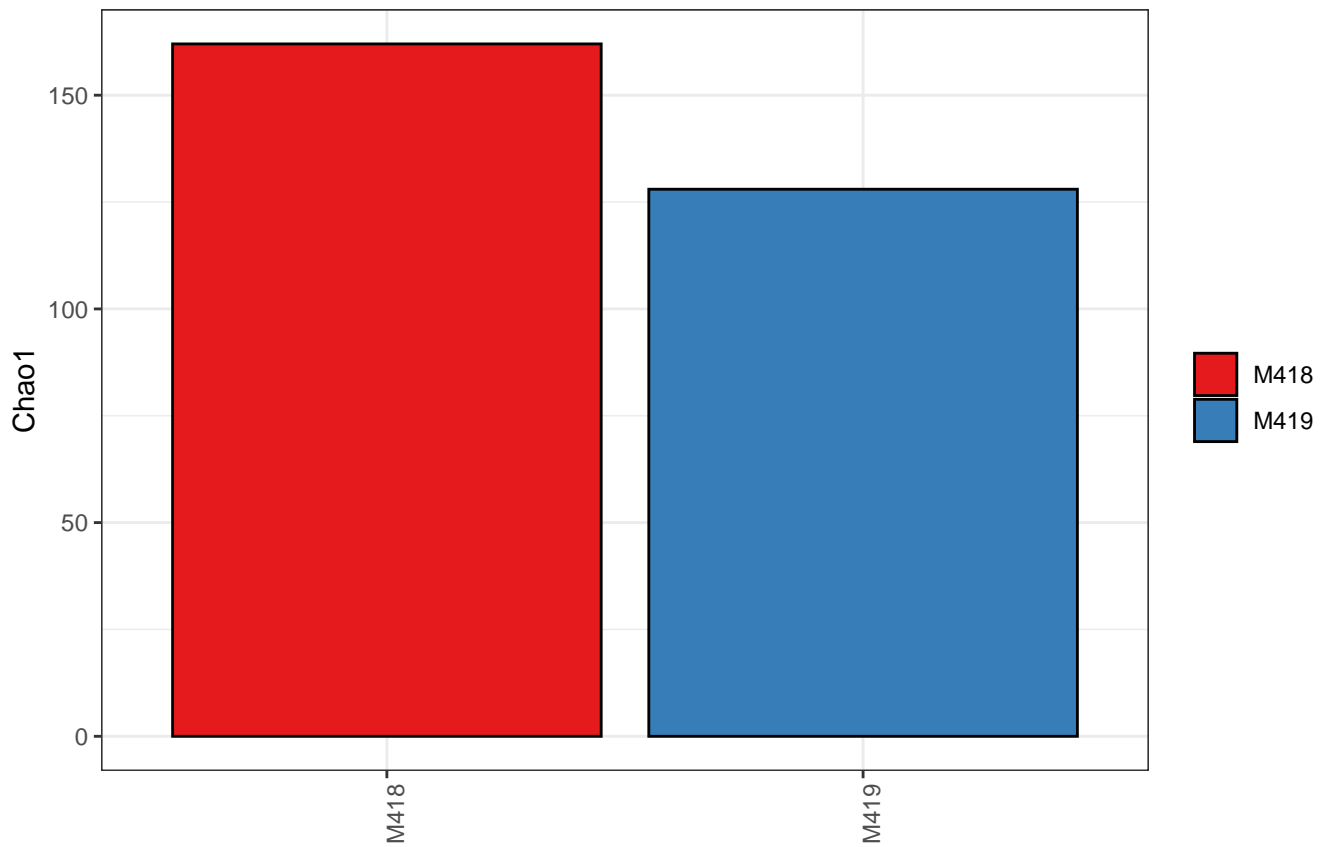


Figure 2: Chao1 Measure of Alpha Diversity- Boxplots-Animal_ID.

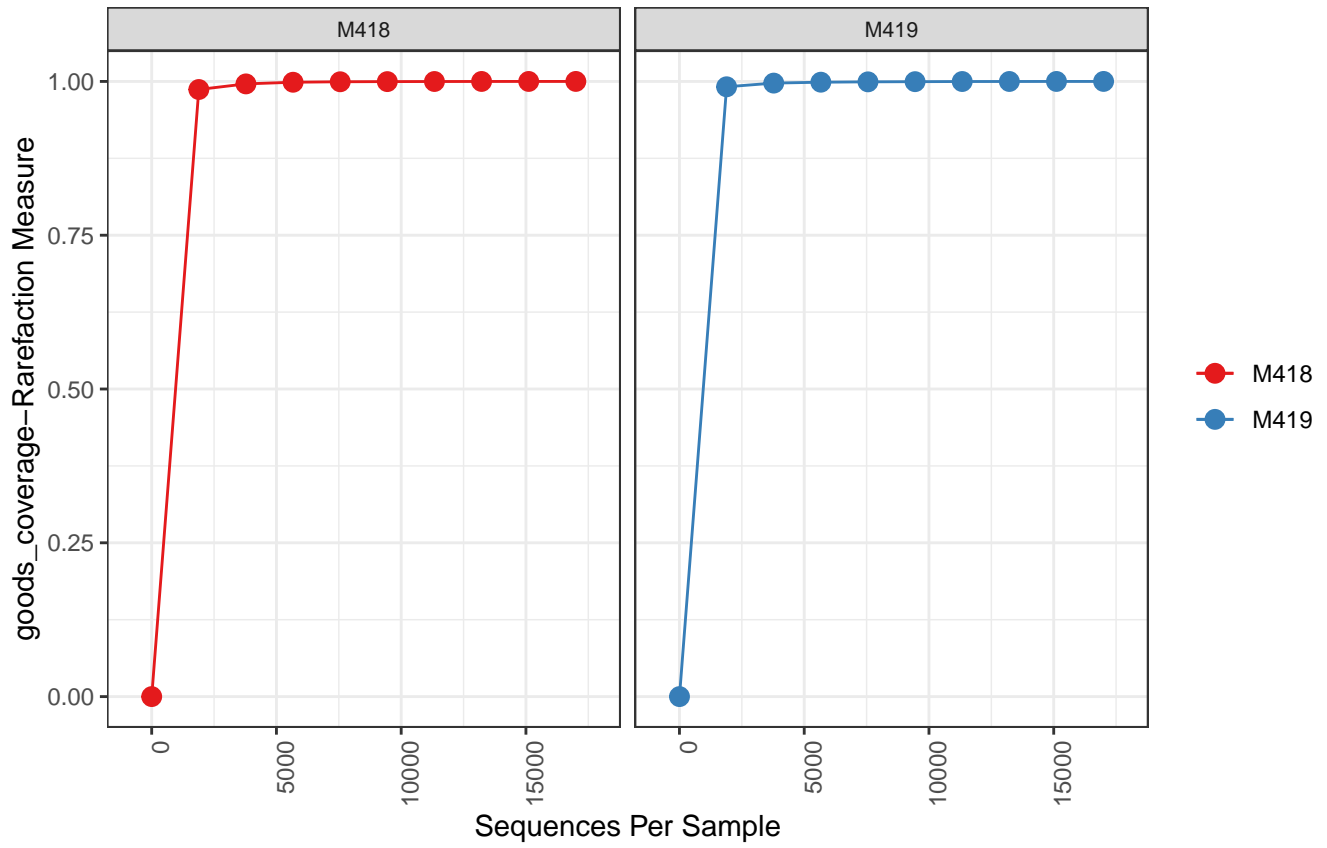


Figure 3: Goods Coverage Measure of Alpha Diversity- Rarefaction curve-Animal_ID.

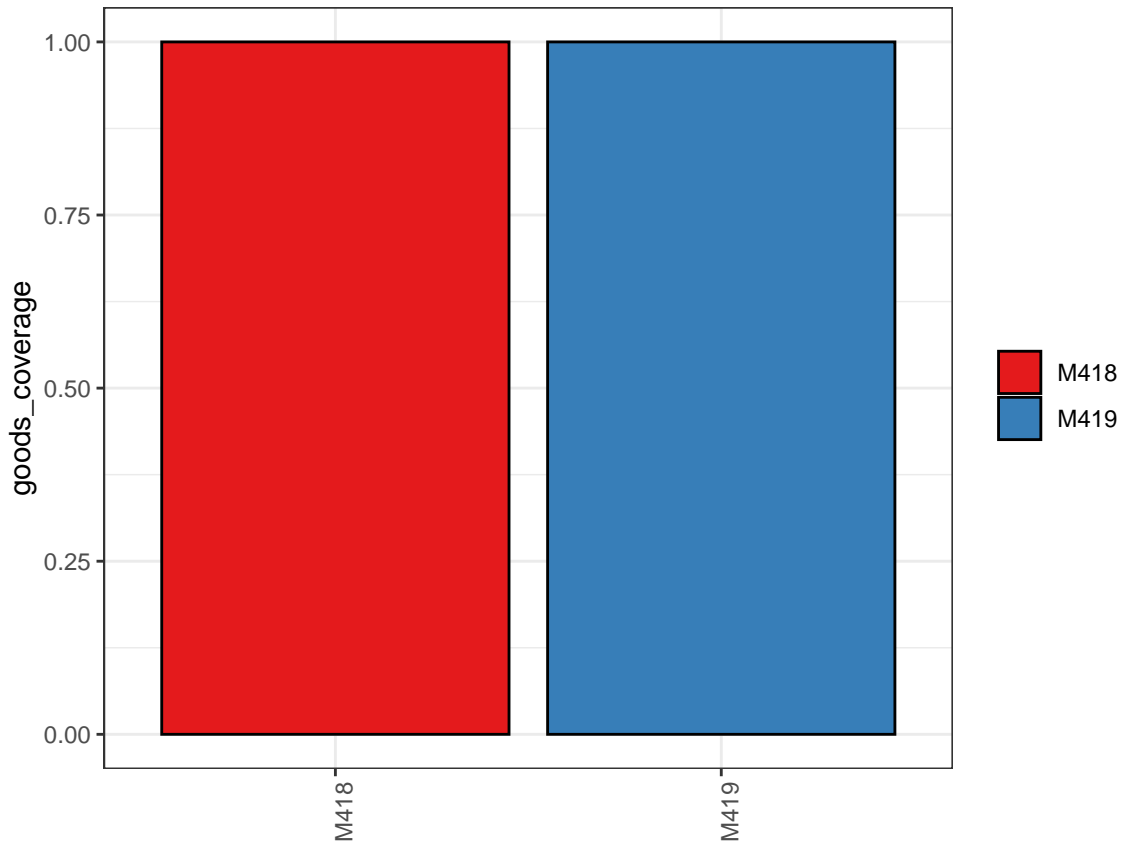


Figure 4: Goods Coverage Measure of Alpha Diversity- Boxplots-Animal_ID.

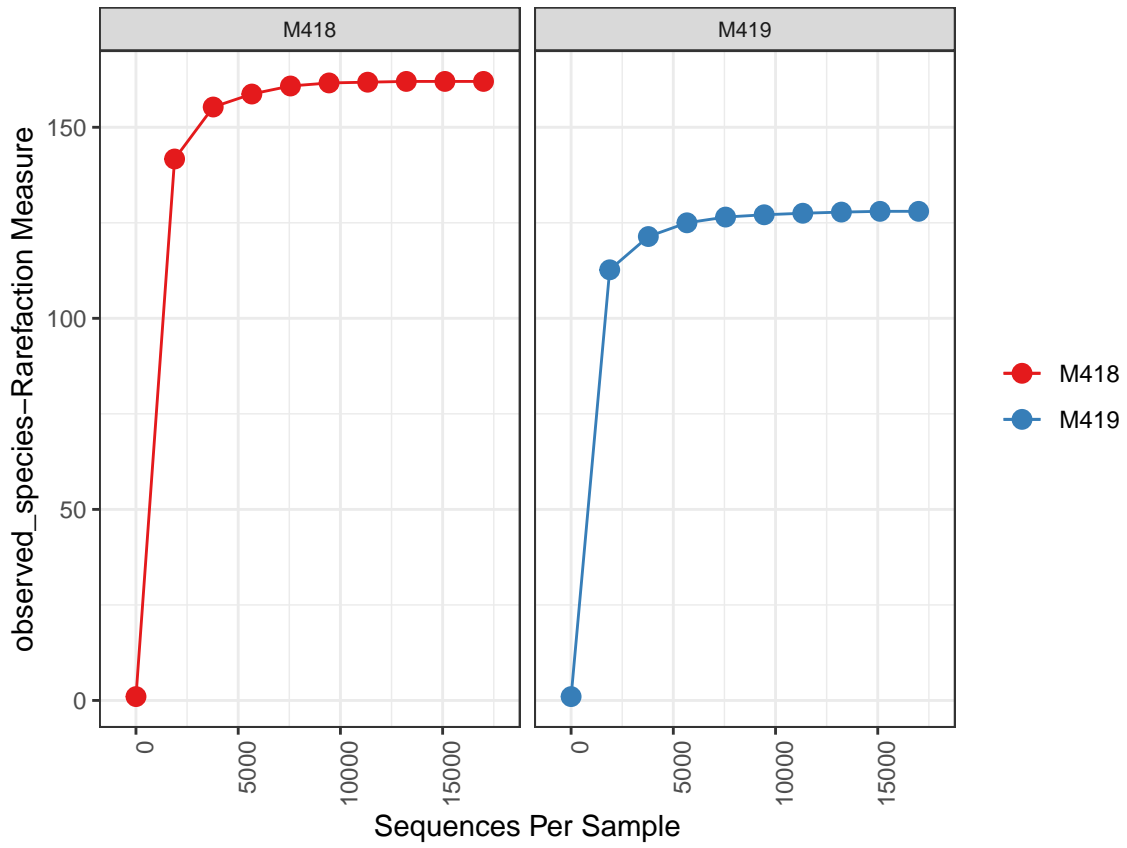


Figure 5: Observed Species Measure of Alpha Diversity- Rarefaction curve-Animal_ID.

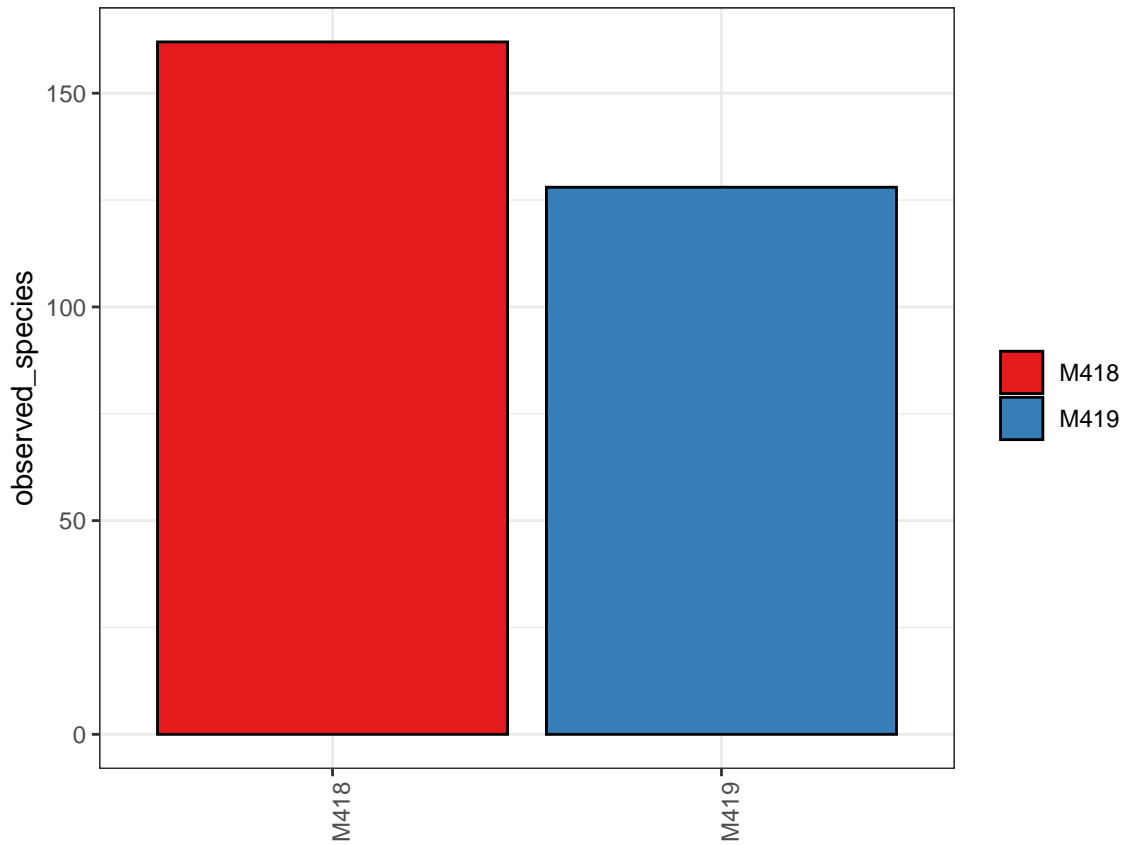


Figure 6: Observed Species Measure of Alpha Diversity- Boxplot-Animal_ID.

