HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH
Microbiome Analysis Core

James Luo Sham Surgery FMT Microbiome 16S

Thomas Kuntz, PhD

June 2021

Contents

1 Data QC **3**

1.1 Raw data processing summary . . . 3

1.2 Metadata group bias . . . 5

1.3 Summary of categorical metadata . . . 7

2 Alpha diversity analysis **8**

2.1 Univariable ANOVA . . . 8

3 Taxonomy visualization **12**

3.1 Taxonomy visualized as stacked barplots . . . 12

3.1.1 Bar plots grouped by metadata . . . 14

3.2 Taxonomy visualized as heatmaps . . . 18

3.2.1 Heatmaps clustered based on similar compositions . . . 18

3.2.2 Heatmaps sorted based on metadata . . . 20

4 Principle coordinates analysis **22**

4.1 PCoA colored by metadata category . . . 22

4.1.1 Bray-Curtis results . . . 22

4.1.2 Weighted UniFrac results . . . 26

4.1.3 Unweighted UniFrac results . . . 30

5 Beta diversity analysis **34**

5.1 Univariable PERMANOVA 34

5.1.1 Bray-Curtis results 34

5.1.2 Weighted UniFrac results 36

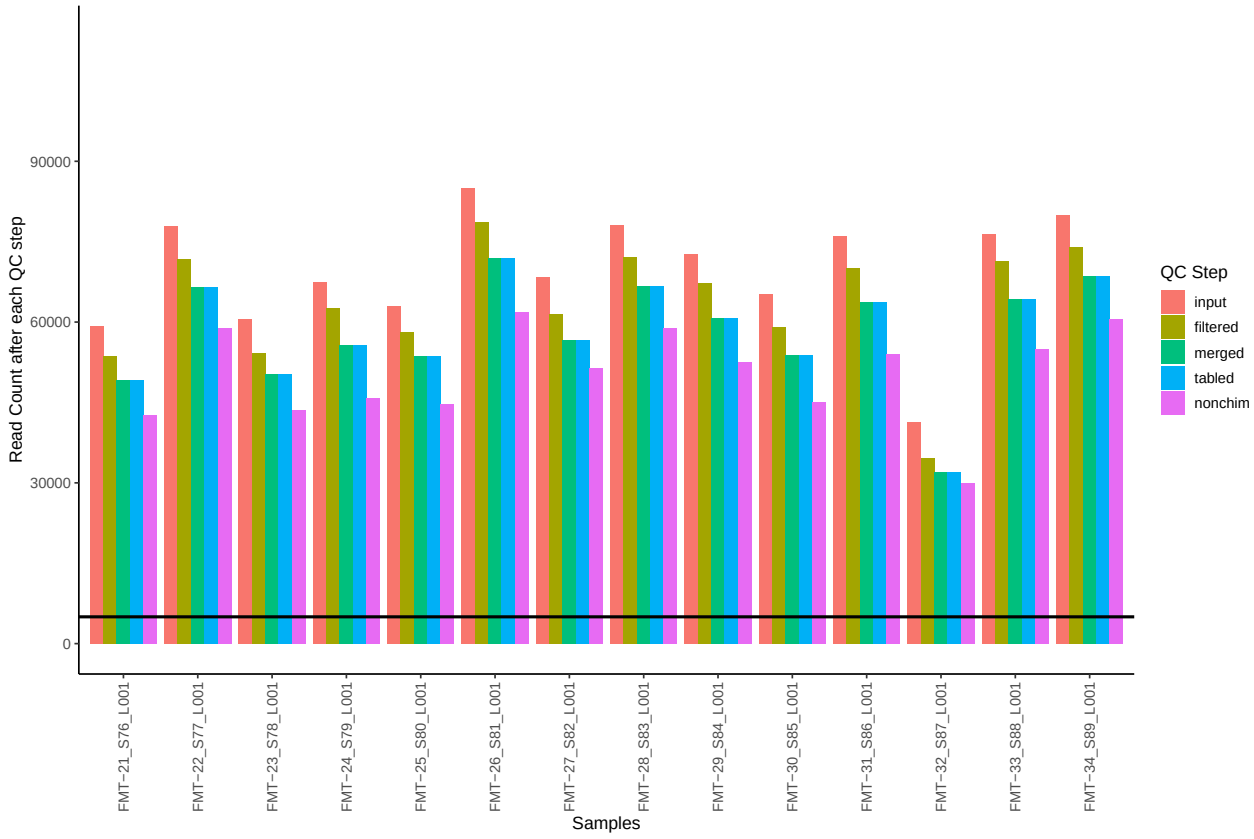
5.1.3 Unweighted UniFrac results 38

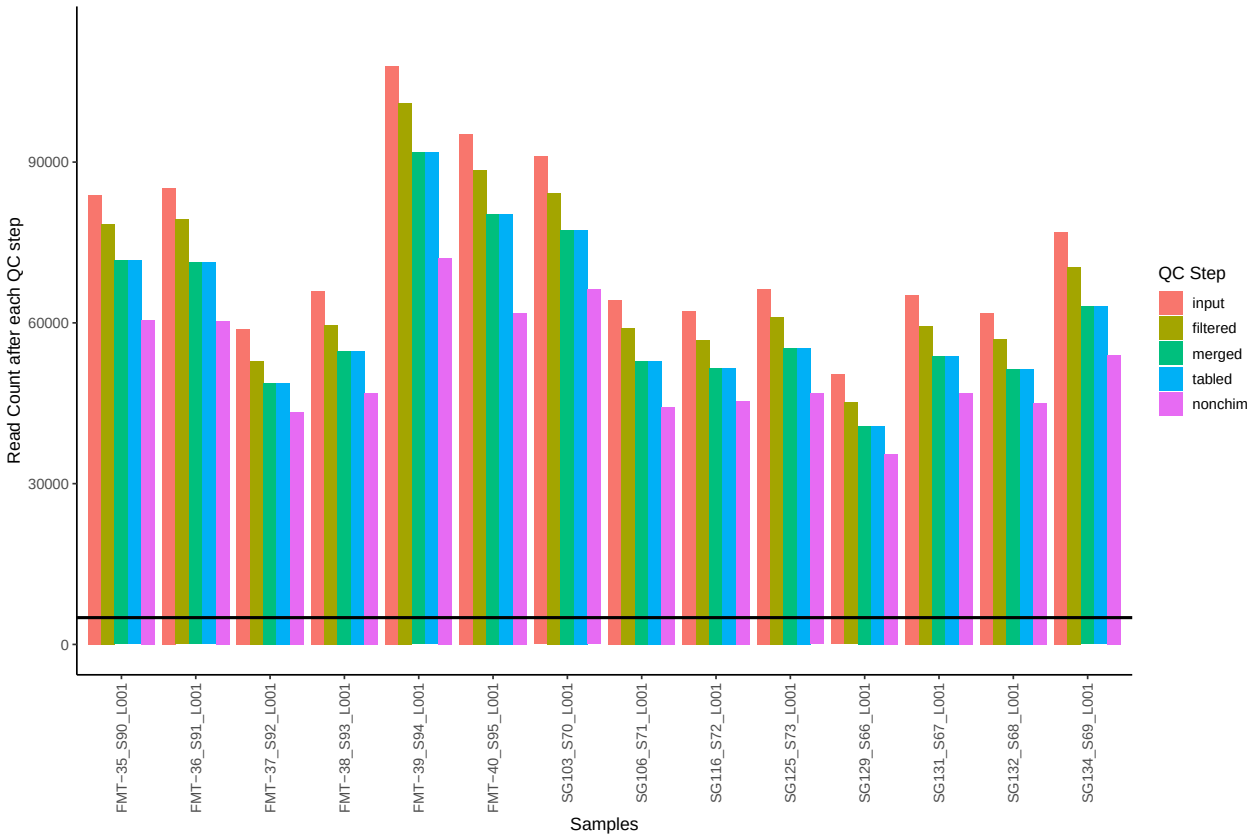
6 Summary of R session **40**

1 Data QC

1.1 Raw data processing summary

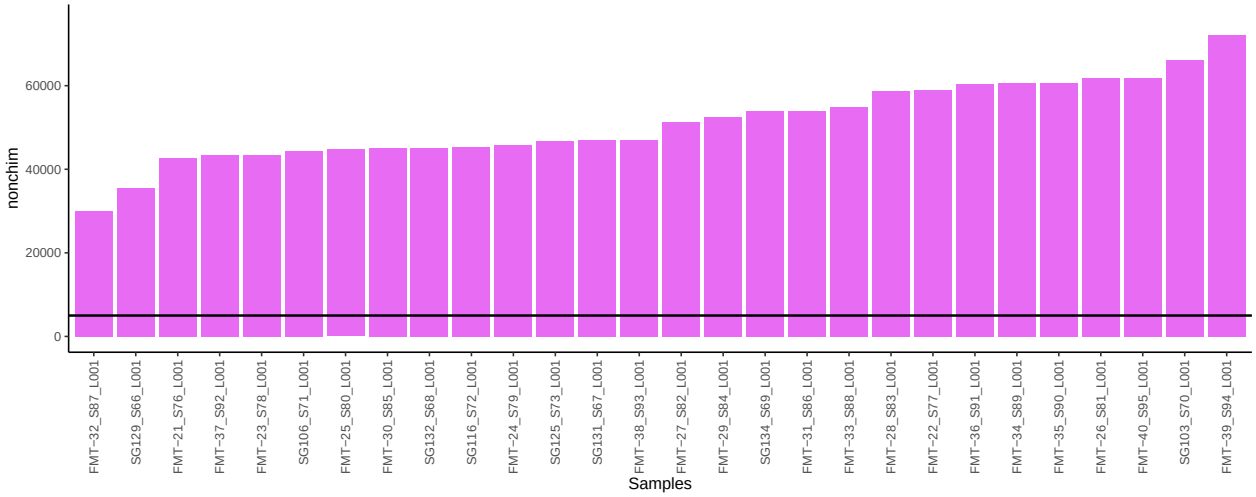
Bar plots of reads per sample at each step of the QC process. Raw reads were processed through the DADA2 pipeline. Missing metadata: POOLED-SG_S75_L001 POOLED-SHAM_S74_L001 removed.



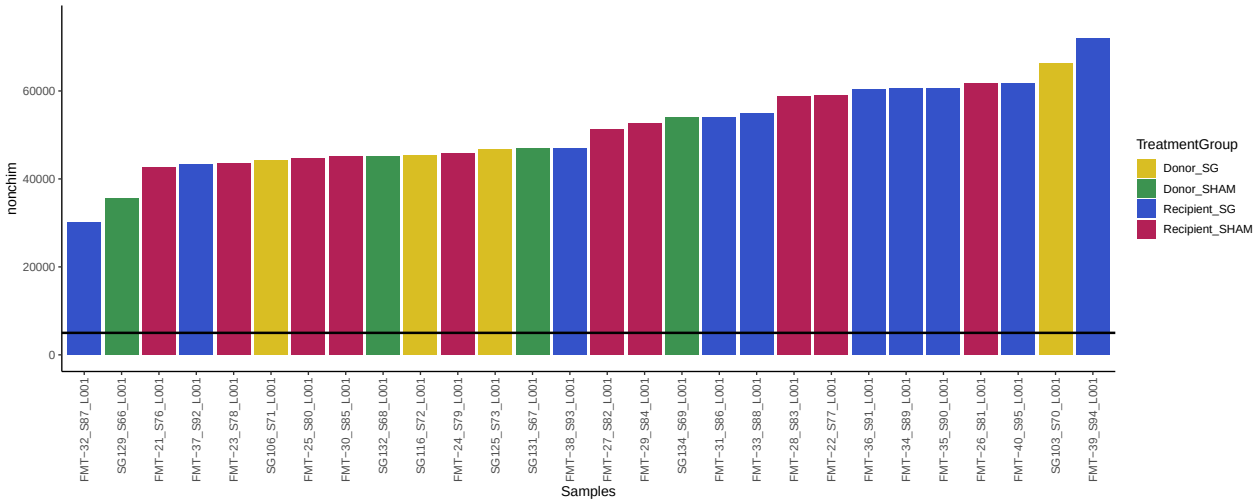

Read Count after each QC step
90000
60000
30000
0
FMT−35_S90_L001
FMT−36_S91_L001
FMT−37_S92_L001
FMT−38_S93_L001
FMT−39_S94_L001
FMT−40_S95_L001
SG103_S70_L001
SG106_S71_L001
SG116_S72_L001
SG125_S73_L001
SG129_S66_L001
SG131_S67_L001
SG132_S68_L001
SG134_S69_L001
Samples
QC Step
input
filtered
merged
tabled
nonchim

1.2 Metadata group bias

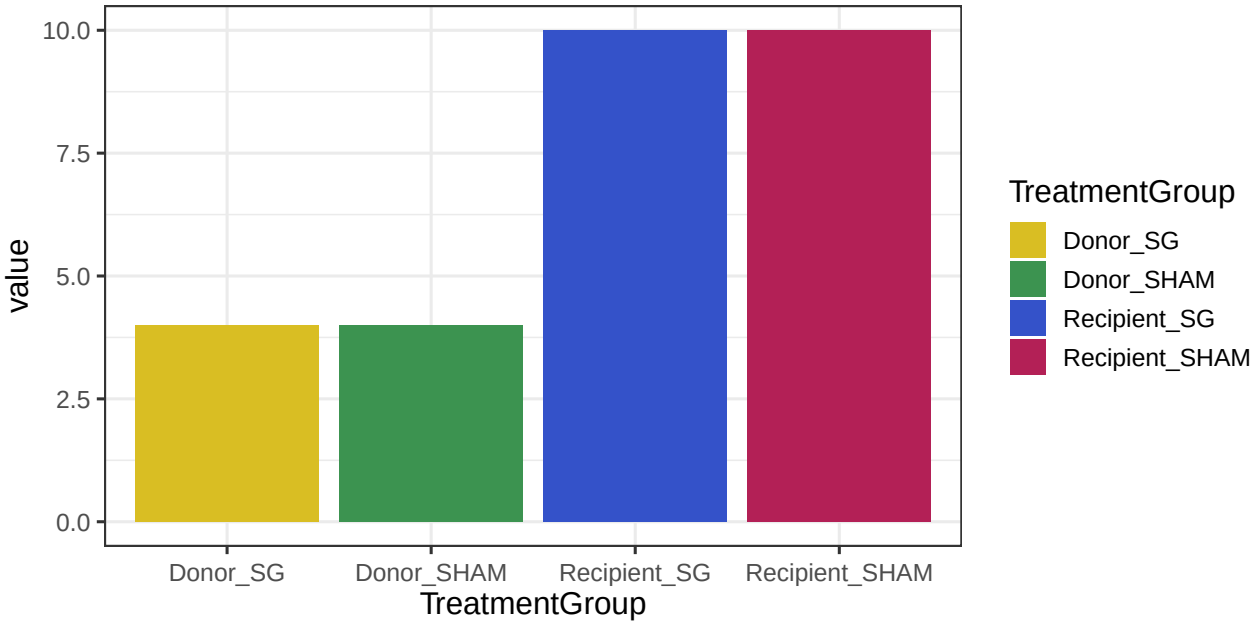
Bar plots of final reads.



Bar plots of final reads colored by metadata grouping.



