

MBVagTrans Results

Simon Reider

Table of contents

1	Introduction	2
2	PBS empty control	2
3	Diversity Analysis	4
3.1	Alpha-diversity	4
3.1.1	Effect of covariates	5
3.2	Beta-diversity	7
4	Differential abundance	11
5	Taxonomic composition	22
5.1	Phylum level	25
5.2	Class level	26
5.3	Order level	26
5.4	Family level	27
5.5	Genus level	28
6	Secondary analyses	28
6.0.1	Duration of menopause	28
6.0.2	Duration of gender-affirming hormone therapy (GAHT)	34
6.0.3	Absolute Height of Testosterone in the trans-men group	43
6.0.4	Duration of GnRH therapy in trans-men group	44
6.0.5	Cycle dependency in pre-menopausal group	45
6.0.6	Influence of sexual activity on microbiome composition	50
6.0.7	Influence of the duration of amenorrhea on microbiome composition	62
6.1	Influence of sex of partner	68

1 Introduction

Include here metadata summaries and summary of phyloseq object

2 PBS empty control

The PBS empty control seems to be contaminated with normal vaginal flora.

```
taxdistri = read_csv2( "results/microbiome/pbs_taxonomic_composition.csv")
kable(mutate(taxdistri, Abundance = round(Abundance, digits = 2)))
```

Taxonomy	Abundance
p_Firmicutes;f_Lactobacillaceae;g_Lactobacillus	30.57
p_Actinobacteriota;f_Bifidobacteriaceae;g_Bifidobacterium	16.69
p_Actinobacteriota;f_Bifidobacteriaceae;g_Gardnerella	12.40
p_Proteobacteria;f_Moraxellaceae;g_Enhydrobacter	5.80
p_Firmicutes;f_Family XI;g_Anaerococcus	4.72
p_Actinobacteriota;f_Bifidobacteriaceae;g_Alloscardovia	3.61
p_Firmicutes;f_Streptococcaceae;g_Lactococcus	3.41
p_Bacteroidota;f_Prevotellaceae;g_Prevotella	3.37
p_Firmicutes;f_Streptococcaceae;g_Streptococcus	3.05
p_Proteobacteria;f_Moraxellaceae;g_Acinetobacter	2.45
p_Actinobacteriota;f_Atopobiaceae;g_Atopobium	2.16
p_Firmicutes;f_Family XI;g_Finegoldia	1.55
p_Firmicutes;f_Staphylococcaceae;g_Staphylococcus	1.31
p_Actinobacteriota;f_Propionibacteriaceae;g_Cutibacterium	1.17
p_Firmicutes;f_Lachnospiraceae;g_Anaerostipes	1.17
p_Firmicutes;f_Ruminococcaceae;g_Faecalibacterium	1.08
p_Actinobacteriota;f_Corynebacteriaceae;g_Corynebacterium	0.92
p_Actinobacteriota;f_Actinomycetaceae;g_NA	0.69
p_Proteobacteria;f_Xanthobacteraceae;g_Bradyrhizobium	0.66
p_Actinobacteriota;f_Actinomycetaceae;g_Varibaculum	0.59
p_Firmicutes;f_Family XI;g_Ezakiella	0.59
p_Firmicutes;f_Enterococcaceae;g_Enterococcus	0.58
p_Actinobacteriota;f_Corynebacteriaceae;g_NA	0.54
p_Actinobacteriota;f_Pseudonocardiaceae;g_Prauserella	0.54
p_Firmicutes;f_Marinococcaceae;g_Alteribacillus	0.39

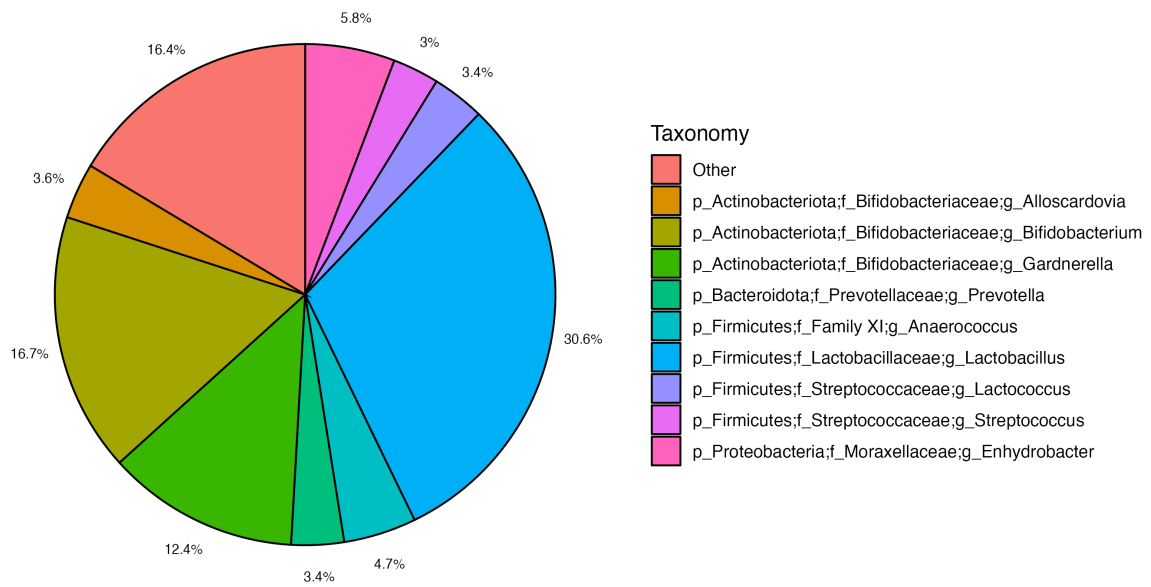


Figure 1: Taxonomic composition of PBS sample

```
ps_samples = readRDS("intermediate/ps_samples.rds")

sample_data(ps_samples)$grp = factor(sample_data(ps_samples)$grp,
                                     labels = c("01_postmenopausal",
                                                "02_premenopausal",
                                                "03_trans-men"))
```

3 Diversity Analysis

3.1 Alpha-diversity

Alpha diversity describes the number and distribution of taxa *within* a sample or group of samples. Different indices exist that capture and describe different aspects of this ecologic measure.

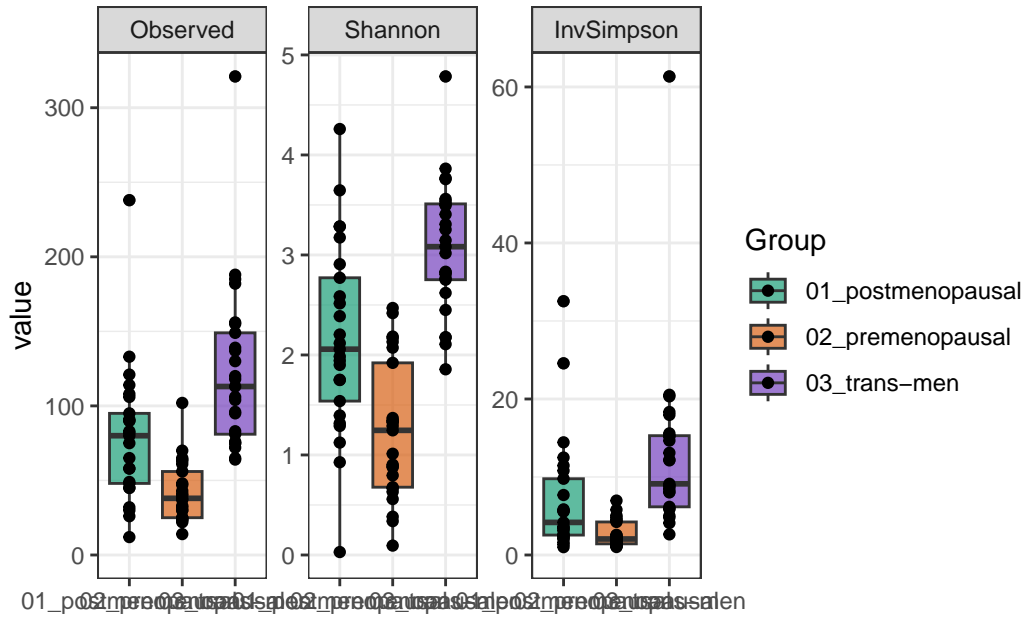
In this dataset, alpha diversity is markedly different between the group of interest (i.e. *03_trans*) and the two control groups. However, the intervention group is more similar to the first control group (i.e. *postmenopausal women*).

```
adiv = plot_richness(ps_samples, x = "grp", measures = c("Observed", "InvSimpson", "Shannon"))

aov = adiv$data %>%
  group_by(variable) %>%
  rstatix::anova_test(value~grp)

tukey = adiv$data %>%
  group_by(variable) %>%
  tukey_hsd(value~grp)

ggplot(adiv$data, aes(x = grp, y = value, fill = grp)) +
  geom_boxplot(alpha = 0.7) +
  geom_point() +
  facet_wrap(~variable, scales = "free_y") +
  labs(x = "", fill = "Group")+
  expand_limits(y=c(0,0,0))+
  scale_fill_manual(values = custom_pal)
```



```
kable(select(tukey, variable, group1,group2,p.adj, p.adj.signif))
```

variable	group1	group2	p.adj	p.adj.signif
Observed	01_postmenopausal	02_premenopausal	1.08e-02	*
Observed	01_postmenopausal	03_trans-men	1.61e-03	**
Observed	02_premenopausal	03_trans-men	0.00e+00	****
Shannon	01_postmenopausal	02_premenopausal	2.35e-04	***
Shannon	01_postmenopausal	03_trans-men	2.88e-04	***
Shannon	02_premenopausal	03_trans-men	0.00e+00	****
InvSimpson	01_postmenopausal	02_premenopausal	1.16e-01	ns
InvSimpson	01_postmenopausal	03_trans-men	3.48e-02	*
InvSimpson	02_premenopausal	03_trans-men	6.14e-05	****

3.1.1 Effect of covariates

To rule out a possible effect of age on alpha diversity, pearson correlation of alpha diversity measures with age are performed within each of the 3 groups separately.

```

adiv$data %>%
  group_by(variable, grp) %>%
  cor_test(value, Age) %>%
  select(grp, variable, cor, statistic, p, conf.low, conf.high) %>%
  kable()

```

grp	variable	cor	statistic	p	conf.low	conf.high
01_postmenopausal	Observed	0.230	1.1510894	0.262	-0.1781717	0.5754530
02_premenopausal	Observed	0.100	0.4925164	0.627	-0.3052951	0.4779954
03_trans-men	Observed	-0.084	-0.4049561	0.689	-0.4638494	0.3216860
01_postmenopausal	Shannon	0.130	0.6463659	0.524	-0.2761356	0.5021954
02_premenopausal	Shannon	-0.076	-0.3639234	0.719	-0.4571294	0.3293109
03_trans-men	Shannon	-0.220	-1.0946590	0.285	-0.5677396	0.1892415
01_postmenopausal	InvSimpson	0.180	0.8633770	0.397	-0.2343598	0.5348641
02_premenopausal	InvSimpson	-0.062	-0.2976233	0.769	-0.4461513	0.3415505
03_trans-men	InvSimpson	-0.110	-0.5522881	0.586	-0.4874977	0.2940179

No comparison is significant.

3.1.1.1 Length of GAHT within patient group

```

adiv$data %>%
  filter(grp == "03_trans-men") %>%
  group_by(variable) %>%
  cor_test(value, GHAT_Length) %>%
  select(variable, cor, statistic, p, conf.low, conf.high) %>%
  kable(digits = 3, caption = "Pearson Correlation of a-diversity measures with length of

```

Table 4: Pearson Correlation of a-diversity measures with length of GAHT in the trans-men patient group

variable	cor	statistic	p	conf.low	conf.high
Observed	-0.15	-0.705	0.488	-0.511	0.265
Shannon	-0.40	-2.102	0.047	-0.688	-0.007
InvSimpson	-0.17	-0.844	0.407	-0.532	0.238

In this analysis, a significant effect on Shannon index with length of hormone replacement therapy could be observed.

3.1.1.2 Influence of Nugent score within patient group

```
adiv$data %>%
  group_by(variable) %>%
  tukey_hsd(value~NugentScore) %>%
  select(variable, group1, group2, estimate, p.adj, p.adj.signif, conf.low, conf.high) %>%
  kable(digits = 3, caption = "Tukey Test on alpha diversity measures by Nugent Score")
```

Table 5: Tukey Test on alpha diversity measures by Nugent Score

variable	group1	group2	estimate	p.adj	p.adj.signif	conf.low	conf.high
Observed	0-3	4-6	26.648	0.342	ns	-15.098	68.393
Observed	0-3	7-10	-18.067	0.929	ns	-96.166	60.033
Observed	0-3	no bacteria	54.183	0.000	***	20.292	88.074
Observed	4-6	7-10	-44.714	0.483	ns	-126.770	37.342
Observed	4-6	no bacteria	27.536	0.323	ns	-14.682	69.753
Observed	7-10	no bacteria	72.250	0.081	ns	-6.102	150.602
Shannon	0-3	4-6	1.167	0.001	***	0.394	1.939
Shannon	0-3	7-10	0.152	0.993	ns	-1.293	1.597
Shannon	0-3	no bacteria	1.250	0.000	****	0.623	1.877
Shannon	4-6	7-10	-1.015	0.302	ns	-2.533	0.504
Shannon	4-6	no bacteria	0.083	0.992	ns	-0.698	0.865
Shannon	7-10	no bacteria	1.098	0.200	ns	-0.352	2.548
InvSimpson	0-3	4-6	5.681	0.167	ns	-1.480	12.842
InvSimpson	0-3	7-10	-0.403	1.000	ns	-13.800	12.994
InvSimpson	0-3	no bacteria	7.012	0.012	*	1.198	12.825
InvSimpson	4-6	7-10	-6.084	0.668	ns	-20.160	7.991
InvSimpson	4-6	no bacteria	1.330	0.963	ns	-5.911	8.572
InvSimpson	7-10	no bacteria	7.415	0.472	ns	-6.025	20.855

Notably, if analysis is split by patient group, none of these differences remain significant. This might be due to a lower sample size.

3.2 Beta-diversity

This diversity metric compares different samples or groups of samples regarding their similarity or difference in community composition. The distance matrix can be visualized using dimension reduction techniques (e.g. PCA) or pairwise distances can be calculated. For this analysis, the main hypothesis is, that overall community composition differs significantly between the patient group (i.e. *03_trans*) and the two control groups. This will be tested using *adonis*.

```
# first, filter out low abundance taxa
lowabund_filter = genefilter_sample(ps_samples, filterfun_sample(function(x) x > 10), A=0.
ps_filtered = prune_taxa(lowabund_filter, ps_samples)
ps_filtered
```

phyloseq-class experiment-level object

```
otu_table() OTU Table: [ 106 taxa and 75 samples ]
sample_data() Sample Data: [ 75 samples by 15 sample variables ]
tax_table() Taxonomy Table: [ 106 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 106 tips and 105 internal nodes ]
refseq() DNASTringSet: [ 106 reference sequences ]
```

```
ps_filtered_rel = transform_sample_counts(ps_filtered, function(x) 100 * x/sum(x))
```

```
ord.bray<- ordinate(ps_filtered_rel, "PCoA", "bray")
ord.wunifrac <- ordinate(ps_filtered_rel, "PCoA", "unifrac", weighted = T)
ord.unifrac <- ordinate(ps_filtered_rel, "PCoA", "unifrac", weighted = F)
```

```
p1 = plot_ordination(ps_filtered_rel, ord.bray, type="samples", color="grp") +
  geom_point(size = 3, alpha = 0.8) +
  scale_color_brewer(type = "qual",palette = "Dark2") +
  stat_ellipse() +
  labs(color = "Group", caption = "Bray-Curtis")
```

```
p2 = plot_ordination(ps_filtered_rel, ord.wunifrac, type="samples", color="grp") +
  geom_point(size = 3, alpha = 0.8) +
  scale_color_brewer(type = "qual",palette = "Dark2") +
  stat_ellipse() +
  labs(color = "Group", caption = "Weighted Unifrac")
```

```
p3 = plot_ordination(ps_filtered_rel, ord.unifrac, type="samples", color="grp") +
  geom_point(size = 3, alpha = 0.8) +
  scale_color_brewer(type = "qual",palette = "Dark2")+
  stat_ellipse() +
  labs(color = "Group", caption = "Unweighted Unifrac")
```

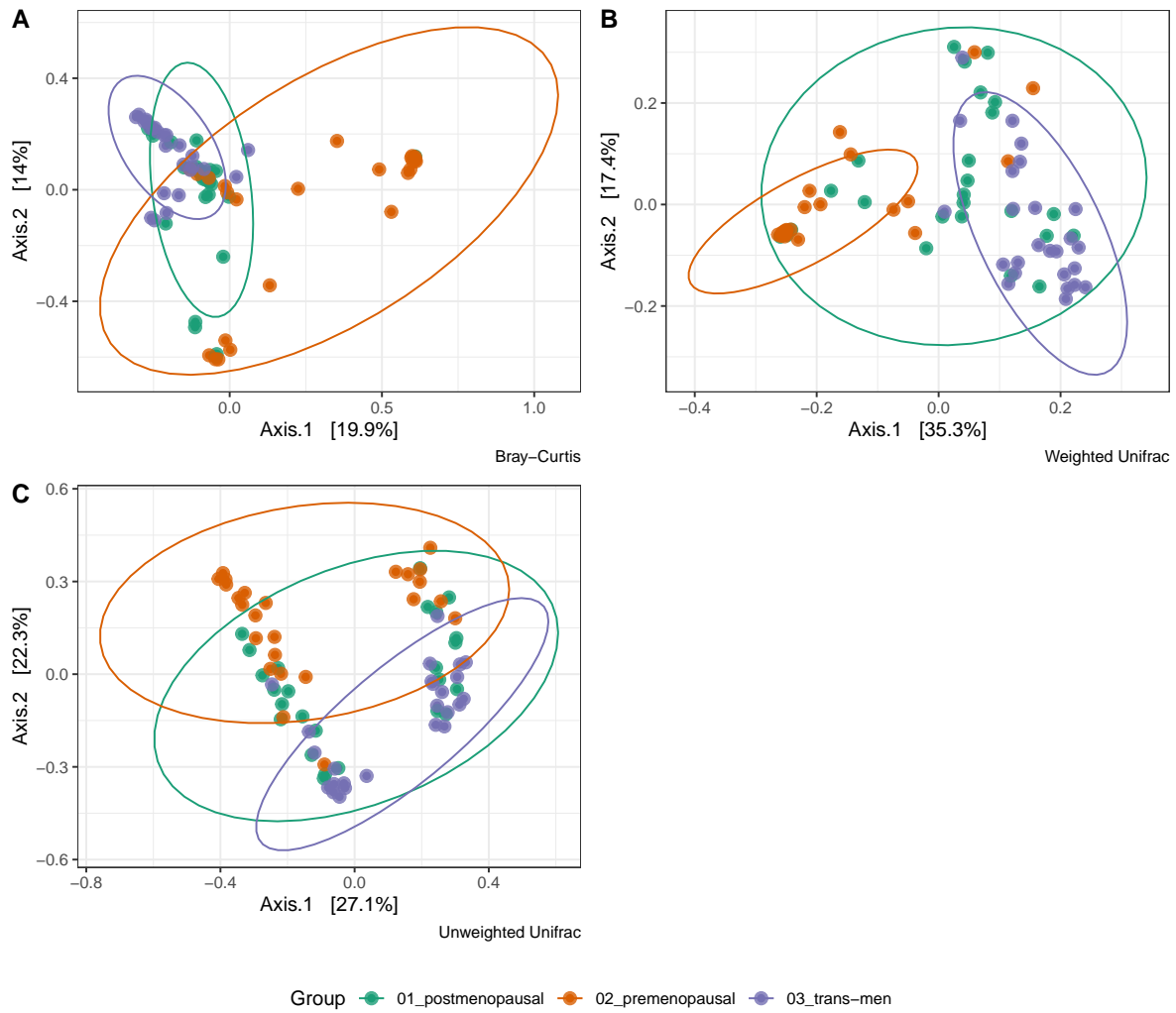
```
# extract the legend from one of the plots
legend_b <- get_legend(
  p1 +
  guides(color = guide_legend(nrow = 1)) +
  theme(legend.position = "bottom")
```



```
)
```

```
plot_ordinations = cowplot::plot_grid(p1 + theme(legend.position="none"),
  p2 + theme(legend.position="none"),
  p3 + theme(legend.position="none"),
  labels = "AUTO", ncol = 2)
```

```
cowplot::plot_grid(plot_ordinations, legend_b, ncol = 1, rel_heights = c(1, .1))
```



```

D_BC <- phyloseq::distance(ps_filtered_rel, "bray")
D_UF <- phyloseq::distance(ps_filtered_rel, "unifrac")
D_wUF <- phyloseq::distance(ps_filtered_rel, "wunifrac")

adonis_bc = vegan::adonis2(D_BC ~ phyloseq::sample_data(ps_filtered_rel)$grp)
adonis_pw_bc = pairwise.adonis(D_BC, phyloseq::sample_data(ps_filtered_rel)$grp)

adonis_uf = vegan::adonis2(D_UF ~ phyloseq::sample_data(ps_filtered_rel)$grp)
adonis_pw_uf = pairwise.adonis(D_UF, phyloseq::sample_data(ps_filtered_rel)$grp)

adonis_wUF = vegan::adonis2(D_wUF ~ phyloseq::sample_data(ps_filtered_rel)$grp)
adonis_pw_wUF = pairwise.adonis(D_wUF, phyloseq::sample_data(ps_filtered_rel)$grp)

kable(adonis_pw_bc, digits = 3, caption = "Bray-Curtis")