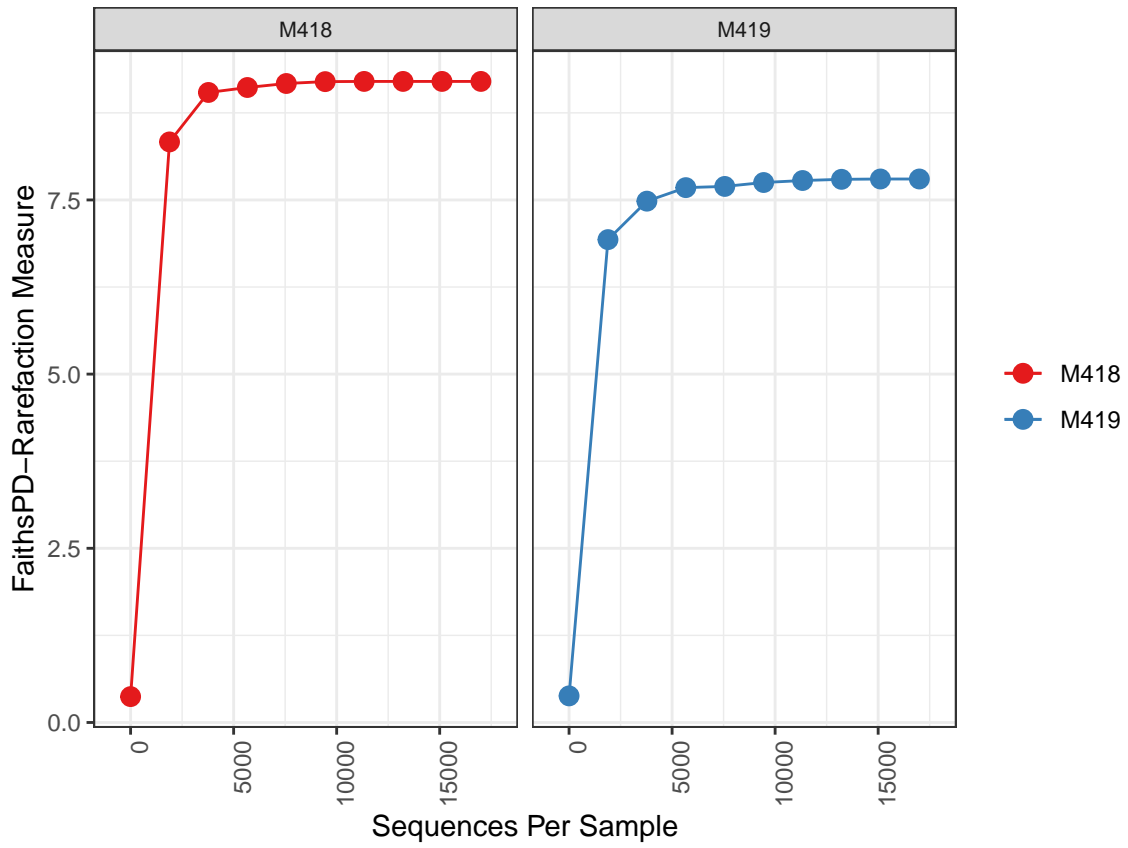


Figure 7: Faith's Phylogenetic Diversity Measure of Alpha Diversity- Rarefaction curve-Animal_ID.

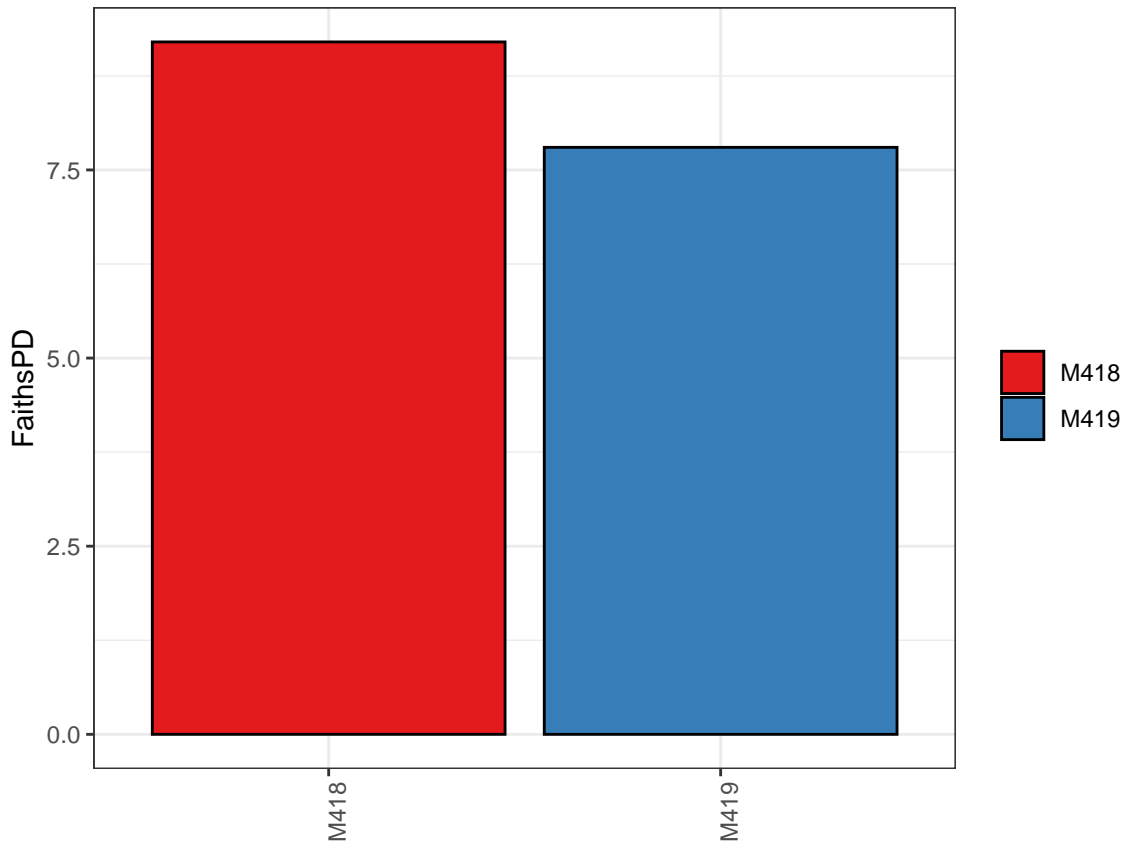


Figure 8: Faith's Phylogenetic Diversity Measure of Alpha Diversity- Boxplots-Animal_ID.

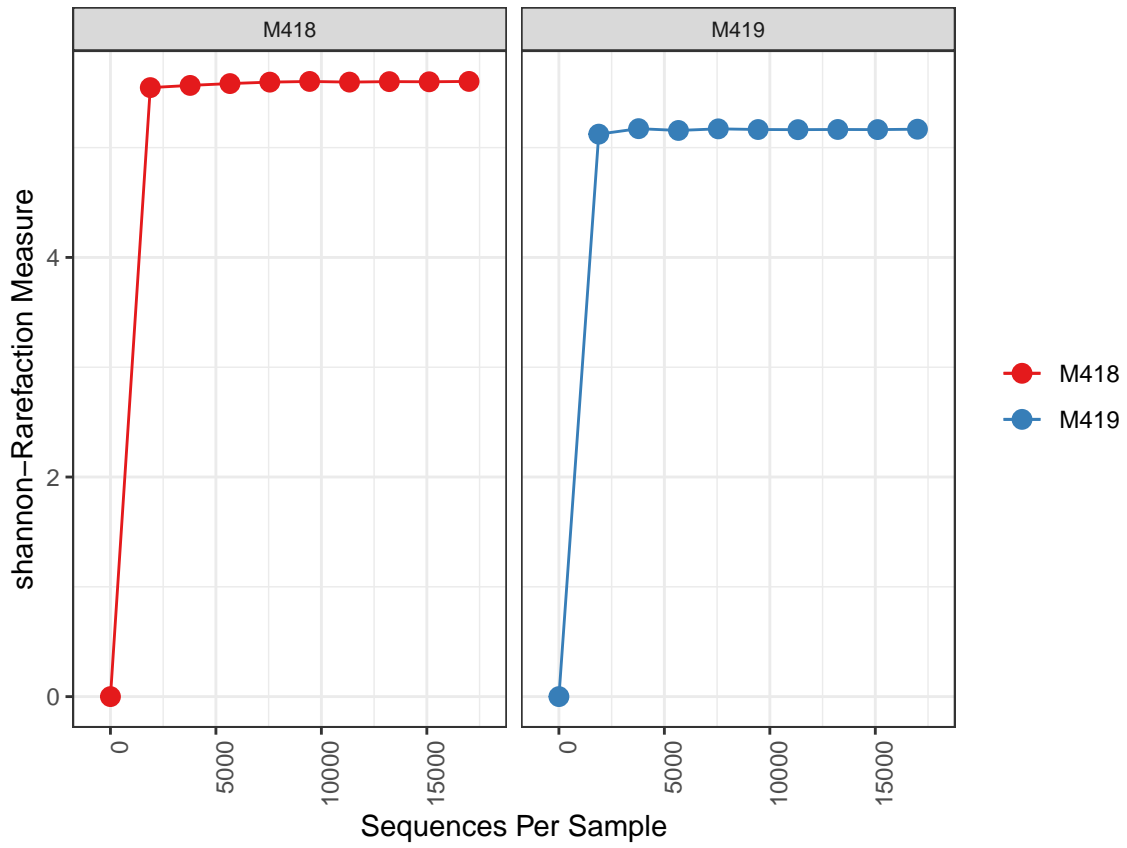


Figure 9: Shannon Measure of Alpha Diversity- Rarefaction curve-Animal_ID.

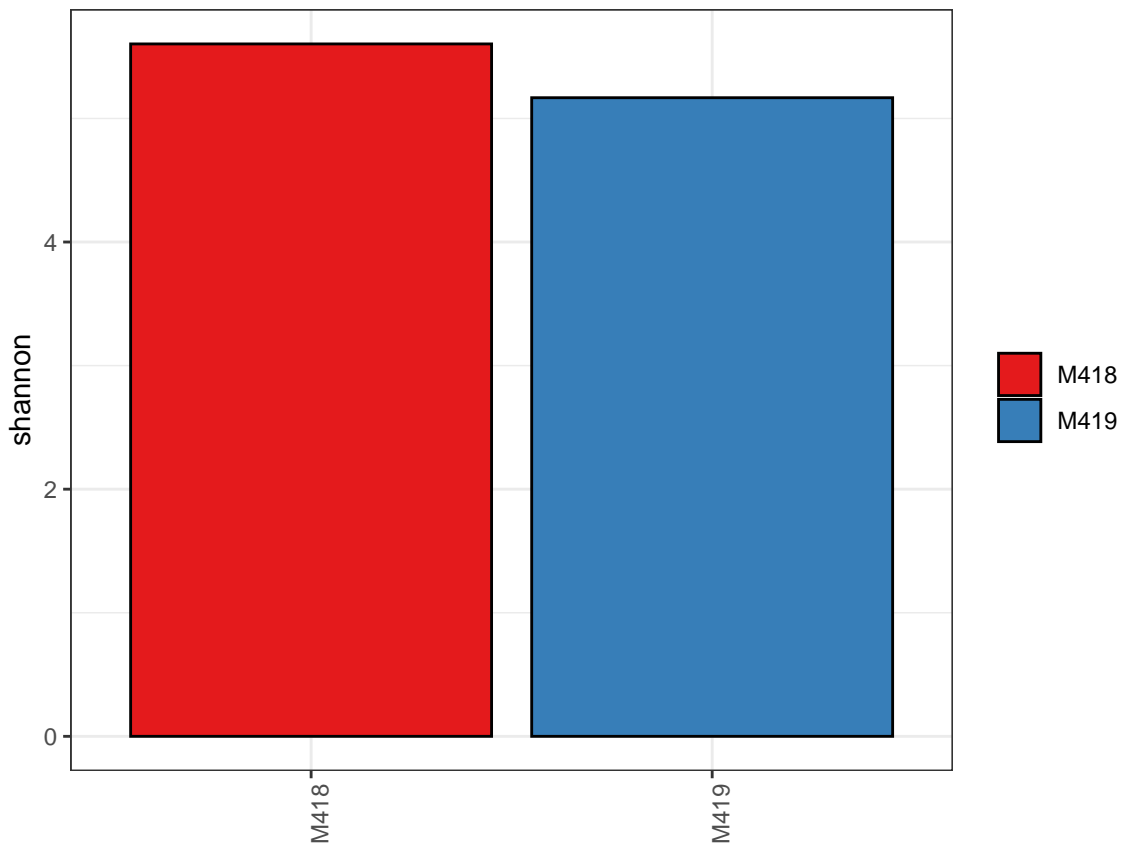


Figure 10: Shannon Measure of Alpha Diversity- Boxplots-Animal_ID.

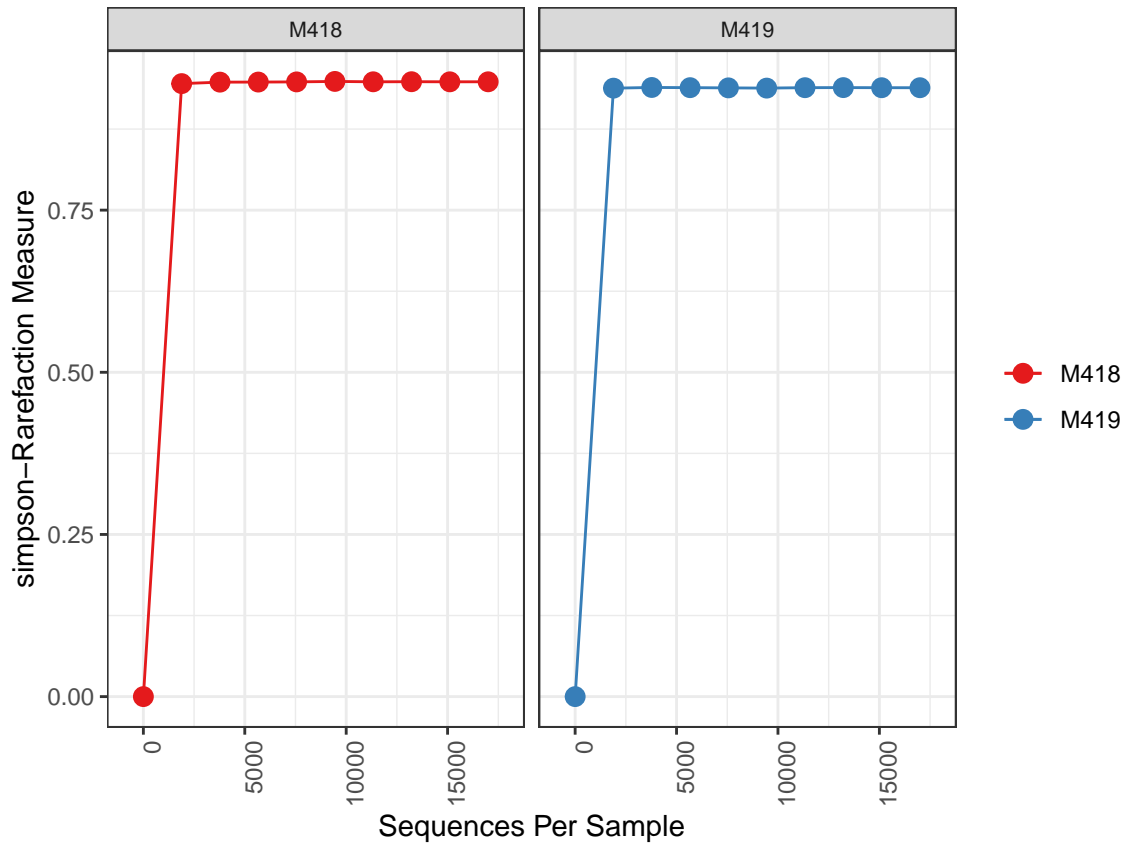


Figure 11: Simpson Measure of Alpha Diversity- Rarefaction curve-Animal_ID.