1.3 Summary of categorical metadata



2 Alpha diversity analysis

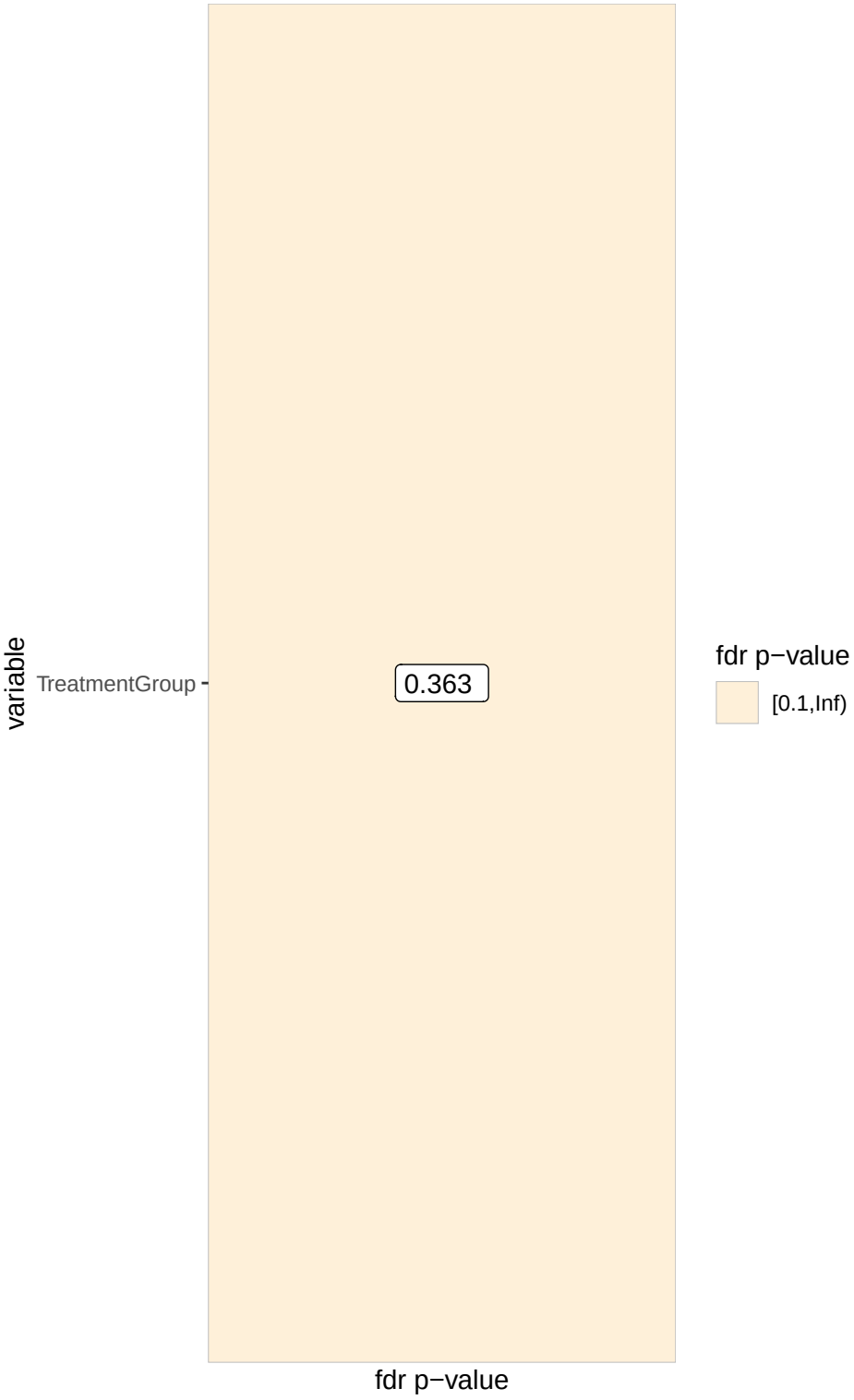
Alpha diversity metrics are performed on the normalized, unfiltered, species level taxonomic table.

2.1 Univariable ANOVA

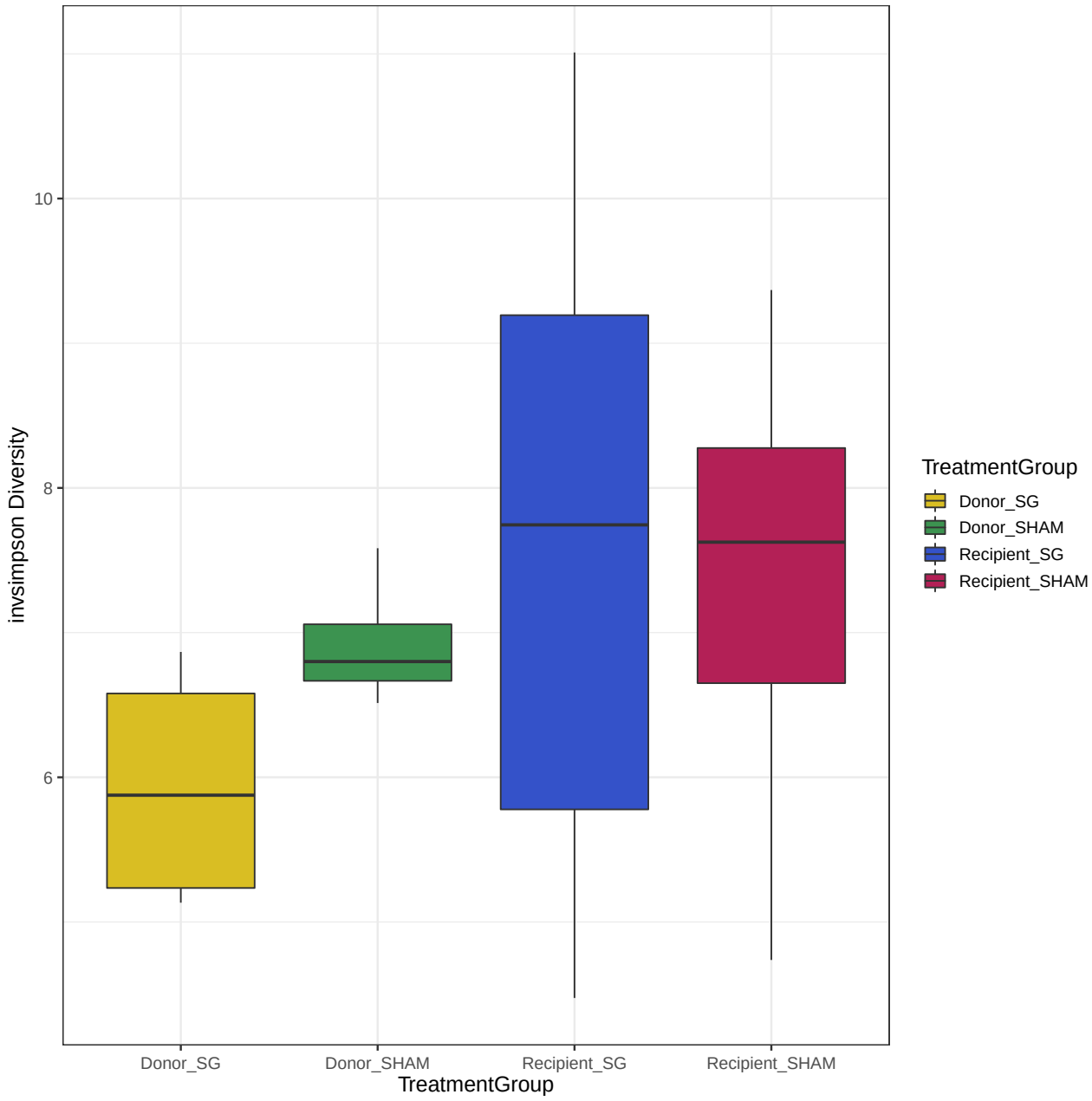
ANOVA on linear model for each independent variable using the Inverse Simpson index.

	Sum Sq	Df	F value	Pr(>F)
TreatmentGroup	9.100	3	1.114	0.363
Residuals	65.351	24	NA	NA

Heatmap of FDR adjusted ANOVA p-values.



Box plots of categorical variable alpha diversity.



Tukey corrected least squares means pairwise comparisons for categorical variables.

Results for TreatmentGroup

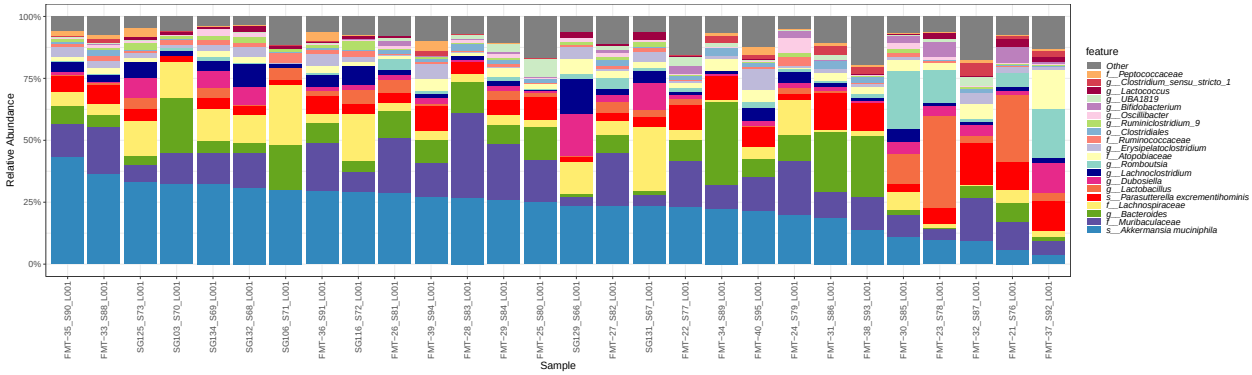
contrast	estimate	SE	df	t.ratio	p.value
Donor_SG - Donor_SHAM	-0.986	1.167	24	-0.845	0.832
Donor_SG - Recipient_SG	-1.725	0.976	24	-1.767	0.313
Donor_SG - Recipient_SHAM	-1.429	0.976	24	-1.464	0.474
Donor_SHAM - Recipient_SG	-0.739	0.976	24	-0.757	0.873
Donor_SHAM - Recipient_SHAM	-0.443	0.976	24	-0.454	0.968
Recipient_SG - Recipient_SHAM	0.296	0.738	24	0.401	0.978

3 Taxonomy visualization

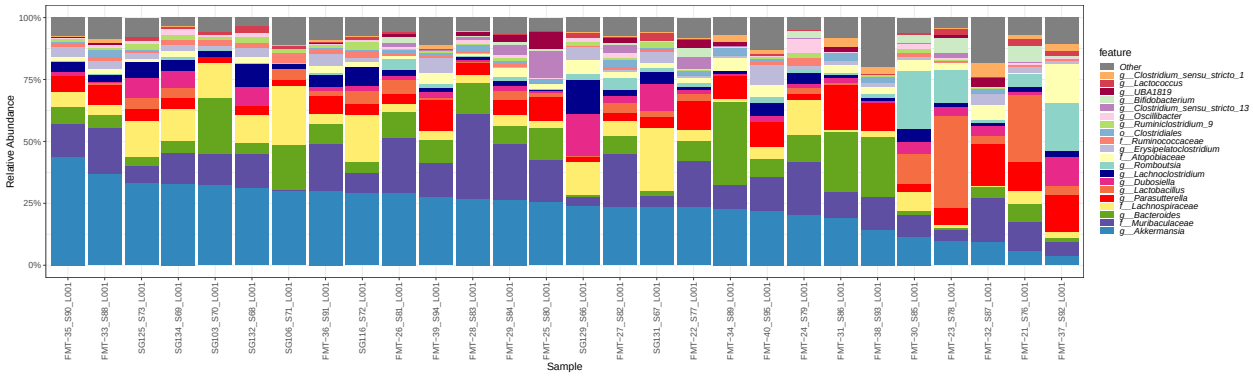
3.1 Taxonomy visualized as stacked barplots

Abundances were passed through a basic filter requiring each species to have at least 0.01% abundance in at least 10% of all samples. A total of 97 species and 90 genera were identified. After basic filtering 60 species and 56 genera remained. The lowest taxonomic rank to be identified above species and genus respectively is displayed.

Bar plot of the top 20 species.

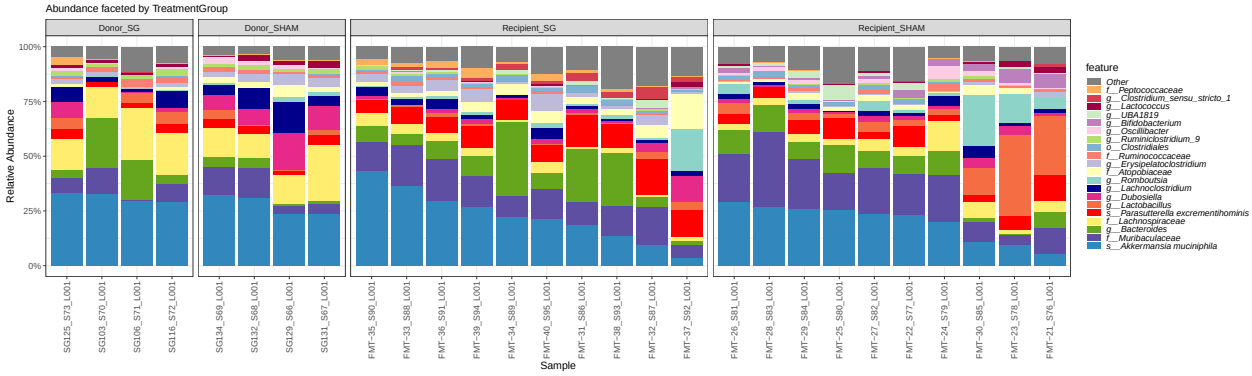


Bar plot of the top 20 genera.