```

Table 6: Bray-Curtis

pairs	Df	SumsOfSqsF	Model	R2	p.value	p.adjusted	sig
01_postmenopausal vs 02_premenopausal	1	1.839	4.972	0.094	0.001	0.003	*
01_postmenopausal vs 03_trans-men	1	1.255	3.474	0.067	0.001	0.003	*
02_premenopausal vs 03_trans-men	1	3.351	10.303	0.177	0.001	0.003	*

```

kable(adonis_pw_uf, digits = 3, caption = "Unweigthed Unifrac")

```

Table 7: Unweigthed Unifrac

pairs	Df	SumsOfSqsF	Model	R2	p.value	p.adjusted	sig
01_postmenopausal vs 02_premenopausal	1	1.104	5.145	0.097	0.001	0.003	*
01_postmenopausal vs 03_trans-men	1	0.704	4.133	0.079	0.002	0.006	*
02_premenopausal vs 03_trans-men	1	2.263	12.762	0.210	0.001	0.003	*

```

kable(adonis_pw_wUF, digits = 3, caption = "Weighted Unifrac")

```

Table 8: Weighted Unifrac

pairs	Df	SumsOfSqsF	Model	R2	p.value	p.adjusted	sig
01_postmenopausal vs 02_premenopausal	1	0.559	8.149	0.145	0.001	0.003	*
01_postmenopausal vs 03_trans-men	1	0.413	5.341	0.100	0.001	0.003	*
02_premenopausal vs 03_trans-men	1	1.369	23.593	0.330	0.001	0.003	*

4 Differential abundance

```
library(DESeq2)

dds = phyloseq_to_deseq2(ps_filtered, ~ grp)
dds$grp<-relevel(dds$grp,ref="03_trans-men")
ds <- estimateSizeFactors(dds, type="poscounts")
ds = DESeq(ds, test="Wald", fitType="parametric")
alpha = 0.01
res_premenopausal = results(ds, contrast = c("grp", "02_premenopausal", "03_trans-men"))
taxa_sig_premenopausal = dplyr::filter(as.data.frame(res_premenopausal), padj < 0.05) %>%
  rownames_to_column("ASV")

taxa_table = tax_table(ps_filtered) %>%
  as.data.frame() %>%
  rownames_to_column("ASV")

diffabund_pre = left_join(taxa_sig_premenopausal, taxa_table, by = "ASV") %>%
  select(-Species) %>%
  write_csv2("results/microbiome/deseq_diff_trans-premenopausal.csv")

kable(diffabund_pre, digits = 3, row.names = F, caption = "Differential abundance between
```

Table 9: Differential abundance between premenopausal women and trans-men

ASV	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	Kingdom	Phylum	Class	Order	Family	Genus
ASV1836.79120	1.386	1.370	0.000	0.000	0.000	0.000	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
ASV2474.42394	1.273	1.925	0.000	0.000	0.000	0.000	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus

ASV	baseMean	2FoldEnrichment	log2FC	negLog10Padj	Kingdom	Phylum	Class	Order	Family	Genus
ASV196.709	-	0.873	-	0.0000.00	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella
3.374		3.867								
ASV23.859	-	2.258	-	0.0130.02	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
5.594		2.478								
ASV178.493	-	1.310	-	0.0000.00	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
ASV97.075	-	1.257	-	0.0000.00	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
5.076		4.037								
ASV145.660	-	1.028	-	0.0000.00	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella
3.843		3.739								
ASV182.749	-	1.052	-	0.0130.02	Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Atopobiaceae	Atopobium
ASV123.599	-	0.716	-	0.0210.04	Bacteria	Firmicutes	Clostridia	Peptostreptococcales	Family	Finegoldia
1.658		2.314						Tissierellales	XI	
ASV17.479	-	1.752	-	0.0160.03	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia-Shigella
ASV65.480	-	1.324	-	0.0020.00	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
4.200		3.172								
ASV161.265	-	1.883	-	0.0010.00	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
ASV80.800	-	1.387	-	0.0040.01	Bacteria	Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Dialister
3.940		2.841						Selenomonadales		
ASV90.251	-	1.783	-	0.0210.04	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
4.108		2.304								
ASV20.863	-	1.255	-	0.0000.00	Bacteria	Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Alloscaclovicia
4.779		3.808								
ASV24.563	-	1.150	-	0.0000.00	Bacteria	Firmicutes	Clostridia	Peptostreptococcales	Family	Fenollaria
6.795		5.909						Tissierellales	XI	
ASV22.469	-	0.675	-	0.0000.00	Bacteria	Firmicutes	Clostridia	Peptostreptococcales	Family	Peptoniphilus
2.718		4.030						Tissierellales	XI	
ASV24.770	-	0.916	-	0.0000.00	Bacteria	Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Nae
3.289		3.590						Selenomonadales		
ASV23.379	-	1.647	-	0.0100.02	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella
4.231		2.569								
ASV46.531	-	1.161	-	0.0040.01	Bacteria	Firmicutes	Clostridia	Peptostreptococcales	Family	Anaerococcus
3.379		2.909						Tissierellales	XI	
ASV30.807	-	1.703	-	0.0240.04	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella
3.842		2.256								
ASV44.188	-	1.023	-	0.0010.00	Bacteria	Campylobacter	Campylobacter	Campylobacter	Campylobacter	Campylobacter
3.255		3.182								
ASV32.757	-	1.086	-	0.0000.00	Bacteria	Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Dialister
5.942		5.472						Selenomonadales		
ASV32.429	-	1.332	-	0.0000.00	Bacteria	Firmicutes	Clostridia	Peptostreptococcales	Family	Anaerococcus
7.577		5.686						Tissierellales	XI	

ASV	base	Meg	2	Fold	CS	Stat	p	val	padj	King	Phylum	Class	Order	Family	Genus
ASV36	919	-	1.452	-	0.00	10.00	Bacte	Bacteroid	Bacteroid	Bacteroidales	Prevotellaceae	Prevotella			
	4.847		3.339												
ASV35	382	-	1.501	-	0.00	0.00	Bacte	Firmicute	Clostridia	Peptostreptococcales	Family	Anaerococcus			
	6.236		4.154										Tissierellales	XI	
ASV32	384	-	1.161	-	0.00	0.00	Bacte	Firmicute	Clostridia	Peptostreptococcales	Family	Anaerococcus			
	4.136		3.564										Tissierellales	XI	
ASV48	977	-	1.463	-	0.00	0.00	Bacte	Actinobacteria	Actinomycetales	Actinomycetaceae	Actinomyces				
	24.332		16.635												
ASV25	838	-	1.108	-	0.00	0.00	Bacte	Firmicute	Clostridia	Peptostreptococcales	Family	Ezakiella			
	5.508		4.972										Tissierellales	XI	
ASV54	315	-	1.934	-	0.00	70.01	Bacte	Bacteroid	Bacteroid	Bacteroidales	Prevotellaceae	Prevotella			
	5.179		2.677												
ASV50	308	4.488	1.696	4.415	0.00	0.00	Bacte	Firmicute	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus			
ASV27	517	-	1.854	-	0.00	0.00	Bacte	Firmicute	Clostridia	Peptostreptococcales	Family	Parvimonas			
	24.675		13.312										Tissierellales	XI	
ASV62	716	-	1.386	-	0.00	80.01	Bacte	Firmicute	Clostridia	Peptostreptococcales	Family	Peptoniphilus			
	3.660		2.641										Tissierellales	XI	
ASV64	221	-	2.204	-	0.00	40.01	Bacte	Firmicute	Negativicutes	Veillonellales	Veillonellaceae	Dialister			
	6.351		2.882										Selenomonadales		
ASV81	391	-	0.983	-	0.00	0.00	Bacte	Firmicute	Clostridia	Lachnospirales	Lachnospiraceae	Howardella			
	4.030		4.099												
ASV23	772	-	1.174	-	0.00	0.00	Bacte	Firmicute	Clostridia	Peptostreptococcales	Family	Peptostreptococcus			
	4.249		3.619										Tissierellales		
ASV34	931	-	1.888	-	0.00	70.01	Bacte	Bacteroid	Bacteroid	Bacteroidales	Prevotellaceae	Prevotella			
	5.092		2.697												
ASV70	488	-	1.847	-	0.00	0.00	Bacte	Bacteroid	Bacteroid	Bacteroidales	Porphyromonadaceae	Porphyromonas			
	6.613		3.580												
ASV56	695	7.955	1.627	4.890	0.00	0.00	Bacte	Firmicute	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus			
ASV88	403	7.726	2.963	2.945	0.00	30.00	Bacte	Firmicute	Bacilli	Mycoplasmatales	Mycoplasmataceae	Mycoplasma			
ASV80	425	-	2.011	-	0.00	80.01	Bacte	Firmicute	Negativicutes	Veillonellales	Veillonellaceae	Dialister			
	5.328		2.649										Selenomonadales		
ASV94	058	-	2.141	-	0.00	0.00	Bacte	Fusobacteria	Fusobacteriales	Fusobacteriaceae	Fusobacterium				
	23.131		10.806												
ASV96	234	-	1.749	-	0.02	00.03	Bacte	Bacteroid	Bacteroid	Bacteroidales	Prevotellaceae	Prevotella			
	4.081		2.333												
ASV97	375	-	2.031	-	0.00	30.00	Bacte	Firmicute	Clostridia	Lachnospirales	Lachnospiraceae	Moryella			
	6.026		2.968												
ASV200	126	6.918	2.032	3.404	0.00	10.00	Bacte	Fusobacteria	Fusobacteriales	Leptotrichaceae	Sarcophaga				
ASV107	60	-	1.790	-	0.00	10.00	Bacte	Firmicute	Negativicutes	Veillonellales	Veillonellaceae	Dialister			
	5.947		3.322										Selenomonadales		

ASV	baseMean	log2FoldChange	lfcSE	negLog10Pvalue	adjPvalue	Kingdom	Phylum	Class	Order	Family	Genus
ASV102223	-	2.185	-	0.0000.000	0.000	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
	24.611		11.264								
ASV10389	-	1.108	-	0.0000.000	0.000	Bacteria	Firmicutes	Clostridia	Peptostreptococcales	Family	Peptoniphilus
	5.048		4.557						Tissierellales	XI	
ASV106386	-	1.796	-	0.0050.012	0.012	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
	5.077		2.827								
ASV183903	-	1.100	-	1.7581.038	0.0000.000	Bacteria	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella
ASV22514	-	1.384	-	1.673.809	0.0000.000	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	HT002
ASV62855	-	1.449	-	0.0000.000	0.000	Bacteria	Actinobacteria	Coriobacteriales	Coriobacteriales	Atopobiaceae	Atopobium
	5.491		3.789								
ASV83747	-	1.848	-	0.0040.01	0.01	Bacteria	Fusobacteriota	Fusobacteriota	Fusobacteriales	Fusobacteriaceae	Fusobacterium
	5.309		2.873								
ASV53748	-	2.268	-	0.0240.045	0.045	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
	5.111		2.254								
ASV45109	-	1.309	-	0.0020.006	0.006	Bacteria	Proteobacteria	Gammaproteobacteria	Burkholderiales	Sutterellaceae	Sutterella
	4.040		3.087								
ASV16334	-	1.919	-	0.0120.027	0.027	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
	4.802		2.503								
ASV4333	-	1.853	-	0.0090.020	0.020	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
	4.835		2.609								

```

# repeat analysis for postmenopausal group
res_postmenopausal = results(ds, contrast = c("grp", "01_postmenopausal", "03_trans-men"))
taxa_sig_postmenopausal = dplyr::filter(as.data.frame(res_postmenopausal), padj < 0.05) %>%
  rownames_to_column("ASV")

diffabund_post = left_join(taxa_sig_postmenopausal, taxa_table, by = "ASV") %>%
  write_csv2("results/microbiome/deseq_diff_trans-postmenopausal.csv")

kable(diffabund_post, digits = 3, row.names = F, caption = "Differential abundance between

```

Table 10: Differential abundance between postmenopausal women and trans-men

ASV	baseMean	log2FoldChange	lfcSE	negLog10Pvalue	adjPvalue	Kingdom	Phylum	Class	Order	Family	Genus	Species
ASV1836	79158	1.386	1.165	0.0000.000	0.000	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
ASV2474	42901	1.272	1.066	0.0020.021	0.021	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
ASV396	709	-	0.873	-	0.0000.000	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotellaceae	Prevotellaceae
	3.282		3.760									

ASV	baseMean	log2FoldChange	lfcSE	negLog10Pvalue	adjPvalue	Kingdom	Phylum	Class	Order	Family	Genus	Species
ASV33	3.859	-	2.269	-	0.0000009	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	NA
	7.935		3.496									
ASV182	7.29839	1.052	2.698	0.007	0.04	Bacteria	Actinobacteria	Coriobacteriales	Coriobacteriales	Actinobacteria	Actinobium	maginae
ASV13	4.75899	1.756	3.359	0.001	0.01	Bacteria	Proteobacteria	Gammaaproteobacteria	Flavobacteriales	Flavobacteriaceae	Escherichia	NA
											Shigella	
ASV14	4.671	-	1.430	-	0.003	0.03	Bacteria	Actinobacteria	Actinobacteriales	Bifidobacteriales	Genus	neoclaginalis
	4.193		2.933									
ASV161	1.265	-	1.927	-	0.000000	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	senii
	22.787		11.827									
ASV80	1.800	-	1.387	-	0.005	0.03	Bacteria	Firmicutes	Negativicoccus	Veillonellales	Veillonellaceae	Dialister NA
	3.860		2.783						Selenomonadales			
ASV33	3.853	3.320	1.994	4.173	0.000000	Bacteria	Actinobacteria	Actinobacteriales	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	
ASV93	3.33	-	1.618	-	0.005	0.03	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
	4.505		2.784									
ASV85	4.494	4.419	2.123	3.484	0.000000	Bacteria	Actinobacteria	Actinobacteriales	Corynebacteriales	Corynebacteriaceae	Corynebacterium	NA
ASV99	4.06	-	1.891	-	0.002	0.02	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella corporis
	5.846		3.091									
ASV45	1.09	-	1.301	-	0.002	0.02	Bacteria	Proteobacteria	Gammaaproteobacteria	Burkholderiales	Sutterella	NA
	4.043		3.107									
ASV45	1.33	-	1.855	-	0.002	0.02	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas
	5.755		3.103									
ASV22	2.26	-	1.645	-	0.005	0.03	Bacteria	Actinobacteria	Coriobacteriales	Coriobacteriales	Eggerthella	NA
	4.593		2.793									

```

asvs = unique(c(diffabund_pre$ASV, diffabund_post$ASV)) # get asvs identified by Deseq2

# next, melt the phyloseq object into a data.frame and filter OTUs for significant ASVs as
ps_sig_melt = psmelt(ps_filtered_rel) %>%
  filter(OTU %in% asvs)

# Summarize Abundance at the genus level for plotting, almost no species are assigned in t
ps_sig_melt2 = ps_sig_melt %>%
  group_by(Sample, Genus, grp) %>%
  summarize(Abundance = sum(Abundance))

# Plot the final figure panel
deseq_res_fig = ggplot(ps_sig_melt2, aes(x = grp, y=Abundance)) +
  geom_jitter(width = 0.3, alpha = 0.6) +
  geom_boxplot(aes(fill = grp), alpha = 0.5, outlier.shape = NA) +

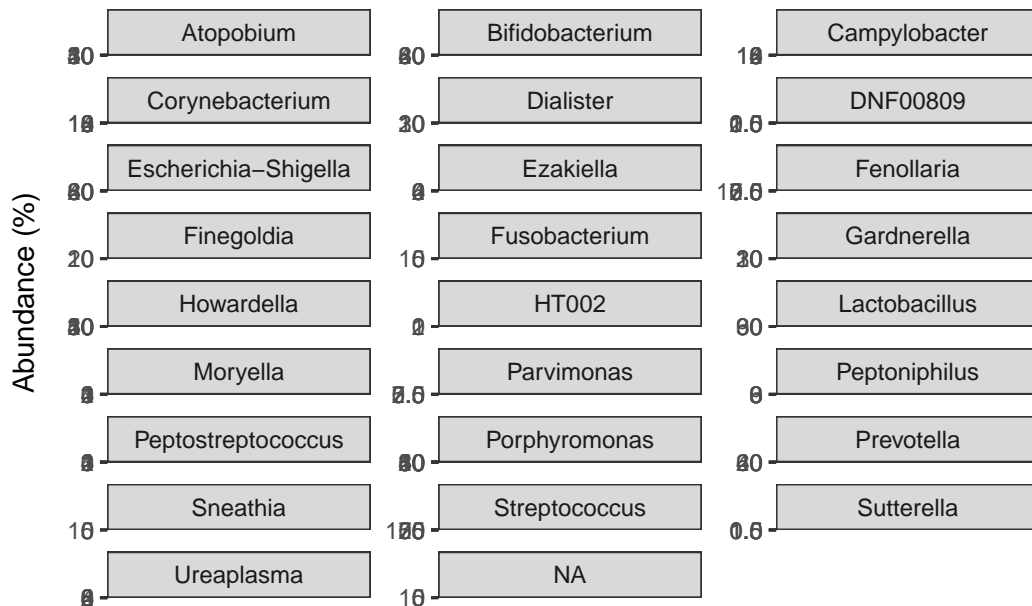
```

```

facet_wrap(~Genus, scales = "free", ncol = 3) +
scale_y_sqrt() +
theme(axis.text.x=element_blank(),
      axis.ticks.x=element_blank()) +
labs(x = "", fill = "Group", y= "Abundance (%)") +
theme(legend.position = "bottom") +
scale_y_continuous(expand = expansion(mult = c(0, 0.2))) +
scale_fill_manual(values = custom_pal)