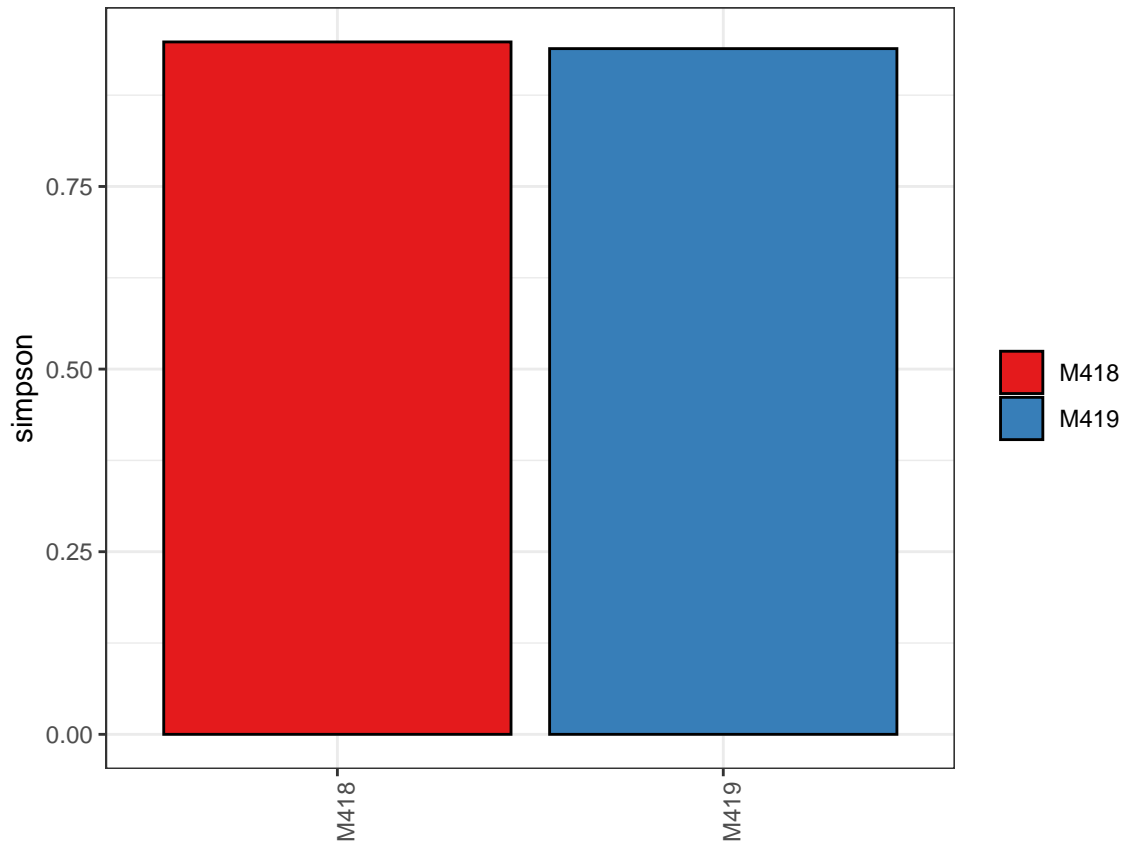


Figure 12: Simpson Measure of Alpha Diversity- Boxplots-Animal_ID.

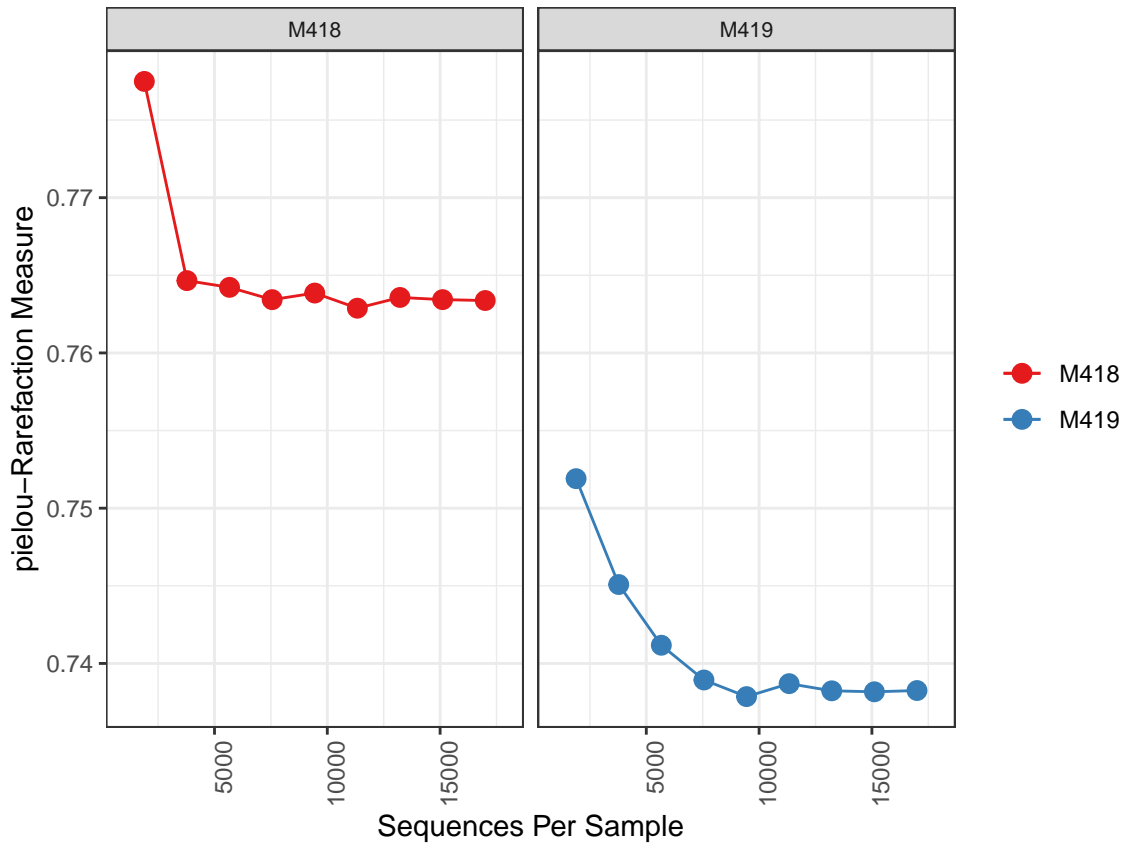


Figure 13: Pielou's Evenness Measure of Alpha Diversity- Rarefaction curve-Animal_ID.

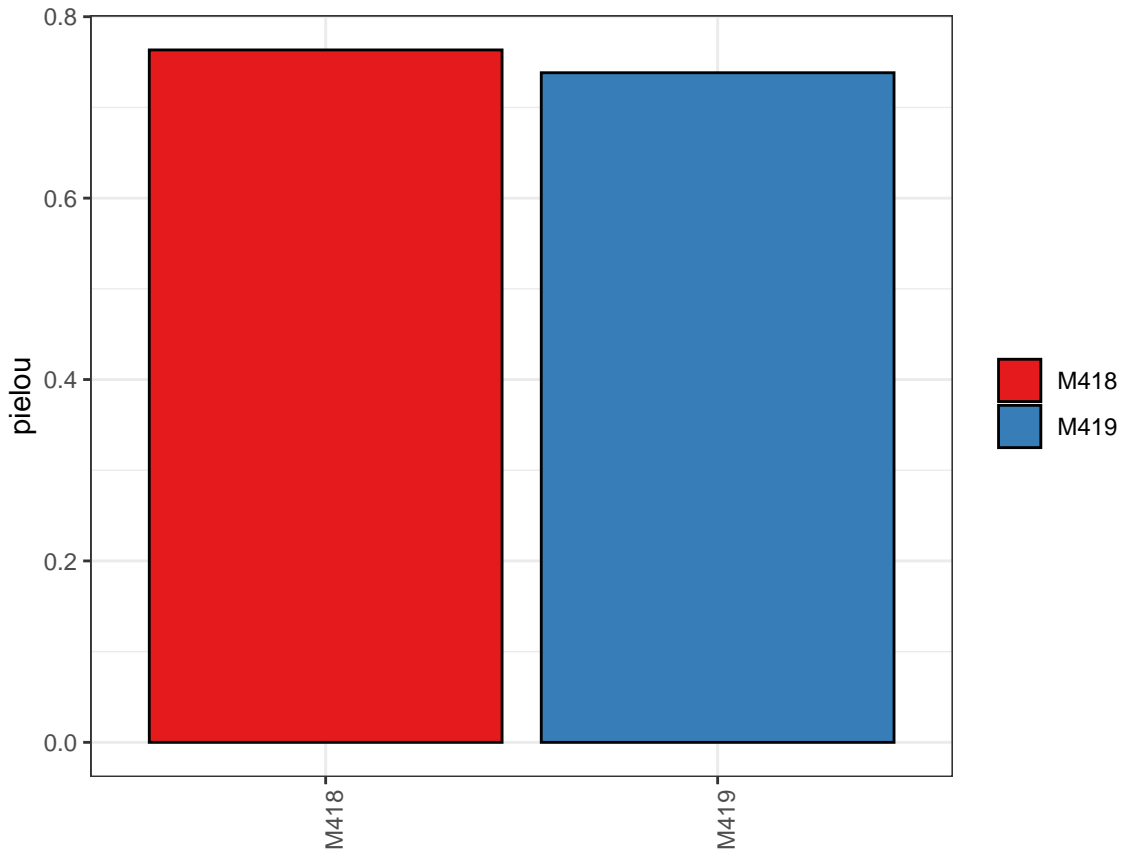


Figure 14: Pielou's Evenness Measure of Alpha Diversity- Boxplots-Animal_ID.

2.2.2 Beta Diversity

Beta diversity is the diversity between different treatment groups. Beta diversity analyses include the phylogenetic measure of unweighted and weighted UniFrac distances and the non-phylogenetic Bray-Curtis and Jaccard methods. Principal coordinates analysis (PCoA) of the first 2 coordinates of the distance matrices are plotted. The shaded ellipse depicts the 95% confidence interval. For more information on the UniFrac distance metric for community comparisons, please see (Lozupone et al., 2011). Comparing the sample beta diversity was performed with PERMANOVA (anderson).

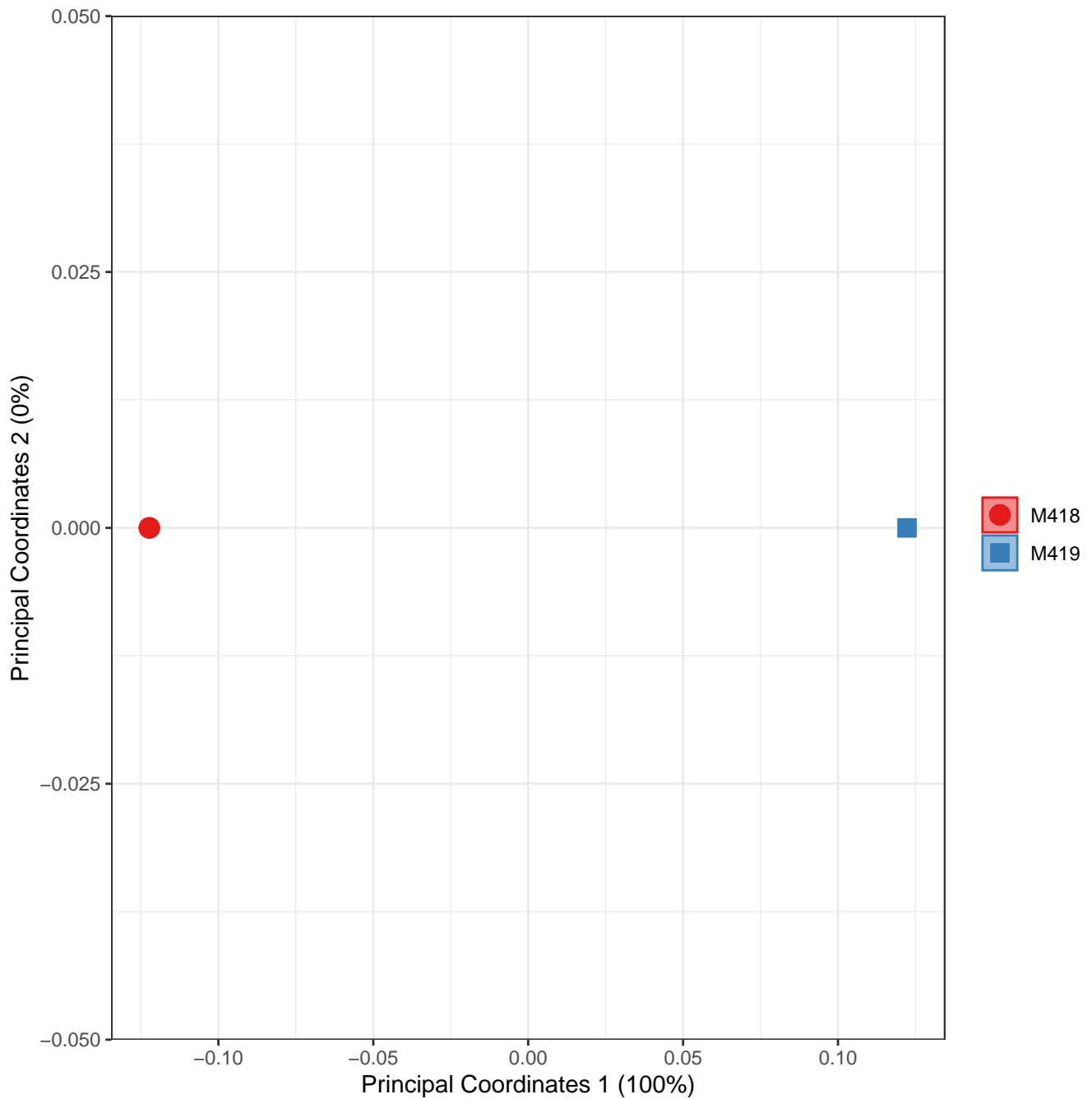


Figure 15: Unweighted Unifrac Measure of Beta Diversity-Animal_ID.

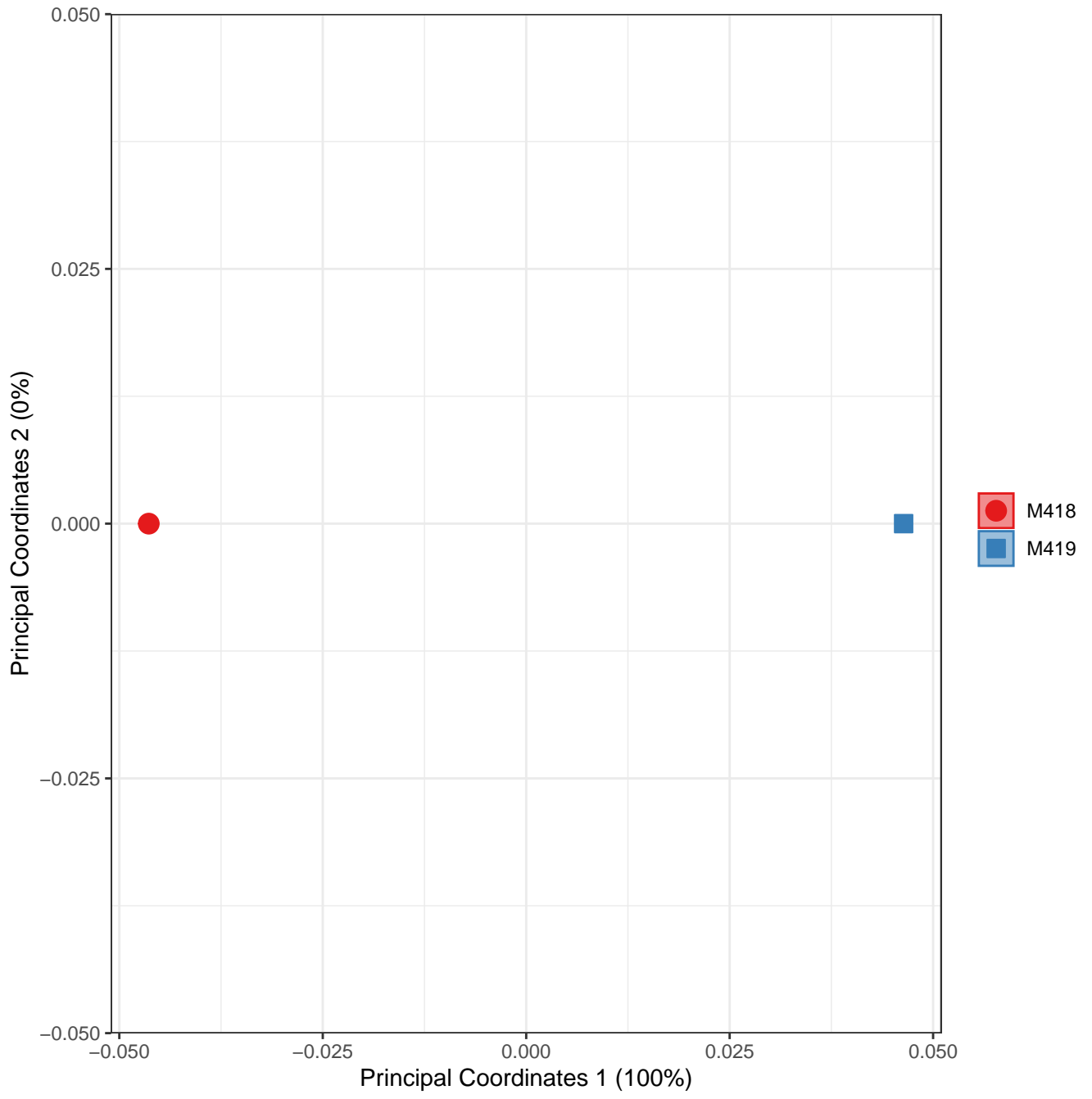


Figure 16: Weighted Unifrac Measure of Beta Diversity-Animal_ID.