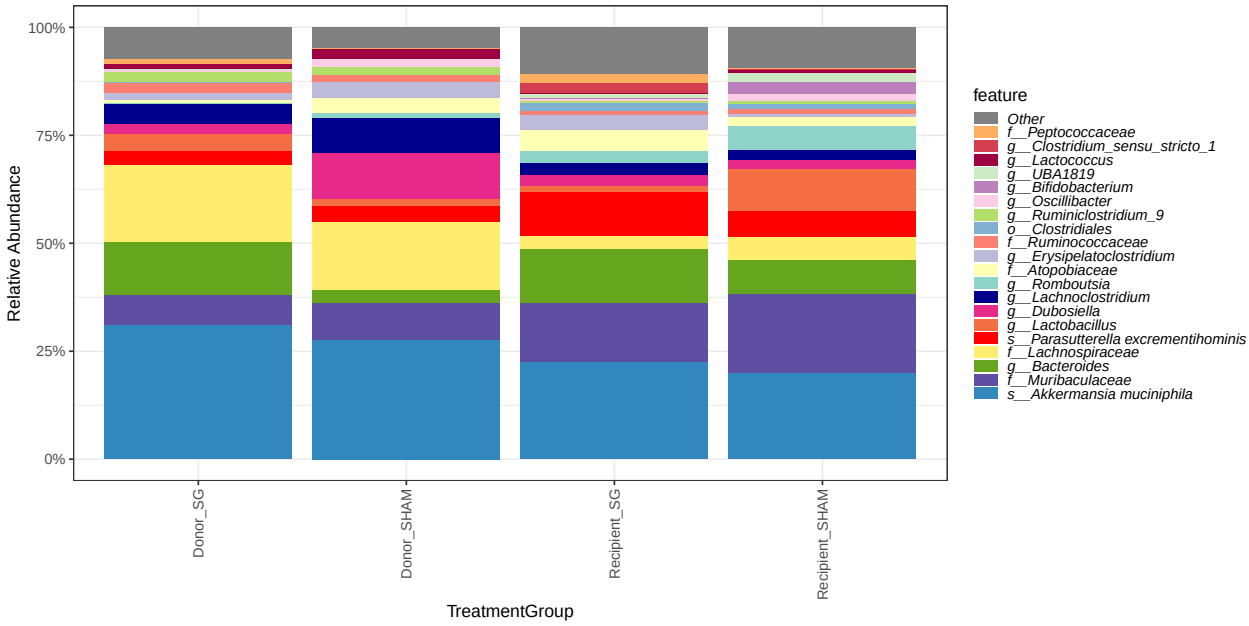


3.1.1 Bar plots grouped by metadata

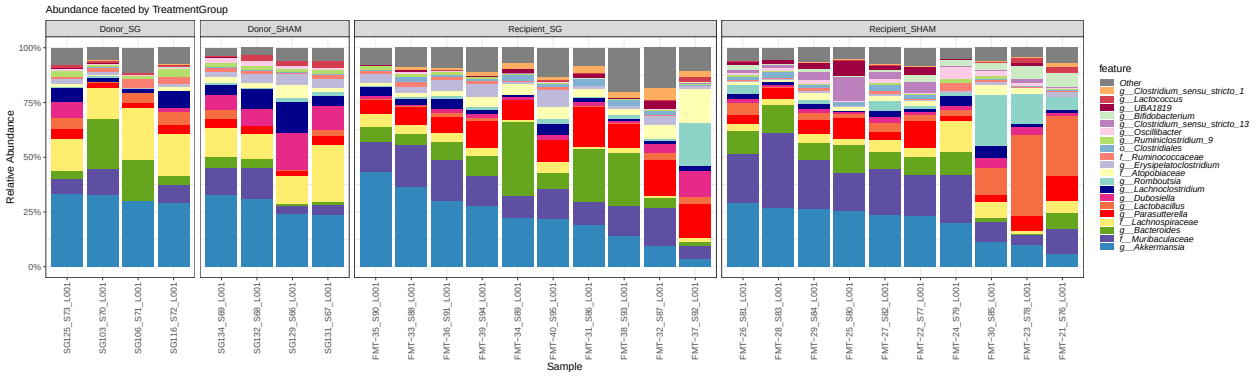
Bar plot of the top 20 species faceted by metadata category.



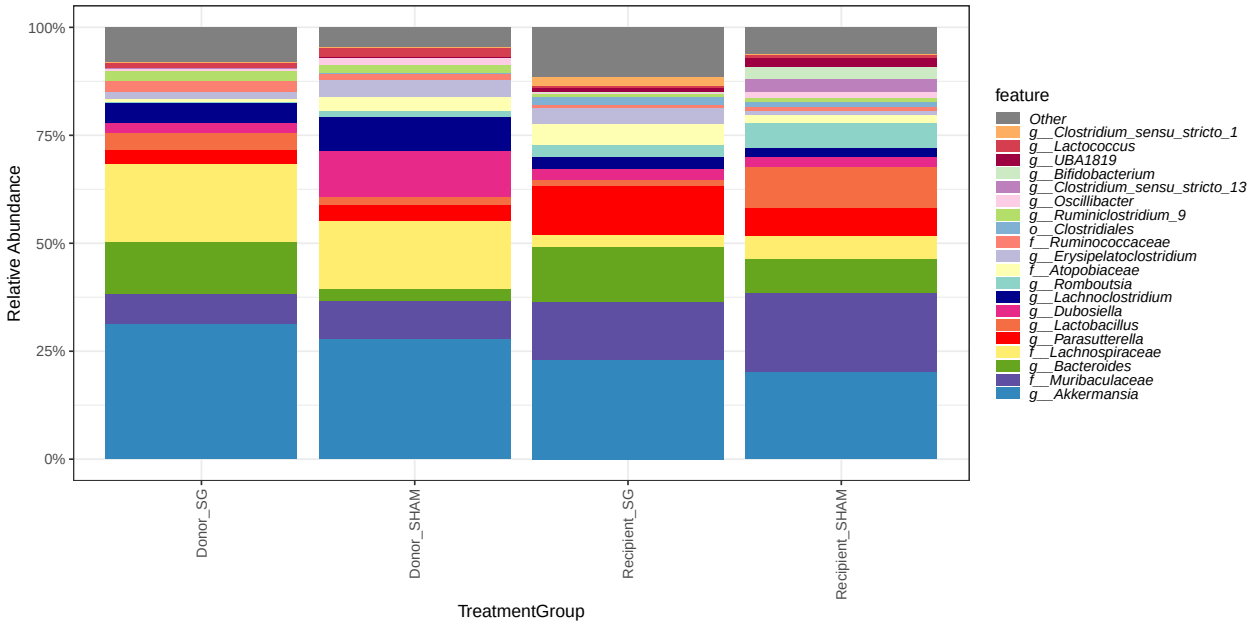
Bar plot of the top 20 species faceted by metadata category and averaged.



Bar plot of the top 20 genera faceted by metadata category.



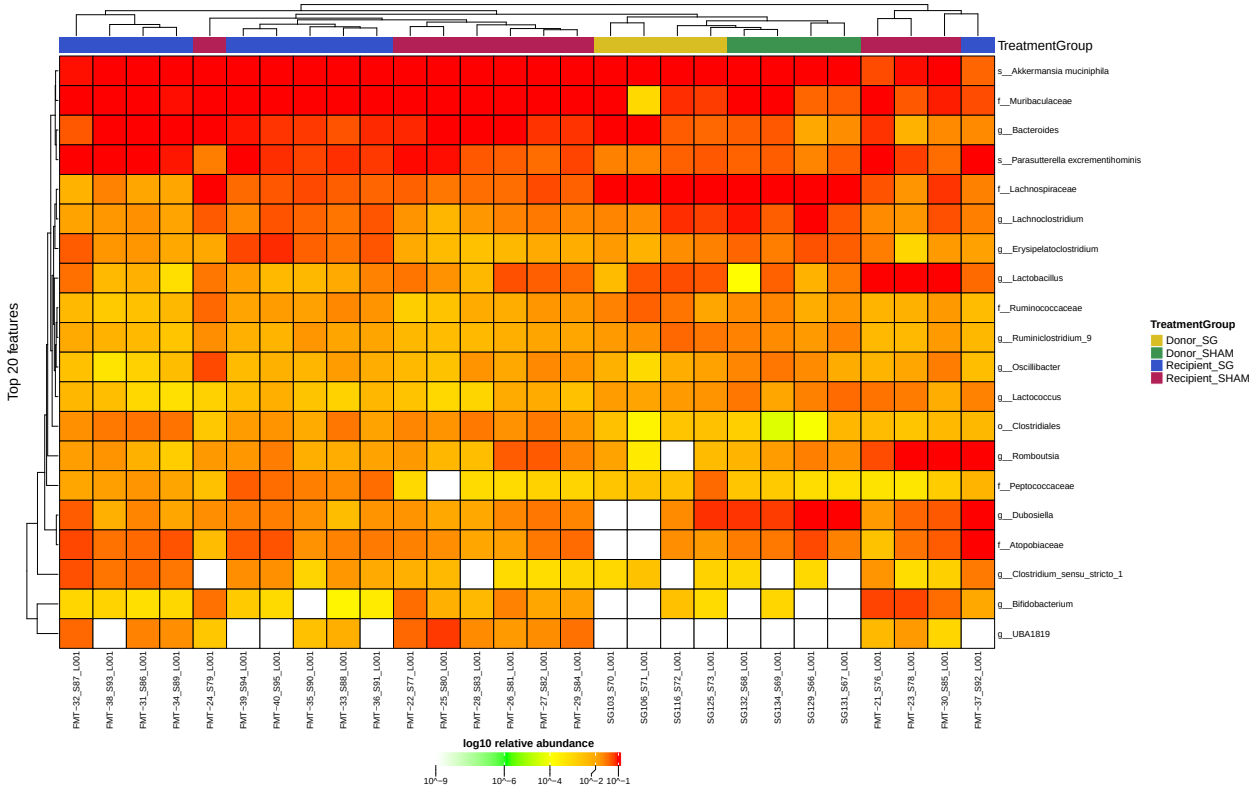
Bar plot of the top 20 genera faceted by metadata and averaged.



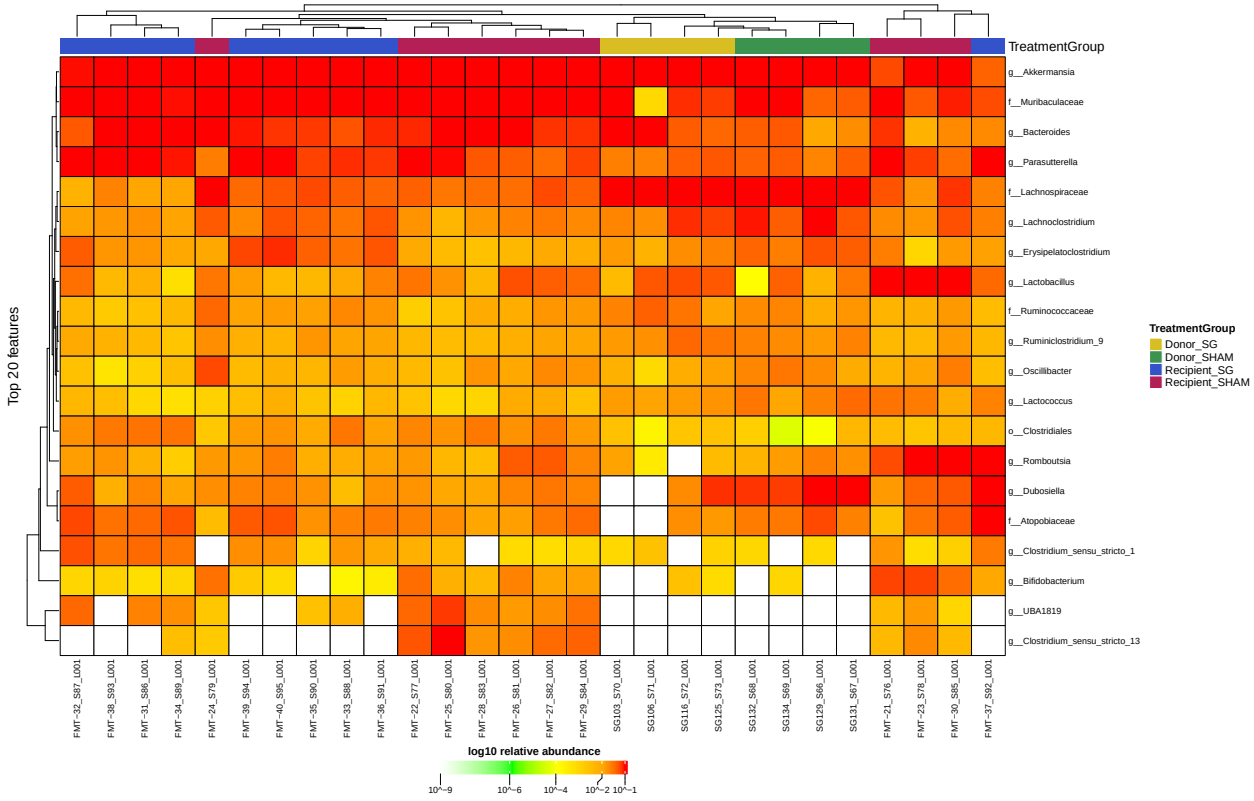
3.2 Taxonomy visualized as heatmaps

3.2.1 Heatmaps clustered based on similar compositions

Heatmap of log10 relative abundances of the species in samples. Samples (columns) and taxa (rows) are UPGMA clustered based on Bray-Curtis dissimilarities and Euclidean distances, respectively. Missing metadata are colored gray. Top 20 most abundant species are displayed.

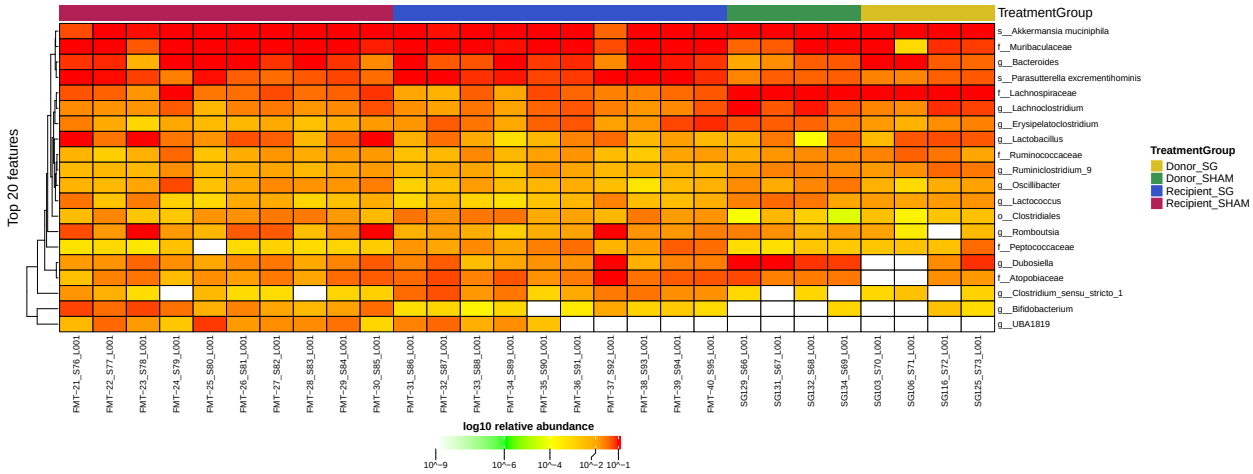


Top 20 most abundant genera are displayed.

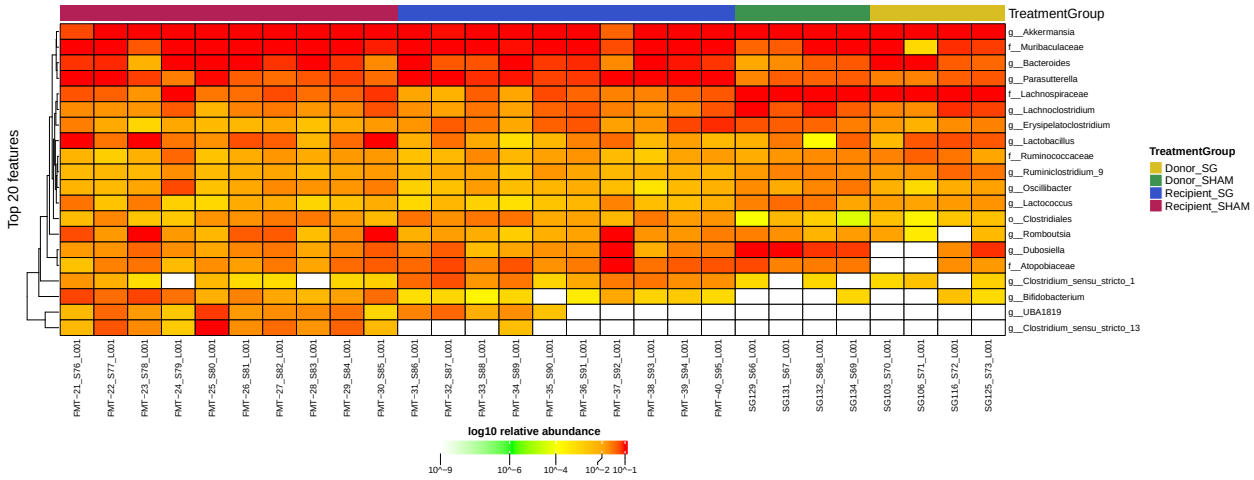


3.2.2 Heatmaps sorted based on metadata

Heatmap of log10 relative abundances of the species in samples. Samples (columns) and taxa (rows) are sorted by metadata and UPGMA clustered, respectively. Missing metadata are colored gray.