```

deseq_res_fig



```

ggsave(filename = "MbVagTrans_Results_files/figure-html/deseq_taxa_final.pdf", plot = dese
        device = "pdf", width = 8, height = 8)

```

```

# summary statistics
summarystat1 = ps_sig_melt2 %>%
  group_by(Genus, grp) %>%
  summarize(mean = mean(Abundance, na.rm=T),
            median = median(Abundance, na.rm =T),
            sd = sd(Abundance, na.rm = T),
            min = min(Abundance, na.rm =T),

```



```

max = max(Abundance, na.rm =T))

write_csv2(x = summarystat1, "results/microbiome/deseq_summarystat_complete.csv")

summarystat2 = ps_sig_melt2 %>%
  group_by(Genus, grp) %>%
  summarize(mean = mean(Abundance, na.rm=T) %>% round(digits = 2),
            sd = sd(Abundance, na.rm = T) %>% round(digits = 2)) %>%
  mutate(stat = paste(mean, "±", sd)) %>%
  select(Genus, grp, stat) %>%
  pivot_wider(id_cols = "Genus", names_from = grp, values_from = stat)

write_csv2(summarystat2, "results/microbiome/deseq_summarystat_complete.csv")

kable(summarystat1)

```

Genus	grp	mean	median	sd	min	max
Actinomyces	01_postmenopausal	0.5713867	0.0000000	1.8727358	0.0000000	9.0728975
Actinomyces	02_premenopausal	1.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Actinomyces	03_trans-men	0.6649058	0.0527025	1.3924048	0.0000000	6.0791468
Alloscardovia	01_postmenopausal	3.9311642	0.2332090	14.0753880	0.0000000	70.4386523
Alloscardovia	02_premenopausal	1.0087547	0.0000000	0.0392597	0.0000000	0.1959104
Alloscardovia	03_trans-men	1.4267694	0.0322511	5.0906189	0.0000000	25.5642757
Anaerococcus	01_postmenopausal	3.0717887	0.5136749	5.5732548	0.0000000	19.8298621
Anaerococcus	02_premenopausal	1.0932654	0.0000000	0.2912320	0.0000000	1.1904090
Anaerococcus	03_trans-men	4.0605333	2.9932469	3.3448785	0.0000000	10.9066824
Atopobium	01_postmenopausal	4.5164924	0.2277015	10.4160035	0.0000000	46.9083969
Atopobium	02_premenopausal	2.3114743	0.0596659	6.5656355	0.0000000	28.4496827
Atopobium	03_trans-men	0.9158061	0.2469136	1.1828644	0.0000000	4.3279139
Bifidobacterium	01_postmenopausal	4.0593030	0.0000000	14.9218728	0.0000000	73.1856134
Bifidobacterium	02_premenopausal	1.4404757	0.0000000	2.1949415	0.0000000	10.9761388
Bifidobacterium	03_trans-men	0.0227102	0.0000000	0.1029739	0.0000000	0.5143203
Campylobacter	01_postmenopausal	6.0349138	0.0000000	1.8448353	0.0000000	6.7958783
Campylobacter	02_premenopausal	1.0562029	0.0000000	0.1934662	0.0000000	0.9248210
Campylobacter	03_trans-men	1.4788043	0.4923414	2.8937352	0.0000000	14.3656167
Corynebacterium	01_postmenopausal	6.8283154	0.0000000	2.8155250	0.0000000	13.6609746
Corynebacterium	02_premenopausal	1.0192636	0.0000000	0.0873295	0.0000000	0.4361759
Corynebacterium	03_trans-men	0.0141055	0.0000000	0.0598928	0.0000000	0.2965044
DNF00809	01_postmenopausal	0.0000000	0.0000000	0.0000000	0.0000000	0.0000000
DNF00809	02_premenopausal	1.0720117	0.0152800	0.2094370	0.0000000	1.0476512
DNF00809	03_trans-men	0.1515091	0.0000000	0.4240207	0.0000000	1.8319212

Genus	grp	mean	median	sd	min	max
Dialister	01_postmenopausal	1.172677	0.1631588	2.1203827	0.0000000	8.7569963
Dialister	02_premenopausal	1.1467705	0.0000000	0.4491541	0.0000000	1.7908018
Dialister	03_trans-men	6.0987898	3.5013155	6.3277216	0.0408521	25.6857604
Escherichia-Shigella	01_postmenopausal	14.422546	0.0000000	14.4460327	0.0000000	71.3925830
Escherichia-Shigella	02_premenopausal	1.5409810	0.0000000	2.0390260	0.0000000	9.9262483
Escherichia-Shigella	03_trans-men	0.0640695	0.0000000	0.1839214	0.0000000	0.8926876
Ezakiella	01_postmenopausal	1.6769803	0.1195764	1.4031535	0.0000000	6.3237888
Ezakiella	02_premenopausal	1.0060341	0.0000000	0.0174443	0.0000000	0.0610501
Ezakiella	03_trans-men	0.8736686	0.2902602	1.2425205	0.0000000	5.0328228
Fenollaria	01_postmenopausal	1.5539208	0.0000000	1.7774051	0.0000000	8.6307504
Fenollaria	02_premenopausal	1.0022957	0.0000000	0.0114784	0.0000000	0.0573921
Fenollaria	03_trans-men	2.2501475	1.3760482	2.5185955	0.0000000	7.2748092
Finegoldia	01_postmenopausal	3.9604420	0.8349390	6.1259312	0.0000000	19.7480620
Finegoldia	02_premenopausal	1.3455590	0.0619963	0.7418217	0.0000000	3.0926810
Finegoldia	03_trans-men	2.2912198	0.9965338	4.7796209	0.0296670	23.8936372
Fusobacterium	01_postmenopausal	1.1976509	0.0000000	1.7826115	0.0000000	8.6828982
Fusobacterium	02_premenopausal	1.0015992	0.0000000	0.0079960	0.0000000	0.0399800
Fusobacterium	03_trans-men	0.9045715	0.0000000	2.4641849	0.0000000	11.9206638
Gardnerella	01_postmenopausal	1.8933718	0.0000000	1.3913979	0.0000000	6.9161111
Gardnerella	02_premenopausal	1.1539334	0.2571324	7.7565603	0.0000000	29.5627645
Gardnerella	03_trans-men	2.3591859	0.0000000	5.5157218	0.0000000	19.5534425
HT002	01_postmenopausal	1.0753210	0.0000000	0.1993038	0.0000000	0.6797486
HT002	02_premenopausal	1.3367638	0.0810235	0.5810368	0.0000000	2.2941247
HT002	03_trans-men	0.0082445	0.0000000	0.0412223	0.0000000	0.2061117
Howardella	01_postmenopausal	2.6314285	0.0000000	10.1000570	0.0000000	49.5850622
Howardella	02_premenopausal	1.0089003	0.0000000	0.0270160	0.0000000	0.1010995
Howardella	03_trans-men	0.4572855	0.2224469	1.0960540	0.0000000	5.6142399
Lactobacillus	01_postmenopausal	21.2775658	6.4315353	34.4814664	0.1284522	99.8202570
Lactobacillus	02_premenopausal	11.6566839	87.7634476	35.7214691	0.5596256	99.5759004
Lactobacillus	03_trans-men	5.9994574	0.2801284	8.9741717	0.0000000	25.9354688
Moryella	01_postmenopausal	1.0215991	0.0000000	0.1079956	0.0000000	0.5399779
Moryella	02_premenopausal	1.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Moryella	03_trans-men	0.4888759	0.0000000	1.0831026	0.0000000	3.8850376
Parvimonas	01_postmenopausal	1.5600192	0.0000000	1.8712302	0.0000000	7.5682018
Parvimonas	02_premenopausal	1.0000000	0.0000000	0.0000000	0.0000000	0.0000000
Parvimonas	03_trans-men	0.7550410	0.1611789	1.4094233	0.0000000	5.5686949
Peptoniphilus	01_postmenopausal	1.7412584	0.3496503	2.7391621	0.0000000	9.2549871
Peptoniphilus	02_premenopausal	1.1094172	0.0337952	0.2215872	0.0000000	0.8850438

Genus	grp	mean	median	sd	min	max
Peptoniphilus	03_trans-men	2.7222443	1.8739690	2.4326037	0.1453488	9.3331711
Peptostreptococcus	01_postmenopausal	0.3240192	0.0492692	0.6665328	0.0000000	2.8736820
Peptostreptococcus	02_premenopausal	0.0224869	0.0000000	0.0885532	0.0000000	0.4375497
Peptostreptococcus	03_trans-men	0.5223149	0.1598837	0.8530107	0.0000000	3.3850214
Porphyromonas	01_postmenopausal	1.2860890	0.0000000	3.3962705	0.0000000	14.8429833
Porphyromonas	02_premenopausal	0.0295355	0.0000000	0.1029816	0.0000000	0.4985755
Porphyromonas	03_trans-men	5.6496421	2.8848984	10.7930069	0.0000000	50.5763397
Prevotella	01_postmenopausal	5.7985547	1.1067842	9.7530935	0.0000000	32.6725746
Prevotella	02_premenopausal	1.1578882	0.2157130	2.8240835	0.0000000	12.5887124
Prevotella	03_trans-men	24.3020650	24.2634315	18.5127750	0.3851765	56.5387016
Sneathia	01_postmenopausal	1.409774	0.0000000	0.6993965	0.0000000	3.4979846
Sneathia	02_premenopausal	1.6402053	0.0000000	2.4028989	0.0000000	11.9972964
Sneathia	03_trans-men	0.0069812	0.0000000	0.0349061	0.0000000	0.1745306
Streptococcus	01_postmenopausal	1.1904312	0.5155545	17.8955941	0.0000000	60.3249549
Streptococcus	02_premenopausal	1.9064624	0.0000000	19.6198466	0.0000000	86.8697860
Streptococcus	03_trans-men	9.6731738	2.0525784	18.5258538	0.0000000	79.4981033
Sutterella	01_postmenopausal	0.0050866	0.0000000	0.0202921	0.0000000	0.0985383
Sutterella	02_premenopausal	0.0027508	0.0000000	0.0083575	0.0000000	0.0340599
Sutterella	03_trans-men	0.1940975	0.0725191	0.3011899	0.0000000	1.2505042
Ureaplasma	01_postmenopausal	0.00654674	0.0000000	0.2859913	0.0000000	1.4312977
Ureaplasma	02_premenopausal	1.5672609	0.0000000	1.7398576	0.0000000	7.2670503
Ureaplasma	03_trans-men	0.0189886	0.0000000	0.0757508	0.0000000	0.3678443
NA	01_postmenopausal	1.8703149	0.2099702	1.9306476	0.0000000	9.2816275
NA	02_premenopausal	1.5669992	0.0000000	1.8913132	0.0000000	8.8465777
NA	03_trans-men	1.5161479	0.8493428	2.3679024	0.0000000	11.9352562

`kable(summarystat2)`

Genus	01_postmenopausal	02_premenopausal	03_trans-men
Actinomyces	0.57 ± 1.87	0 ± 0	0.66 ± 1.39
Alloscardovia	3.93 ± 14.08	0.01 ± 0.04	1.43 ± 5.09
Anaerococcus	3.07 ± 5.57	0.09 ± 0.29	4.06 ± 3.34
Atopobium	4.52 ± 10.42	2.31 ± 6.57	0.92 ± 1.18
Bifidobacterium	4.06 ± 14.92	0.44 ± 2.19	0.02 ± 0.1
Campylobacter	0.93 ± 1.84	0.06 ± 0.19	1.48 ± 2.89
Corynebacterium	0.83 ± 2.82	0.02 ± 0.09	0.01 ± 0.06
DNF00809	0 ± 0	0.07 ± 0.21	0.15 ± 0.42
Dialister	1.12 ± 2.12	0.15 ± 0.45	6.1 ± 6.33
Escherichia-Shigella	4.44 ± 14.45	0.54 ± 2.04	0.06 ± 0.18

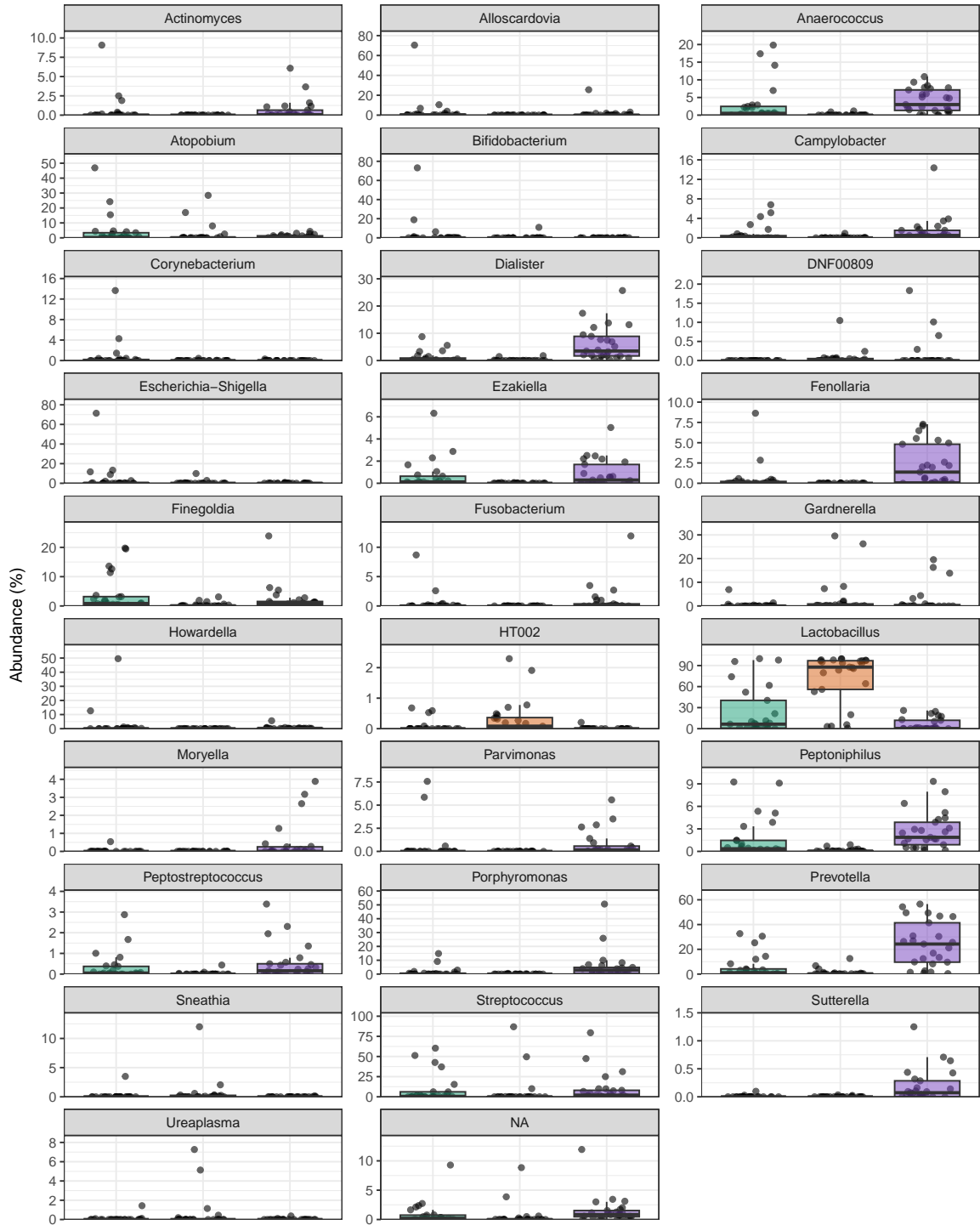
Genus	01_postmenopausal	02_premenopausal	03_trans-men
Ezakiella	0.68 ± 1.4	0.01 ± 0.02	0.87 ± 1.24
Fenollaria	0.55 ± 1.78	0 ± 0.01	2.25 ± 2.52
Finegoldia	3.96 ± 6.13	0.35 ± 0.74	2.29 ± 4.78
Fusobacterium	0.5 ± 1.78	0 ± 0.01	0.9 ± 2.46
Gardnerella	0.39 ± 1.39	3.15 ± 7.76	2.36 ± 5.52
HT002	0.08 ± 0.2	0.34 ± 0.58	0.01 ± 0.04
Howardella	2.63 ± 10.1	0.01 ± 0.03	0.46 ± 1.1
Lactobacillus	24.28 ± 34.48	71.66 ± 35.72	6 ± 8.97
Moryella	0.02 ± 0.11	0 ± 0	0.49 ± 1.08
Parvimonas	0.56 ± 1.87	0 ± 0	0.76 ± 1.41
Peptoniphilus	1.74 ± 2.74	0.11 ± 0.22	2.72 ± 2.43
Peptostreptococcus	0.32 ± 0.67	0.02 ± 0.09	0.52 ± 0.85
Porphyromonas	1.29 ± 3.4	0.03 ± 0.1	5.65 ± 10.79
Prevotella	5.8 ± 9.75	1.16 ± 2.82	24.3 ± 18.51
Sneathia	0.14 ± 0.7	0.64 ± 2.4	0.01 ± 0.03
Streptococcus	9.19 ± 17.9	5.91 ± 19.62	9.67 ± 18.53
Sutterella	0.01 ± 0.02	0 ± 0.01	0.19 ± 0.3
Ureaplasma	0.07 ± 0.29	0.57 ± 1.74	0.02 ± 0.08
NA	0.87 ± 1.93	0.57 ± 1.89	1.52 ± 2.37

This analysis reveals, that the *Trans-men* group is characterized by a loss of *Lactobacillus* and an increase in bacteria more commonly associated with the intestinal flora (e.g. *Campylobacter*, *Anaerococcus*, *Dialister*, *Prevotella*) compared to the premenopausal group.

Comparing the *Trans-men* to the *postmenopausal* group only revealed an even larger reduction in *Lactobacillus*.