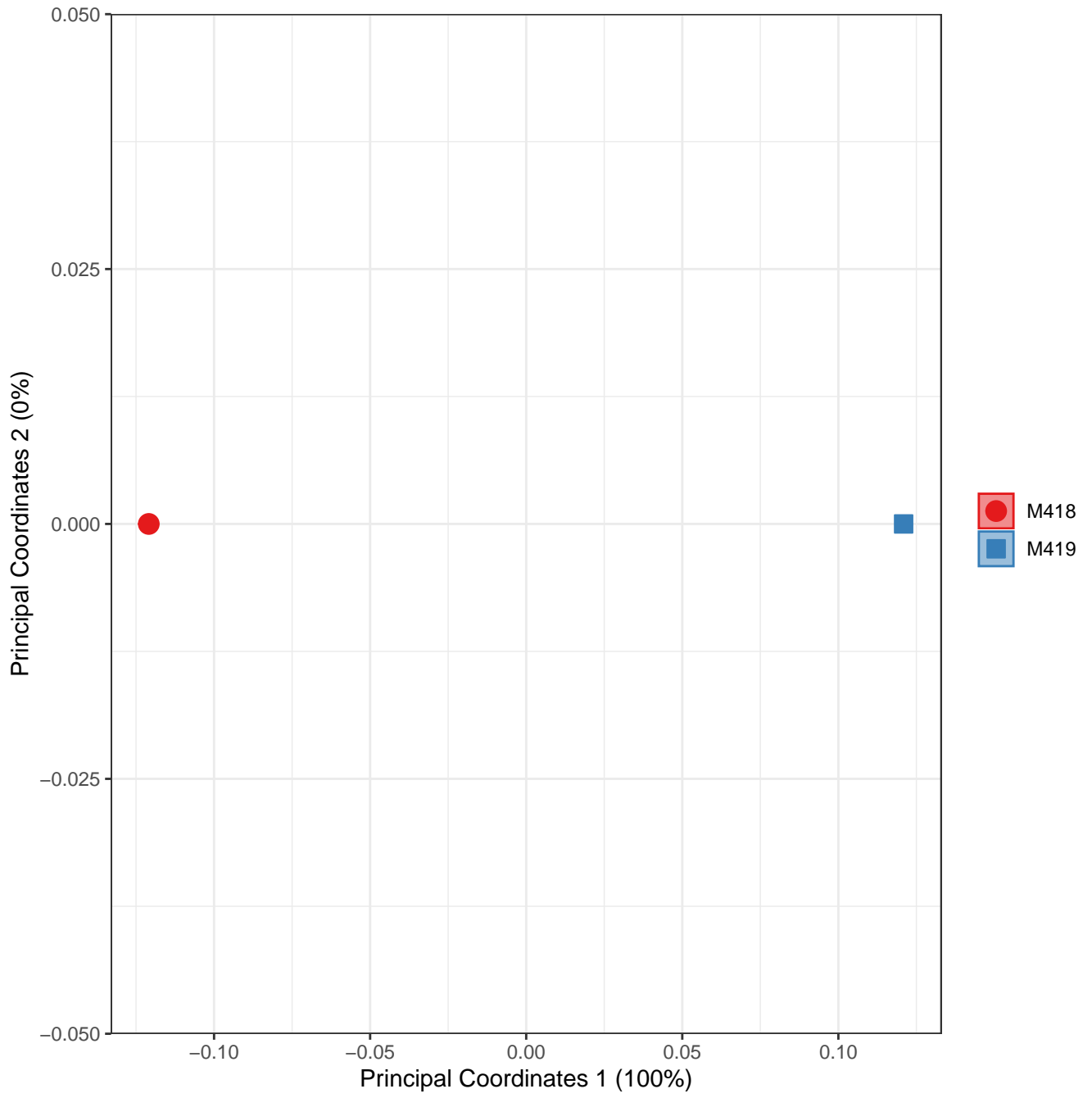


Figure 17: Bray-Curtis Measure of Beta Diversity-Animal_ID.

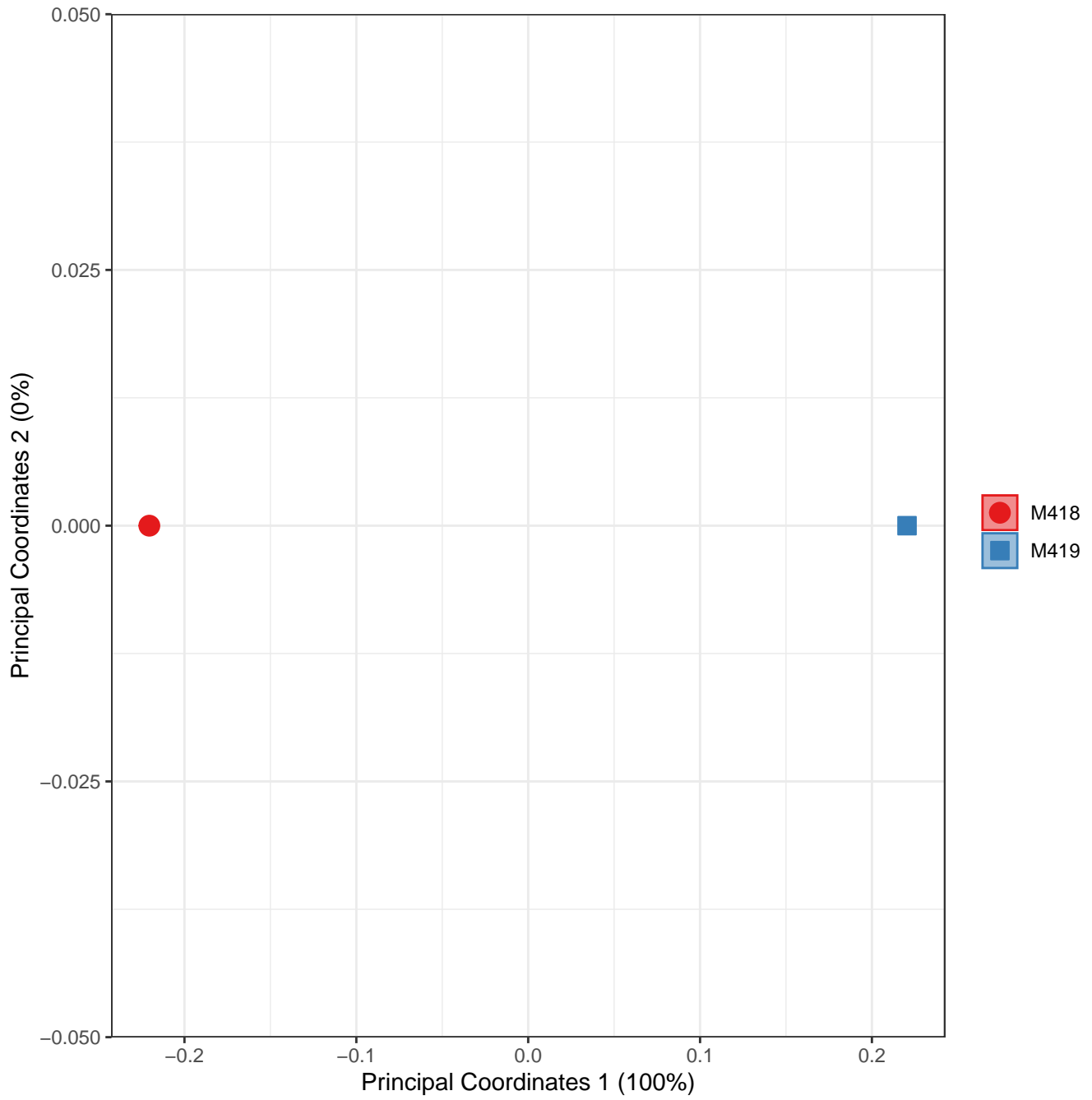


Figure 18: Jaccard Measure of Beta Diversity-Animal_ID.

2.3 Data analysis using taxa abundance data

-> -> -> -> -> -> -> -> -> -> ->

-> ->

->

2.3.1 Stacked bar graphs of Taxa abundances at each level

Sequence counts are converted to relative abundance (% abundance). Taxa abundance data is presented in 3 ways: unfiltered, filtered, and filtered/rescaled. "Unfiltered" is the unmanipulated taxa abundance data. "Filtered" data removes low abundance taxa (mean of any group < 0.05%) and genera not present in at least 50% of samples within an experimental group. Filtering out taxa will then reduce the overall abundance to less than 100%. "Filtered /Rescaled" renormalizes the filtered data back to 100% after the low abundance taxa are removed.