Top 20 most abundant genera are displayed.

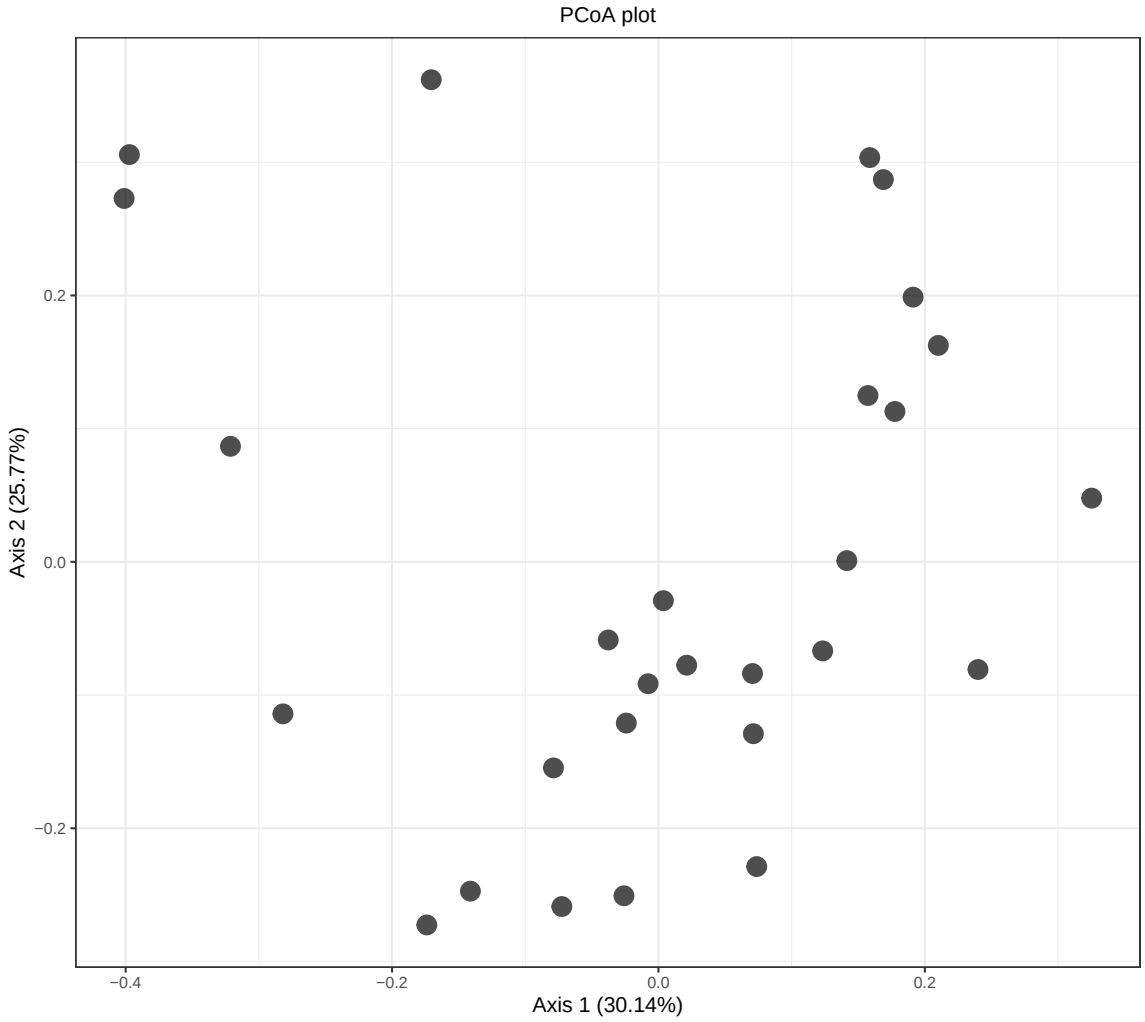


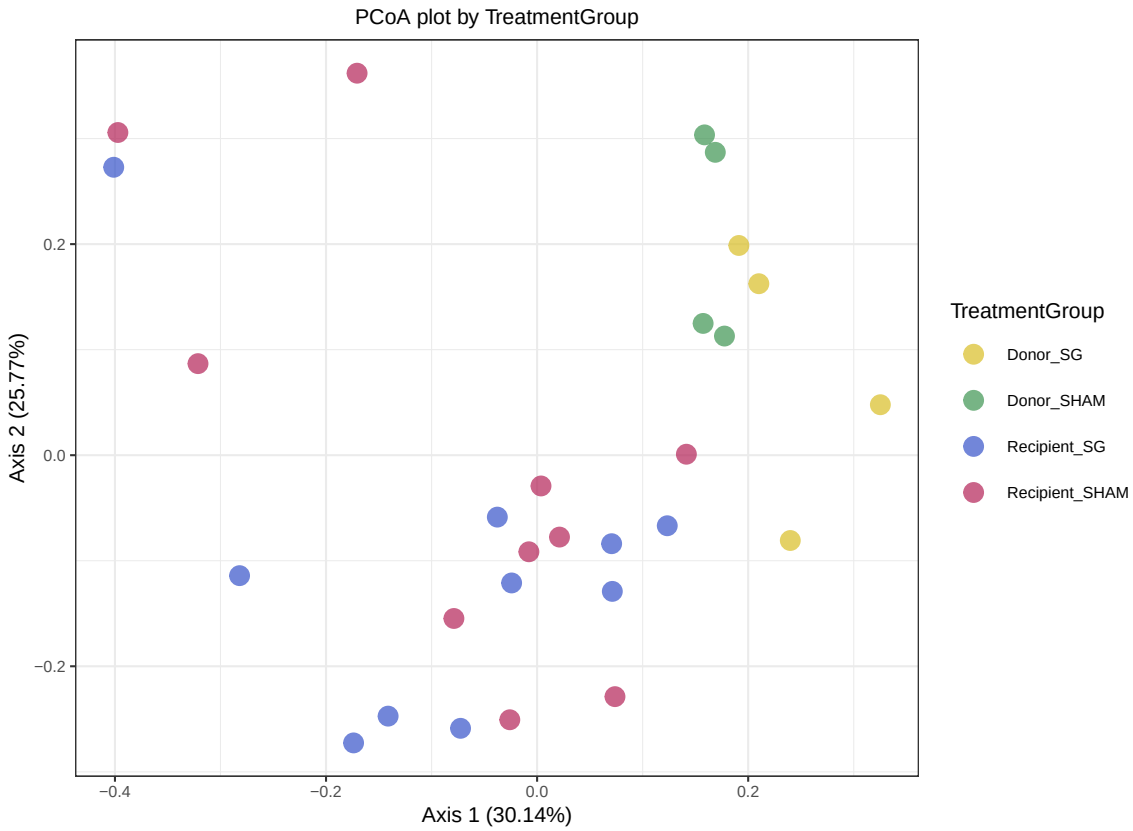
4 Principle coordinates analysis

4.1 PCoA colored by metadata category

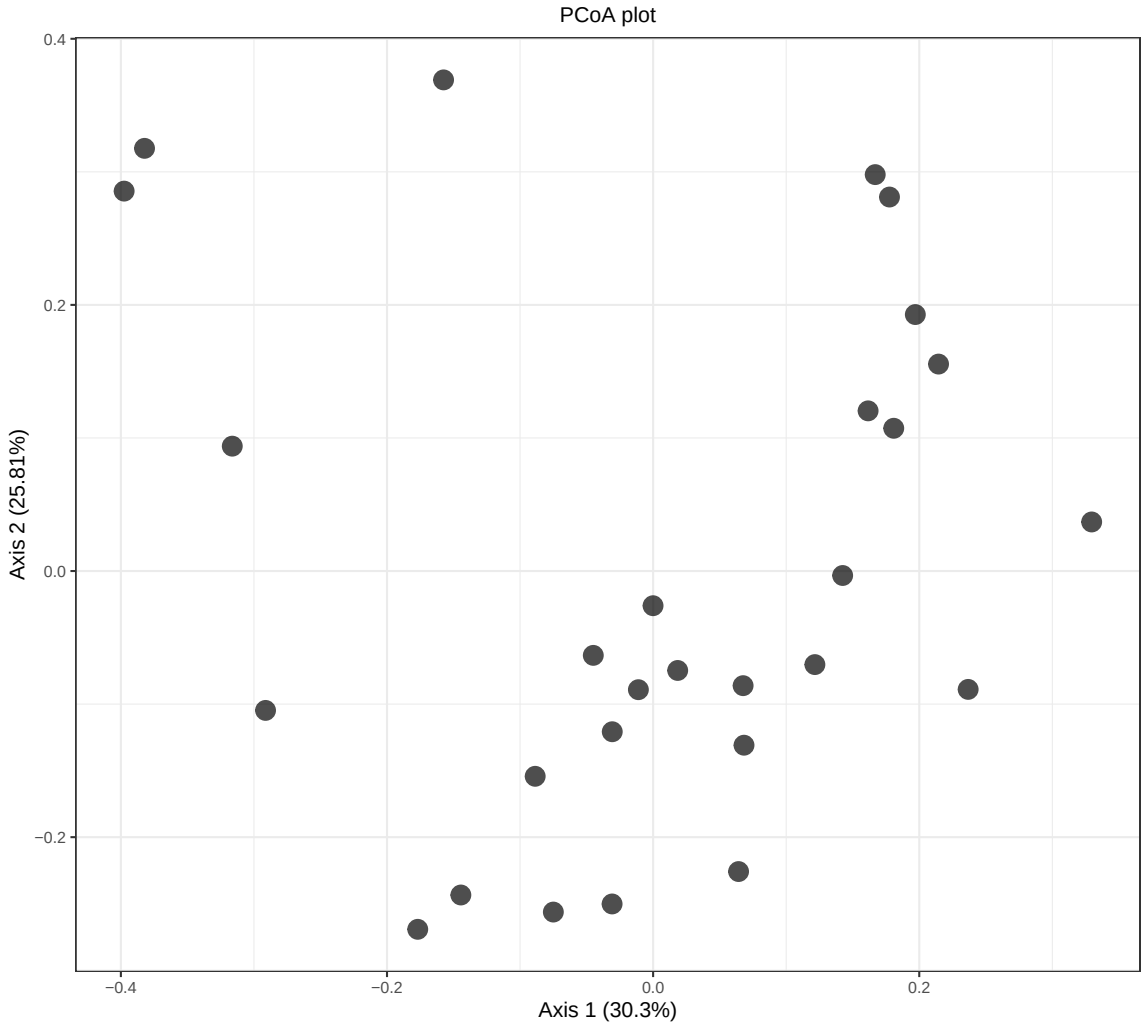
4.1.1 Bray-Curtis results

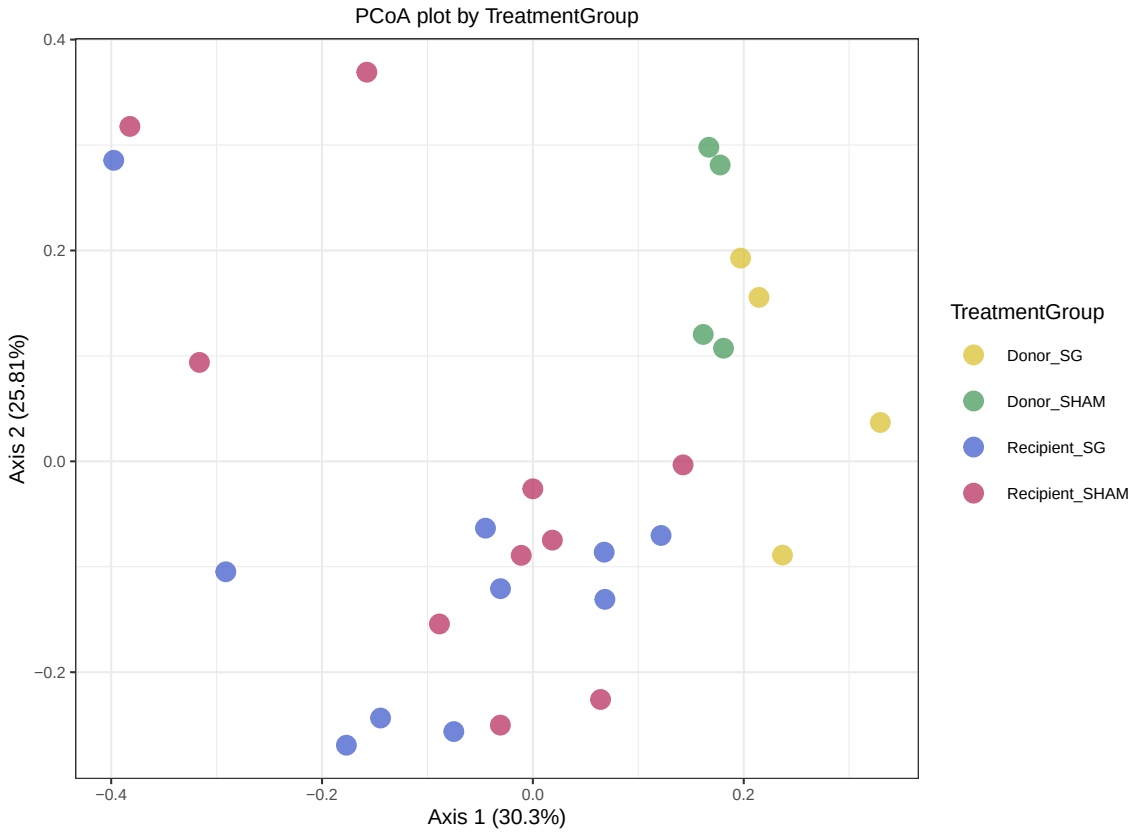
Species level plots.