Thus, overall the *trans-men* group is strikingly similar in its taxonomic composition compared to postmenopausal women.

deseq_res_fig



```
ggsave(deseq_res_fig, width = 7, height = 10, filename = "results/microbiome/deseq_asv_fig")
ggsave(deseq_res_fig, width = 7, height = 8, filename = "results/microbiome/deseq_asv_fig")
```

5 Taxonomic composition

The following plots illustrate taxonomic composition at all levels of the taxonomy, The taxon table is filtered to only differentiate the most abundant taxa at each level (depending on the level).

```
ps.ns = prune_taxa(taxa_sums(ps_samples) > 1, ps_samples) # remove singleton ASVs
ps_rel = transform_sample_counts(ps.ns, function(x) x/sum(x) * 100) # transform to rel counts

# Generate agglomerated phyloseq objects
genus_rel = tax_glom(ps_rel, taxrank="Genus", NArm = F)
genus_rel # 383 genera
```

phyloseq-class experiment-level object

```
otu_table() OTU Table: [ 398 taxa and 75 samples ]
sample_data() Sample Data: [ 75 samples by 15 sample variables ]
tax_table() Taxonomy Table: [ 398 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 398 tips and 397 internal nodes ]
refseq() DNASTringSet: [ 398 reference sequences ]
```

```
genus = psmelt(genus_rel) %>%
  mutate(Taxonomy = paste0("p_", Phylum, ";c_", Class, ";o_", Order, ";f_", Family, ";g_", Genus))
  select(Sample, Taxonomy, grp, Age:Estradiole, Abundance)
genus$Taxonomy[genus$Abundance < 15] <- "< 15 % abund."

genusplot = ggplot(genus, aes(x = grp, y = Abundance/25, fill = Taxonomy)) +
  geom_col(position = "stack") +
  theme(legend.position = "right") +
  labs(y = "Abundance (%)", x = "", caption = "taxonomic level: genus") +
  rotate_x_text()

# Family Level
family_rel = tax_glom(ps_rel, taxrank="Family", NArm = F)
family_rel # 383 genera
```

phyloseq-class experiment-level object

```
otu_table() OTU Table: [ 152 taxa and 75 samples ]
sample_data() Sample Data: [ 75 samples by 15 sample variables ]
tax_table() Taxonomy Table: [ 152 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 152 tips and 151 internal nodes ]
refseq() DNASTringSet: [ 152 reference sequences ]
```

```
family = psmelt(family_rel) %>%
  mutate(Taxonomy = paste0("p_", Phylum, ";c_", Class, ";o_", Order, ";f_", Family)) %>%
  select(Sample, Taxonomy, grp, Age:Estradiole, Abundance)
family$Taxonomy[family$Abundance < 15] <- "< 15 % abund."

familyplot = ggplot(family, aes(x = grp, y = Abundance/25, fill = Taxonomy)) +
  geom_col(position = "stack") +
  theme(legend.position = "right") +
  labs(y = "Abundance (%)", x = "", caption = "taxonomic level: family") +
  rotate_x_text()

# Order Level
order_rel = tax_glom(ps_rel, taxrank="Order", NArm = F)
order_rel # 75 taxa
```

phyloseq-class experiment-level object

```
otu_table() OTU Table: [ 78 taxa and 75 samples ]
sample_data() Sample Data: [ 75 samples by 15 sample variables ]
tax_table() Taxonomy Table: [ 78 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 78 tips and 77 internal nodes ]
refseq() DNASTringSet: [ 78 reference sequences ]
```

```
order = psmelt(order_rel) %>%
  mutate(Taxonomy = paste0("p_", Phylum, ";c_", Class, ";o_", Order)) %>%
  select(Sample, Taxonomy, grp, Age:Estradiole, Abundance)
order$Taxonomy[order$Abundance < 20] <- "< 20 % abund."

orderplot = ggplot(order, aes(x = grp, y = Abundance/25, fill = Taxonomy)) +
  geom_col(position = "stack") +
  theme(legend.position = "right") +
  labs(y = "Abundance (%)", x = "", caption = "taxonomic level: order") +
  rotate_x_text()

# Class Level
```

```
class_rel = tax_glom(ps_rel, taxrank="Class", NArm = F)
class_rel # 30 taxa
```

phyloseq-class experiment-level object

```
otu_table() OTU Table: [ 32 taxa and 75 samples ]
sample_data() Sample Data: [ 75 samples by 15 sample variables ]
tax_table() Taxonomy Table: [ 32 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 32 tips and 31 internal nodes ]
refseq() DNASTringSet: [ 32 reference sequences ]
```

```
class = psmelt(class_rel) %>%
  mutate(Taxonomy = paste0("p_", Phylum, ";c_", Class)) %>%
  select(Sample, Taxonomy, grp, Age:Estradiole, Abundance)
class$Taxonomy[class$Abundance < 25] <- "< 25 % abund."

classplot = ggplot(class, aes(x = grp, y = Abundance/25, fill = Taxonomy)) +
  geom_col(position = "stack") +
  theme(legend.position = "right") +
  labs(y = "Abundance (%)", x = "", caption = "taxonomic level: class") +
  rotate_x_text()

# Phylum Level
phylum_rel = tax_glom(ps_rel, taxrank="Phylum", NArm = F)
phylum_rel # 18 taxa
```

phyloseq-class experiment-level object

```
otu_table() OTU Table: [ 19 taxa and 75 samples ]
sample_data() Sample Data: [ 75 samples by 15 sample variables ]
tax_table() Taxonomy Table: [ 19 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 19 tips and 18 internal nodes ]
refseq() DNASTringSet: [ 19 reference sequences ]
```

```
phylum = psmelt(phylum_rel) %>%
  mutate(Taxonomy = paste0("p_", Phylum)) %>%
  select(Sample, Taxonomy, grp, Age:Estradiole, Abundance)
phylum$Taxonomy[phylum$Abundance < 25] <- "< 25 % abund."

phylumplot = ggplot(phylum, aes(x = grp, y = Abundance/25, fill = Taxonomy)) +
  geom_col(position = "stack") +
```

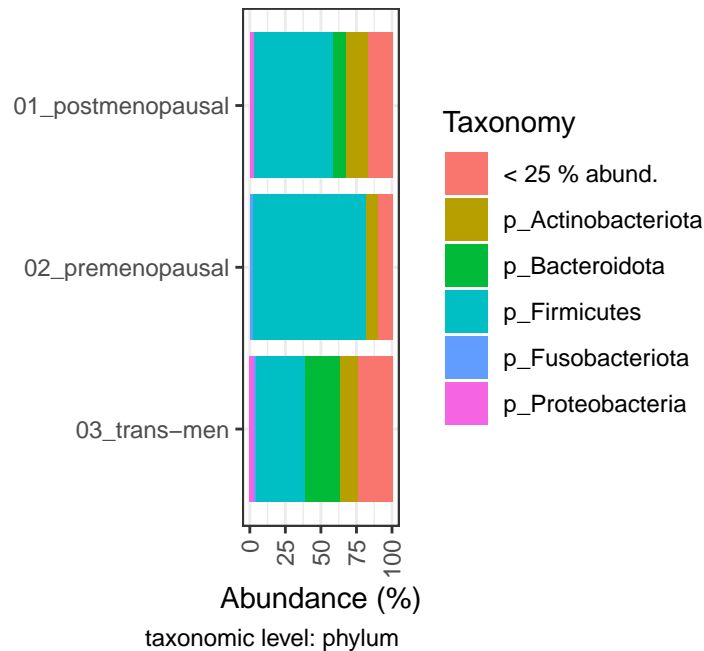


```
theme(legend.position = "right") +
labs(y = "Abundance (%)", x = "", caption = "taxonomic level: phylum") +
rotate_x_text()
```

5.1 Phylum level

On the phylum level, the taxonomic distribution is dominated by *Firmicutes* (these correspond to *Lactobacilli*, see below). The *Trans-men* group also shows an increase in *Bacteroidota* compared to the other 2 groups.

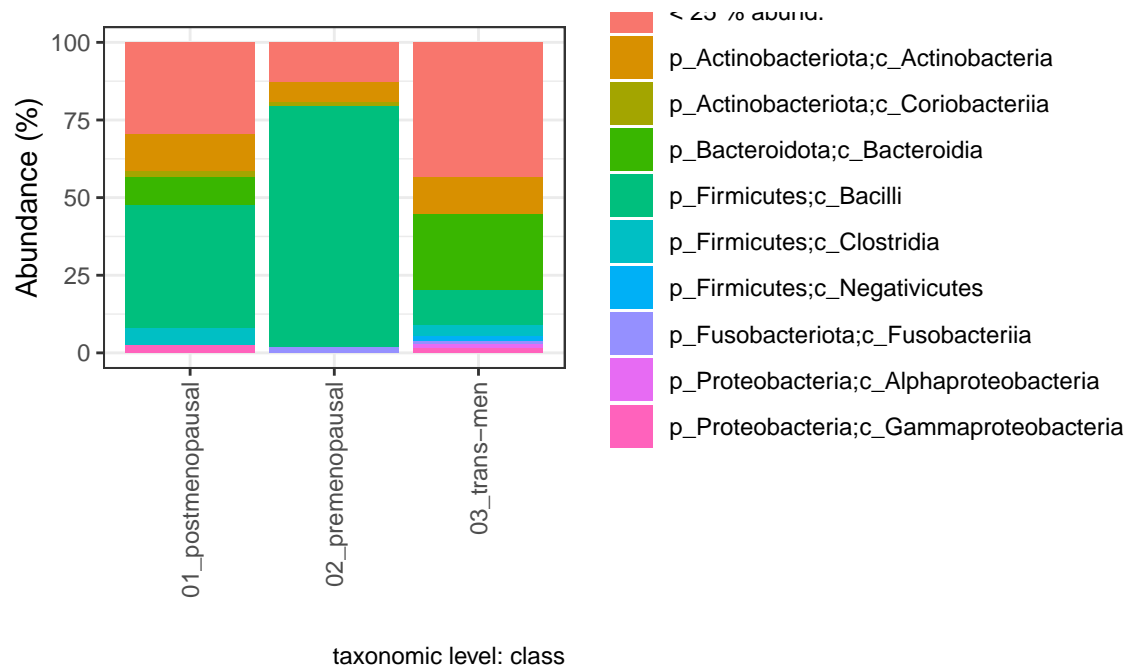
```
phylumplot = phylumplot + scale_x_discrete(limits = rev(levels(phylum$grp))) + coord_flip(
phylumplot
```



```
ggsave(plot = phylumplot, filename = "results/microbiome/taxplots_phylum.pdf", width = 7,
```

5.2 Class level

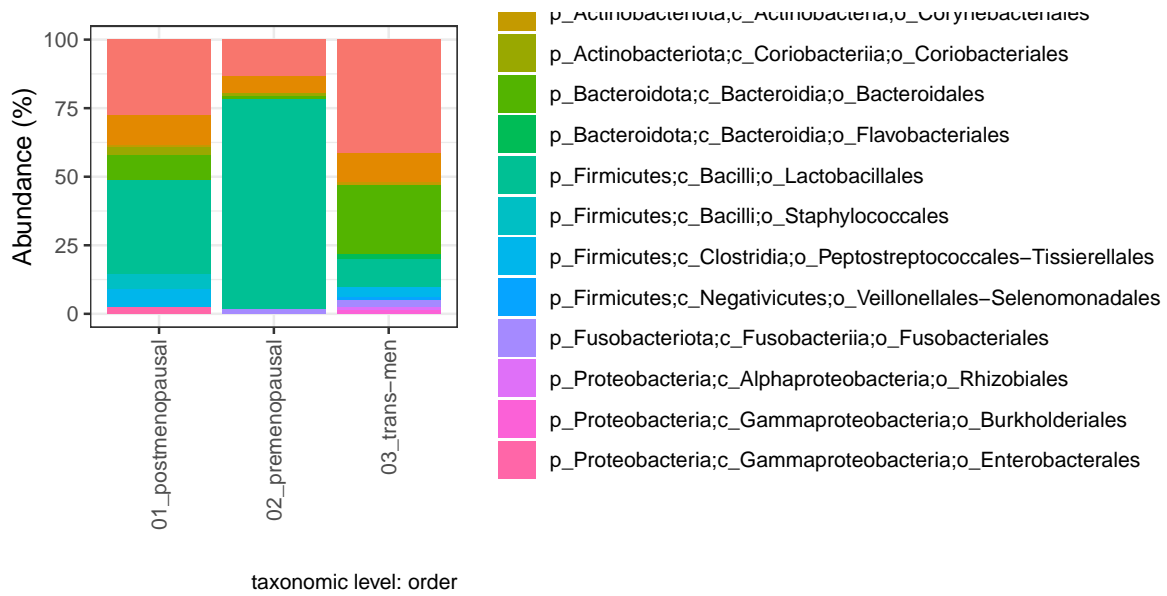
`classplot`



```
ggsave(plot = classplot, filename = "results/microbiome/taxplots_class.pdf", width = 6, height = 6)
ggsave(plot = classplot, filename = "results/microbiome/taxplots_class.png", width = 6, height = 6)
```

5.3 Order level

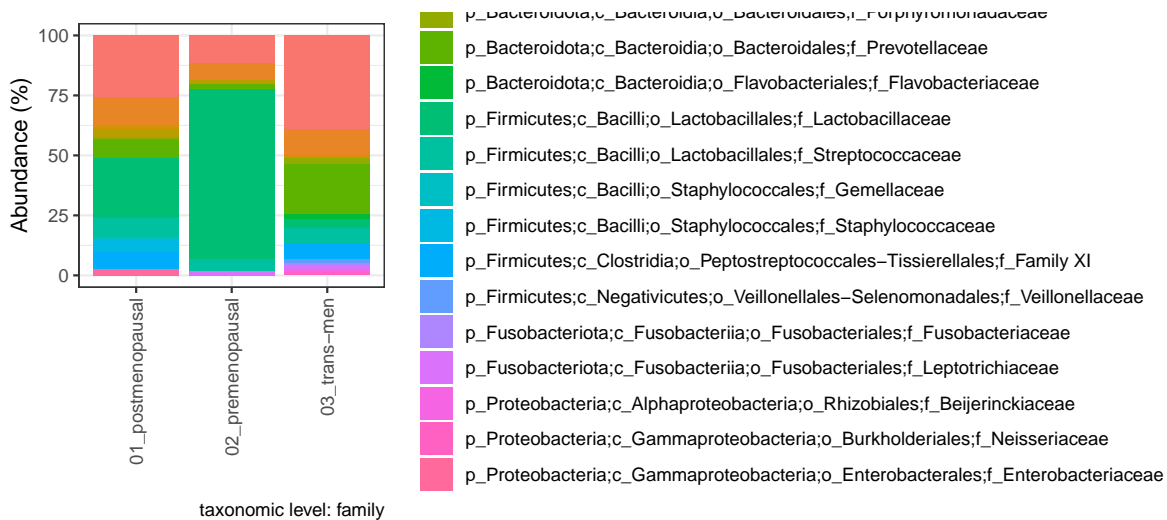
`orderplot`



```
ggsave(plot = orderplot, filename = "results/microbiome/taxplots_order.pdf", width = 7, height = 7)
ggsave(plot = orderplot, filename = "results/microbiome/taxplots_order.png", width = 7, height = 7)
```

5.4 Family level

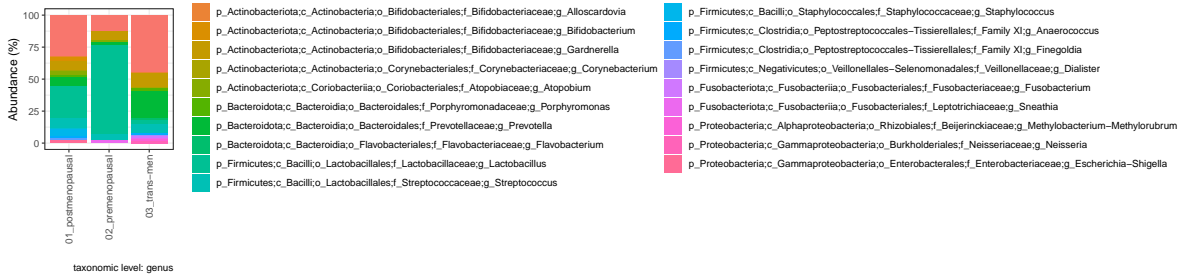
familyplot



```
ggsave(plot = familyplot, filename = "results/microbiome/taxplots_family.pdf", width = 8,
ggsave(plot = familyplot, filename = "results/microbiome/taxplots_family.png", width = 8,
```

5.5 Genus level

genusplot



```
ggsave(plot = genusplot, filename = "results/microbiome/taxplots_genus.pdf", width = 15, h
ggsave(plot = genusplot, filename = "results/microbiome/taxplots_family.png", width = 15,
```

6 Secondary analyses

6.0.1 Duration of menopause

This analysis is only performed in the postmenopausal group

```
head(sample_data(ps_samples))
```

	CID	grp	Age	GHAT_Length	GNRHA	Virgo	SexuallyActive
22010-0001	101	01_postmenopausal	64	0	<NA>	no	yes
22010-0002	102	01_postmenopausal	49	0	<NA>	no	yes
22010-0003	103	01_postmenopausal	61	0	<NA>	no	yes
22010-0004	104	01_postmenopausal	70	0	<NA>	no	yes
22010-0005	105	01_postmenopausal	66	0	<NA>	no	yes
22010-0006	106	01_postmenopausal	71	0	<NA>	no	no
	SexOfPartner	NugentScore	Testosterone	Estradiole	DurationMenopause		
22010-0001	M	7-10	NA	NA			11
22010-0002	M	0-3	NA	NA			2

22010-0003	M	7-10	NA	NA	10
22010-0004	M	4-6	NA	NA	22
22010-0005	M no bacteria		NA	NA	12
22010-0006	M	0-3	NA	NA	22
	DurationAmenorrhea	CycleDaySampling	Duration_GNRH		
22010-0001	NA		NA	NA	
22010-0002	NA		NA	NA	
22010-0003	NA		NA	NA	
22010-0004	NA		NA	NA	
22010-0005	NA		NA	NA	
22010-0006	NA		NA	NA	

```
# for menopause duration, use maaslin2
library(Maaslin2)

# only 01_postmenopausal samples
ps_menopause = subset_samples(ps_samples, grp == "01_postmenopausal") %>%
  transform_sample_counts(function(x) x/sum(x) * 100) # relative abundances are necessary

asv_durationMenopause = Maaslin2(input_data = otu_table(ps_menopause),
  input_metadata = as(sample_data(ps_menopause), "data.frame"),
  output = "results/maaslin2/asv_menopause-duration",
  min_abundance = 2.5,
  min_prevalence = 0.1,
  max_significance = 0.2,
  fixed_effects = "DurationMenopause")
```

```
2024-03-06 21:55:08.026172 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/asv_menopause-duration/maaslin2.1
2024-03-06 21:55:09.750488 INFO::Writing function arguments to log file
2024-03-06 21:55:09.75613 INFO::Verifying options selected are valid
2024-03-06 21:55:09.778012 INFO::Determining format of input files
2024-03-06 21:55:09.77853 INFO::Input format is data samples as rows and metadata samples as
2024-03-06 21:55:09.788965 INFO::Formula for fixed effects: expr ~ DurationMenopause
2024-03-06 21:55:09.789506 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:09.789774 INFO::Total samples in data: 25
2024-03-06 21:55:09.790004 INFO::Min samples required with min abundance for a feature not t
2024-03-06 21:55:09.800948 INFO::Total filtered features: 2545
2024-03-06 21:55:09.801775 INFO::Filtered feature names from abundance and prevalence filter
2024-03-06 21:55:09.802742 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:09.802992 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:09.803211 INFO::Running selected normalization method: TSS
```

```

2024-03-06 21:55:09.812711 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:09.815964 INFO::Running selected transform method: LOG
2024-03-06 21:55:09.816617 INFO::Running selected analysis method: LM
2024-03-06 21:55:09.820323 INFO::Fitting model to feature number 1, ASV1
2024-03-06 21:55:09.822208 INFO::Fitting model to feature number 2, ASV2
2024-03-06 21:55:09.823396 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:09.824553 INFO::Fitting model to feature number 4, ASV7
2024-03-06 21:55:09.8255 INFO::Fitting model to feature number 5, ASV8
2024-03-06 21:55:09.82638 INFO::Fitting model to feature number 6, ASV9
2024-03-06 21:55:09.82732 INFO::Fitting model to feature number 7, ASV10
2024-03-06 21:55:09.828487 INFO::Fitting model to feature number 8, ASV11
2024-03-06 21:55:09.829825 INFO::Fitting model to feature number 9, ASV12
2024-03-06 21:55:09.830961 INFO::Fitting model to feature number 10, ASV13
2024-03-06 21:55:09.832183 INFO::Fitting model to feature number 11, ASV21
2024-03-06 21:55:09.833212 INFO::Fitting model to feature number 12, ASV24
2024-03-06 21:55:09.834122 INFO::Fitting model to feature number 13, ASV38
2024-03-06 21:55:09.835367 INFO::Fitting model to feature number 14, ASV60
2024-03-06 21:55:09.836343 INFO::Fitting model to feature number 15, ASV94
2024-03-06 21:55:09.839294 INFO::Counting total values for each feature
2024-03-06 21:55:09.84007 INFO::Writing filtered data to file results/maaslin2/asv_menopause-
2024-03-06 21:55:09.841251 INFO::Writing filtered, normalized data to file results/maaslin2/
2024-03-06 21:55:09.842003 INFO::Writing filtered, normalized, transformed data to file resu
2024-03-06 21:55:09.842822 WARNING::Deleting existing residuals file: results/maaslin2/asv_m
2024-03-06 21:55:09.843227 INFO::Writing residuals to file results/maaslin2/asv_menopause-du
2024-03-06 21:55:09.843668 WARNING::Deleting existing fitted file: results/maaslin2/asv_menop
2024-03-06 21:55:09.844 INFO::Writing fitted values to file results/maaslin2/asv_menopause-du
2024-03-06 21:55:09.844405 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:09.845021 INFO::Writing the significant results (those which are less than c
2024-03-06 21:55:09.845373 INFO::Writing heatmap of significant results to file: results/maas
[1] "There are no associations to plot!"
2024-03-06 21:55:09.846217 INFO::Writing association plots (one for each significant associat
[1] "There are no associations to plot!"