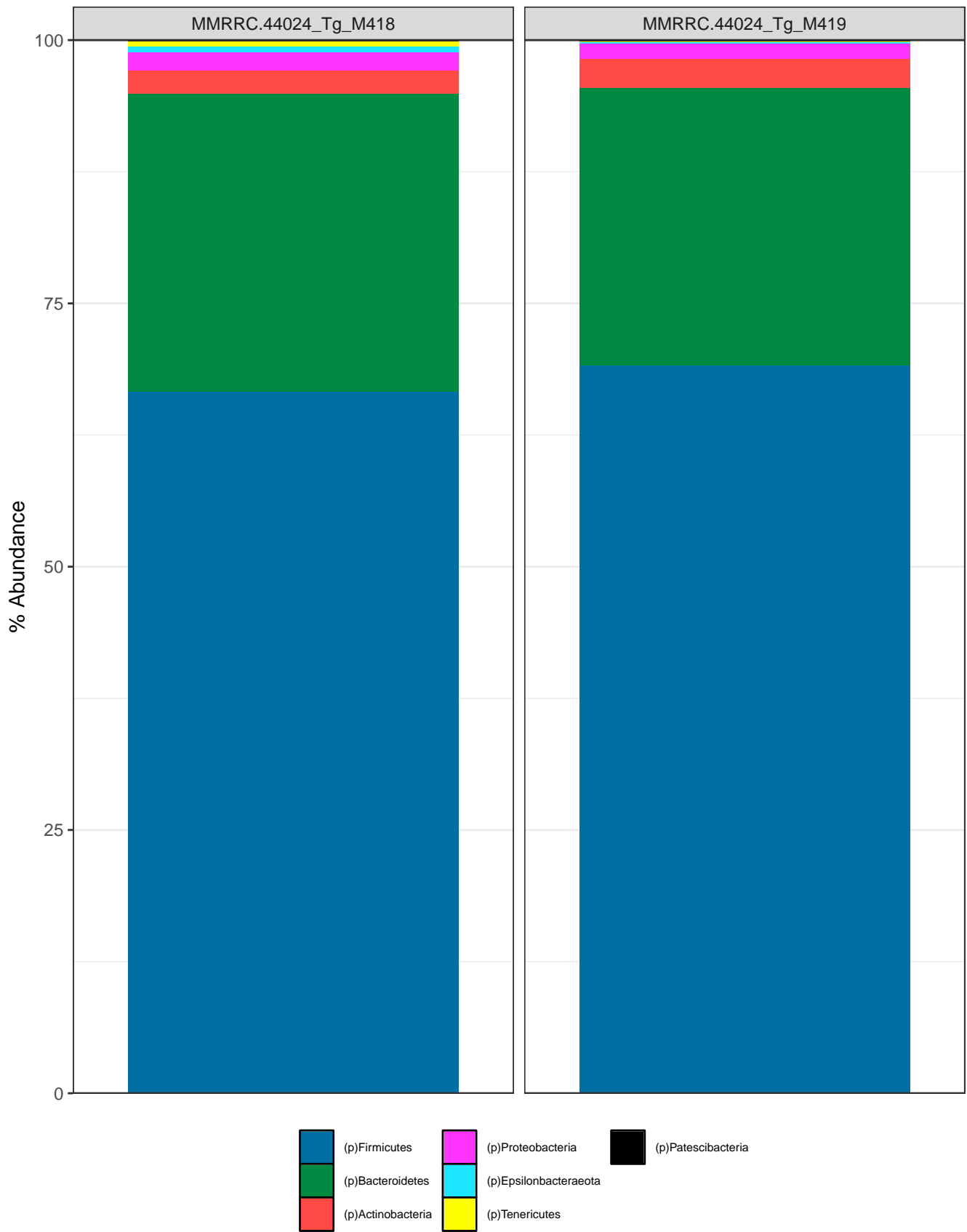


Figure 19: Phylum level abundances-UNFILTERED

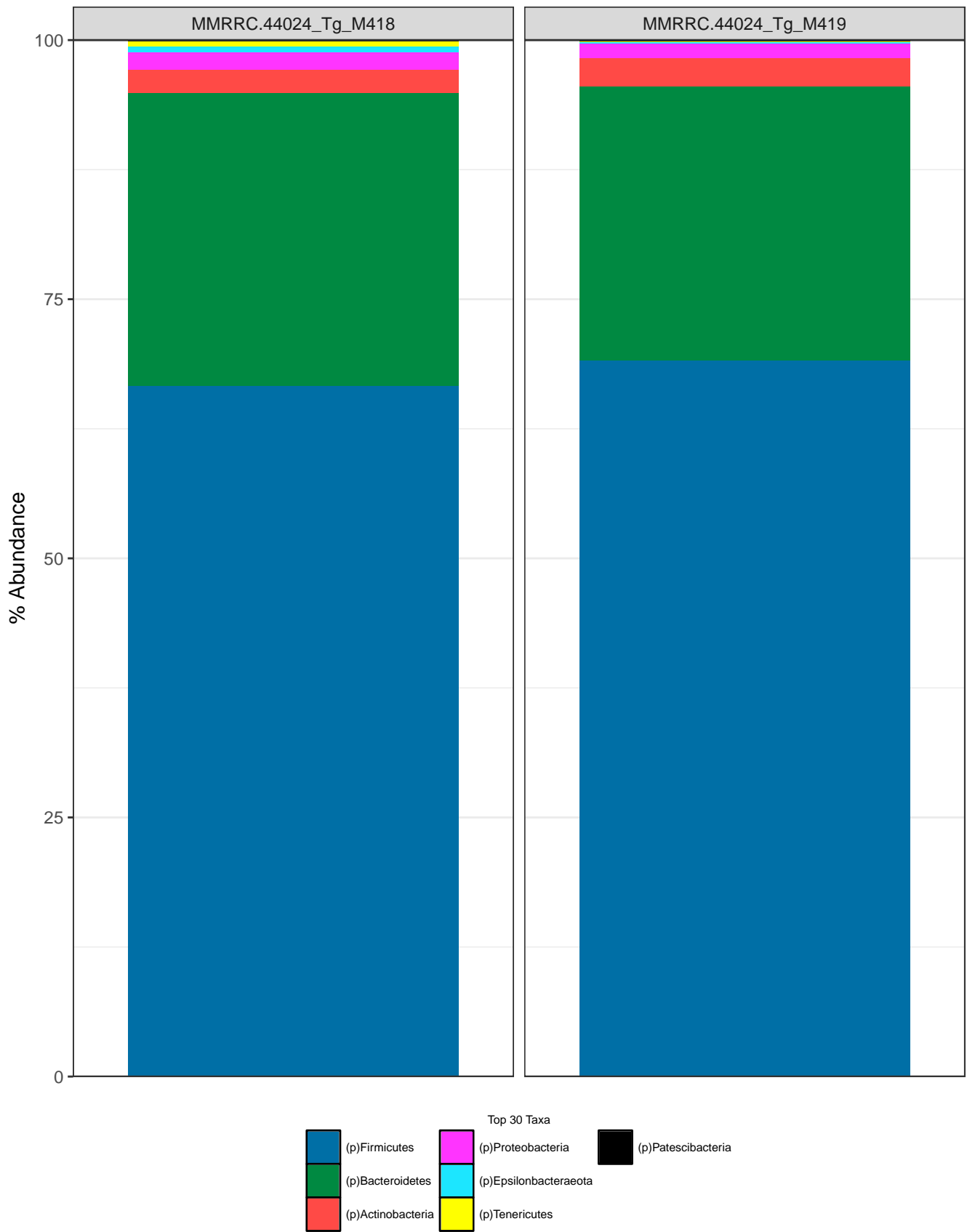


Figure 20: Phylum level abundances-FILTERED: 0 phylum level taxa removed

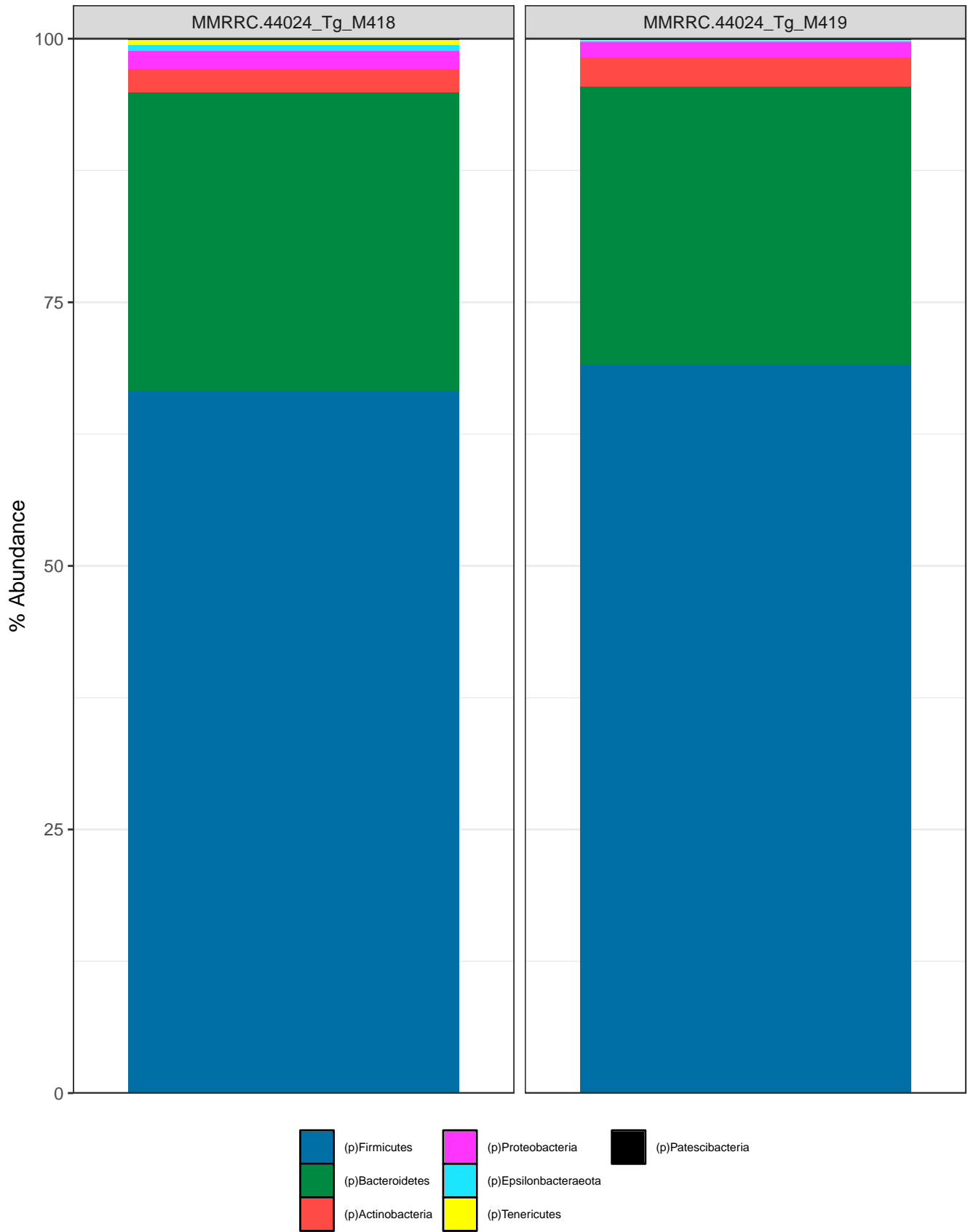


Figure 21: Phylum level abundances-FILTERED & RESCALED

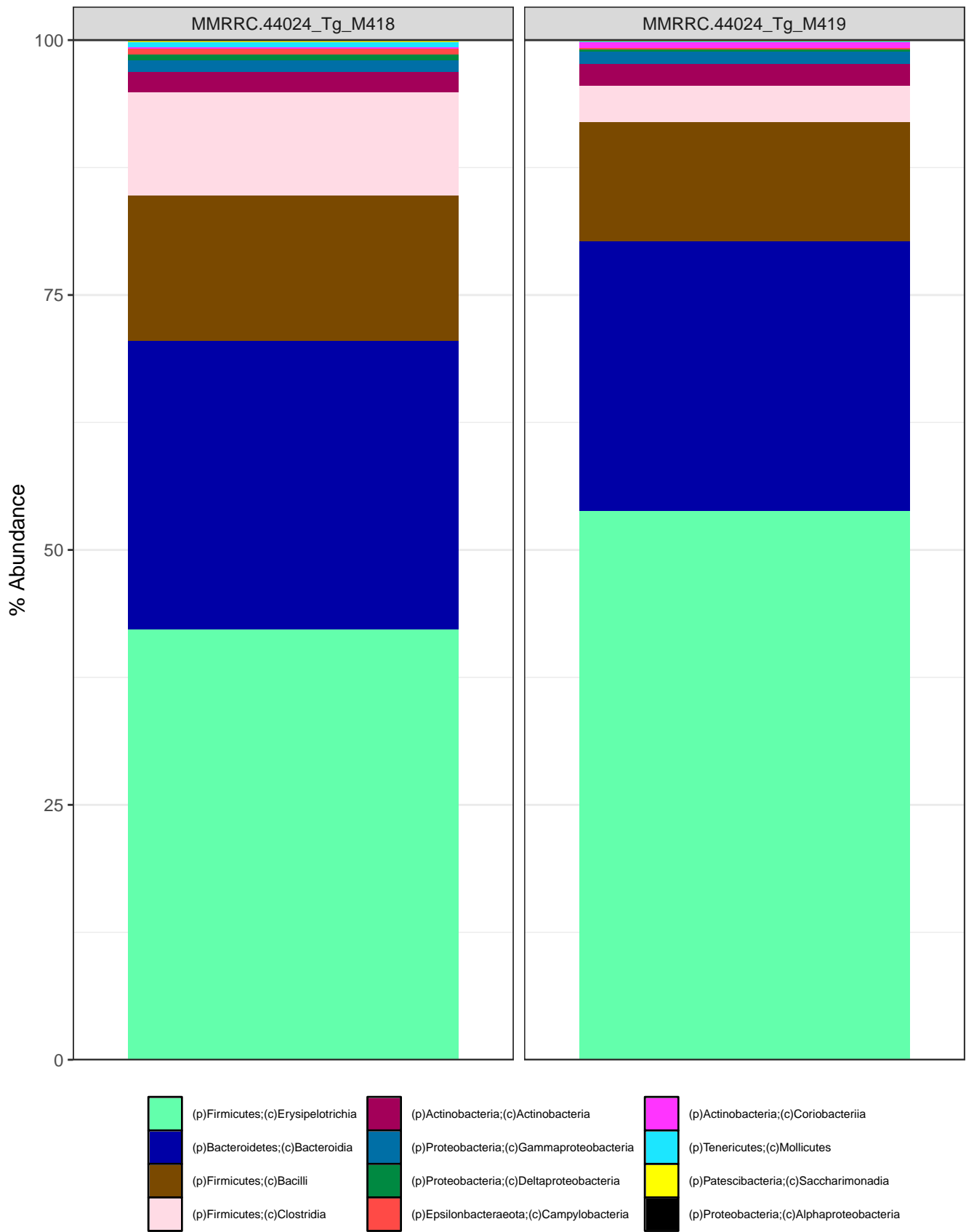


Figure 22: Class level abundances-UNFILTERED

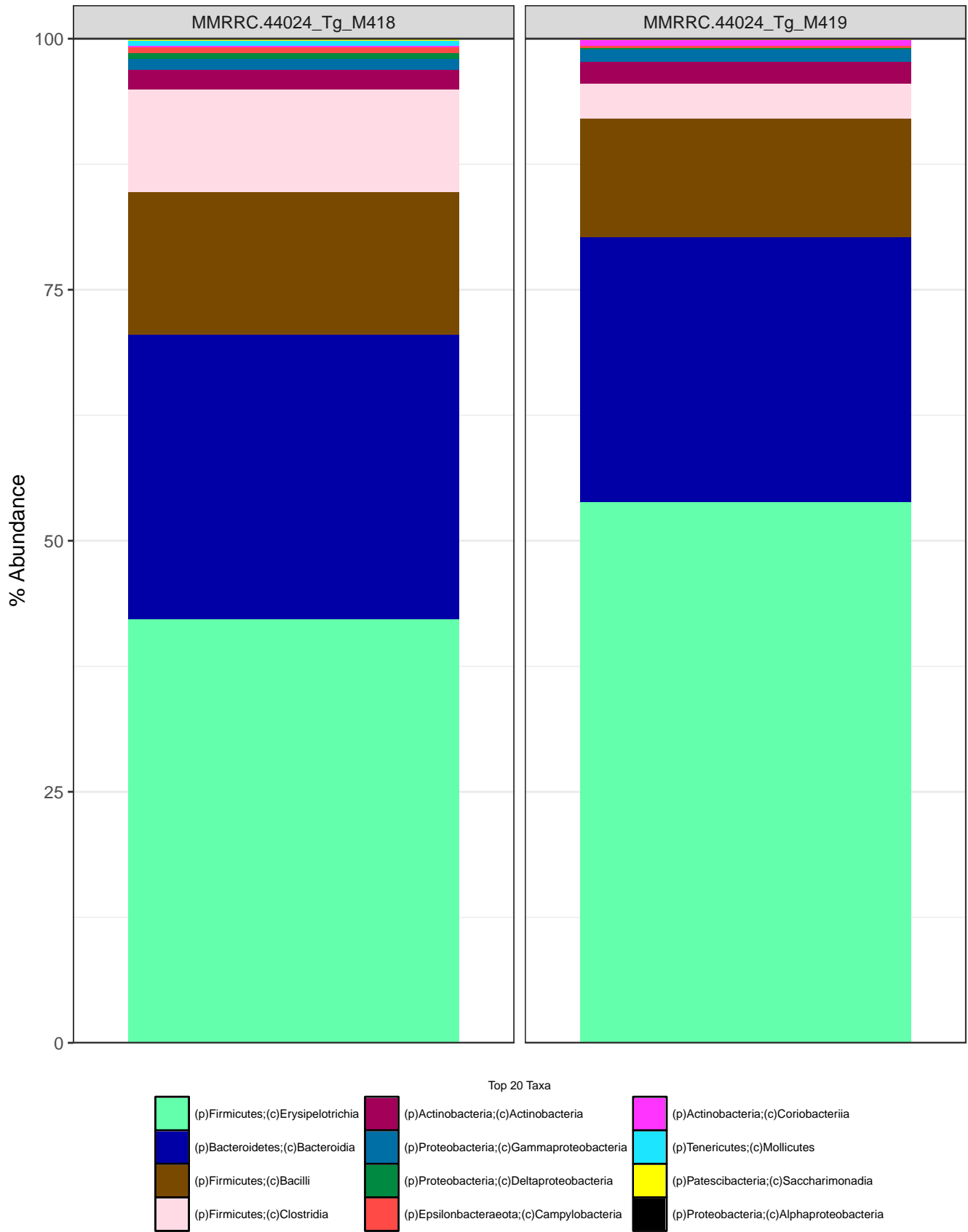


Figure 23: Class level abundances-FILTERED: 0 class level taxa removed

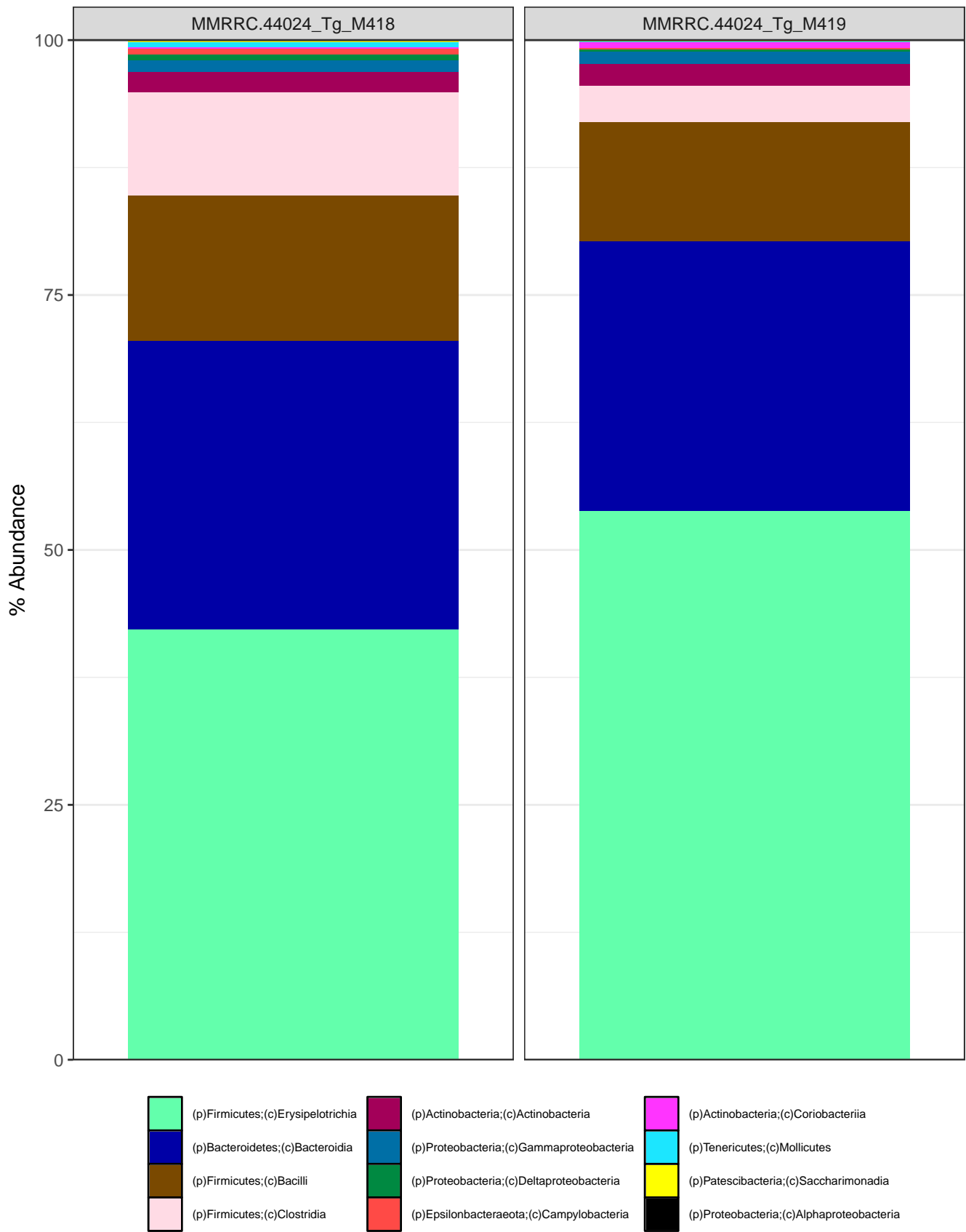