PCoA plot by TreatmentGroup
Axis 2 (25.77%)
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
Axis 1 (30.14%)
TreatmentGroup
Donor_SG
Donor_SHAM
Recipient_SG
Recipient_SHAM

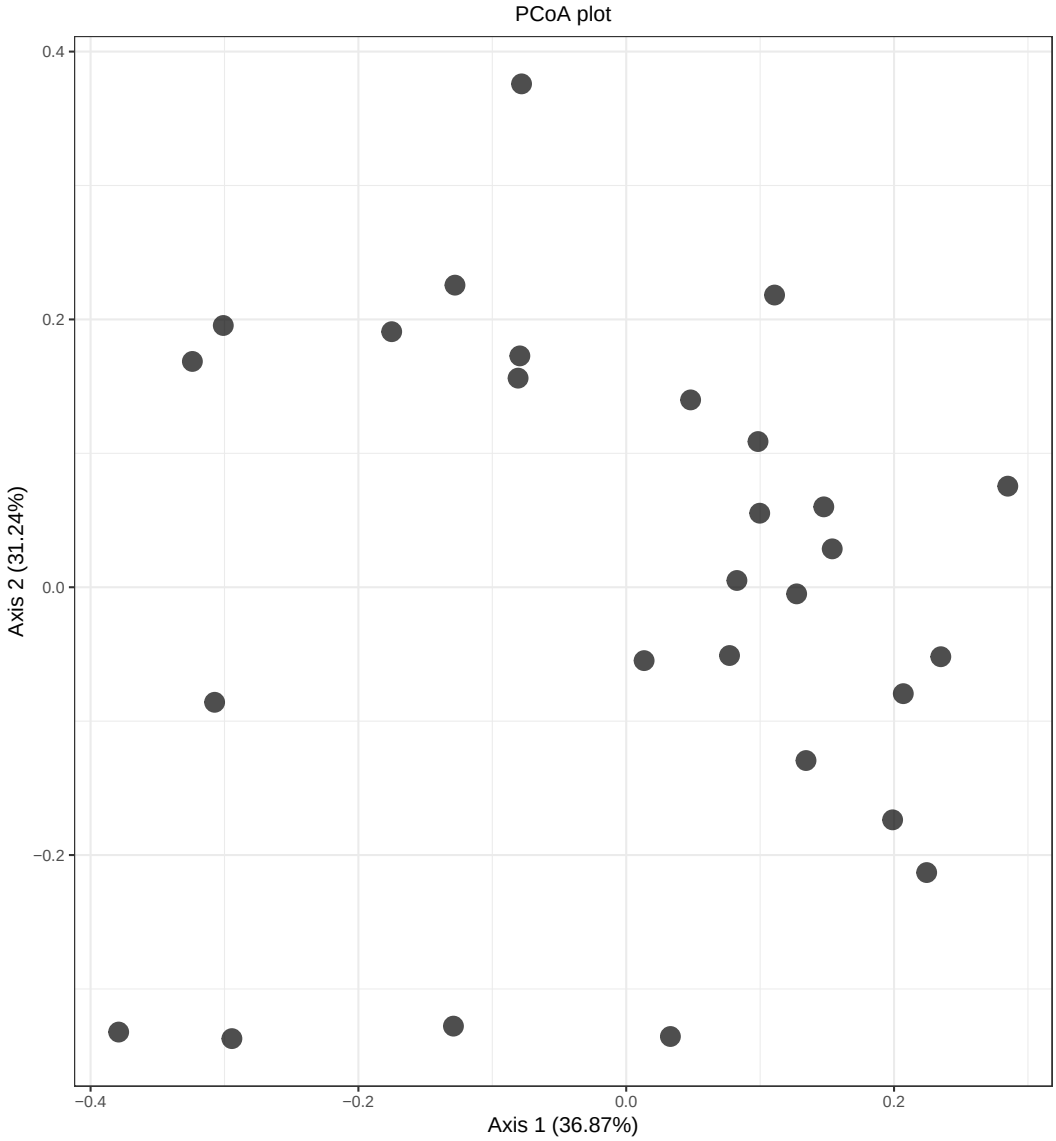
Genus level plots.

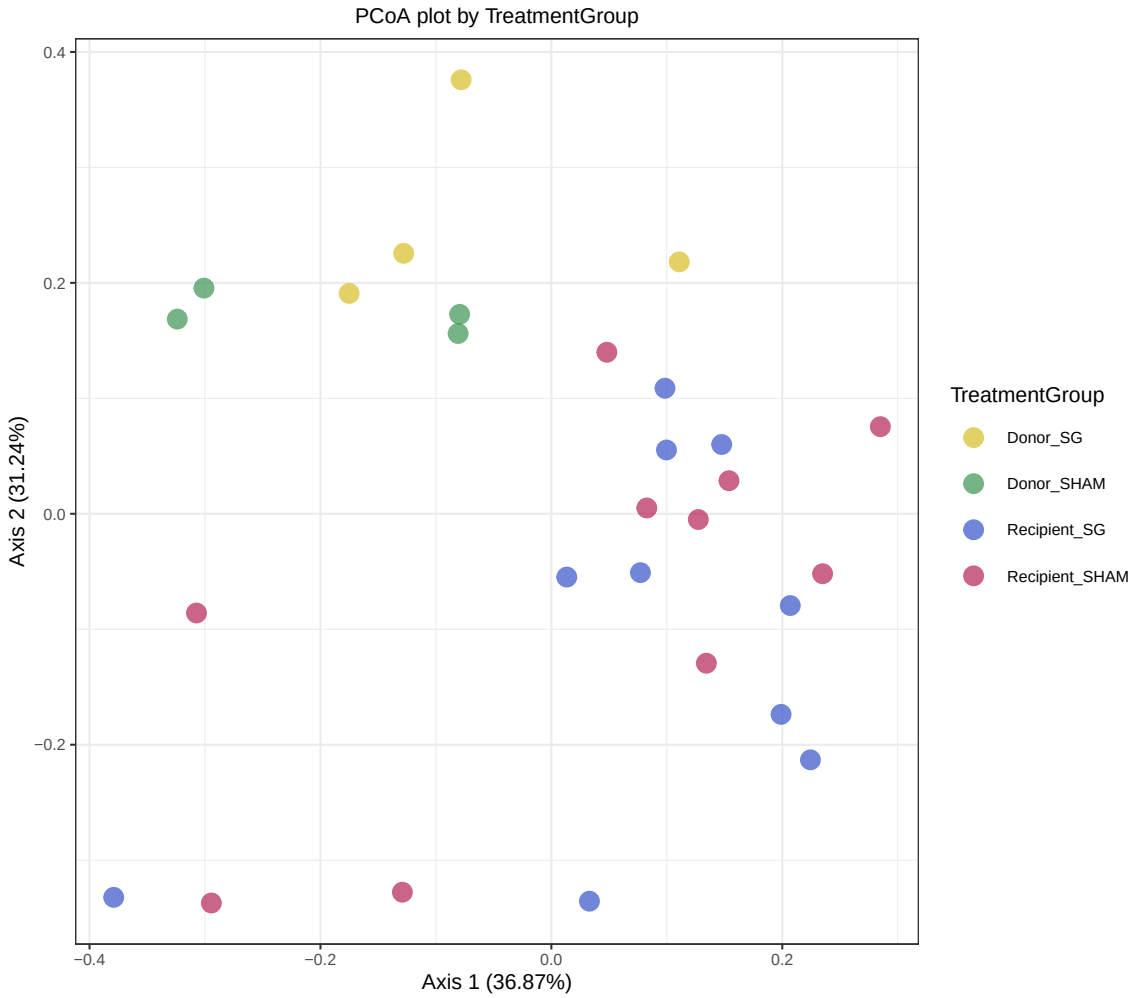



PCoA plot by TreatmentGroup
Axis 2 (25.81%)
Axis 1 (30.3%)
0.4
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
TreatmentGroup
Donor_SG
Donor_SHAM
Recipient_SG
Recipient_SHAM

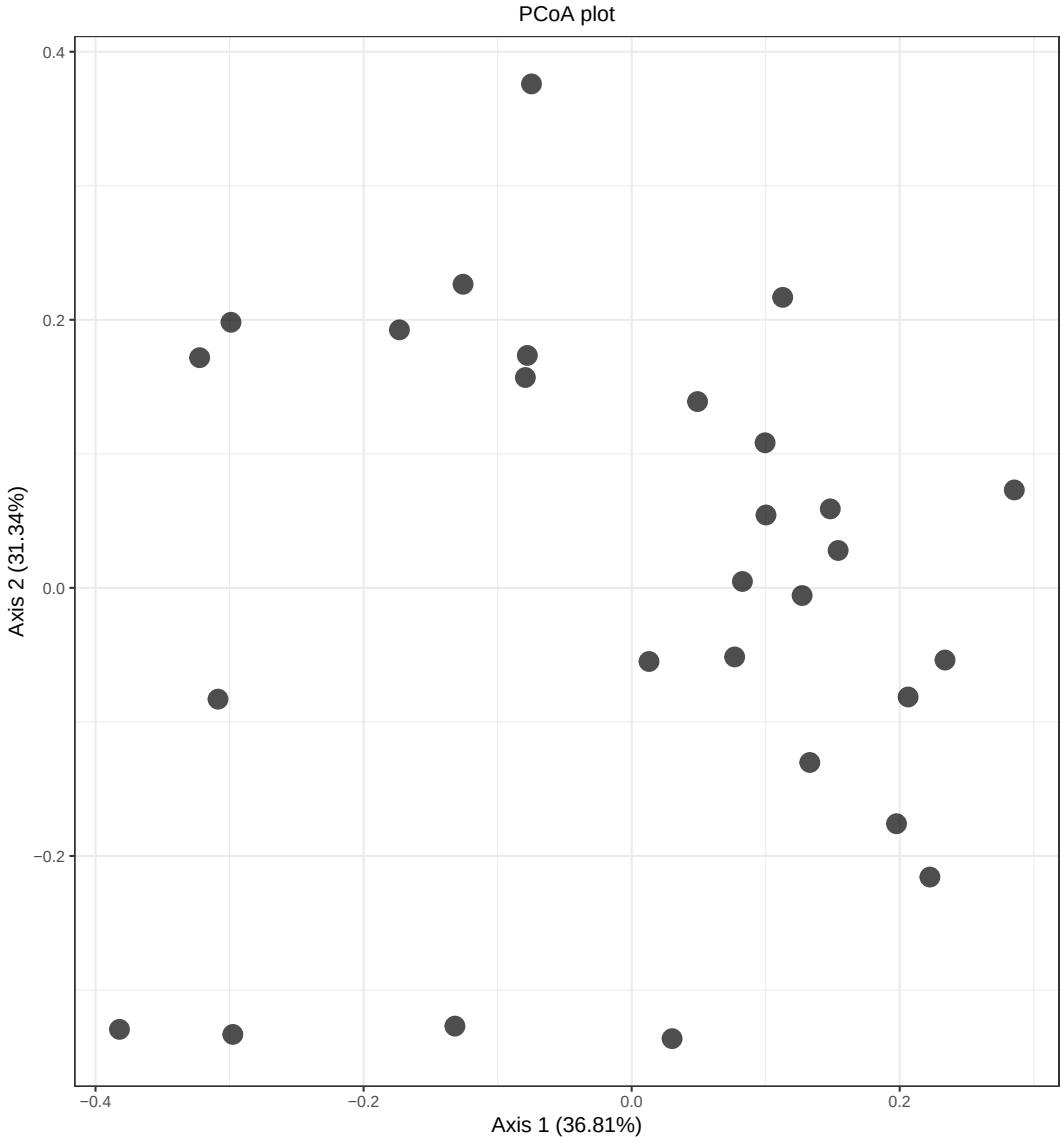
4.1.2 Weighted UniFrac results

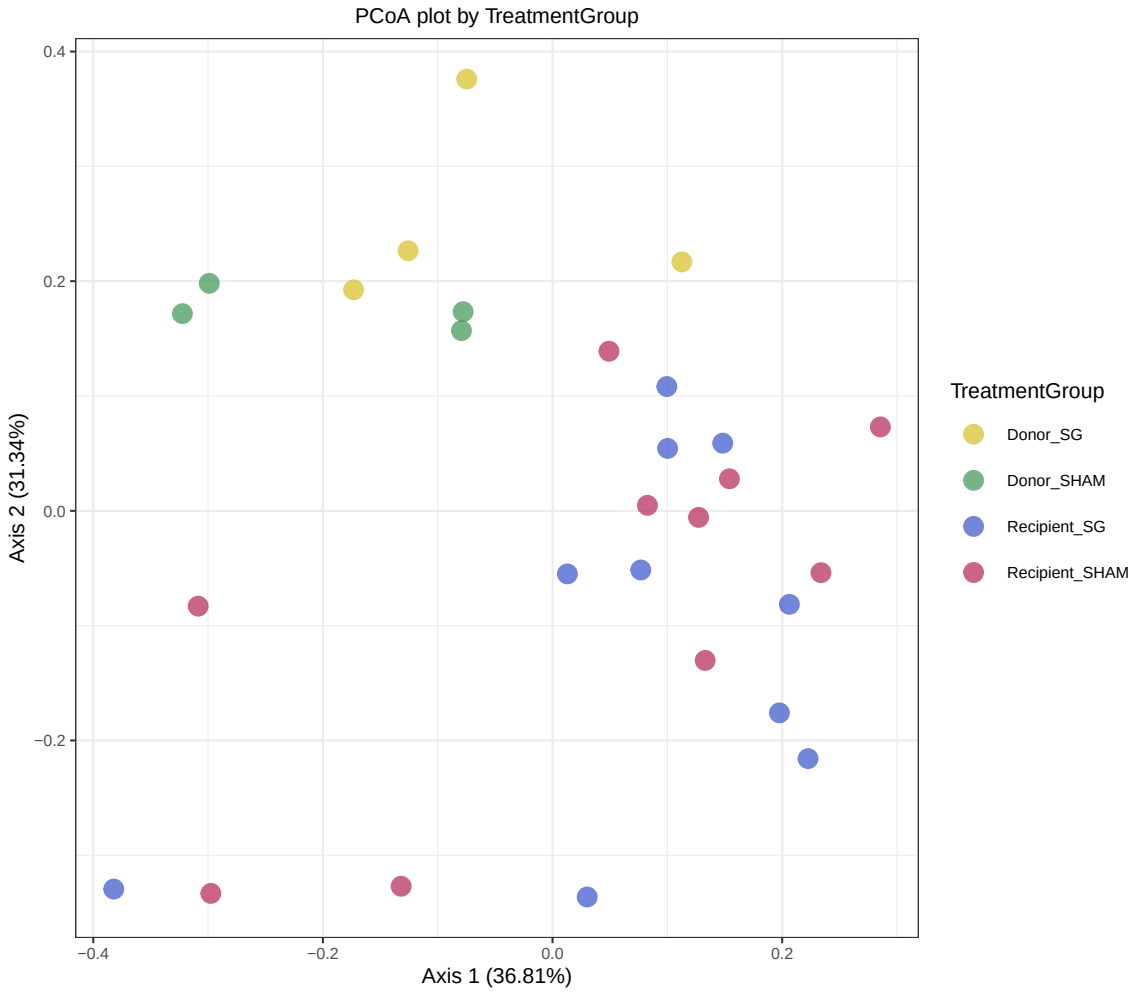
Species level plots.