```

```

# no significant effect on ASV level, try glom at genuslevel
ps_menopause_genus = tax_glom(ps_menopause, taxrank = "Genus")

genus_durationMenopause = Maaslin2(input_data = otu_table(ps_menopause_genus),
  input_metadata = as(sample_data(ps_menopause_genus), "data.frame"),
  output = "results/maaslin2/genus_menopause-duration",
  min_abundance = 5,
  min_prevalence = 0.2,

```

```

max_significance = 0.2,
fixed_effects = "DurationMenopause")

```

```

2024-03-06 21:55:13.678241 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_menopause-duration/maaslin2
2024-03-06 21:55:13.867157 INFO::Writing function arguments to log file
2024-03-06 21:55:13.869957 INFO::Verifying options selected are valid
2024-03-06 21:55:13.870258 INFO::Determining format of input files
2024-03-06 21:55:13.870503 INFO::Input format is data samples as rows and metadata samples as
2024-03-06 21:55:13.872347 INFO::Formula for fixed effects: expr ~ DurationMenopause
2024-03-06 21:55:13.872606 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:13.872818 INFO::Total samples in data: 25
2024-03-06 21:55:13.873017 INFO::Min samples required with min abundance for a feature not to
2024-03-06 21:55:13.874522 INFO::Total filtered features: 316
2024-03-06 21:55:13.874839 INFO::Filtered feature names from abundance and prevalence filteri
2024-03-06 21:55:13.875312 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:13.875545 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:13.875756 INFO::Running selected normalization method: TSS
2024-03-06 21:55:13.876114 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:13.878492 INFO::Running selected transform method: LOG
2024-03-06 21:55:13.878858 INFO::Running selected analysis method: LM
2024-03-06 21:55:13.879116 INFO::Fitting model to feature number 1, ASV2
2024-03-06 21:55:13.880164 INFO::Fitting model to feature number 2, ASV7
2024-03-06 21:55:13.881041 INFO::Fitting model to feature number 3, ASV18
2024-03-06 21:55:13.881921 INFO::Fitting model to feature number 4, ASV33
2024-03-06 21:55:13.88448 INFO::Counting total values for each feature
2024-03-06 21:55:13.885095 INFO::Writing filtered data to file results/maaslin2/genus_menopa
2024-03-06 21:55:13.885975 INFO::Writing filtered, normalized data to file results/maaslin2/g
2024-03-06 21:55:13.886587 INFO::Writing filtered, normalized, transformed data to file resu
2024-03-06 21:55:13.887173 WARNING::Deleting existing residuals file: results/maaslin2/genus
2024-03-06 21:55:13.887621 INFO::Writing residuals to file results/maaslin2/genus_menopause-c
2024-03-06 21:55:13.888046 WARNING::Deleting existing fitted file: results/maaslin2/genus_mer
2024-03-06 21:55:13.888395 INFO::Writing fitted values to file results/maaslin2/genus_menopa
2024-03-06 21:55:13.888765 INFO::Writing all results to file (ordered by increasing q-values
2024-03-06 21:55:13.889314 INFO::Writing the significant results (those which are less than c
2024-03-06 21:55:13.889658 INFO::Writing heatmap of significant results to file: results/maas
[1] "There are no associations to plot!"
2024-03-06 21:55:13.890126 INFO::Writing association plots (one for each significant associat
[1] "There are no associations to plot!"

```

```
# no significant effect on genus level, try on family level
ps_menopause_family = tax_glom(ps_menopause, taxrank = "Family")

family_durationMenopause = Maaslin2(input_data = otu_table(ps_menopause_family),
    input_metadata = as(sample_data(ps_menopause_family), "data.frame"),
    output = "results/maaslin2/family_menopause-duration",
    min_abundance = 10,
    min_prevalence = 0.3,
    max_significance = 0.2,
    fixed_effects = "DurationMenopause")
```

```
2024-03-06 21:55:15.79665 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/family_menopause-duration/maaslin2-
2024-03-06 21:55:15.863116 INFO::Writing function arguments to log file
2024-03-06 21:55:15.865145 INFO::Verifying options selected are valid
2024-03-06 21:55:15.865387 INFO::Determining format of input files
2024-03-06 21:55:15.865649 INFO::Input format is data samples as rows and metadata samples as
2024-03-06 21:55:15.86679 INFO::Formula for fixed effects: expr ~ DurationMenopause
2024-03-06 21:55:15.867063 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:15.867279 INFO::Total samples in data: 25
2024-03-06 21:55:15.867484 INFO::Min samples required with min abundance for a feature not to
2024-03-06 21:55:15.868252 INFO::Total filtered features: 117
2024-03-06 21:55:15.868496 INFO::Filtered feature names from abundance and prevalence filter
2024-03-06 21:55:15.868846 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:15.869072 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:15.869274 INFO::Running selected normalization method: TSS
2024-03-06 21:55:15.869644 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:15.872274 INFO::Running selected transform method: LOG
2024-03-06 21:55:15.872699 INFO::Running selected analysis method: LM
2024-03-06 21:55:15.872975 INFO::Fitting model to feature number 1, ASV2
2024-03-06 21:55:15.874094 INFO::Fitting model to feature number 2, ASV7
2024-03-06 21:55:15.87501 INFO::Fitting model to feature number 3, ASV12
2024-03-06 21:55:15.875848 INFO::Fitting model to feature number 4, ASV18
2024-03-06 21:55:15.886075 INFO::Counting total values for each feature
2024-03-06 21:55:15.886744 INFO::Writing filtered data to file results/maaslin2/family_menop
2024-03-06 21:55:15.887341 INFO::Writing filtered, normalized data to file results/maaslin2/
2024-03-06 21:55:15.887895 INFO::Writing filtered, normalized, transformed data to file resu
2024-03-06 21:55:15.888429 WARNING::Deleting existing residuals file: results/maaslin2/family
2024-03-06 21:55:15.88912 INFO::Writing residuals to file results/maaslin2/family_menopause-
2024-03-06 21:55:15.889829 WARNING::Deleting existing fitted file: results/maaslin2/family_m
2024-03-06 21:55:15.890341 INFO::Writing fitted values to file results/maaslin2/family_menop
2024-03-06 21:55:15.890791 INFO::Writing all results to file (ordered by increasing q-values)
```



```

2024-03-06 21:55:15.891422 INFO::Writing the significant results (those which are less than c
2024-03-06 21:55:15.891786 INFO::Writing heatmap of significant results to file: results/maas
[1] "There are no associations to plot!"
2024-03-06 21:55:15.892249 INFO::Writing association plots (one for each significant associat
[1] "There are no associations to plot!"

```

```

## Alpha diversity by duration of menopause
ps_menopause_abs = subset_samples(ps_samples, grp == "01_postmenopausal")

adiv_menopause = plot_richness(ps_menopause_abs, measures = c("Observed", "Shannon", "InvS

adiv_menopause$data %>%
  select(DurationMenopause, variable, value) %>%
  group_by(variable) %>%
  cor_test(value, DurationMenopause) %>%
  kable(digits = 3)

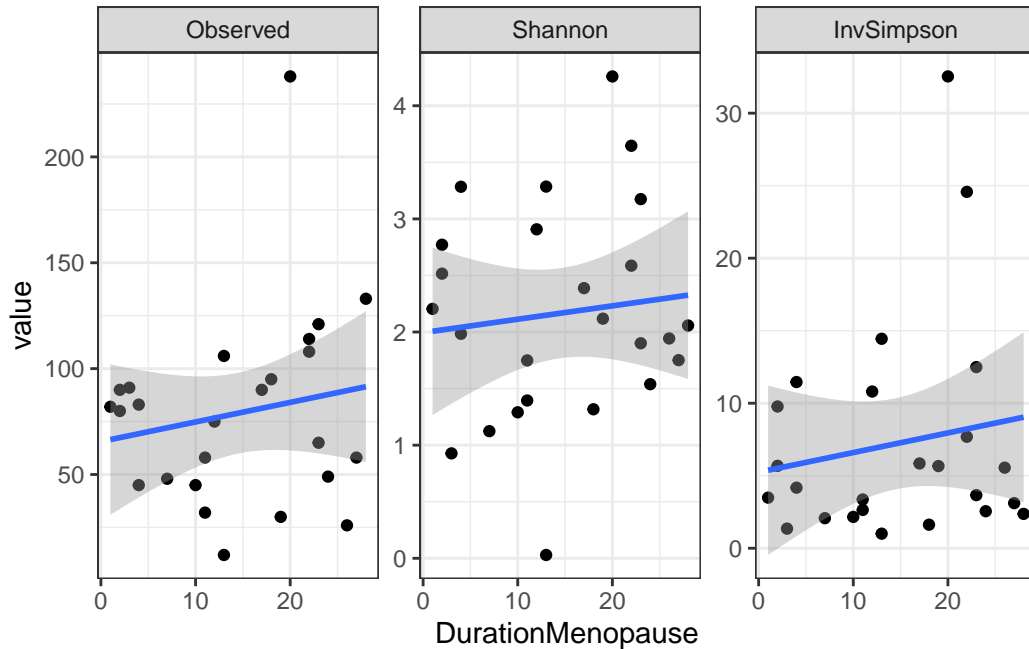
```

variable	var1	var2	cor	statistic	p	conf.low	conf.high	method
Observed	value	DurationMenopause	0.18	0.861	0.398	-0.235	0.534	Pearson
Shannon	value	DurationMenopause	0.11	0.528	0.602	-0.299	0.484	Pearson
InvSimpson	value	DurationMenopause	0.16	0.767	0.451	-0.253	0.521	Pearson

```

adiv_menopause$data %>%
  ggplot(aes(x = DurationMenopause, y = value)) +
  geom_point()+
  geom_smooth(method = "lm") +
  facet_wrap(~variable, scales = "free_y")

```



```
BC_menopause_genus = vegan::vegdist(otu_table(ps_menopause_genus))

menopausduration_adonis_bc = vegan::adonis2(BC_menopause_genus ~ phyloseq::sample_data(ps_menopause_genus)$DurationMenopause)

# no significant effect on the beta diversity level as well

# Summary statistics of duration menopause
sample_data(ps_menopause_genus)$DurationMenopause %>%
  summary()
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	7.00	13.00	14.48	22.00	28.00

There is no significant correlation of duration of menopause (years) with alpha diversity of vaginal microbiomes and overall community composition (beta-diversity). Duration of menopause in this sample is rather varied ranging from 1 to 28 years (mean 14.5 years).

6.0.2 Duration of gender-affirming hormone therapy (GAHT)

This analysis is only performed in the trans-men group.

```

# only 03_trans samples
ps_trans = subset_samples(ps_samples, grp == "03_trans-men") %>%
  transform_sample_counts(function(x) x/sum(x) * 100) # relative abundances are necessary

asv_GHAT = Maaslin2(input_data = otu_table(ps_trans),
  input_metadata = as(sample_data(ps_trans), "data.frame"),
  output = "results/maaslin2/asv_GHAT_Length",
  min_abundance = 1,
  min_prevalence = 0.1,
  max_significance = 0.2,
  fixed_effects = "GHAT_Length")

```

```

2024-03-06 21:55:16.167914 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/asv_GHAT_Length/maaslin2.log"
2024-03-06 21:55:17.598958 INFO::Writing function arguments to log file
2024-03-06 21:55:17.602204 INFO::Verifying options selected are valid
2024-03-06 21:55:17.602528 INFO::Determining format of input files
2024-03-06 21:55:17.602787 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:17.612953 INFO::Formula for fixed effects: expr ~ GHAT_Length
2024-03-06 21:55:17.613485 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:17.613746 INFO::Total samples in data: 25
2024-03-06 21:55:17.61397 INFO::Min samples required with min abundance for a feature not to be filtered: 1
2024-03-06 21:55:17.62434 INFO::Total filtered features: 2511
2024-03-06 21:55:17.625139 INFO::Filtered feature names from abundance and prevalence filtering: 2511
2024-03-06 21:55:17.626284 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:17.626529 INFO::Filtered feature names from variance filtering: 0
2024-03-06 21:55:17.626742 INFO::Running selected normalization method: TSS
2024-03-06 21:55:17.627265 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:17.629637 INFO::Running selected transform method: LOG
2024-03-06 21:55:17.630213 INFO::Running selected analysis method: LM
2024-03-06 21:55:17.630469 INFO::Fitting model to feature number 1, ASV2
2024-03-06 21:55:17.63148 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:17.632356 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:17.633211 INFO::Fitting model to feature number 4, ASV6
2024-03-06 21:55:17.634082 INFO::Fitting model to feature number 5, ASV7
2024-03-06 21:55:17.634906 INFO::Fitting model to feature number 6, ASV9
2024-03-06 21:55:17.63573 INFO::Fitting model to feature number 7, ASV10
2024-03-06 21:55:17.636539 INFO::Fitting model to feature number 8, ASV11
2024-03-06 21:55:17.637337 INFO::Fitting model to feature number 9, ASV12
2024-03-06 21:55:17.638141 INFO::Fitting model to feature number 10, ASV14
2024-03-06 21:55:17.638975 INFO::Fitting model to feature number 11, ASV15
2024-03-06 21:55:17.639823 INFO::Fitting model to feature number 12, ASV17

```

2024-03-06 21:55:17.640626 INFO::Fitting model to feature number 13, ASV18
2024-03-06 21:55:17.64143 INFO::Fitting model to feature number 14, ASV21
2024-03-06 21:55:17.64224 INFO::Fitting model to feature number 15, ASV23
2024-03-06 21:55:17.643043 INFO::Fitting model to feature number 16, ASV24
2024-03-06 21:55:17.643937 INFO::Fitting model to feature number 17, ASV25
2024-03-06 21:55:17.64486 INFO::Fitting model to feature number 18, ASV27
2024-03-06 21:55:17.645707 INFO::Fitting model to feature number 19, ASV28
2024-03-06 21:55:17.646529 INFO::Fitting model to feature number 20, ASV29
2024-03-06 21:55:17.64735 INFO::Fitting model to feature number 21, ASV31
2024-03-06 21:55:17.648163 INFO::Fitting model to feature number 22, ASV32
2024-03-06 21:55:17.649055 INFO::Fitting model to feature number 23, ASV34
2024-03-06 21:55:17.649988 INFO::Fitting model to feature number 24, ASV35
2024-03-06 21:55:17.650891 INFO::Fitting model to feature number 25, ASV36
2024-03-06 21:55:17.651731 INFO::Fitting model to feature number 26, ASV39
2024-03-06 21:55:17.652556 INFO::Fitting model to feature number 27, ASV40
2024-03-06 21:55:17.653353 INFO::Fitting model to feature number 28, ASV45
2024-03-06 21:55:17.654162 INFO::Fitting model to feature number 29, ASV46
2024-03-06 21:55:17.654976 INFO::Fitting model to feature number 30, ASV47
2024-03-06 21:55:17.655849 INFO::Fitting model to feature number 31, ASV48
2024-03-06 21:55:17.65676 INFO::Fitting model to feature number 32, ASV49
2024-03-06 21:55:17.657683 INFO::Fitting model to feature number 33, ASV54
2024-03-06 21:55:17.658587 INFO::Fitting model to feature number 34, ASV57
2024-03-06 21:55:17.659479 INFO::Fitting model to feature number 35, ASV62
2024-03-06 21:55:17.660359 INFO::Fitting model to feature number 36, ASV64
2024-03-06 21:55:17.661179 INFO::Fitting model to feature number 37, ASV66
2024-03-06 21:55:17.661999 INFO::Fitting model to feature number 38, ASV70
2024-03-06 21:55:17.662799 INFO::Fitting model to feature number 39, ASV73
2024-03-06 21:55:17.66358 INFO::Fitting model to feature number 40, ASV74
2024-03-06 21:55:17.664363 INFO::Fitting model to feature number 41, ASV76
2024-03-06 21:55:17.665162 INFO::Fitting model to feature number 42, ASV83
2024-03-06 21:55:17.665989 INFO::Fitting model to feature number 43, ASV91
2024-03-06 21:55:17.666813 INFO::Fitting model to feature number 44, ASV92
2024-03-06 21:55:17.667679 INFO::Fitting model to feature number 45, ASV97
2024-03-06 21:55:17.668543 INFO::Fitting model to feature number 46, ASV102
2024-03-06 21:55:17.669439 INFO::Fitting model to feature number 47, ASV104
2024-03-06 21:55:17.670251 INFO::Fitting model to feature number 48, ASV116
2024-03-06 21:55:17.671095 INFO::Fitting model to feature number 49, ASV144
2024-03-06 21:55:17.675124 INFO::Counting total values for each feature
2024-03-06 21:55:17.676304 INFO::Writing filtered data to file results/maaslin2/asv_GHAT_Length/
2024-03-06 21:55:17.677503 INFO::Writing filtered, normalized data to file results/maaslin2/
2024-03-06 21:55:17.678528 INFO::Writing filtered, normalized, transformed data to file resu
2024-03-06 21:55:17.679758 WARNING::Deleting existing residuals file: results/maaslin2/asv_GH
2024-03-06 21:55:17.68016 INFO::Writing residuals to file results/maaslin2/asv_GHAT_Length/f

```

2024-03-06 21:55:17.680791 WARNING::Deleting existing fitted file: results/maaslin2/asv_GHAT_Length
2024-03-06 21:55:17.681117 INFO::Writing fitted values to file results/maaslin2/asv_GHAT_Length
2024-03-06 21:55:17.681739 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:17.682586 INFO::Writing the significant results (those which are less than 0.05)
2024-03-06 21:55:17.682961 INFO::Writing heatmap of significant results to file: results/maaslin2/asv_GHAT_Length
[1] "There are no associations to plot!"
2024-03-06 21:55:17.683419 INFO::Writing association plots (one for each significant association)
[1] "There are no associations to plot!"

```

```

tax_table(ps_trans) %>%
  as.data.frame() %>%
  rownames_to_column("ASV") %>%
  filter(ASV %in% filter(asv_GHAT$results, qval < 0.2)$feature)

```

```

[1] ASV      Kingdom Phylum Class Order Family Genus Species
<0 rows> (or 0-length row.names)

```