Figure 24: Class level abundances-FILTERED & RESCALED

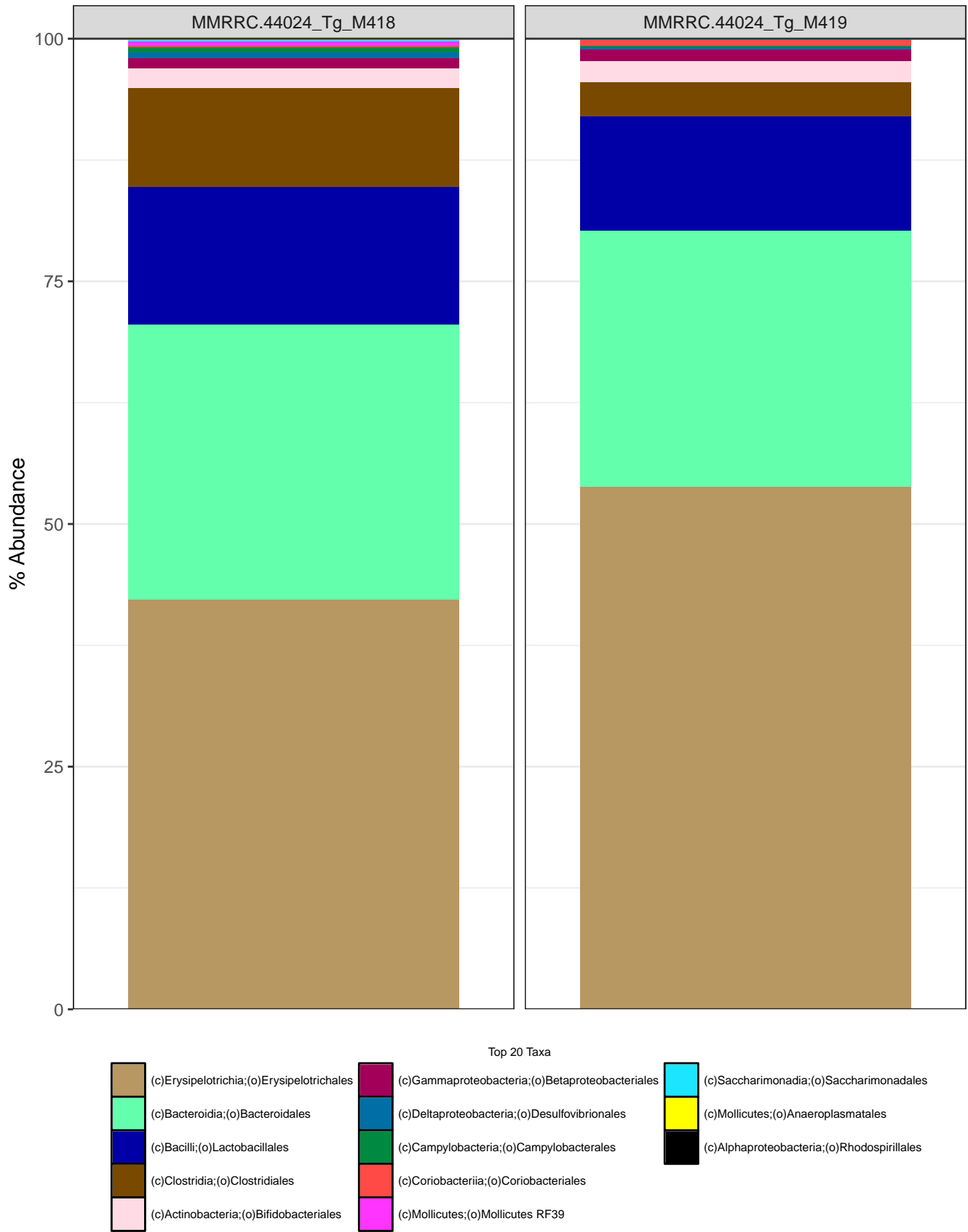


Figure 25: Order level abundances-UNFILTERED

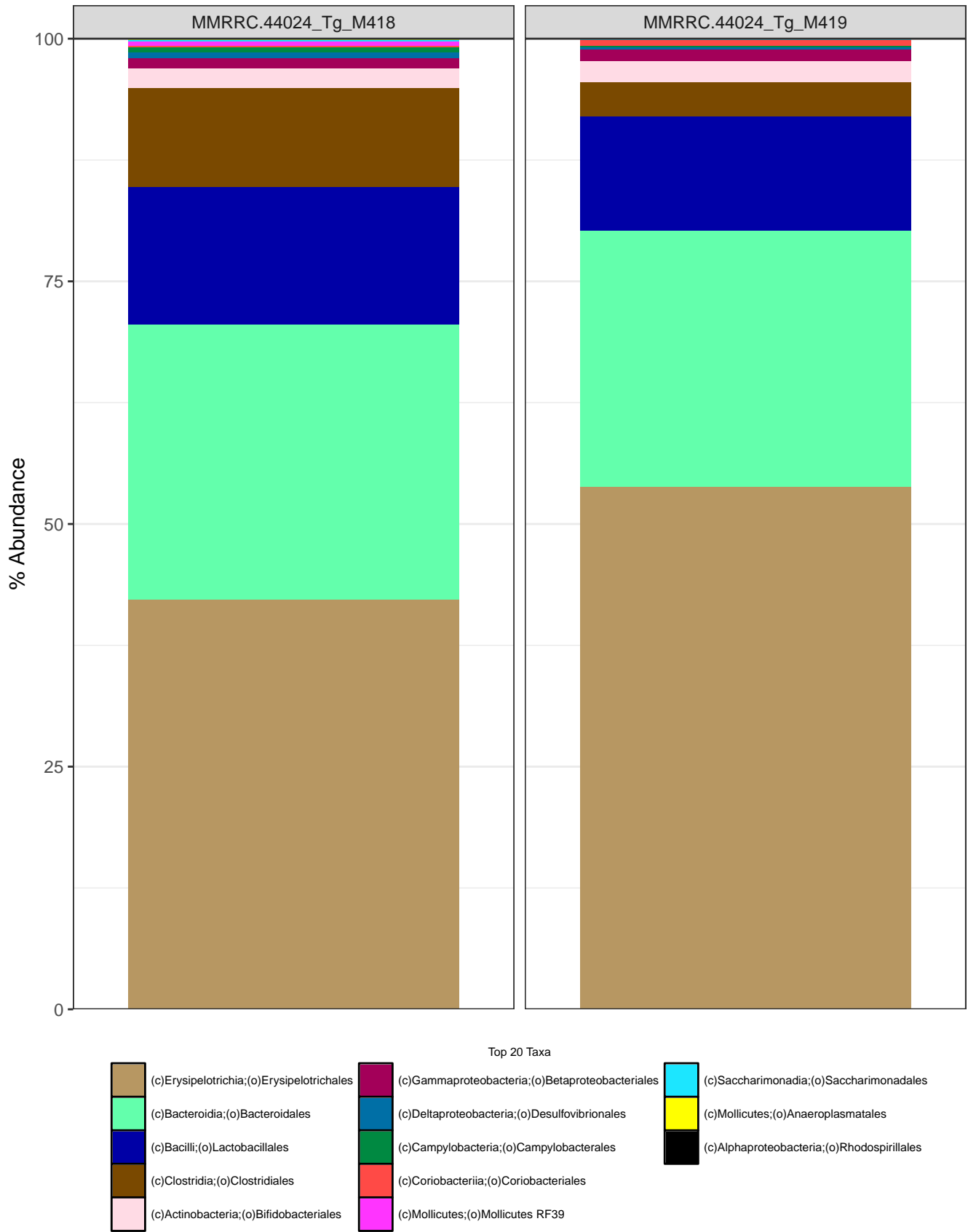


Figure 26: Order level abundances-FILTERED: 0 order level taxa removed

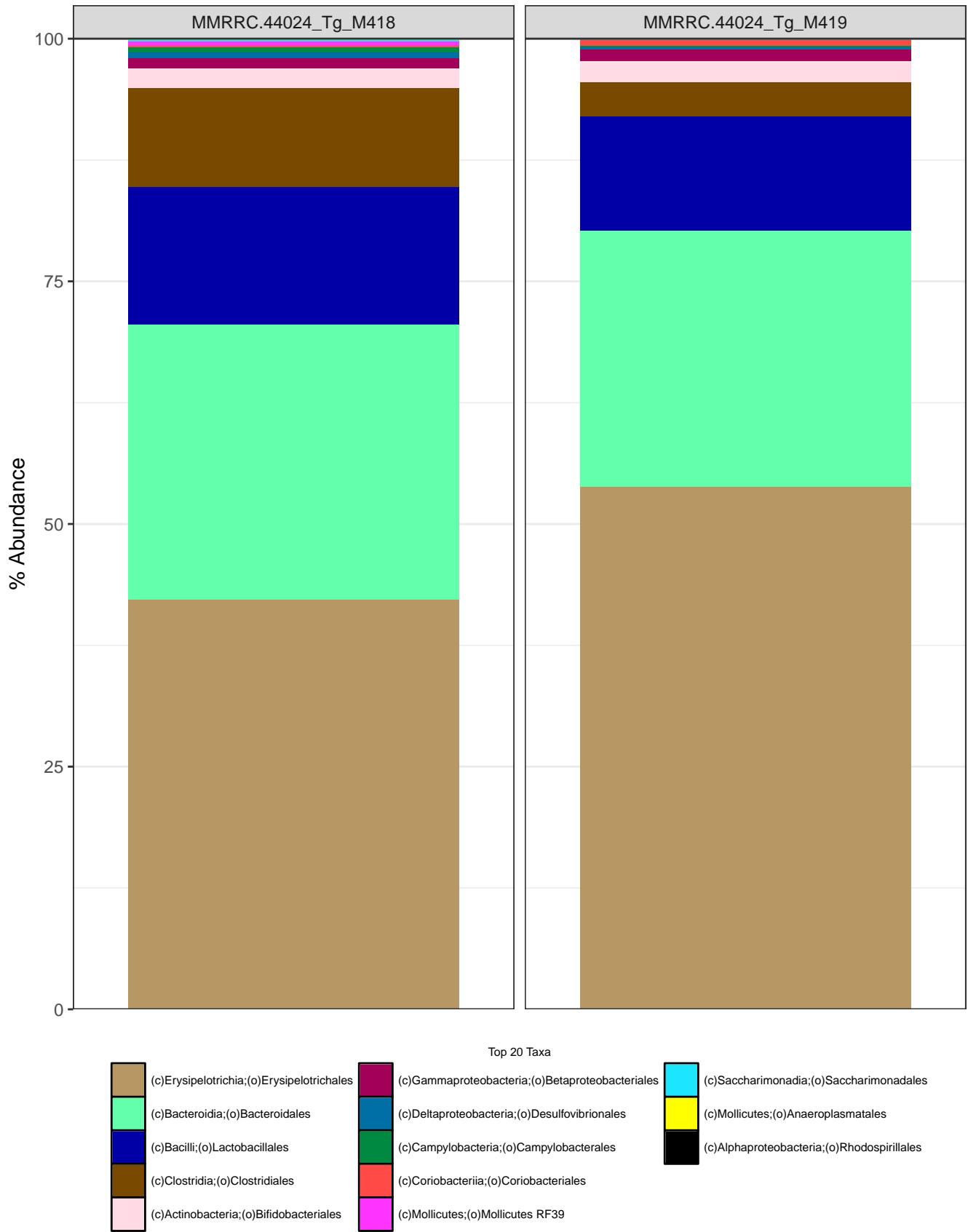


Figure 27: Order level abundances-FILTERED & RESCALED

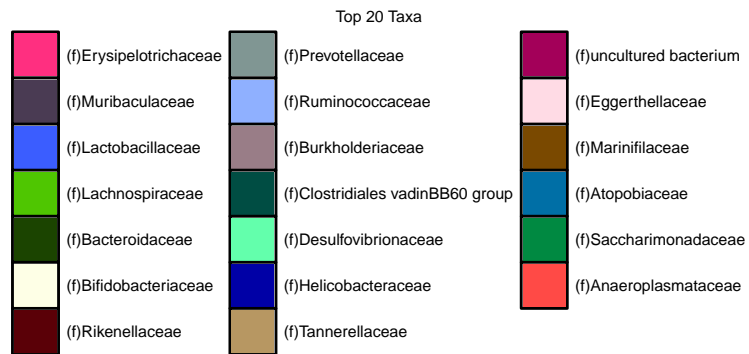
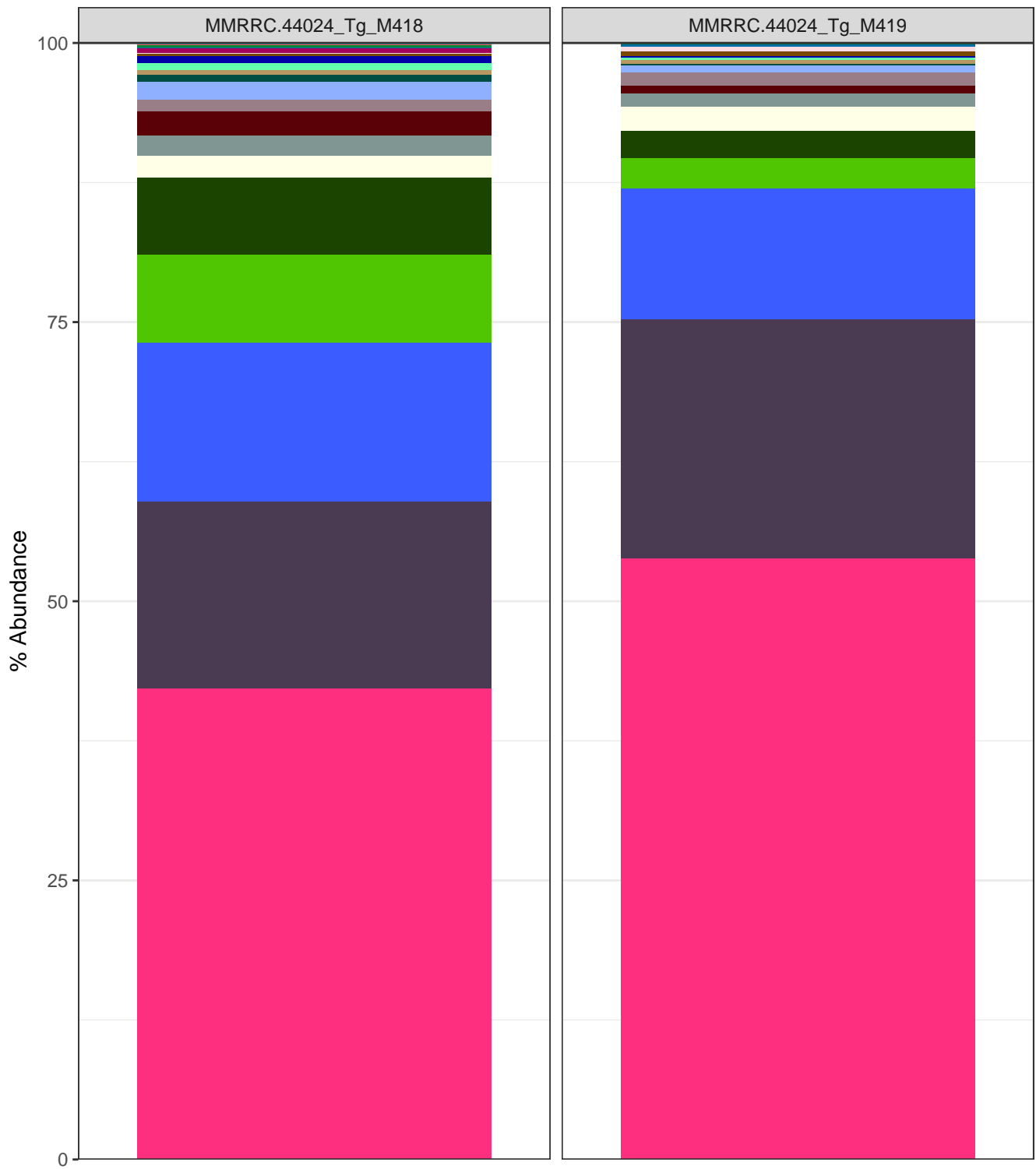


Figure 28: Family level abundances-UNFILTERED

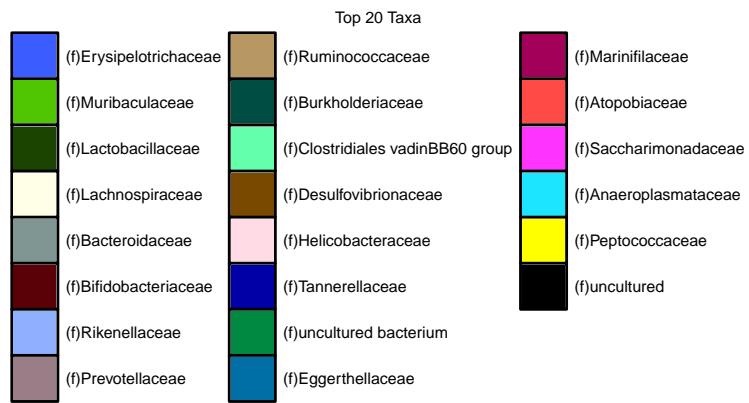
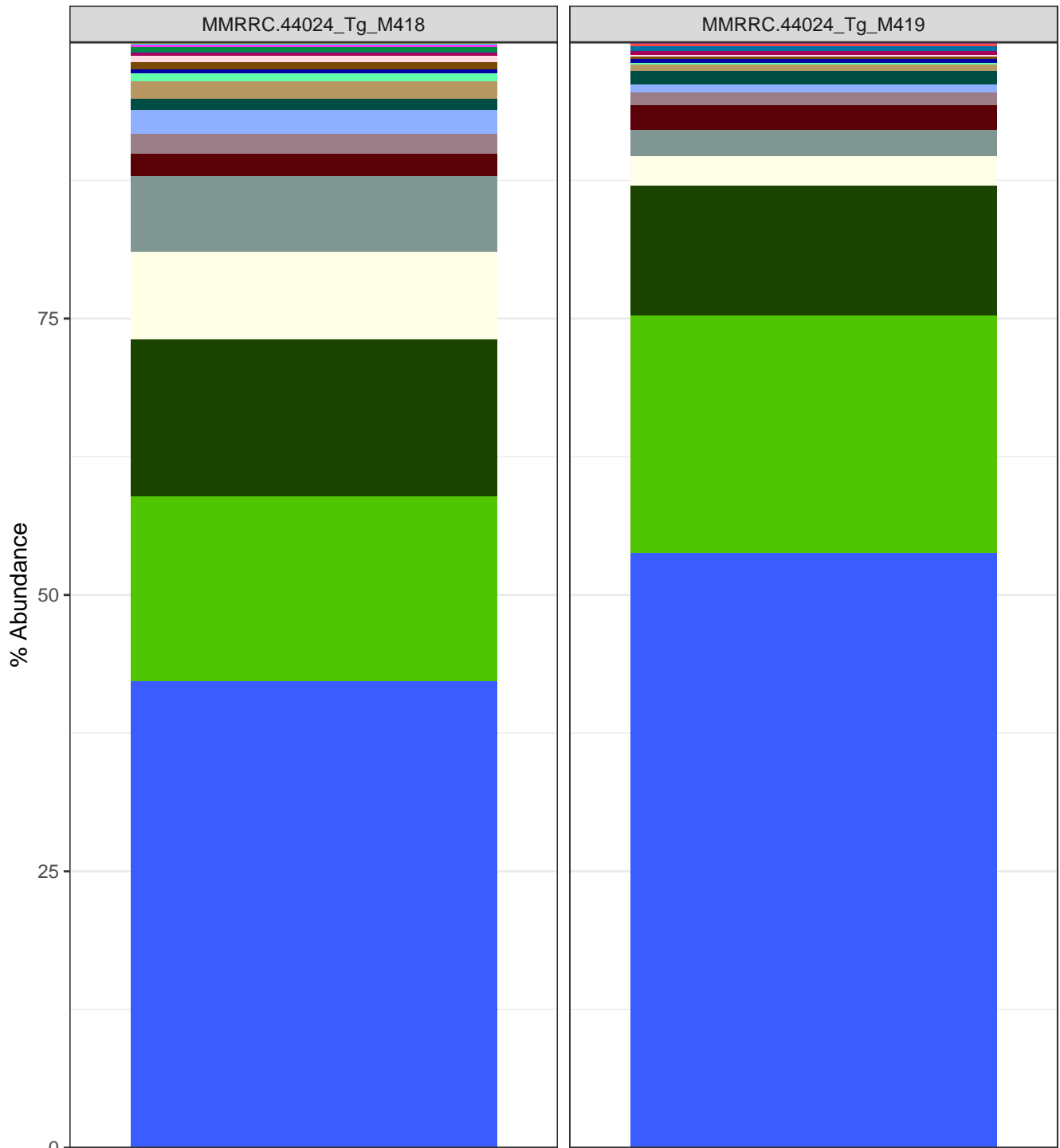