PCoA plot by TreatmentGroup
Axis 2 (31.24%)
Axis 1 (36.87%)
0.4
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
TreatmentGroup
Donor_SG
Donor_SHAM
Recipient_SG
Recipient_SHAM

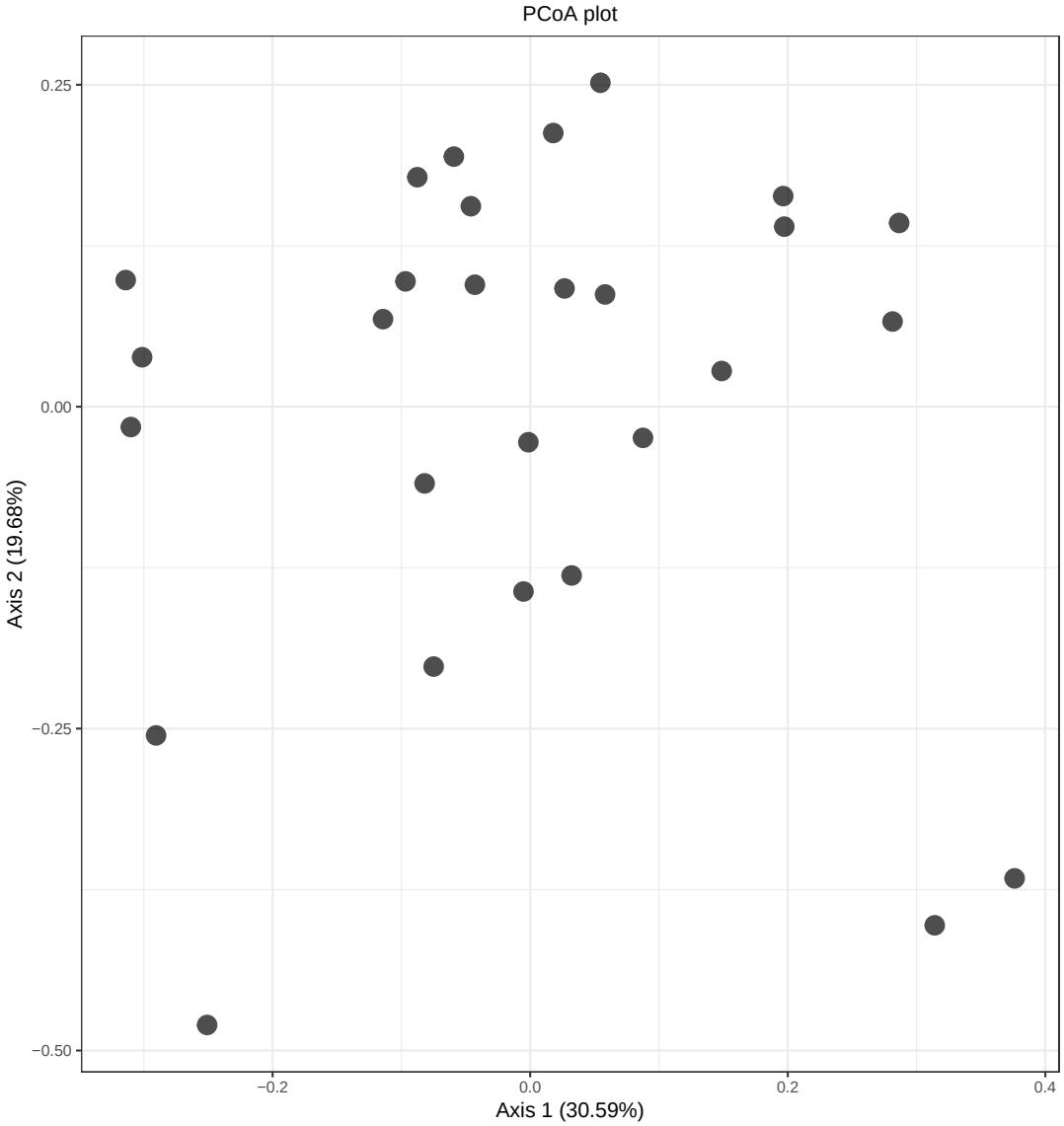
Genus level plots.

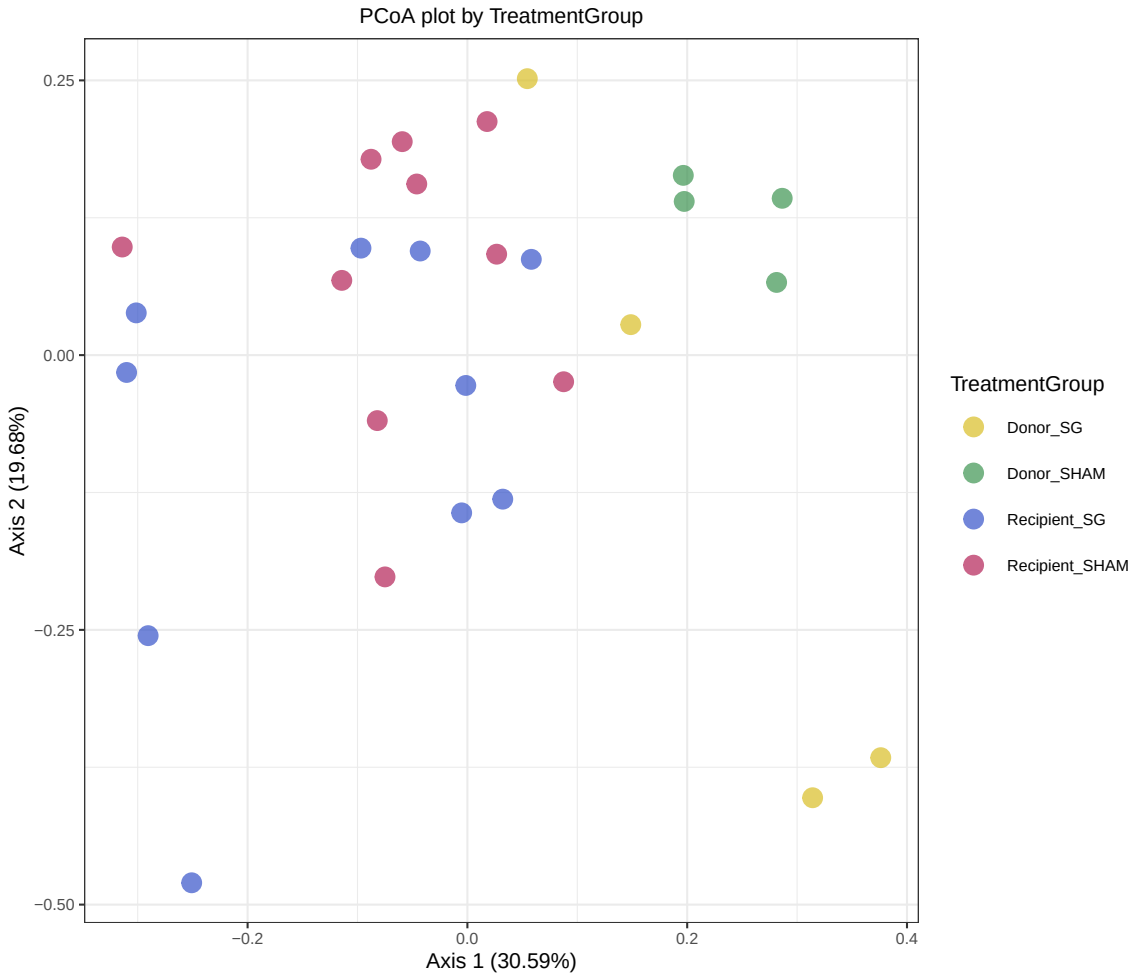



PCoA plot by TreatmentGroup
Axis 2 (31.34%)
Axis 1 (36.81%)
0.4
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
TreatmentGroup
Donor_SG
Donor_SHAM
Recipient_SG
Recipient_SHAM

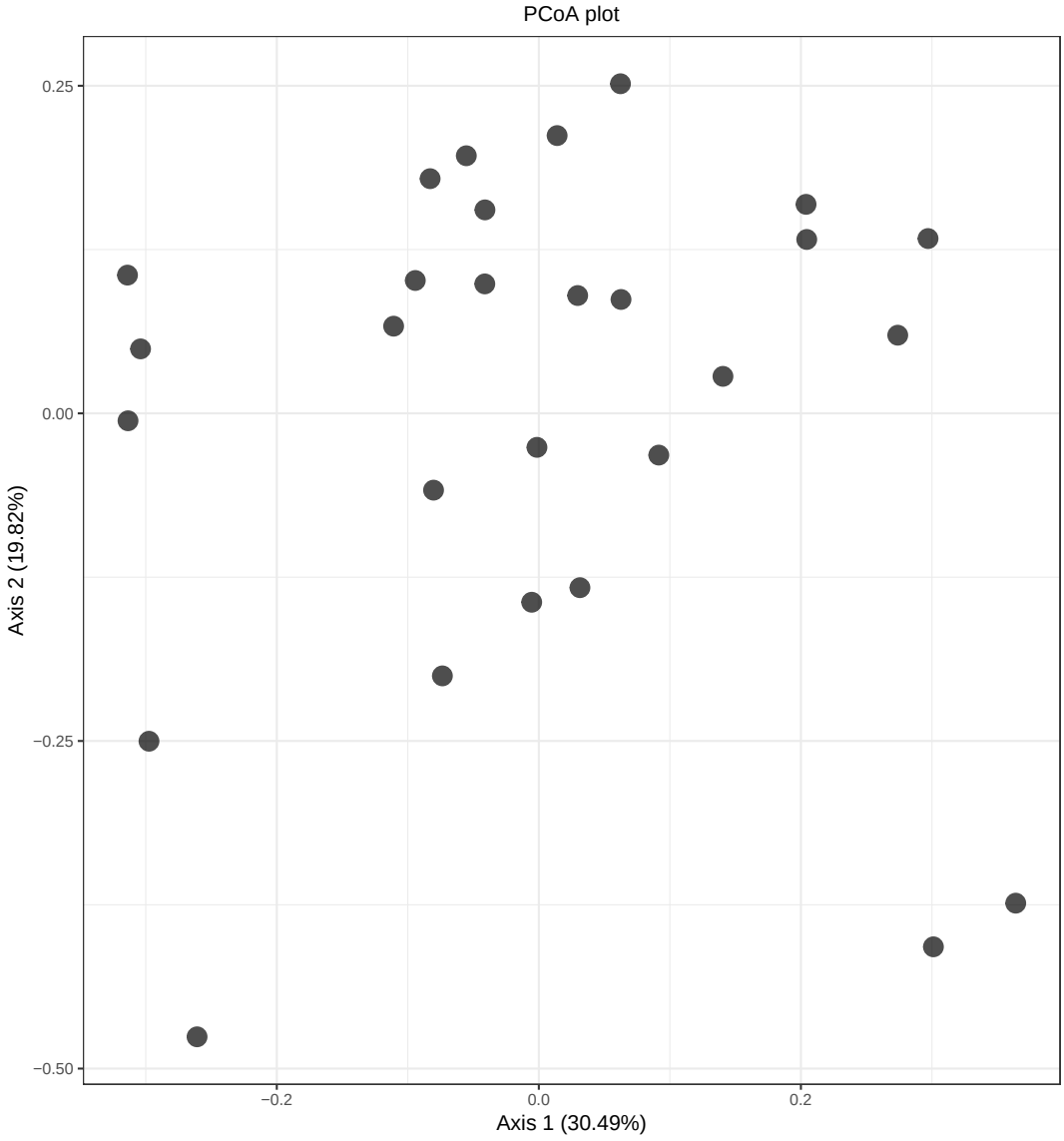
4.1.3 Unweighted UniFrac results

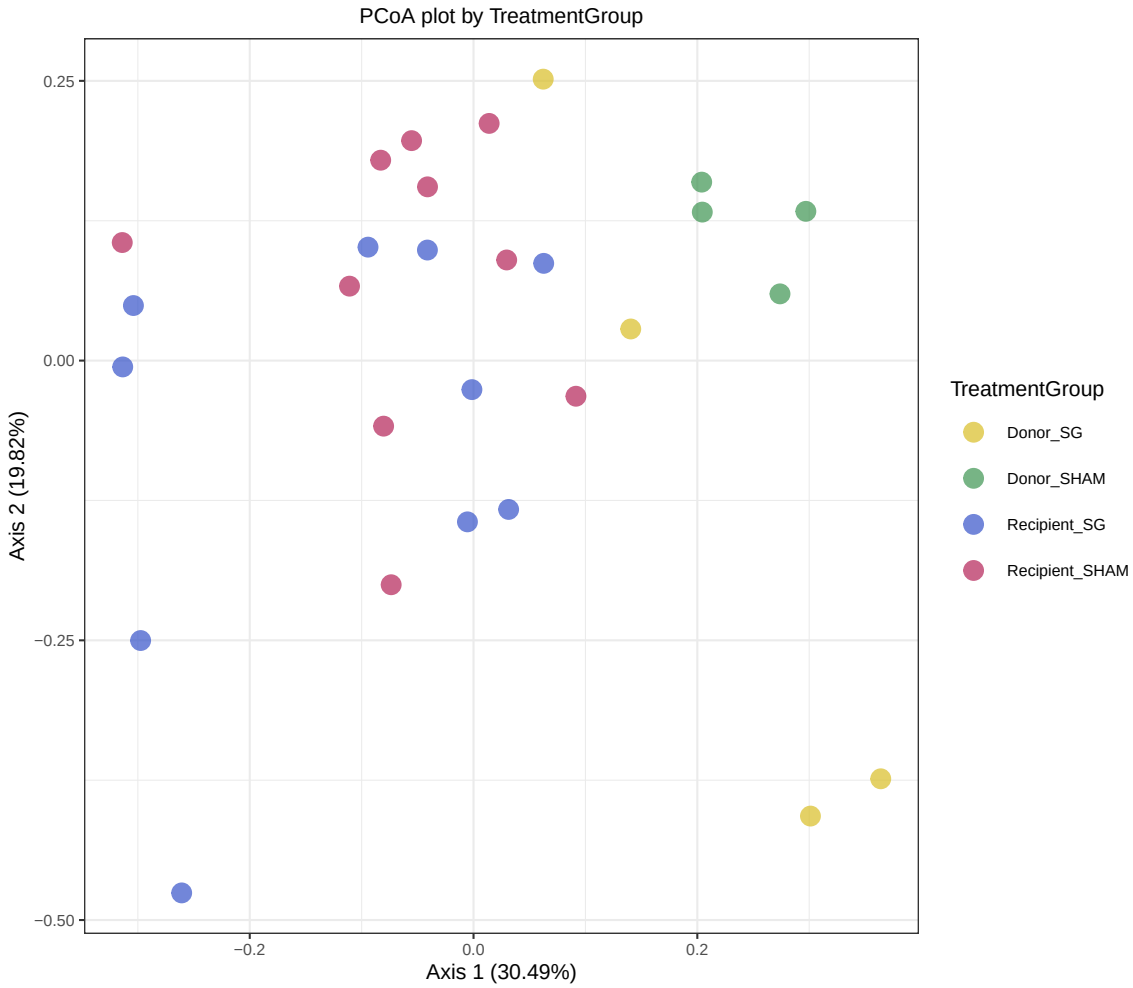
Species level plots.




PCoA plot by TreatmentGroup
0.25
0.00
−0.25
−0.50
Axis 2 (19.68%)
−0.2
0.0
0.2
0.4
Axis 1 (30.59%)
TreatmentGroup
Donor_SG
Donor_SHAM
Recipient_SG
Recipient_SHAM

Genus level plots.