```

ps_trans_genus = tax_glom(ps_trans, taxrank = "Genus")
genus_GHAT = Maaslin2(input_data = otu_table(ps_trans_genus),
  input_metadata = as(sample_data(ps_trans_genus), "data.frame"),
  output = "results/maaslin2/genus_GHAT_Length",
  min_abundance = 5,
  min_prevalence = 0.1,
  max_significance = 0.2,
  fixed_effects = "GHAT_Length")

```

```

2024-03-06 21:55:21.545683 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_GHAT_Length/maaslin2.log"
2024-03-06 21:55:21.717396 INFO::Writing function arguments to log file
2024-03-06 21:55:21.720259 INFO::Verifying options selected are valid
2024-03-06 21:55:21.720535 INFO::Determining format of input files
2024-03-06 21:55:21.720788 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:21.72265 INFO::Formula for fixed effects: expr ~ GHAT_Length
2024-03-06 21:55:21.722909 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:21.723128 INFO::Total samples in data: 25
2024-03-06 21:55:21.72333 INFO::Min samples required with min abundance for a feature not to be filtered: 5
2024-03-06 21:55:21.724905 INFO::Total filtered features: 310
2024-03-06 21:55:21.72519 INFO::Filtered feature names from abundance and prevalence filtering: 310
2024-03-06 21:55:21.725672 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:21.725907 INFO::Filtered feature names from variance filtering: 0

```

```

2024-03-06 21:55:21.726115 INFO::Running selected normalization method: TSS
2024-03-06 21:55:21.72652 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:21.728888 INFO::Running selected transform method: LOG
2024-03-06 21:55:21.729296 INFO::Running selected analysis method: LM
2024-03-06 21:55:21.729591 INFO::Fitting model to feature number 1, ASV2
2024-03-06 21:55:21.730633 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:21.73148 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:21.73231 INFO::Fitting model to feature number 4, ASV6
2024-03-06 21:55:21.733131 INFO::Fitting model to feature number 5, ASV17
2024-03-06 21:55:21.733946 INFO::Fitting model to feature number 6, ASV20
2024-03-06 21:55:21.734767 INFO::Fitting model to feature number 7, ASV24
2024-03-06 21:55:21.735581 INFO::Fitting model to feature number 8, ASV31
2024-03-06 21:55:21.736393 INFO::Fitting model to feature number 9, ASV70
2024-03-06 21:55:21.737229 INFO::Fitting model to feature number 10, ASV105
2024-03-06 21:55:21.739873 INFO::Counting total values for each feature
2024-03-06 21:55:21.740554 INFO::Writing filtered data to file results/maaslin2/genus_GHAT_Length
2024-03-06 21:55:21.74122 INFO::Writing filtered, normalized data to file results/maaslin2/genus_GHAT_Length
2024-03-06 21:55:21.741955 INFO::Writing filtered, normalized, transformed data to file results/maaslin2/genus_GHAT_Length
2024-03-06 21:55:21.742694 WARNING::Deleting existing residuals file: results/maaslin2/genus_GHAT_Length
2024-03-06 21:55:21.74336 INFO::Writing residuals to file results/maaslin2/genus_GHAT_Length
2024-03-06 21:55:21.743905 WARNING::Deleting existing fitted file: results/maaslin2/genus_GHAT_Length
2024-03-06 21:55:21.74429 INFO::Writing fitted values to file results/maaslin2/genus_GHAT_Length
2024-03-06 21:55:21.744678 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:21.74527 INFO::Writing the significant results (those which are less than 0.05)
2024-03-06 21:55:21.745656 INFO::Writing heatmap of significant results to file: results/maaslin2/genus_GHAT_Length
[1] "There is not enough metadata in the associations to create a heatmap plot. Please review the metadata."
2024-03-06 21:55:21.746187 INFO::Writing association plots (one for each significant association)
2024-03-06 21:55:21.74673 INFO::Plotting associations from most to least significant, grouped by metadata number
2024-03-06 21:55:21.746997 INFO::Plotting data for metadata number 1, GHAT_Length
2024-03-06 21:55:21.747849 INFO::Creating scatter plot for continuous data, GHAT_Length vs ASV2

```

Warning: Removed 1 rows containing missing values (`geom_point()`).

```

2024-03-06 21:55:21.813431 INFO::Creating scatter plot for continuous data, GHAT_Length vs ASV3

```

Warning: Removed 2 rows containing missing values (`geom_point()`).

Removed 1 rows containing missing values (`geom_point()`).

Warning: Removed 2 rows containing missing values (`geom_point()`).

```
tax_table(ps_trans_genus) %>%
  as.data.frame() %>%
  rownames_to_column("ASV") %>%
  filter(ASV %in% filter(genus_GHAT$results, qval < 0.2)$feature)
```

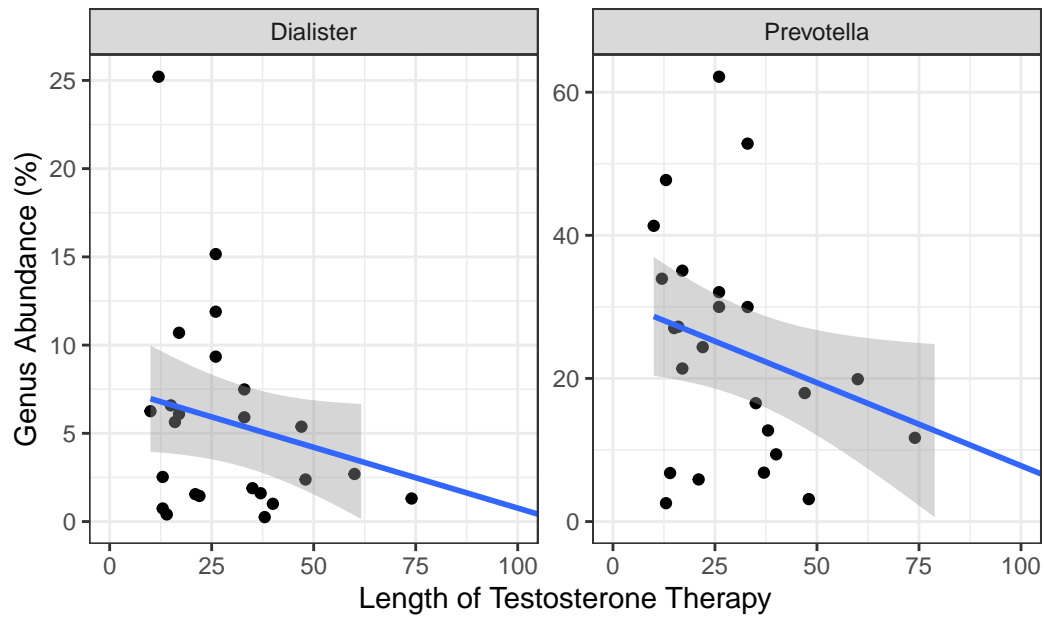
	ASV	Kingdom	Phylum	Class	Order
1	ASV3	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales
2	ASV17	Bacteria	Firmicutes	Negativicutes	Veillonellales-Selenomonadales
		Family	Genus	Species	
1	Prevotellaceae	Prevotella	<NA>		
2	Veillonellaceae	Dialister	<NA>		

```
# plot correlation for these 2 genera
psmelt_trans_genus = psmelt(ps_trans_genus)
maaslin_res_genus_ghat = ggplot(filter(psmelt_trans_genus, OTU %in% filter(genus_GHAT$resu
                                aes(x = GHAT_Length, y = Abundance)) +

  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~Genus, scales = "free") +
  scale_y_continuous(limits = function(x){c(0, max(0.1, x))}) +
  coord_cartesian(xlim = c(0,100)) +
  labs(x = "Length of Testosterone Therapy",
       y = "Genus Abundance (%)",
       caption = "Maaslin2 results")

maaslin_res_genus_ghat
```

Warning: Removed 29 rows containing missing values (`geom_smooth()`).



Maaslin2 results

```
ggsave(maaslin_res_genus_ghat, filename = "results/maaslin2/genus_GHAT_Length_panelplot.pdf",
       height = 4, width = 6)
```

Warning: Removed 29 rows containing missing values (`geom_smooth()`).

```
## Alpha diversity by duration of gaht
ps_trans_abs = subset_samples(ps_samples, grp == "03_trans-men")
ps_trans_genus = tax_glom(ps_trans_abs, taxrank = "Genus") %>%
  transform_sample_counts(function(x) x/sum(x) * 100)

adiv_trans = plot_richness(ps_trans_abs, measures = c("Observed", "Shannon", "InvSimpson"))

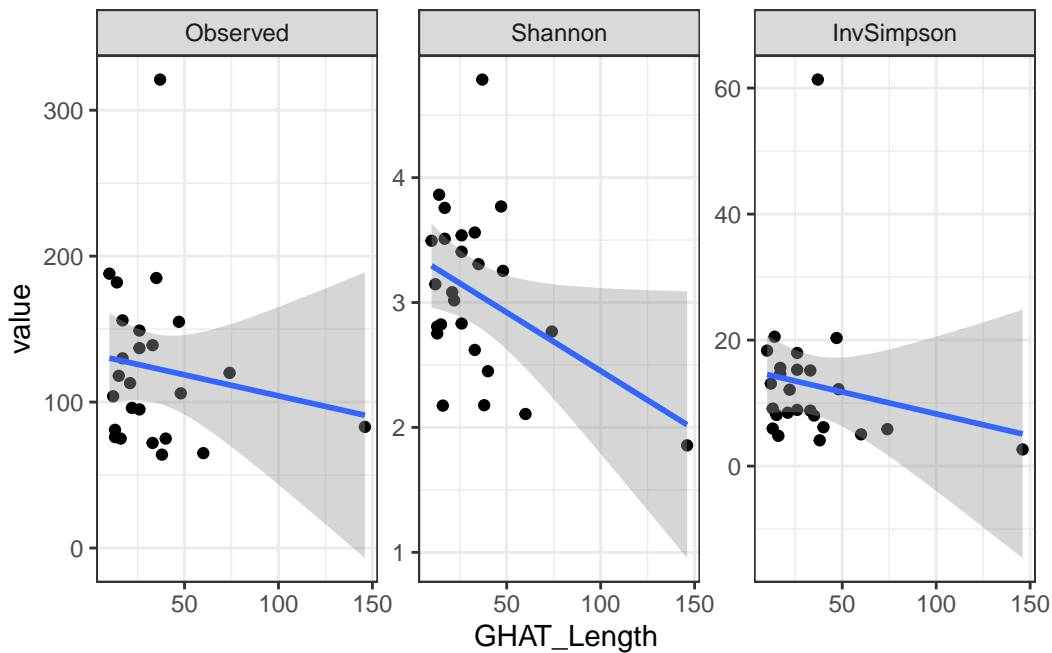
adiv_trans$data %>%
  select(GHAT_Length, variable, value) %>%
  group_by(variable) %>%
  cor_test(value, GHAT_Length) %>%
  kable(digits = 3)
```


variable	var1	var2	cor	statistic	p	conf.low	conf.high	method
Observed	value	GHAT_Length	0.15	-0.705	0.488	-0.511	0.265	Pearson
Shannon	value	GHAT_Length	0.40	-2.102	0.047	-0.688	-0.007	Pearson
InvSimpson	value	GHAT_Length	0.17	-0.844	0.407	-0.532	0.238	Pearson

```

adiv_trans$data %>%
  ggplot(aes(x = GHAT_Length, y = value)) +
  geom_point()+
  geom_smooth(method = "lm") +
  facet_wrap(~variable, scales = "free_y")

```



```

BC_trans_genus = vegan::vegdist(otu_table(ps_trans_genus))
ord.bc = ordinate(ps_trans_genus, method = "PCoA")

gaht_adonis_bc = vegan::adonis2(BC_trans_genus ~ phyloseq::sample_data(ps_trans_genus)$GHA

#This is highly significant on the genus level!
kable(gaht_adonis_bc)

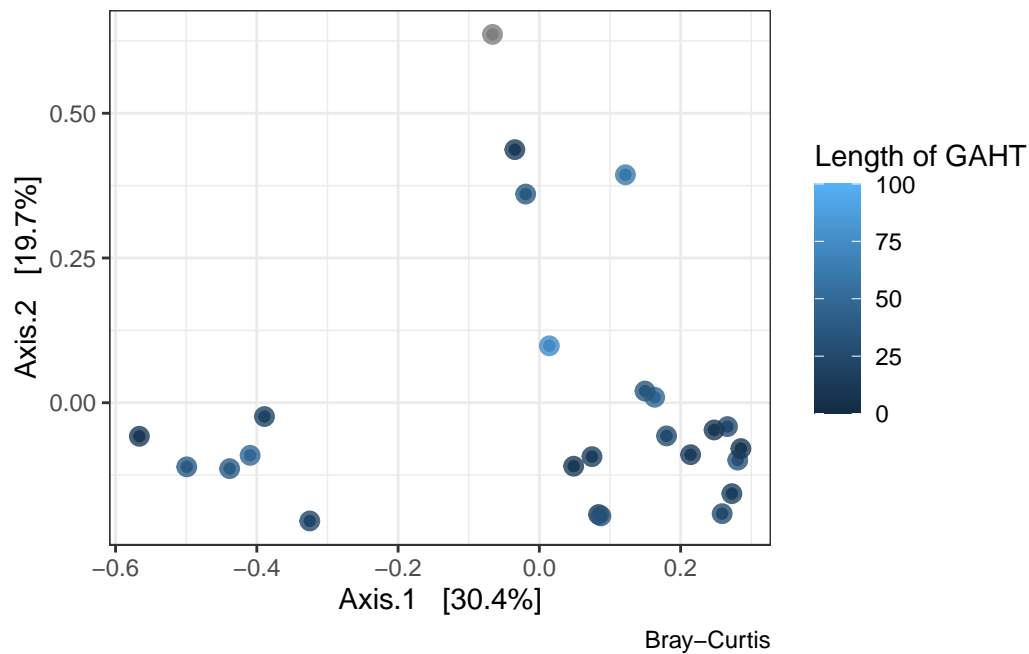
```

	Df	SumOfSqs	R2	F	Pr(>F)
phyloseq::sample_data(ps_trans_genus)\$GHAT_Length	65	0.00245	0.1102164	2.848982	0.007
Residual	23	5.2476854	0.8897836	NA	NA
Total	24	5.8977098	1.0000000	NA	NA

```
# Summary statistics of duration menopause
sample_data(ps_trans_genus)$GHAT_Length %>%
  summary()
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
10.00 16.00 26.00 33.56 38.00 146.00
```

```
p_gaht_beta = plot_ordination(ps_trans_genus, ord.bc, type="samples", color="GHAT_Length") +
  geom_point(size = 3, alpha = 0.8) +
  labs(color = "Length of GAHT", caption = "Bray-Curtis") +
  scale_color_continuous(limits = c(0,100))
p_gaht_beta
```



There is a significant effect of duration of GAHT on the shannon index value in the trans group and on overall community composition although this seems to depend on a single outlier with

a duration of 150 of hormone therapy. On the genus level, there are 2 taxa (Dialister and Prevotella) who are inversely correlated to duration of testosterone supplementation.

6.0.3 Absolute Height of Testosterone in the trans-men group

```
# Height of testosterone
genus_testosterone = Maaslin2(input_data = otu_table(ps_trans_genus),
    input_metadata = as(sample_data(ps_trans_genus), "data.frame"),
    output = "results/maaslin2/genus_Testosterone",
    min_abundance = 5,
    min_prevalence = 0.3,
    max_significance = 0.2,
    fixed_effects = "Testosterone") # no significant effect
```

```
2024-03-06 21:55:26.228165 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_Testosterone/maaslin2.log"
2024-03-06 21:55:26.401316 INFO::Writing function arguments to log file
2024-03-06 21:55:26.403962 INFO::Verifying options selected are valid
2024-03-06 21:55:26.404253 INFO::Determining format of input files
2024-03-06 21:55:26.404516 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:26.414598 INFO::Formula for fixed effects: expr ~ Testosterone
2024-03-06 21:55:26.415075 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:26.415325 INFO::Total samples in data: 25
2024-03-06 21:55:26.415547 INFO::Min samples required with min abundance for a feature not to be filtered: 5
2024-03-06 21:55:26.417071 INFO::Total filtered features: 314
2024-03-06 21:55:26.417358 INFO::Filtered feature names from abundance and prevalence filtering: 314
2024-03-06 21:55:26.417811 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:26.418036 INFO::Filtered feature names from variance filtering: 0
2024-03-06 21:55:26.418244 INFO::Running selected normalization method: TSS
2024-03-06 21:55:26.418624 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:26.421014 INFO::Running selected transform method: LOG
2024-03-06 21:55:26.421428 INFO::Running selected analysis method: LM
2024-03-06 21:55:26.421704 INFO::Fitting model to feature number 1, ASV2
2024-03-06 21:55:26.422909 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:26.423844 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:26.424763 INFO::Fitting model to feature number 4, ASV6
2024-03-06 21:55:26.425616 INFO::Fitting model to feature number 5, ASV17
2024-03-06 21:55:26.42662 INFO::Fitting model to feature number 6, ASV31
2024-03-06 21:55:26.429313 INFO::Counting total values for each feature
2024-03-06 21:55:26.429992 INFO::Writing filtered data to file results/maaslin2/genus_Testosterone/maaslin2_filtered_data.csv
2024-03-06 21:55:26.43064 INFO::Writing filtered, normalized data to file results/maaslin2/genus_Testosterone/maaslin2_normalized_data.csv
```

```

2024-03-06 21:55:26.431215 INFO::Writing filtered, normalized, transformed data to file results/maaslin2/genus_Testosterone
2024-03-06 21:55:26.43181 WARNING::Deleting existing residuals file: results/maaslin2/genus_Testosterone_residuals
2024-03-06 21:55:26.432248 INFO::Writing residuals to file results/maaslin2/genus_Testosterone_residuals
2024-03-06 21:55:26.432665 WARNING::Deleting existing fitted file: results/maaslin2/genus_Testosterone_fitted
2024-03-06 21:55:26.433018 INFO::Writing fitted values to file results/maaslin2/genus_Testosterone_fitted
2024-03-06 21:55:26.43339 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:26.433946 INFO::Writing the significant results (those which are less than 0.05)
2024-03-06 21:55:26.434276 INFO::Writing heatmap of significant results to file: results/maaslin2/genus_Testosterone_heatmap
[1] "There are no associations to plot!"
2024-03-06 21:55:26.434713 INFO::Writing association plots (one for each significant association)
[1] "There are no associations to plot!"

```

No significant associations.

6.0.4 Duration of GnRH therapy in trans-men group

```
summary(sample_data(ps_trans_genus)$Duration_GNRH) # only 10 values!
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
12.0	18.0	36.0	37.6	56.0	74.0	15

```

genus_gnrh= Maaslin2(input_data = otu_table(ps_trans_genus),
  input_metadata = as(sample_data(ps_trans_genus), "data.frame"),
  output = "results/maaslin2/genus_DurationGNRH",
  min_abundance = 5,
  min_prevalence = 0.3,
  max_significance = 0.2,
  fixed_effects = "Duration_GNRH") # no significant effect

```

```

2024-03-06 21:55:26.440831 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_DurationGNRH/maaslin2.log"
2024-03-06 21:55:26.610636 INFO::Writing function arguments to log file
2024-03-06 21:55:26.613425 INFO::Verifying options selected are valid
2024-03-06 21:55:26.613727 INFO::Determining format of input files
2024-03-06 21:55:26.613957 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:26.615761 INFO::Formula for fixed effects: expr ~ Duration_GNRH
2024-03-06 21:55:26.616037 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:26.61625 INFO::Total samples in data: 25
2024-03-06 21:55:26.616457 INFO::Min samples required with min abundance for a feature not to be dropped: 10

```

```

2024-03-06 21:55:26.617956 INFO::Total filtered features: 314
2024-03-06 21:55:26.618233 INFO::Filtered feature names from abundance and prevalence filter
2024-03-06 21:55:26.618668 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:26.618897 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:26.619124 INFO::Running selected normalization method: TSS
2024-03-06 21:55:26.619514 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:26.621917 INFO::Running selected transform method: LOG
2024-03-06 21:55:26.622293 INFO::Running selected analysis method: LM
2024-03-06 21:55:26.622564 INFO::Fitting model to feature number 1, ASV2
2024-03-06 21:55:26.623597 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:26.624509 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:26.625441 INFO::Fitting model to feature number 4, ASV6
2024-03-06 21:55:26.626364 INFO::Fitting model to feature number 5, ASV17
2024-03-06 21:55:26.627283 INFO::Fitting model to feature number 6, ASV31
2024-03-06 21:55:26.630051 INFO::Counting total values for each feature
2024-03-06 21:55:26.630757 INFO::Writing filtered data to file results/maaslin2/genus_Duration
2024-03-06 21:55:26.631638 INFO::Writing filtered, normalized data to file results/maaslin2/g
2024-03-06 21:55:26.632221 INFO::Writing filtered, normalized, transformed data to file resu
2024-03-06 21:55:26.632804 WARNING::Deleting existing residuals file: results/maaslin2/genus
2024-03-06 21:55:26.633242 INFO::Writing residuals to file results/maaslin2/genus_DurationGNI
2024-03-06 21:55:26.633658 WARNING::Deleting existing fitted file: results/maaslin2/genus_Dur
2024-03-06 21:55:26.63417 INFO::Writing fitted values to file results/maaslin2/genus_Duration
2024-03-06 21:55:26.634673 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:26.635313 INFO::Writing the significant results (those which are less than 0
2024-03-06 21:55:26.635689 INFO::Writing heatmap of significant results to file: results/maas
[1] "There are no associations to plot!"
2024-03-06 21:55:26.636162 INFO::Writing association plots (one for each significant associat
[1] "There are no associations to plot!"