Figure 29: Family level abundances-FILTERED: 2 family level taxa removed

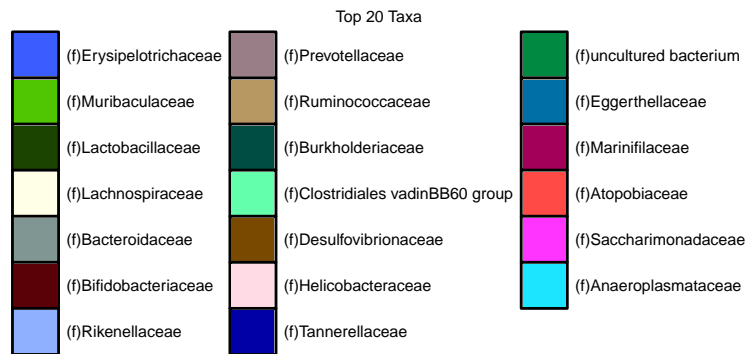
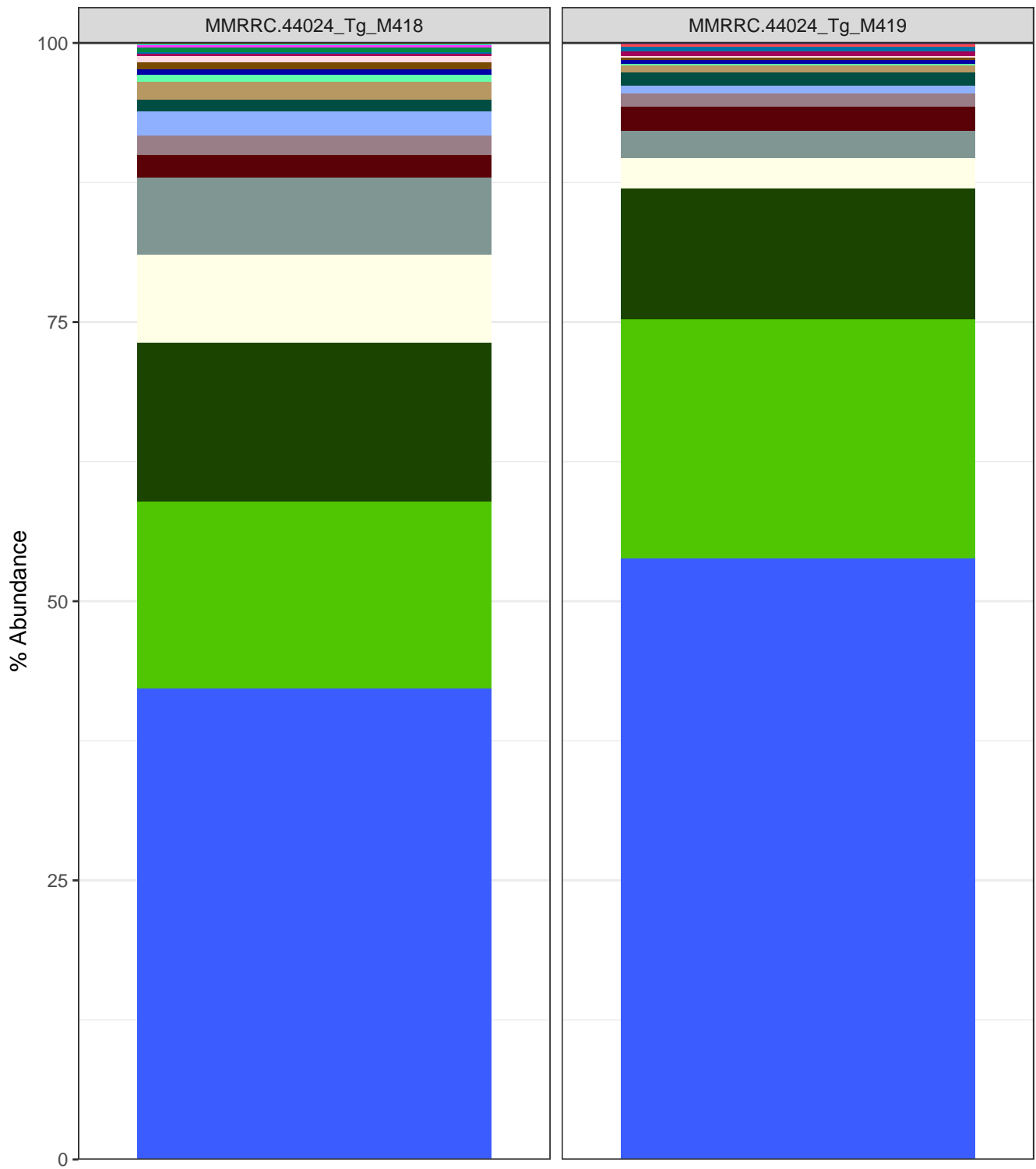


Figure 30: Family level abundances-FILTERED & RESCALED

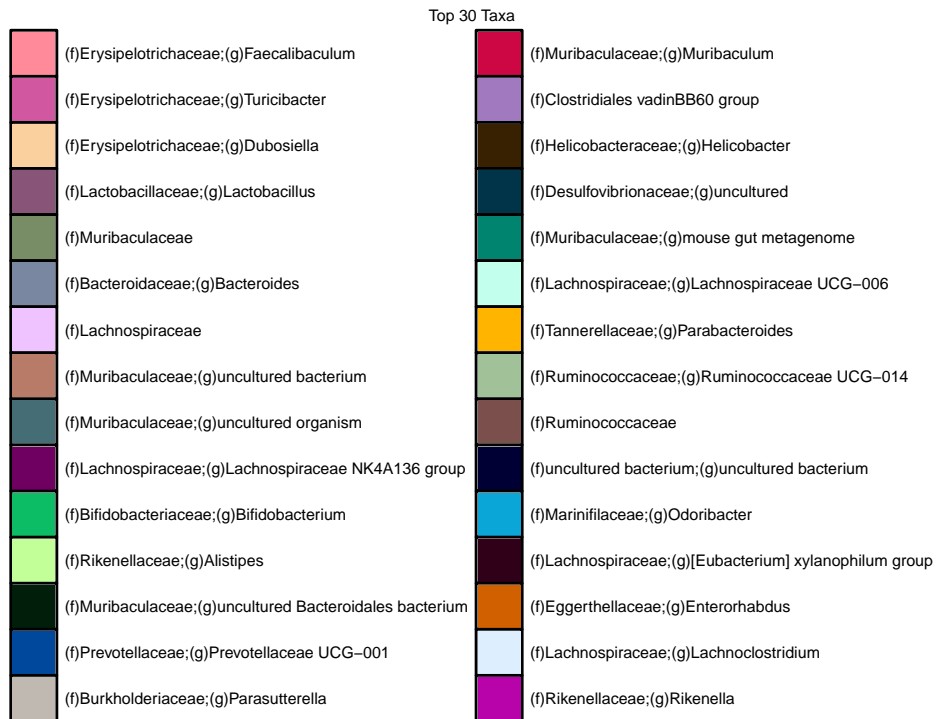
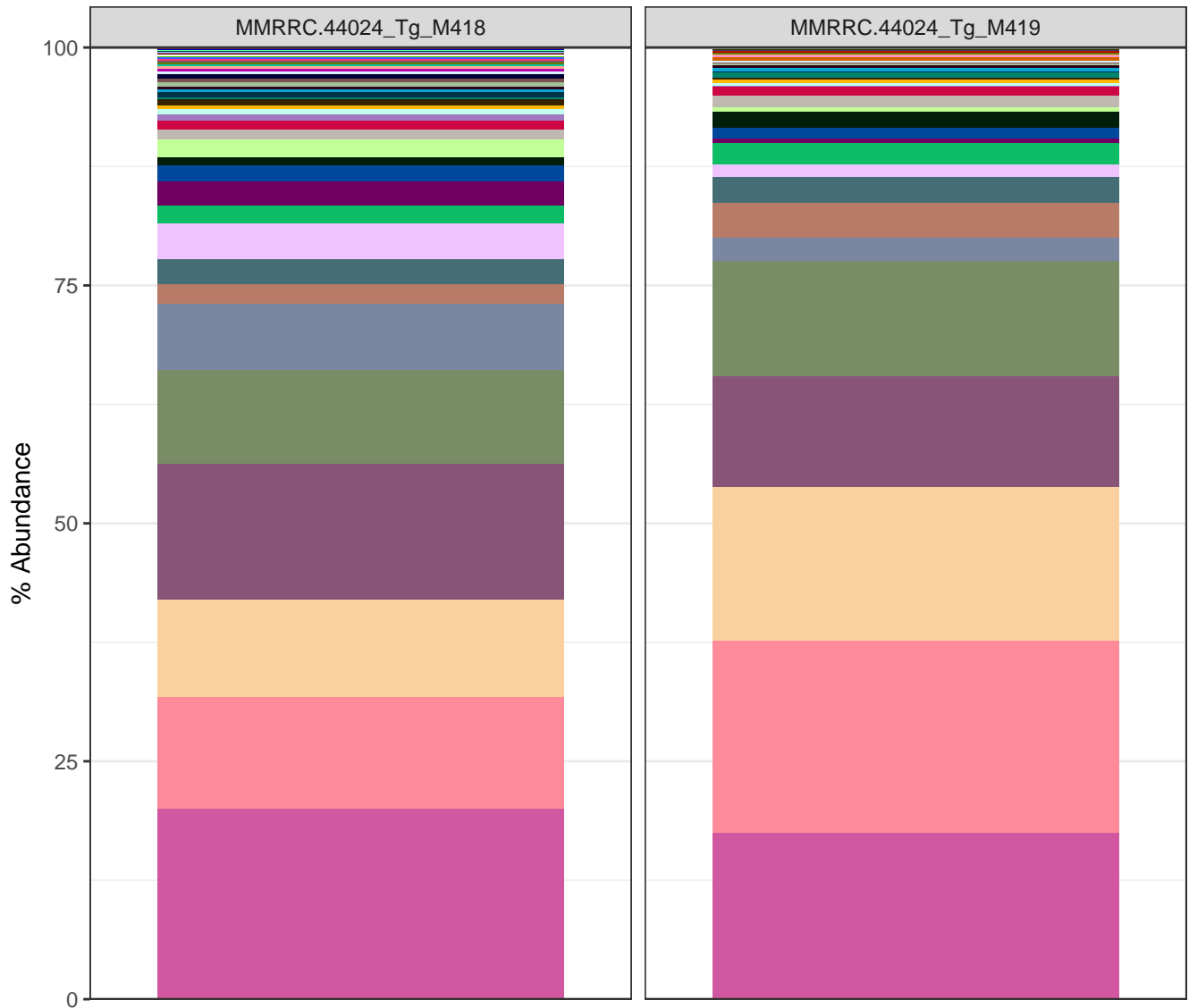


Figure 31: Genus level abundances-UNFILTERED

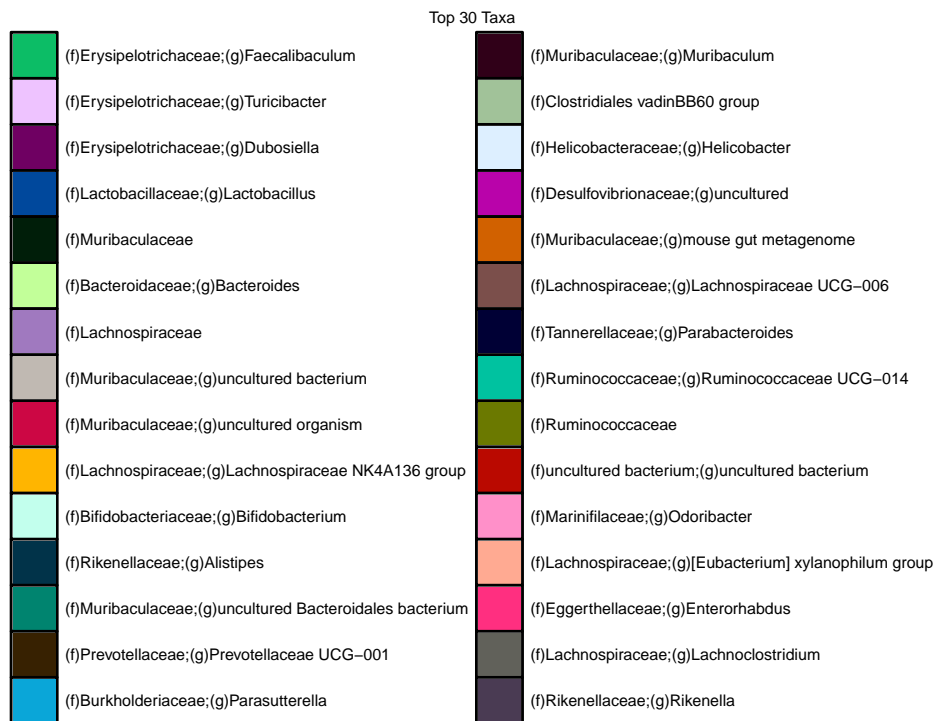
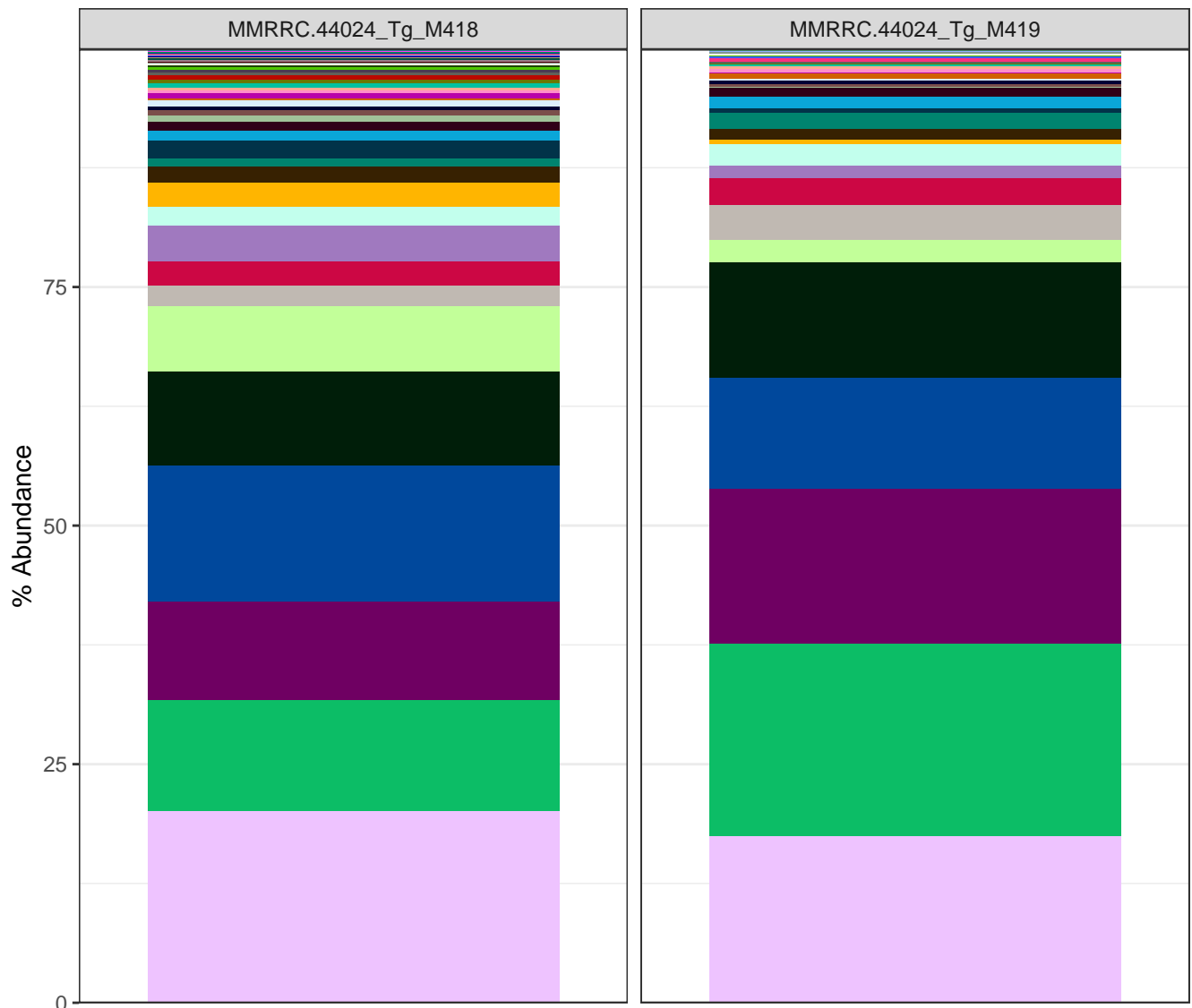