PCoA plot by TreatmentGroup
0.25
0.00
−0.25
−0.50
Axis 2 (19.82%)
−0.2
0.0
0.2
Axis 1 (30.49%)
TreatmentGroup
Donor_SG
Donor_SHAM
Recipient_SG
Recipient_SHAM

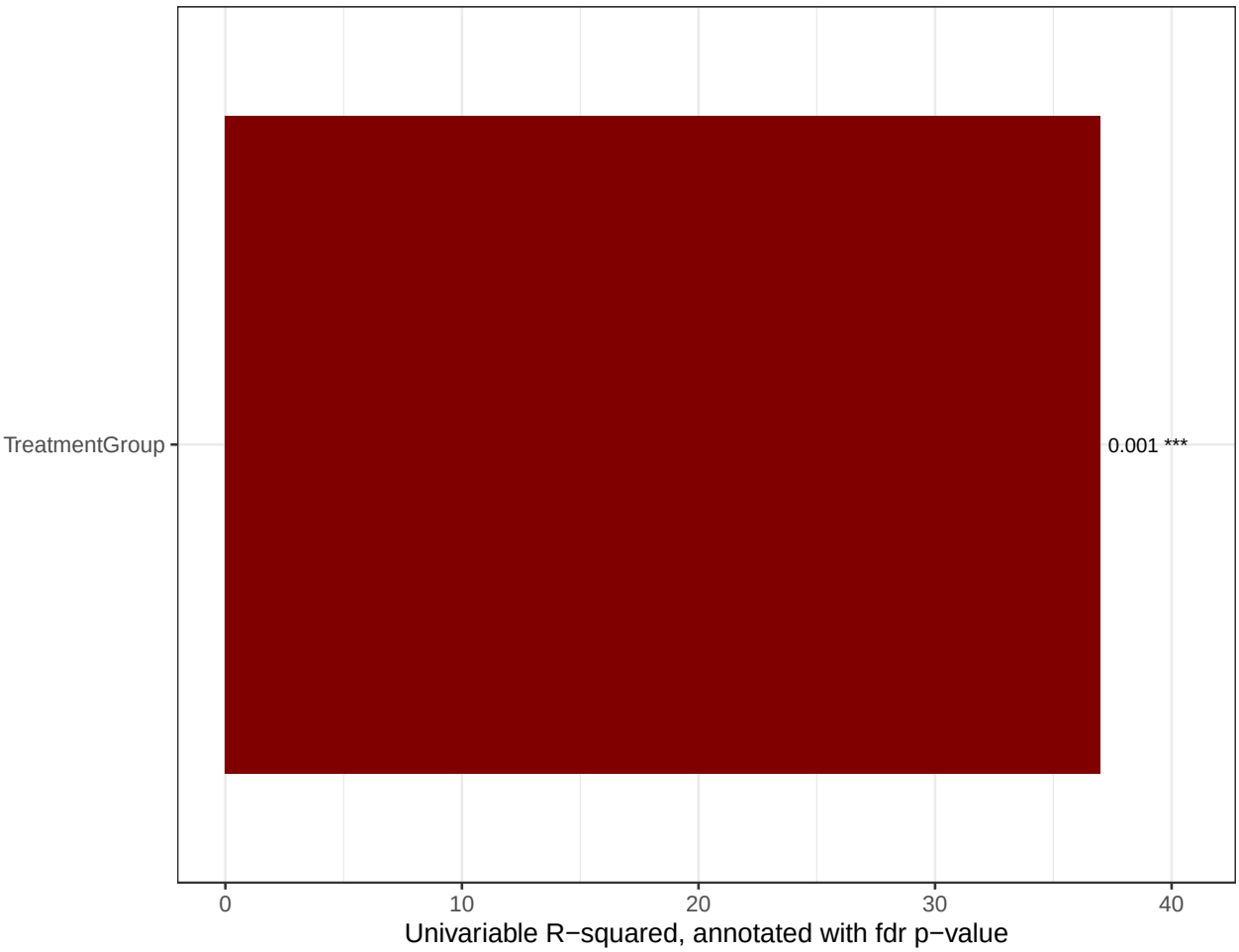
5 Beta diversity analysis

5.1 Univariable PERMANOVA

5.1.1 Bray-Curtis results

PERMANOVA on Bray-Curtis of Species level relative abundances with 999 permutations.

	Df	SumOfSqs	R2	F	Pr(>F)
TreatmentGroup	3	1.034	0.37	4.695	0.001
Residual	24	1.762	0.63	NA	NA
Total	27	2.797	1.00	NA	NA

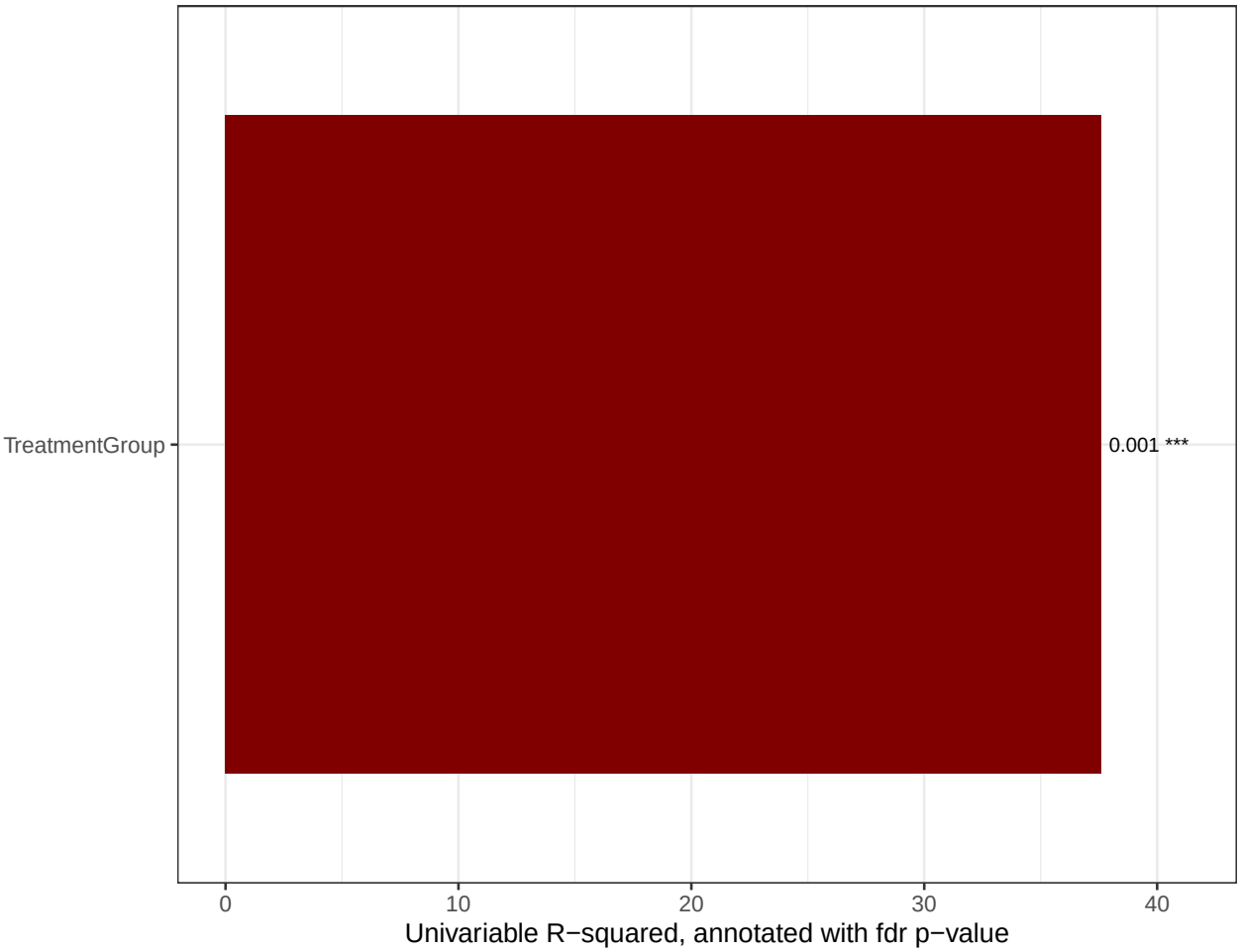


Pairwise PERMANOVA on TreatmentGroup with FDR adjusted p-values.

	R2	PValue	p_adj
Recipient_SHAM - Recipient_SG	0.163	0.003	0.006
Recipient_SHAM - Donor_SG	0.292	0.007	0.008
Recipient_SHAM - Donor_SHAM	0.291	0.005	0.007
Recipient_SG - Donor_SG	0.333	0.003	0.006
Recipient_SG - Donor_SHAM	0.323	0.003	0.006
Donor_SG - Donor_SHAM	0.326	0.059	0.059

PERMANOVA on Bray-Curtis of Genus level relative abundances with 999 permutations.

	Df	SumOfSqs	R2	F	Pr(>F)
TreatmentGroup	3	1.042	0.376	4.819	0.001
Residual	24	1.730	0.624	NA	NA
Total	27	2.773	1.000	NA	NA



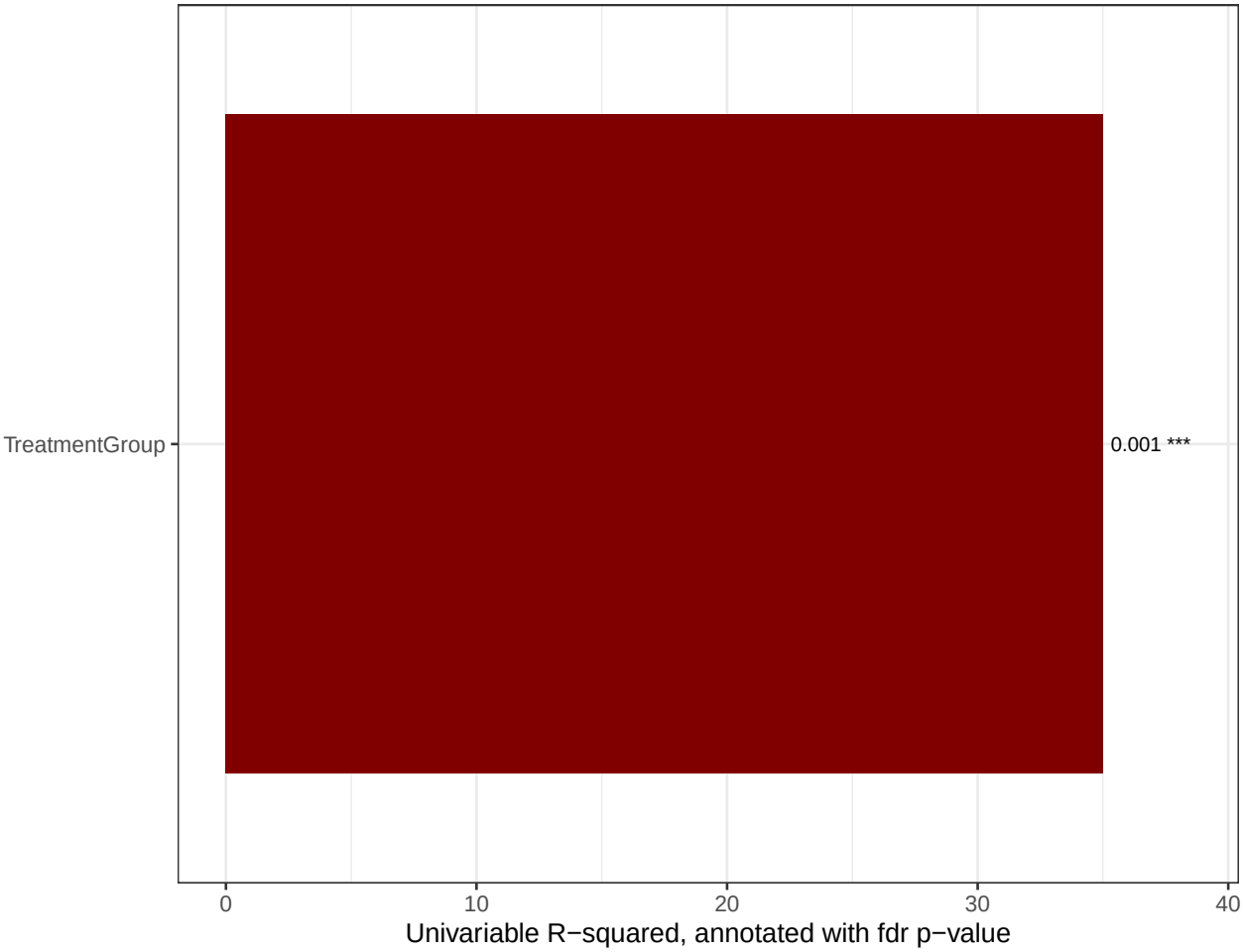
Pairwise PERMANOVA on TreatmentGroup with FDR adjusted p-values.

	R2	PValue	p_adj
Recipient_SHAM - Recipient_SG	0.169	0.003	0.005
Recipient_SHAM - Donor_SG	0.296	0.007	0.008
Recipient_SHAM - Donor_SHAM	0.296	0.003	0.005
Recipient_SG - Donor_SG	0.338	0.003	0.005
Recipient_SG - Donor_SHAM	0.329	0.003	0.005
Donor_SG - Donor_SHAM	0.325	0.059	0.059

5.1.2 Weighted UniFrac results

PERMANOVA on Weighted UniFrac of Species level relative abundances with 999 permutations.

	Df	SumOfSqs	R2	F	Pr(>F)
TreatmentGroup	3	0.431	0.35	4.307	0.001
Residual	24	0.800	0.65	NA	NA
Total	27	1.231	1.00	NA	NA

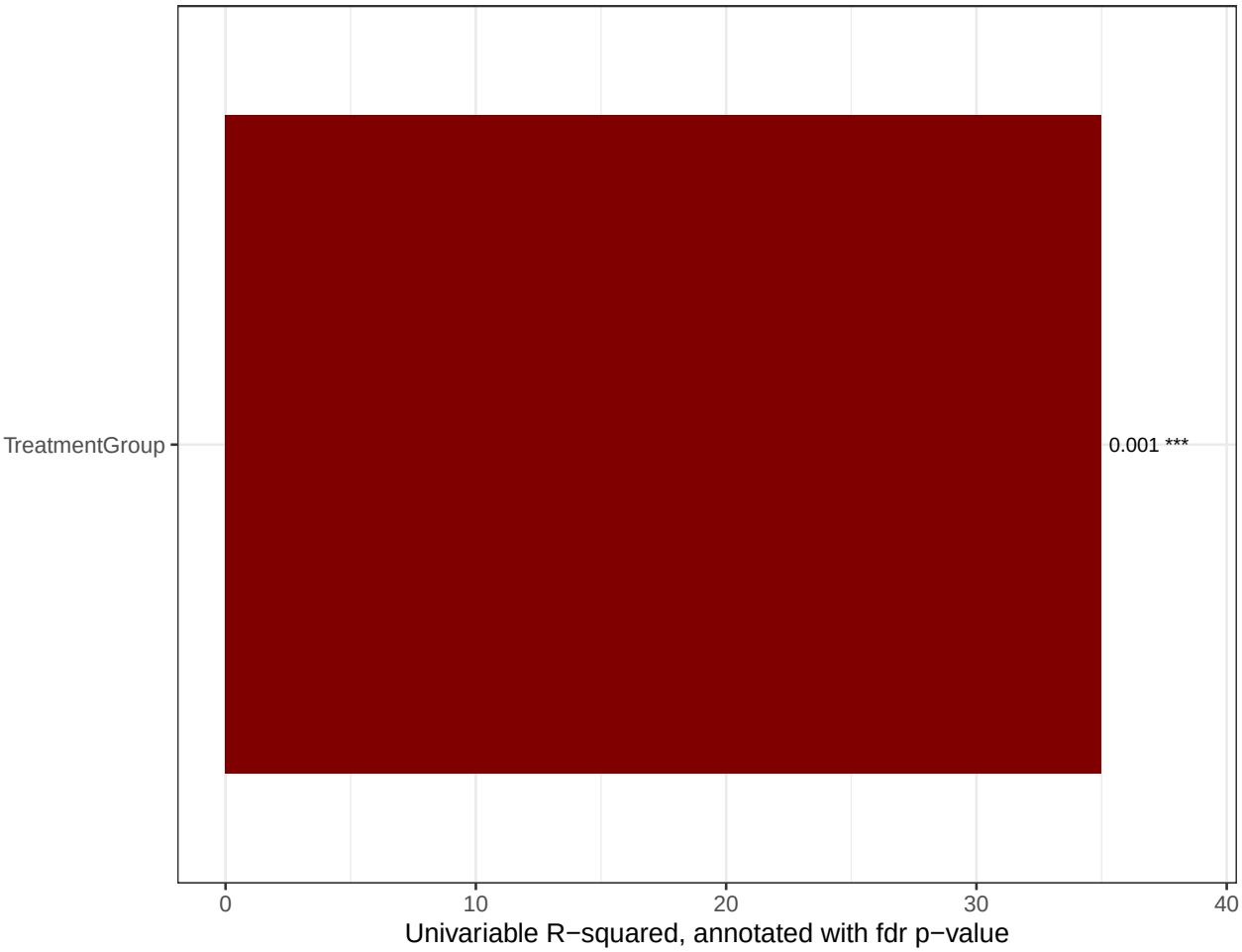


Pairwise PERMANOVA on TreatmentGroup with FDR adjusted p-values.