```

```
# no significant associations
```

No significant associations.

6.0.5 Cycle dependency in pre-menopausal group

```

ps_pre = subset_samples(ps_samples, grp == "02_premenopausal") %>%
  transform_sample_counts(function(x) x/sum(x) * 100) # relative abundances are necessary

asv_cycleday = Maaslin2(input_data = otu_table(ps_pre),
  input_metadata = as(sample_data(ps_pre), "data.frame"),

```

```

output = "results/maaslin2/asv_cycle_premeno",
min_abundance = 3,
min_prevalence = 0.1,
max_significance = 0.2,
fixed_effects = "CycleDaySampling") # no significant effect

```

```

2024-03-06 21:55:26.669099 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/asv_cycle_premeno/maaslin2.log"
2024-03-06 21:55:28.054718 INFO::Writing function arguments to log file
2024-03-06 21:55:28.057068 INFO::Verifying options selected are valid
2024-03-06 21:55:28.057314 INFO::Determining format of input files
2024-03-06 21:55:28.057537 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:28.067359 INFO::Formula for fixed effects: expr ~ CycleDaySampling
2024-03-06 21:55:28.067808 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:28.068049 INFO::Total samples in data: 25
2024-03-06 21:55:28.068256 INFO::Min samples required with min abundance for a feature not to be filtered: 3
2024-03-06 21:55:28.078422 INFO::Total filtered features: 2551
2024-03-06 21:55:28.079126 INFO::Filtered feature names from abundance and prevalence filtering:
2024-03-06 21:55:28.079991 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:28.08023 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:28.080445 INFO::Running selected normalization method: TSS
2024-03-06 21:55:28.080833 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:28.08328 INFO::Running selected transform method: LOG
2024-03-06 21:55:28.08373 INFO::Running selected analysis method: LM
2024-03-06 21:55:28.084009 INFO::Fitting model to feature number 1, ASV1
2024-03-06 21:55:28.085195 INFO::Fitting model to feature number 2, ASV2
2024-03-06 21:55:28.086225 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:28.087095 INFO::Fitting model to feature number 4, ASV7
2024-03-06 21:55:28.087934 INFO::Fitting model to feature number 5, ASV8
2024-03-06 21:55:28.088801 INFO::Fitting model to feature number 6, ASV11
2024-03-06 21:55:28.089656 INFO::Fitting model to feature number 7, ASV14
2024-03-06 21:55:28.09052 INFO::Fitting model to feature number 8, ASV16
2024-03-06 21:55:28.091348 INFO::Fitting model to feature number 9, ASV79
2024-03-06 21:55:28.094142 INFO::Counting total values for each feature
2024-03-06 21:55:28.09483 INFO::Writing filtered data to file results/maaslin2/asv_cycle_premeno/maaslin2_filtered.csv
2024-03-06 21:55:28.095487 INFO::Writing filtered, normalized data to file results/maaslin2/asv_cycle_premeno/maaslin2_normalized.csv
2024-03-06 21:55:28.096046 INFO::Writing filtered, normalized, transformed data to file results/maaslin2/asv_cycle_premeno/maaslin2_transformed.csv
2024-03-06 21:55:28.096802 WARNING::Deleting existing residuals file: results/maaslin2/asv_cycle_premeno/residuals.csv
2024-03-06 21:55:28.097209 INFO::Writing residuals to file results/maaslin2/asv_cycle_premeno/residuals.csv
2024-03-06 21:55:28.09762 WARNING::Deleting existing fitted file: results/maaslin2/asv_cycle_premeno/fitted.csv
2024-03-06 21:55:28.097947 INFO::Writing fitted values to file results/maaslin2/asv_cycle_premeno/fitted.csv
2024-03-06 21:55:28.098306 INFO::Writing all results to file (ordered by increasing q-values): results/maaslin2/asv_cycle_premeno/maaslin2_results.csv

```

```

2024-03-06 21:55:28.098855 INFO::Writing the significant results (those which are less than 0.05)
2024-03-06 21:55:28.099252 INFO::Writing heatmap of significant results to file: results/maaslin2/maaslin2_heatmap.png
[1] "There is not enough metadata in the associations to create a heatmap plot. Please review the metadata."
2024-03-06 21:55:28.099809 INFO::Writing association plots (one for each significant association)
2024-03-06 21:55:28.100372 INFO::Plotting associations from most to least significant, grouped by metadata number
2024-03-06 21:55:28.100666 INFO::Plotting data for metadata number 1, CycleDaySampling
2024-03-06 21:55:28.102588 INFO::Creating scatter plot for continuous data, CycleDaySampling

```

```

2024-03-06 21:55:28.164456 INFO::Creating scatter plot for continuous data, CycleDaySampling

```

```

Warning: Removed 1 rows containing missing values (`geom_point()`).

```

```

2024-03-06 21:55:28.240402 INFO::Creating scatter plot for continuous data, CycleDaySampling

```

```

Warning: Removed 1 rows containing missing values (`geom_point()`).

```

```

Removed 1 rows containing missing values (`geom_point()`).

```

```

Removed 1 rows containing missing values (`geom_point()`).

```

```

ps_pre_genus = tax_glom(ps_pre, "Genus")

genus_cycleday = Maaslin2(input_data = otu_table(ps_pre_genus),
  input_metadata = as(sample_data(ps_pre_genus), "data.frame"),
  output = "results/maaslin2/genus_cycle_premeno",
  min_abundance = 5,
  min_prevalence = 0.3,
  max_significance = 0.2,
  fixed_effects = "CycleDaySampling")

```

```

2024-03-06 21:55:32.44491 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_cycle_premeno/maaslin2.log"
2024-03-06 21:55:32.630249 INFO::Writing function arguments to log file
2024-03-06 21:55:32.632428 INFO::Verifying options selected are valid
2024-03-06 21:55:32.632687 INFO::Determining format of input files
2024-03-06 21:55:32.632912 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:32.634715 INFO::Formula for fixed effects: expr ~ CycleDaySampling
2024-03-06 21:55:32.634992 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:32.635205 INFO::Total samples in data: 25
2024-03-06 21:55:32.635407 INFO::Min samples required with min abundance for a feature not to be filtered: 5
2024-03-06 21:55:32.636906 INFO::Total filtered features: 318
2024-03-06 21:55:32.637185 INFO::Filtered feature names from abundance and prevalence filter:

```

```

2024-03-06 21:55:32.63761 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:32.637839 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:32.63807 INFO::Running selected normalization method: TSS
2024-03-06 21:55:32.63843 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:32.640737 INFO::Running selected transform method: LOG
2024-03-06 21:55:32.641087 INFO::Running selected analysis method: LM
2024-03-06 21:55:32.641348 INFO::Fitting model to feature number 1, ASV1
2024-03-06 21:55:32.642362 INFO::Fitting model to feature number 2, ASV4
2024-03-06 21:55:32.644782 INFO::Counting total values for each feature
2024-03-06 21:55:32.645379 INFO::Writing filtered data to file results/maaslin2/genus_cycle_1
2024-03-06 21:55:32.646027 INFO::Writing filtered, normalized data to file results/maaslin2/genus_cycle_1
2024-03-06 21:55:32.646571 INFO::Writing filtered, normalized, transformed data to file results/maaslin2/genus_cycle_1
2024-03-06 21:55:32.647074 WARNING::Deleting existing residuals file: results/maaslin2/genus_cycle_1
2024-03-06 21:55:32.647491 INFO::Writing residuals to file results/maaslin2/genus_cycle_premaaslin2
2024-03-06 21:55:32.647907 WARNING::Deleting existing fitted file: results/maaslin2/genus_cycle_premaaslin2
2024-03-06 21:55:32.648248 INFO::Writing fitted values to file results/maaslin2/genus_cycle_premaaslin2
2024-03-06 21:55:32.648589 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:32.649106 INFO::Writing the significant results (those which are less than 0.2)
2024-03-06 21:55:32.649475 INFO::Writing heatmap of significant results to file: results/maaslin2/genus_cycle_premaaslin2
[1] "There is not enough metadata in the associations to create a heatmap plot. Please review the metadata."
2024-03-06 21:55:32.649998 INFO::Writing association plots (one for each significant association)
2024-03-06 21:55:32.650497 INFO::Plotting associations from most to least significant, grouped by metadata number
2024-03-06 21:55:32.650773 INFO::Plotting data for metadata number 1, CycleDaySampling
2024-03-06 21:55:32.651284 INFO::Creating scatter plot for continuous data, CycleDaySampling

```

Warning: Removed 1 rows containing missing values (`geom_point()`).

```

2024-03-06 21:55:32.715025 INFO::Creating scatter plot for continuous data, CycleDaySampling

```

Warning: Removed 1 rows containing missing values (`geom_point()`).

Removed 1 rows containing missing values (`geom_point()`).

Removed 1 rows containing missing values (`geom_point()`).

```

tax_table(ps_pre_genus) %>%
  as.data.frame() %>%
  rownames_to_column("ASV") %>%
  filter(ASV %in% filter(genus_cycleday$results, qval < 0.2)$feature) %>%
  kable()

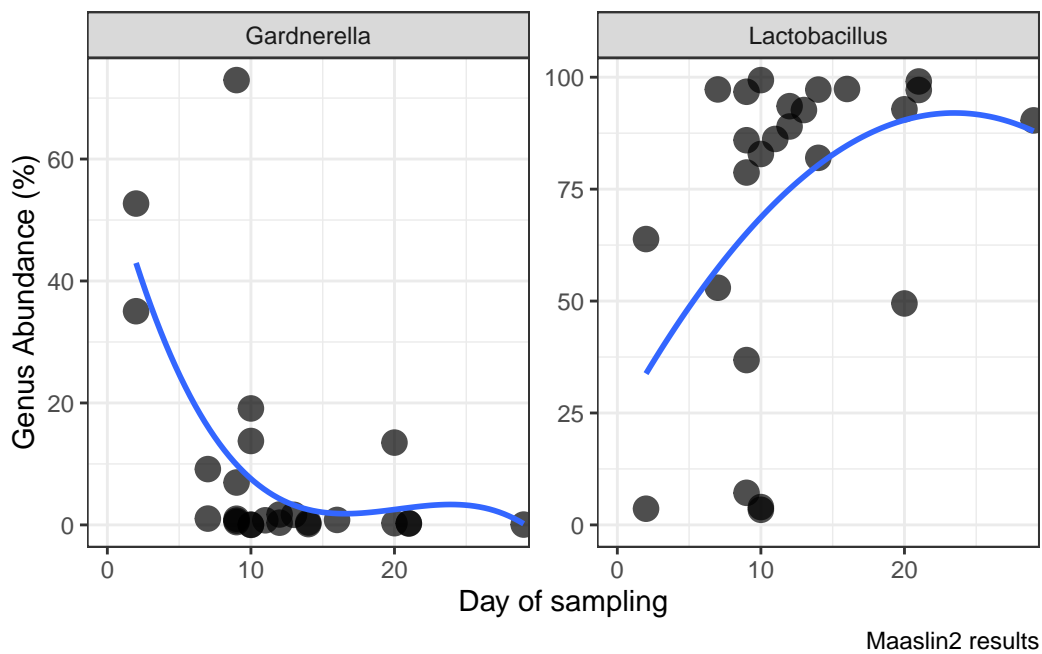
```


ASV	Kingdom	Phylum	Class	Order	Family	Genus	Species
ASV1	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	NA
ASV4	Bacteria	Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella	NA

```
psmelt_pre_genus = psmelt(ps_pre_genus)

maaslin_res_genus_cycleday = ggplot(filter(psmelt_pre_genus, OTU %in% filter(genus_cycleday,
aes(x = CycleDaySampling, y = Abundance)) +
  geom_point(size = 4, alpha = 0.7) +
  facet_wrap(~Genus, scales = "free") +
  geom_smooth(method = "lm", se = F, formula = y ~poly(x, 3, raw =T)) +
  scale_y_continuous(limits = function(x){c(0, max(0.1, x))}) +
  coord_cartesian(xlim = c(0,28)) +
  labs(x = "Day of sampling",
       y = "Genus Abundance (%)",
       caption = "Maaslin2 results")
```

```
maaslin_res_genus_cycleday
```



```
ggsave(maaslin_res_genus_cycleday, filename = "results/maaslin2/genus_CycleDay_panelplot.pdf",
       height = 5, width = 6)
```

Gardnerella is highest in the early days of the menstrual cycle while *Lactobacillus* during the middle of the cycle. This seems to correspond to *Vomstein et al 2022*.

6.0.6 Influence of sexual activity on microbiome composition

```
# Sexual activity
ps_rel_genus = tax_glom(ps_samples, taxrank = "Genus") %>%
  transform_sample_counts(function(x) x/sum(x) * 100)
ps_rel_genus_sexuallyactive = subset_samples(ps_rel_genus, sample_data(ps_rel_genus)$SexuallyActive == "Active")

ps_abs_genus = tax_glom(ps_samples, taxrank = "Genus")
ps_abs_genus_sexuallyactive = subset_samples(ps_abs_genus, sample_data(ps_abs_genus)$SexuallyActive == "Active")

ps_rel_genus_sexuallyactive
```

phyloseq-class experiment-level object

```
otu_table() OTU Table: [ 320 taxa and 72 samples ]
sample_data() Sample Data: [ 72 samples by 15 sample variables ]
tax_table() Taxonomy Table: [ 320 taxa by 7 taxonomic ranks ]
phy_tree() Phylogenetic Tree: [ 320 tips and 319 internal nodes ]
refseq() DNASTringSet: [ 320 reference sequences ]
```

```
# Via Maaslin2
genus_SexuallyActive = Maaslin2(input_data = otu_table(ps_rel_genus_sexuallyactive),
                                input_metadata = as(sample_data(ps_rel_genus_sexuallyactive), "data.frame"),
                                output = "results/maaslin2/genus_SexuallyActive",
                                min_abundance = 5,
                                min_prevalence = 0.2,
                                max_significance = 0.2,
                                fixed_effects = "SexuallyActive")
```

```
2024-03-06 21:55:42.065702 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_SexuallyActive/maaslin2.log"
2024-03-06 21:55:42.239165 INFO::Writing function arguments to log file
2024-03-06 21:55:42.241843 INFO::Verifying options selected are valid
2024-03-06 21:55:42.242105 INFO::Determining format of input files
```

```

2024-03-06 21:55:42.242343 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:42.244332 INFO::Formula for fixed effects: expr ~ SexuallyActive
2024-03-06 21:55:42.244648 INFO::Factor detected for categorical metadata 'SexuallyActive'. Prevalence filter applied
2024-03-06 21:55:42.244867 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:42.245081 INFO::Total samples in data: 72
2024-03-06 21:55:42.24528 INFO::Min samples required with min abundance for a feature not to be filtered: 1
2024-03-06 21:55:42.246855 INFO::Total filtered features: 316
2024-03-06 21:55:42.247142 INFO::Filtered feature names from abundance and prevalence filtering: 316
2024-03-06 21:55:42.247557 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:42.247783 INFO::Filtered feature names from variance filtering: 0
2024-03-06 21:55:42.24799 INFO::Running selected normalization method: TSS
2024-03-06 21:55:42.248381 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:42.250527 INFO::Running selected transform method: LOG
2024-03-06 21:55:42.250893 INFO::Running selected analysis method: LM
2024-03-06 21:55:42.25115 INFO::Fitting model to feature number 1, ASV1
2024-03-06 21:55:42.25236 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:42.253393 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:42.254371 INFO::Fitting model to feature number 4, ASV6
2024-03-06 21:55:42.257269 INFO::Counting total values for each feature
2024-03-06 21:55:42.257975 INFO::Writing filtered data to file results/maaslin2/genus_SexuallyActive/filtered_data.csv
2024-03-06 21:55:42.259249 INFO::Writing filtered, normalized data to file results/maaslin2/genus_SexuallyActive/normalized_data.csv
2024-03-06 21:55:42.260266 INFO::Writing filtered, normalized, transformed data to file results/maaslin2/genus_SexuallyActive/transformed_data.csv
2024-03-06 21:55:42.261304 WARNING::Deleting existing residuals file: results/maaslin2/genus_SexuallyActive/residuals.csv
2024-03-06 21:55:42.261869 INFO::Writing residuals to file results/maaslin2/genus_SexuallyActive/residuals.csv
2024-03-06 21:55:42.26235 WARNING::Deleting existing fitted file: results/maaslin2/genus_SexuallyActive/fitted_values.csv
2024-03-06 21:55:42.262927 INFO::Writing fitted values to file results/maaslin2/genus_SexuallyActive/fitted_values.csv
2024-03-06 21:55:42.263578 INFO::Writing all results to file (ordered by increasing q-values): results/maaslin2/genus_SexuallyActive/all_results.csv
2024-03-06 21:55:42.264242 INFO::Writing the significant results (those which are less than 0.2) to file: results/maaslin2/genus_SexuallyActive/significant_results.csv
2024-03-06 21:55:42.26464 INFO::Writing heatmap of significant results to file: results/maaslin2/genus_SexuallyActive/heatmap.csv
[1] "There is not enough metadata in the associations to create a heatmap plot. Please review the metadata and associations."
2024-03-06 21:55:42.265181 INFO::Writing association plots (one for each significant association) to file: results/maaslin2/genus_SexuallyActive/association_plots/
2024-03-06 21:55:42.265778 INFO::Plotting associations from most to least significant, grouped by metadata
2024-03-06 21:55:42.266072 INFO::Plotting data for metadata number 1, SexuallyActive
2024-03-06 21:55:42.266629 INFO::Creating boxplot for categorical data, SexuallyActive vs ASV1
2024-03-06 21:55:42.331569 INFO::Creating boxplot for categorical data, SexuallyActive vs ASV3