Figure 32: Genus level abundances-FILTERED: 9 genus level taxa removed

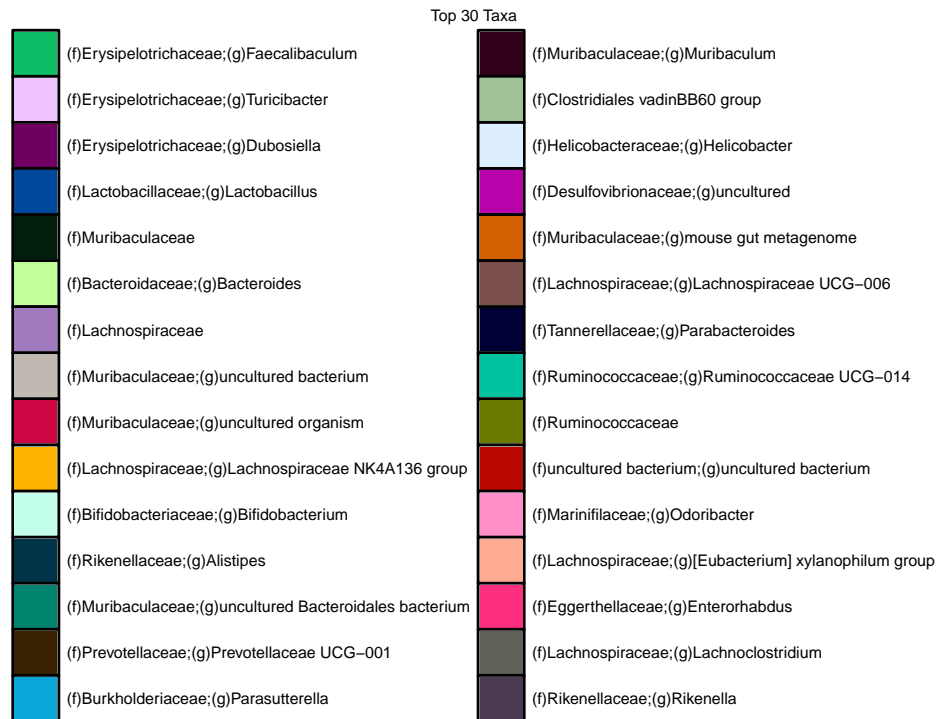
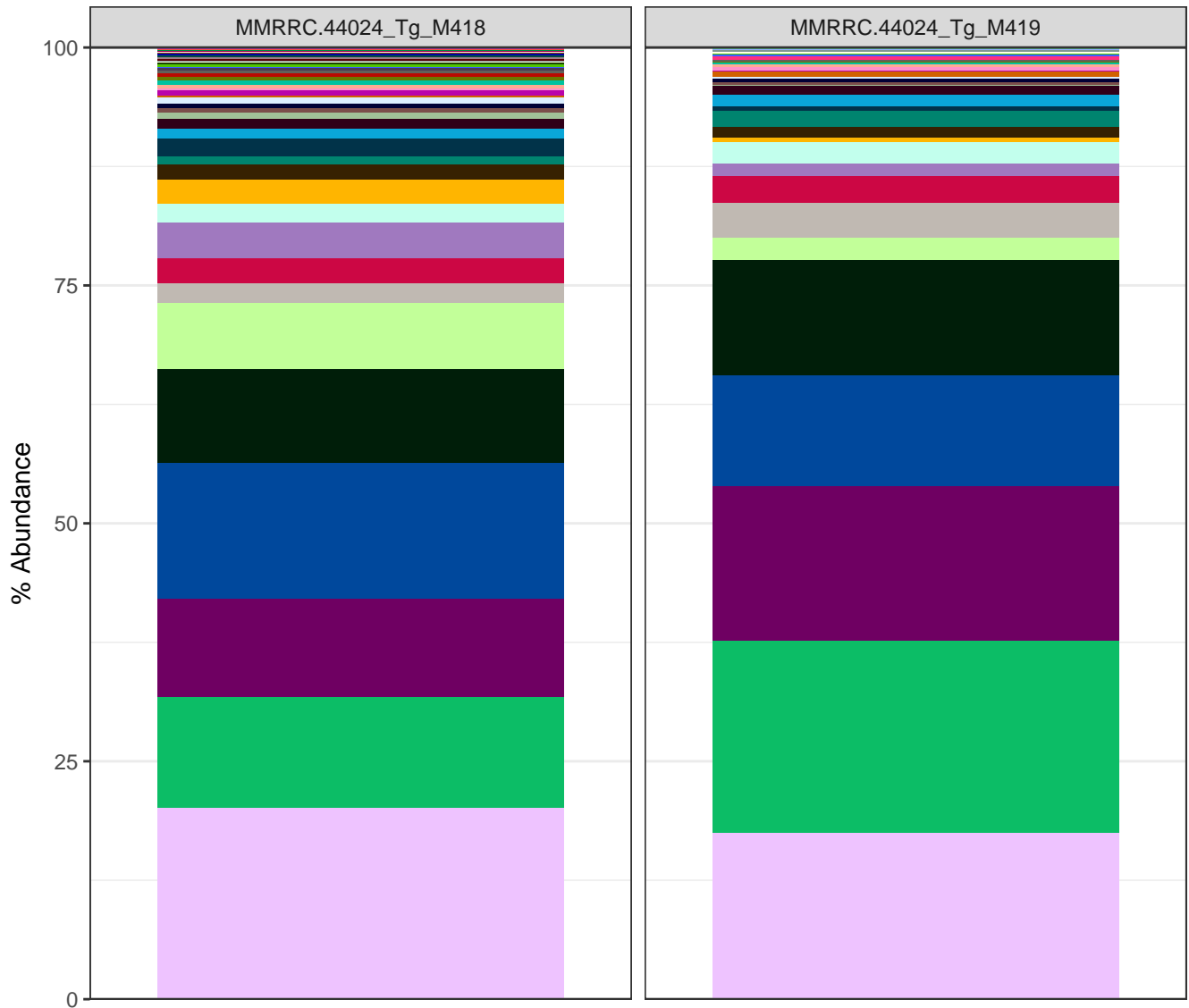


Figure 33: Genus level abundances-FILTERED & RESCALED

A Appendix 1 (Taxa Abundance Tables)

Table 2: Phylum (Filtered) level taxa group mean percent abundances

Taxa	MMRRC.44024_Tg_M418	MMRRC.44024_Tg_M419
(p)Actinobacteria	2.18%	2.75%
(p)Bacteroidetes	28.32%	26.43%
(p)Epsilonbacteraeota	0.57%	0.19%
(p)Firmicutes	66.58%	69.06%
(p)Patescibacteria	0.15%	0.1%
(p)Proteobacteria	1.73%	1.39%
(p)Tenericutes	0.47%	0.07%
NA	100	100

Table 3: Class (Filtered) level taxa group mean percent abundances

Taxa	MMRRC.44024_Tg_M418	MMRRC.44024_Tg_M419
(p)Actinobacteria;(c)Actinobacteria	1.98%	2.2%
(p)Actinobacteria;(c)Coriobacteriia	0.2%	0.56%
(p)Bacteroidetes;(c)Bacteroidia	28.32%	26.43%
(p)Epsilonbacteraeota;(c)Campylobacteria	0.57%	0.19%
(p)Firmicutes;(c)Bacilli	14.21%	11.75%
(p)Firmicutes;(c)Clostridia	10.19%	3.5%
(p)Firmicutes;(c)Erysipelotrichia	42.18%	53.8%
(p)Patescibacteria;(c)Saccharimonadia	0.15%	0.1%
(p)Proteobacteria;(c)Alphaproteobacteria	0.06%	0%
(p)Proteobacteria;(c)Deltaproteobacteria	0.59%	0.21%
(p)Proteobacteria;(c)Gammaproteobacteria	1.08%	1.18%
(p)Tenericutes;(c)Mollicutes	0.47%	0.07%
NA	100	100

Table 4: Order (Filtered) level taxa group mean percent abundances

Taxa	MMRRC.44024_Tg_M418	MMRRC.44024_Tg_M419
(c)Actinobacteria;(o)Bifidobacteriales	1.98%	2.2%
(c)Alphaproteobacteria;(o)Rhodospirillales	0.06%	0%
(c)Bacilli;(o)Lactobacillales	14.21%	11.75%
(c)Bacteroidia;(o)Bacteroidales	28.32%	26.43%
(c)Campylobacteria;(o)Campylobacterales	0.57%	0.19%
(c)Clostridia;(o)Clostridiales	10.19%	3.5%
(c)Coriobacteriia;(o)Coriobacteriales	0.2%	0.56%
(c)Deltaproteobacteria;(o)Desulfovibrionales	0.59%	0.21%
(c)Erysipelotrichia;(o)Erysipelotrichales	42.18%	53.8%
(c)Gammaproteobacteria;(o)Betaproteobacteriales	1.08%	1.18%
(c)Mollicutes;(o)Anaeroplasmatales	0.09%	0%
(c)Mollicutes;(o)Mollicutes RF39	0.38%	0.07%
(c)Saccharimonadia;(o)Saccharimonadales	0.15%	0.1%
NA	100	100

Table 5: Family (Filtered) level taxa group mean percent abundances

Taxa	MMRRC.44024_Tg_M418	MMRRC.44024_Tg_M419
(f)Anaeroplasmataceae	0.09%	0%
(f)Atopobiaceae	0.09%	0.18%
(f)Bacteroidaceae	6.89%	2.44%
(f)Bifidobacteriaceae	1.98%	2.2%
(f)Burkholderiaceae	1.08%	1.18%
(f)Clostridiales vadinBB60 group	0.65%	0.18%
(f)Desulfovibrionaceae	0.59%	0.21%
(f)Eggerthellaceae	0.11%	0.37%
(f)Erysipelotrichaceae	42.18%	53.8%
(f)Helicobacteraceae	0.57%	0.19%
(f)Lachnospiraceae	7.91%	2.68%
(f)Lactobacillaceae	14.21%	11.75%
(f)Marinifilaceae	0.26%	0.37%
(f)Muribaculaceae	16.72%	21.41%
(f)Peptococcaceae	0.06%	0.02%
(f)Prevotellaceae	1.84%	1.15%
(f)Rikenellaceae	2.13%	0.74%
(f)Ruminococcaceae	1.54%	0.61%
(f)Saccharimonadaceae	0.15%	0.1%
(f)Tannerellaceae	0.49%	0.32%
(f)uncultured	0.06%	0%
(f)uncultured bacterium	0.38%	0.07%
NA	99.971	99.983

Table 6: Genus (Filtered) level taxa group mean percent abundances

Taxa	MMRRC.44024_Tg_M418	MMRRC.44024_Tg_M419
(f) Anaeroplasmataceae;(g) Anaeroplasma	0.09%	0%
(f) Atopobiaceae	0.09%	0.18%
(f) Bacteroidaceae;(g) Bacteroides	6.89%	2.44%
(f) Bifidobacteriaceae;(g) Bifidobacterium	1.98%	2.2%
(f) Burkholderiaceae;(g) Parasutterella	1.08%	1.18%
(f) Clostridiales vadinBB60 group	0.65%	0.18%
(f) Desulfovibrionaceae;(g) Desulfovibrio	0.05%	0.06%
(f) Desulfovibrionaceae;(g) uncultured	0.54%	0.15%
(f) Eggerthellaceae	0%	0.06%
(f) Eggerthellaceae;(g) Enterorhabdus	0.08%	0.31%
(f) Erysipelotrichaceae	0.06%	0%
(f) Erysipelotrichaceae;(g) Dubosiella	10.31%	16.15%
(f) Erysipelotrichaceae;(g) Faecalibaculum	11.71%	20.14%
(f) Erysipelotrichaceae;(g) Turicibacter	20.01%	17.48%
(f) Erysipelotrichaceae;(g) uncultured	0.08%	0.03%
(f) Helicobacteraceae;(g) Helicobacter	0.57%	0.19%
(f) Lachnospiraceae	3.7%	1.38%
(f) Lachnospiraceae;(g) [Eubacterium] xylanophilum group	0.35%	0.25%
(f) Lachnospiraceae;(g) ASF356	0.13%	0.08%
(f) Lachnospiraceae;(g) GCA-900066575	0.11%	0%
(f) Lachnospiraceae;(g) Lachnoclostridium	0.29%	0.13%
(f) Lachnospiraceae;(g) Lachnospiraceae NK4A136 group	2.52%	0.52%
(f) Lachnospiraceae;(g) Lachnospiraceae UCG-006	0.5%	0.31%
(f) Lachnospiraceae;(g) Marvinbryantia	0.09%	0%
(f) Lachnospiraceae;(g) uncultured	0.17%	0.01%
(f) Lactobacillaceae;(g) Lactobacillus	14.21%	11.75%
(f) Marinifilaceae;(g) Odoribacter	0.26%	0.37%
(f) Muribaculaceae	9.87%	12.04%
(f) Muribaculaceae;(g) mouse gut metagenome	0.24%	0.51%
(f) Muribaculaceae;(g) Muribaculum	0.98%	0.86%
(f) Muribaculaceae;(g) uncultured bacterium	2.17%	3.58%
(f) Muribaculaceae;(g) uncultured Bacteroidales bacterium	0.88%	1.63%
(f) Muribaculaceae;(g) uncultured organism	2.58%	2.78%
(f) Peptococcaceae;(g) uncultured	0.06%	0.02%
(f) Prevotellaceae;(g) Paraprevotella	0.22%	0.05%
(f) Prevotellaceae;(g) Prevotellaceae UCG-001	1.62%	1.1%
(f) Rikenellaceae;(g) Alistipes	1.82%	0.57%
(f) Rikenellaceae;(g) Millionella	0.05%	0.07%
(f) Rikenellaceae;(g) Rikenella	0.25%	0.1%
(f) Ruminococcaceae	0.39%	0.14%
(f) Ruminococcaceae;(g) Anaerotruncus	0.08%	0%
(f) Ruminococcaceae;(g) Angelakisella	0.06%	0.02%
(f) Ruminococcaceae;(g) Butyricoccus	0.1%	0%
(f) Ruminococcaceae;(g) Oscillibacter	0.19%	0.08%
(f) Ruminococcaceae;(g) Ruminiclostridium 5	0.06%	0%
(f) Ruminococcaceae;(g) Ruminiclostridium 9	0.09%	0%
(f) Ruminococcaceae;(g) Ruminococcaceae UCG-013	0.12%	0.12%
(f) Ruminococcaceae;(g) Ruminococcaceae UCG-014	0.4%	0.17%
(f) Saccharimonadaceae;(g) Candidatus Saccharimonas	0.15%	0.1%
(f) Tannerellaceae;(g) Parabacteroides	0.49%	0.32%

Table 6: Genus (Filtered) level taxa group mean percent abundances (*continued*)

Taxa	MMRRC.44024_Tg_M418	MMRRC.44024_Tg_M419
(f)uncultured bacterium;(g)uncultured bacterium	0.38%	0.07%
(f)uncultured;(g)Azospirillum sp 47_25	0.06%	0%
NA	99.854	99.903

B Appendix 2 (Taxa removed from filtered datasets at each level)

Table 7: Phylum level taxa removed via filtering- group mean abundances

X.Total...Removed.	X.0.	X.0..1
Total % Removed	0	0

Table 8: Class level taxa removed via filtering- group mean abundances

X.Total...Removed.	X.0.	X.0..1
Total % Removed	0	0

Table 9: Order level taxa removed via filtering- group mean abundances

X.Total...Removed.	X.0.	X.0..1
Total % Removed	0	0

Table 10: Family level taxa removed via filtering- group mean abundances

Taxa	MMRRC.44024_Tg_M418	MMRRC.44024_Tg_M419
(f)Christensenellaceae	0.0000	0.0172
(f)Family XIII	0.0293	0.0000
Total % Removed	0.0293	0.0172

Table 11: Genus level taxa removed via filtering- group mean abundances

Taxa	MMRRC.44024_Tg_M418	MMRRC.44024_Tg_M419
(f)Christensenellaceae;(g)uncultured	0.0000	0.0172
(f)Eggerthellaceae;(g)Gordonibacter	0.0293	0.0000
(f)Family XIII;(g)[Eubacterium] nodatum group	0.0293	0.0000
(f)Lachnospiraceae;(g)Roseburia	0.0469	0.0000
(f)Ruminococcaceae;(g)Candidatus Soleaferrea	0.0000	0.0229
(f)Ruminococcaceae;(g)Ruminiclostridium	0.0410	0.0000
(f)Ruminococcaceae;(g)Ruminococcaceae NK4A214 group	0.0000	0.0343
(f)Ruminococcaceae;(g)uncultured	0.0000	0.0229
Total % Removed	0.1464	0.0973

Important: Please remember the policy to acknowledge the UC Davis MMPC in your publications or presentations of the work that was generated using UC Davis MMPC services. Please include a statement in the acknowledgement section: Research was supported by NIH grant U24-DK092993 (MMPC-University of California Davis Microbiome and Host Response Core, RRID:SCR_015361).

References:

- Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. Johnson, and S. P. Holmes. 2016. “DADA2: High-Resolution Sample Inference from Illumina Amplicon Data.” Journal Article. *Nat Methods* 13 (7): 581–3. <https://doi.org/10.1038/nmeth.3869>.
- Katoh, K., and D. M. Standley. 2013. “MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability.” Journal Article. *Mol Biol Evol* 30 (4): 772–80. <https://doi.org/10.1093/molbev/mst010>.
- Lozupone, C., M. E. Lladser, D. Knights, J. Stombaugh, and R. Knight. 2011. “UniFrac: An Effective Distance Metric for Microbial Community Comparison.” Journal Article. *ISME J* 5 (2): 169–72. <https://doi.org/10.1038/ismej.2010.133>.
- Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. “FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments.” Journal Article. *PLoS One* 5 (3): e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- QIIME 2 Development Team. 2017. “QIIME 2 Docs.” Web Page. <https://docs.qiime2.org/2017.9/tutorials/moving-pictures/>.
- Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner. 2013. “The Silva Ribosomal Rna Gene Database Project: Improved Data Processing and Web-Based Tools.” Journal Article. *Nucleic Acids Res* 41: D590–6. <https://doi.org/10.1093/nar/gks1219>.