	R2	PValue	p_adj
Recipient_SHAM - Recipient_SG	0.108	0.085	0.102
Recipient_SHAM - Donor_SG	0.281	0.009	0.013
Recipient_SHAM - Donor_SHAM	0.292	0.006	0.012
Recipient_SG - Donor_SG	0.332	0.003	0.012
Recipient_SG - Donor_SHAM	0.336	0.005	0.012
Donor_SG - Donor_SHAM	0.272	0.192	0.192

PERMANOVA on Weighted UniFrac of Genus level relative abundances with 999 permutations.

	Df	SumOfSqs	R2	F	Pr(>F)
TreatmentGroup	3	0.430	0.35	4.304	0.001
Residual	24	0.799	0.65	NA	NA
Total	27	1.229	1.00	NA	NA



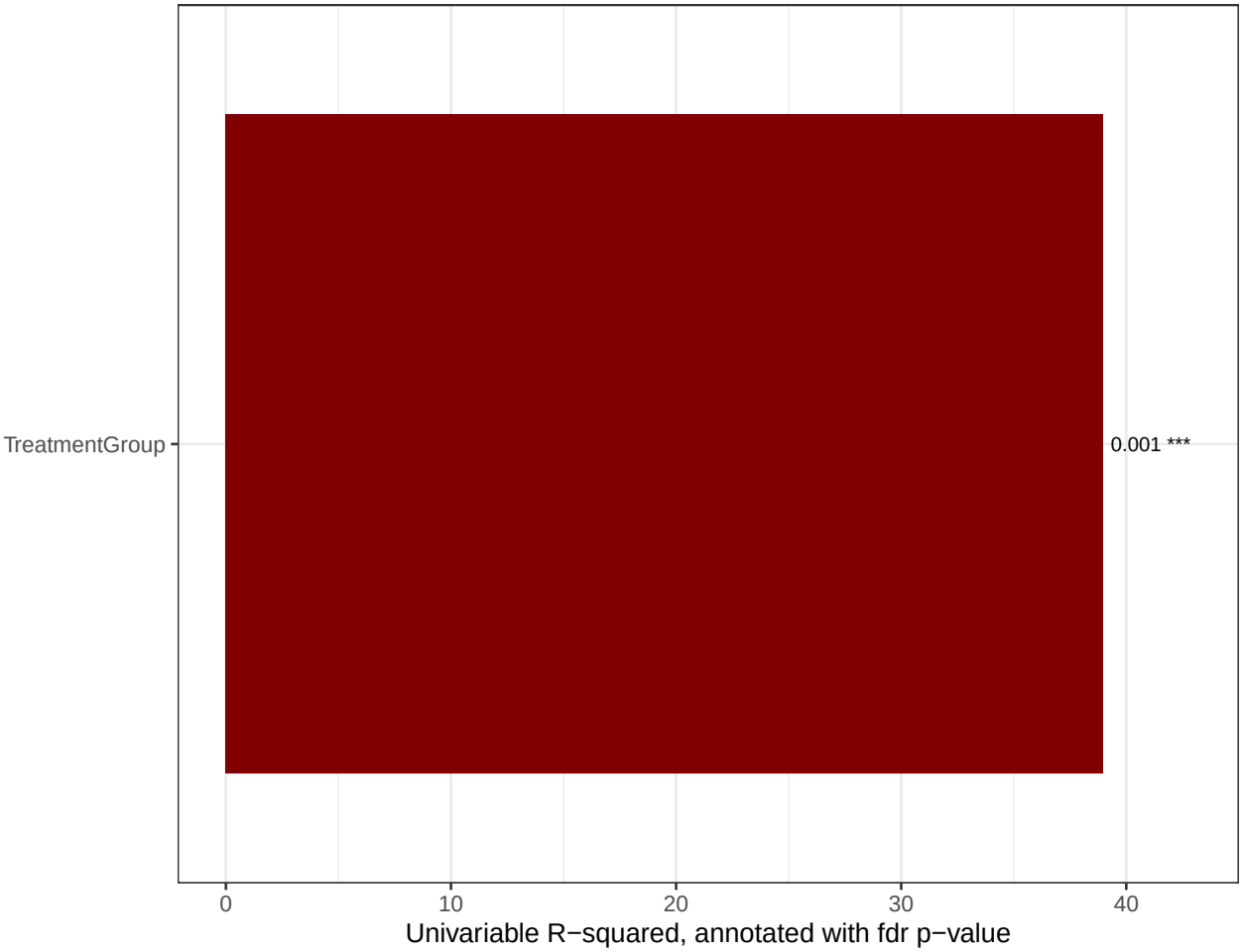
Pairwise PERMANOVA on TreatmentGroup with FDR adjusted p-values.

	R2	PValue	p_adj
Recipient_SHAM - Recipient_SG	0.108	0.085	0.102
Recipient_SHAM - Donor_SG	0.281	0.009	0.013
Recipient_SHAM - Donor_SHAM	0.292	0.006	0.012
Recipient_SG - Donor_SG	0.332	0.003	0.012
Recipient_SG - Donor_SHAM	0.336	0.005	0.012
Donor_SG - Donor_SHAM	0.272	0.192	0.192

5.1.3 Unweighted UniFrac results

PERMANOVA on Unweighted UniFrac of Species level relative abundances with 999 permutations.

	Df	SumOfSqs	R2	F	Pr(>F)
TreatmentGroup	3	0.256	0.39	5.106	0.001
Residual	24	0.401	0.61	NA	NA
Total	27	0.656	1.00	NA	NA

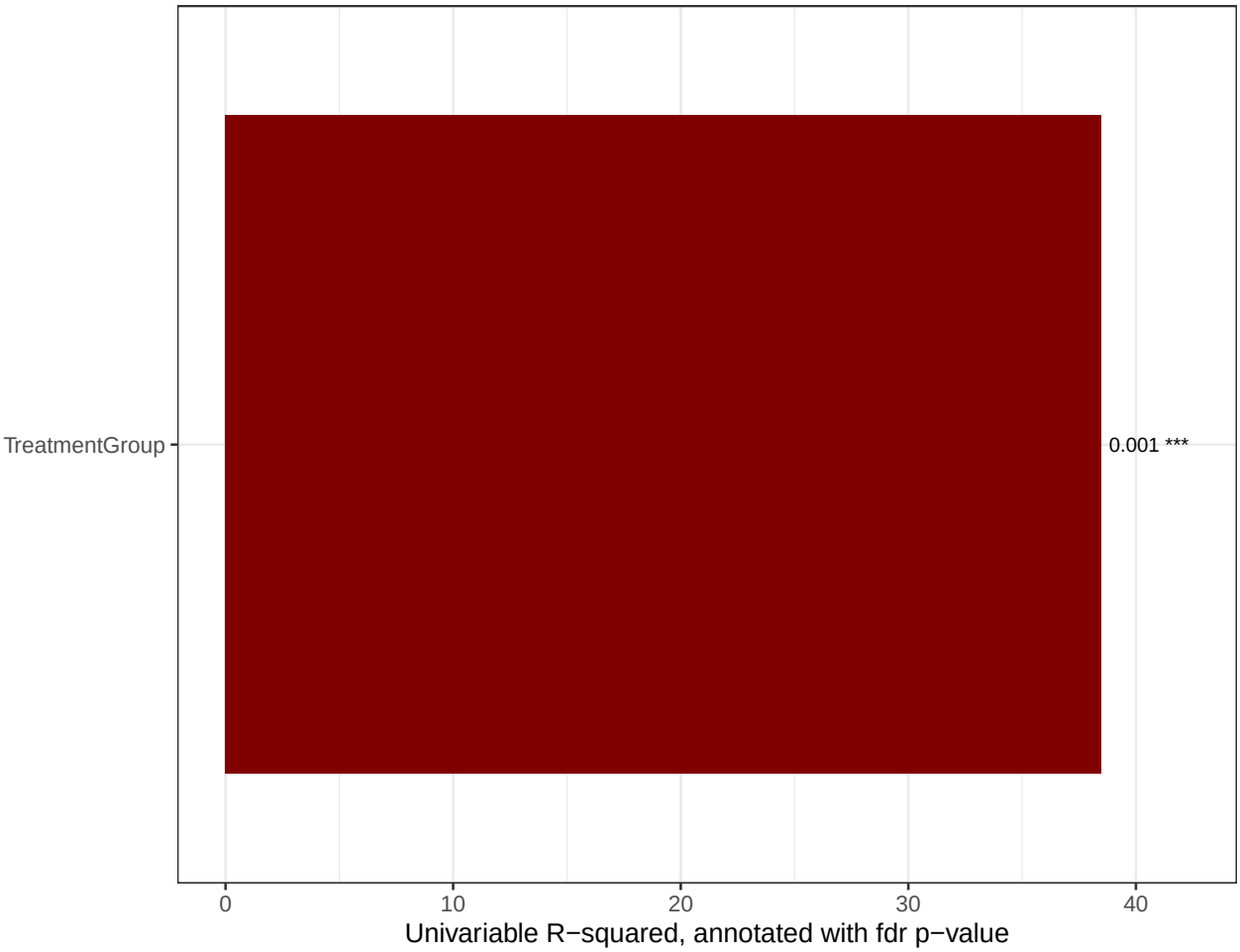


Pairwise PERMANOVA on TreatmentGroup with FDR adjusted p-values.

	R2	PValue	p_adj
Recipient_SHAM - Recipient_SG	0.182	0.002	0.009
Recipient_SHAM - Donor_SG	0.299	0.007	0.011
Recipient_SHAM - Donor_SHAM	0.405	0.003	0.009
Recipient_SG - Donor_SG	0.266	0.011	0.013
Recipient_SG - Donor_SHAM	0.377	0.005	0.010
Donor_SG - Donor_SHAM	0.254	0.220	0.220

PERMANOVA on Unweighted UniFrac of Genus level relative abundances with 999 permutations.

	Df	SumOfSqs	R2	F	Pr(>F)
TreatmentGroup	3	0.249	0.385	5.002	0.001
Residual	24	0.399	0.615	NA	NA
Total	27	0.648	1.000	NA	NA



Pairwise PERMANOVA on TreatmentGroup with FDR adjusted p-values.

	R2	PValue	p_adj
Recipient_SHAM - Recipient_SG	0.177	0.002	0.009
Recipient_SHAM - Donor_SG	0.292	0.007	0.011
Recipient_SHAM - Donor_SHAM	0.405	0.003	0.009
Recipient_SG - Donor_SG	0.258	0.011	0.013
Recipient_SG - Donor_SHAM	0.376	0.005	0.010
Donor_SG - Donor_SHAM	0.252	0.220	0.220

6 Summary of R session

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] grid      stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] pracma_2.3.3              forcats_0.5.1
##  [3] stringr_1.4.0             dplyr_1.0.6
##  [5] purrr_0.3.4               readr_1.4.0
##  [7] tidyr_1.1.3               tibble_3.1.1
##  [9] ggplot2_3.3.3             tidyverse_1.3.1
## [11] circlize_0.4.12           seriation_1.2-9
## [13] ComplexHeatmap_2.7.8.1000 reshape2_1.4.4
## [15] scales_1.1.1              RColorBrewer_1.1-2
## [17] emmeans_1.6.0             car_3.0-10
## [19] carData_3.0-4             vegan_2.5-7
## [21] lattice_0.20-44           permute_0.9-5
## [23] phytools_0.7-70           maps_3.3.0
## [25] ape_5.5                   phyloseq_1.34.0
##
## loaded via a namespace (and not attached):
##   [1] readxl_1.3.1            backports_1.2.1         fastmatch_1.1-0
##   [4] plyr_1.8.6              igraph_1.2.6            splines_4.0.3
##   [7] TH.data_1.0-10          digest_0.6.27           foreach_1.5.1
##  [10] htmltools_0.5.1.1       fansi_0.4.2             magrittr_2.0.1
##  [13] cluster_2.1.2           doParallel_1.0.16       openxlsx_4.2.3
##  [16] Biostrings_2.58.0       modelr_0.1.8            matrixStats_0.58.0
##  [19] sandwich_3.0-0          prettyunits_1.1.1       colorspace_2.0-1
##  [22] rvest_1.0.0             haven_2.4.1             xfun_0.22
##  [25] crayon_1.4.1            jsonlite_1.7.2          survival_3.2-11
##  [28] zoo_1.8-9               phangorn_2.7.0          iterators_1.0.13
##  [31] glue_1.4.2              registry_0.5-1          gtable_0.3.0
##  [34] zlibbioc_1.36.0         XVector_0.30.0          GetoptLong_1.0.5
##  [37] Rhdf5lib_1.12.1         shape_1.4.5             BiocGenerics_0.36.1
##  [40] abind_1.4-5             mvtnorm_1.1-1           DBI_1.1.1
##  [43] Rcpp_1.0.6              plotrix_3.8-1           xtable_1.8-4
##  [46] progress_1.2.2          tmvnsim_1.0-2           clue_0.3-59
##  [49] foreign_0.8-81          stats4_4.0.3            httr_1.4.2
##  [52] ellipsis_0.3.2          farver_2.1.0            pkgconfig_2.0.3
##  [55] dbplyr_2.1.1            utf8_1.2.1              labeling_0.4.2
##  [58] tidyselect_1.1.1        rlang_0.4.11            munsell_0.5.0
##  [61] cellranger_1.1.0        tools_4.0.3             cli_2.5.0
```

```
##  [64] generics_0.1.0          ade4_1.7-16              broom_0.7.6
##  [67] evaluate_0.14           biomformat_1.18.0        yaml_2.2.1
##  [70] knitr_1.33              fs_1.5.0                 zip_2.1.1
##  [73] nlme_3.1-152            xml2_1.3.2               rstudioapi_0.13
##  [76] compiler_4.0.3          curl_4.3.1               png_0.1-7
##  [79] reprex_2.0.0            clusterGeneration_1.3.7  stringi_1.6.1
##  [82] Matrix_1.3-3            multtest_2.46.0          vctrs_0.3.8
##  [85] pillar_1.6.0            lifecycle_1.0.0          rhdf5filters_1.2.1
##  [88] combinat_0.0-8          GlobalOptions_0.1.2      estimability_1.3
##  [91] data.table_1.14.0       R6_2.5.0                 TSP_1.1-10
##  [94] rio_0.5.26              IRanges_2.24.1           codetools_0.2-18
##  [97] MASS_7.3-54             gtools_3.8.2             assertthat_0.2.1
## [100] rhdf5_2.34.0            rjson_0.2.20             withr_2.4.2
## [103] mnormt_2.0.2            multcomp_1.4-17          S4Vectors_0.28.1
## [106] mgcv_1.8-35             expm_0.999-6             parallel_4.0.3
## [109] hms_1.0.0               quadprog_1.5-8           coda_0.19-4
## [112] rmarkdown_2.8           Cairo_1.5-12.2           numDeriv_2016.8-1.1
## [115] scatterplot3d_0.3-41    Biobase_2.50.0           lubridate_1.7.10
```