```

```

tax_table(ps_rel_genus_sexuallyactive) %>%
  as.data.frame() %>%
  rownames_to_column("ASV") %>%
  filter(ASV %in% filter(genus_SexuallyActive$results, qval < 0.2)$feature)

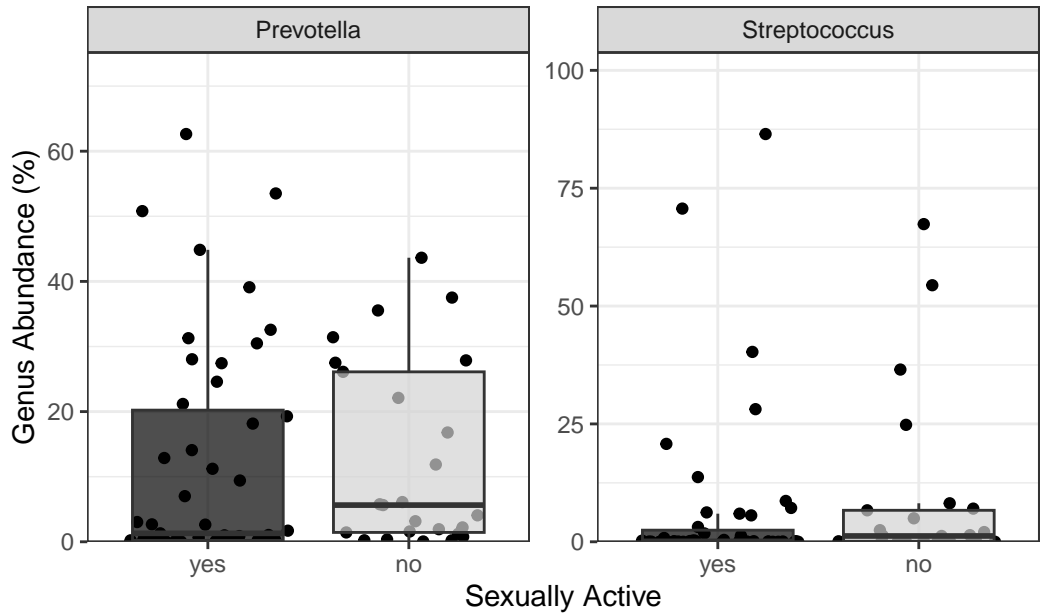
```

	ASV	Kingdom	Phylum	Class	Order	Family
1	ASV3	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae
2	ASV6	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae
		Genus	Species			
1		Prevotella	<NA>			
2		Streptococcus	<NA>			

```
# plot sexuallyActive vs these 2 genera
psmelt_rel_genus_sexuallyactive = psmelt(ps_rel_genus_sexuallyactive)
maaslin_res_genus_sexactive = ggplot(filter(psmelt_rel_genus_sexuallyactive, OTU %in% filt
                                     aes(x = SexuallyActive, y = Abundance, fill = SexuallyActiv

geom_jitter() +
geom_boxplot(alpha = 0.7, outlier.shape = NA) +
facet_wrap(~Genus, scales = "free") +
scale_y_continuous(limits = function(x){c(0, max(0.1, x))}) +
theme(legend.position = "none")+
labs(x = "Sexually Active",
     y = "Genus Abundance (%)",
     caption = "Maaslin2 results")+
scale_y_continuous(expand = expansion(mult = c(0, 0.2))) +
scale_fill_manual(values = c("black", "lightgrey"))

maaslin_res_genus_sexactive
```



Maaslin2 results

```
ggsave(maaslin_res_genus_sexactive, filename = "results/maaslin2/genus_SexuallyActive_pane
      height = 3, width = 4)
```

```
# Via Deseq2
sexactive_ds = phyloseq_to_deseq2(ps_abs_genus_sexuallyactive, ~SexuallyActive)
cts = counts(sexactive_ds)
geoMeans = apply(cts, 1, function(row) if (all(row == 0)) 0 else exp(mean(log(row[row != 0])))
sexactive_ds = estimateSizeFactors(sexactive_ds, geoMeans=geoMeans)
sexactive_ds = DESeq(sexactive_ds, test = "Wald", fitType = "parametric")
res = results(sexactive_ds, cooksCutoff = FALSE)
alpha = 0.1
sigtab = res[which(res$padj < alpha), ]
sigtab = cbind(as(sigtab, "data.frame"), as(tax_table(ps_abs_genus_sexuallyactive)[rownames(sigtab), ]))
sigtab = as.data.frame(sigtab) %>%
  rownames_to_column("feature")
```

```
sigtab
```

	feature	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
1	ASV119	38.25813	-4.983385	1.344519	-3.706444	0.0002101895	0.0540187
	Kingdom	Phylum	Class	Order	Family	Genus	Species
1	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	HT002	<NA>

```

# Wilcoxon
psmelt_rel_genus_sexuallyactive %>%
  group_by(OTU) %>%
  rstatix::t_test(data = ., Abundance~SexuallyActive) %>%
  rstatix::adjust_pvalue(method = "fdr") %>%
  filter(p.adj < 0.1)

# A tibble: 0 x 10
# i 10 variables: OTU <chr>, .y. <chr>, group1 <chr>, group2 <chr>, n1 <int>,
#   n2 <int>, statistic <dbl>, df <dbl>, p <dbl>, p.adj <dbl>

# using simple parametric statistical test (t-test) no significant diff.

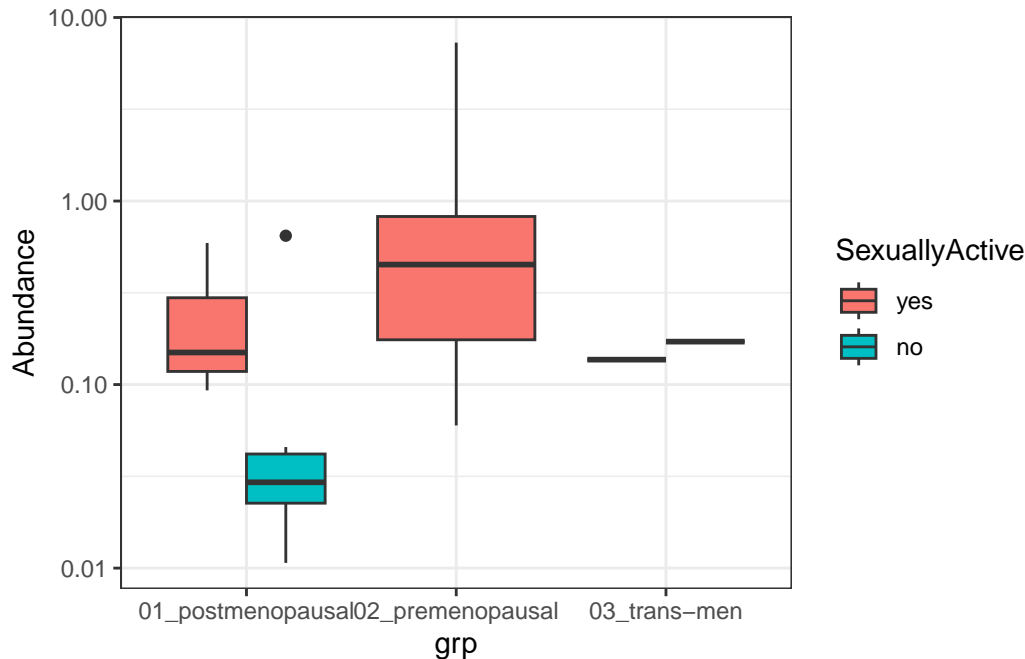
# generate a plot of HT002 abundance per group
data_rel = transform_sample_counts(ps_abs_genus_sexuallyactive, function(x) x/sum(x)*100)
data_psmelt = psmelt(data_rel)

filter(data_psmelt, OTU %in% sigtab$feature) %>%
  ggplot(aes(x = grp, y = Abundance, fill = SexuallyActive)) +
  geom_boxplot(position = position_dodge()) +
  scale_y_log10()

```

Warning: Transformation introduced infinite values in continuous y-axis

Warning: Removed 44 rows containing non-finite values (`stat_boxplot()`).



Although Maaslin2 detects significant changes in some taxa, this might not be statistically relevant due to the extremely skewed distribution of data and non-suitability of a linear model for comparison of 2 groups. Likewise, using a simple parametric statistical test such as the t-test is sub-optimal. This approach fails to detect any significantly altered taxa.

However, Genus *HT002* of family Lactobacillaceae reaches borderline significance ($q = 0.054$). Some studies use a p-value cutoff of $q < 0.01$ for these analyses. However, this might be confounded by all patients in the premenopausal group being sexually active (i.e. association with premenopausal group per se and not with sexual activity). Next, we analyse and describe the taxonomic composition within the *trans-men* and control groups by sexual activity. There are no significant statistical differences and the sample size is small. Therefore, the top 10 abundant taxa are listed for sexually active and non-active subgroups:

```
# Trans Group taxa selection
ps_rel_genus_sexuallyactive_trans = subset_samples(ps_rel_genus_sexuallyactive, grp == "03_
top <- names(sort(taxa_sums(ps_rel_genus_sexuallyactive_trans), decreasing = TRUE))[1:20]
psmelt_rel_genus_sexuallyactive_trans.top <- prune_taxa(top, psmelt_rel_genus_sexuallyactive_t
psmelt() %>%
mutate(tax = paste(Phylum, Class, Order, Family, Genus, sep = "|")) %>%
group_by(tax, SexuallyActive)

# Plot
```

```

genus_sexactive_trans_top20 = ggplot(psmelt_rel_genus_sexuallyactive_trans.top, aes(x = tax,
  geom_boxplot(position="dodge", alpha = 0.5) +
  geom_point(aes(color = SexuallyActive), position = position_dodge(width = 0.7), alpha =
  scale_fill_manual(values = c("black", "lightgrey")) +
  scale_color_manual(values = c("black", "lightgrey")) +
  coord_flip() +
  theme_bw() +
  lims(y = c(0,100)) +
  labs(title = "Trans-men group", y = "Abundance (%)", x = "", color = "Sexually Active",

# Summarize
Genus_Sexuallyactive_Trans_top20 = psmelt_rel_genus_sexuallyactive_trans.top %>%
  summarize(mean = mean(Abundance, na.rm=T),
    median = median(Abundance, na.rm=T),
    sd = sd(Abundance, na.rm=T),
    IQR = IQR(Abundance ,na.rm=T))

knitr::kable(Genus_Sexuallyactive_Trans_top20, digits = 2)

```

tax				Sexually Active	mean	median	sd	IQR
Actinobacteriota	Actinobacteria	Actinomycetales	Actinomycetales	Actinomyces	0.89	0.22	1.58	0.78
Actinobacteriota	Actinobacteria	Actinomycetales	Actinomycetales	Actinomyces	0.66	0.24	1.09	0.60
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriales	Allosciuria	1.78	0.12	5.92	0.12
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriales	Allosciuria	0.52	0.13	0.77	0.82
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriales	Gardnerella	16.16	1.80	21.54	28.99
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriales	Gardnerella	7.75	1.08	14.65	4.95
Actinobacteriota	Actinobacteria	Corynebacteriales	Corynebacteriales	Corynebacterium	1.98	0.32	3.48	1.38
Actinobacteriota	Actinobacteria	Corynebacteriales	Corynebacteriales	Corynebacterium	0.91	0.06	2.84	5.17
Bacteroidota	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	4.02	2.37	6.20	3.88
Bacteroidota	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	5.72	2.18	11.20	2.15
Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	25.68	24.59	19.97	23.16
Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	23.56	24.81	13.84	21.41
Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	0.02	0.00	0.05	0.00
Bacteroidota	Bacteroidia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	4.37	0.00	13.82	0.00
Campylobacterota	Campylobacteria	Campylobacterales	Campylobacteraceae	Campylobacter	0.68	0.28	10.88	4.41
Campylobacterota	Campylobacteria	Campylobacterales	Campylobacteraceae	Campylobacter	2.03	0.83	13.72	0.87
Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	4.85	0.25	7.10	7.75
Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	8.24	0.24	16.71	7.14
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	9.01	1.30	19.43	6.11
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	2.50	0.80	3.36	5.32

tax	SexuallyActive	mean	median	sd	IQR
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Anaerococcus	yes	3.35	2.51	2.84	3.61
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Anaerococcus	no	6.36	4.88	6.74	5.60
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Fenollaria	yes	1.61	0.45	2.40	2.10
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Fenollaria	no	2.65	2.64	2.24	3.37
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Finegoldia	yes	0.91	0.52	1.13	0.63
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Finegoldia	no	1.17	0.65	1.62	0.76
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Peptoniphilus	yes	2.38	2.26	2.26	2.56
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Peptoniphilus	no	3.06	3.01	1.97	3.48
Firmicutes Negativicutes Veillonellales-Selenomonadales Veillonellaceae Dialister	yes	4.49	2.69	4.52	5.67
Firmicutes Negativicutes Veillonellales-Selenomonadales Veillonellaceae Dialister	no	7.70	6.00	8.07	7.75
Firmicutes Negativicutes Veillonellales-Selenomonadales Veillonellaceae Veillonella	yes	1.55	0.26	2.90	1.68
Firmicutes Negativicutes Veillonellales-Selenomonadales Veillonellaceae Veillonella	no	0.18	0.00	0.38	0.21
Fusobacteriota Fusobacteriia Fusobacteriales Fusobacteriaceae Fusobacterium	yes	1.32	0.32	1.72	2.01
Fusobacteriota Fusobacteriia Fusobacteriales Fusobacteriaceae Fusobacterium	no	3.07	0.11	7.55	1.56
Fusobacteriota Fusobacteriia Fusobacteriales Leptotrichiaceae Synechthia	yes	4.28	0.12	9.65	3.56
Fusobacteriota Fusobacteriia Fusobacteriales Leptotrichiaceae Synechthia	no	0.79	0.00	1.71	0.00
Proteobacteria Alphaproteobacteria Rhizobiales Beijerinckiaceae Methylobacterium	yes	0.01	0.00	0.00	0.00
Proteobacteria Alphaproteobacteria Rhizobiales Beijerinckiaceae Methylobacterium	no	2.92	0.00	8.26	0.00
Proteobacteria Gammaproteobacteria Burkholderiales Neisseriaceae Neisseria	yes	3.13	0.00	11.12	0.06
Proteobacteria Gammaproteobacteria Burkholderiales Neisseriaceae Neisseria	no	0.17	0.00	0.19	0.01

```
# Postmenopausal group
```

```
ps_rel_genus_sexuallyactive_post = subset_samples(ps_rel_genus_sexuallyactive, grp == "01_p
top2 <- names(sort(taxa_sums(ps_rel_genus_sexuallyactive_post), decreasing = TRUE))[1:20]
psmelt_rel_genus_sexuallyactive_post.top <- prune_taxa(top2, ps_rel_genus_sexuallyactive_p
```

```

psmelt() %>%
mutate(tax = paste(Phylum, Class, Order, Family, Genus, sep = "|")) %>%
group_by(tax, SexuallyActive)

# Plot
genus_sexactive_post_top20 = ggplot(psmelt_rel_genus_sexuallyactive_post.top, aes(x = tax,
geom_boxplot(position="dodge", alpha = 0.5) +
geom_point(aes(color = SexuallyActive), position = position_dodge(width = 0.7), alpha =
scale_fill_manual(values = c("black", "lightgrey"))) +
scale_color_manual(values = c("black", "lightgrey"))) +
coord_flip() +
theme_bw() +
lims(y = c(0,100)) +
labs(title = "Postmenopausal group", y = "Abundance (%)", x = "", color = "Sexually Active"))

ggsave(filename = "results/Genus_SexuallyActive_Descriptive_Postmenopausal.pdf", plot = genus_sexactive_post_top20,
height = 6, width = 10)
ggsave(filename = "results/Genus_SexuallyActive_Descriptive_Trans.pdf", plot = genus_sexactive_post_top20,
height = 6, width = 10)

# Summarize
Genus_Sexuallyactive_Post_top20 = psmelt_rel_genus_sexuallyactive_post.top %>%
summarize(mean = mean(Abundance, na.rm=T),
median = median(Abundance, na.rm=T),
sd = sd(Abundance, na.rm=T),
IQR = IQR(Abundance ,na.rm=T))

knitr::kable(Genus_Sexuallyactive_Post_top20, digits = 2)

```

tax				SexuallyActive	mean	median	sd	IQR
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Allosciadon	147	10.20	3.02	1.74
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Allosciadon	319	0.06	6.33	0.32
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	633	0.10	15.73	0.18
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	35	0.19	5.76	0.26
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella	1570	1.17	25.04	14.28
Actinobacteriota	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	Gardnerella	256	0.26	4.73	1.79
Actinobacteriota	Actinobacteria	Corynebacteriales	Corynebacteriaceae	Corynebacterium	81	0.53	0.93	0.90
Actinobacteriota	Actinobacteria	Corynebacteriales	Corynebacteriaceae	Corynebacterium	10	0.17	7.84	5.04
Actinobacteriota	Coriobacteriia	Coriobacteriales	Atopobiaceae	Atopobium	96	1.33	7.27	4.11
Actinobacteriota	Coriobacteriia	Coriobacteriales	Atopobiaceae	Atopobium	46	0.23	11.00	1.47
Bacteroidota	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	388	0.06	6.98	4.58

tax	Sexually Active				mean	median	sd	IQR
Bacteroidota Bacteroidia Bacteroidales Porphyromonadaceae Porphyromonas	yes	0.48	0.14	0.83	0.50			
Bacteroidota Bacteroidia Bacteroidales Prevotellaceae Prevotella	yes	17.77	11.21	18.03	30.23			
Bacteroidota Bacteroidia Bacteroidales Prevotellaceae Prevotella	no	5.30	1.58	9.00	4.29			
Bacteroidota Bacteroidia Bacteroidales Prevotellaceae Prevotella_9	yes	0.00	0.00	0.00	0.00			
Bacteroidota Bacteroidia Bacteroidales Prevotellaceae Prevotella_9	no	0.98	0.00	3.15	0.00			
Campylobacterota Campylobacteria Campylobacterales Campylobacteraceae Campylobacter	yes	5.51	0.00	2.44	2.32			
Campylobacterota Campylobacteria Campylobacterales Campylobacteraceae Campylobacter	no	0.27	0.00	0.58	0.40			
Firmicutes Bacilli Lactobacillales Lactobacillaceae Lactobacillus	yes	23.50	6.49	33.92	36.76			
Firmicutes Bacilli Lactobacillales Lactobacillaceae Lactobacillus	no	30.61	10.26	36.18	55.70			
Firmicutes Bacilli Lactobacillales Streptococcaceae Streptococcus	yes	5.67	0.81	9.56	5.90			
Firmicutes Bacilli Lactobacillales Streptococcaceae Streptococcus	no	12.99	1.23	22.29	14.78			
Firmicutes Bacilli Staphylococcales Gemellaceae Gemella	yes	0.06	0.00	0.14	0.05			
Firmicutes Bacilli Staphylococcales Gemellaceae Gemella	no	1.90	0.08	4.82	0.34			
Firmicutes Bacilli Staphylococcales Staphylococcaceae Staphylococcus	yes	0.07	0.06	0.08	0.10			
Firmicutes Bacilli Staphylococcales Staphylococcaceae Staphylococcus	no	9.34	0.05	24.53	0.22			
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Anaerococcus	yes	3.43	2.06	5.29	2.71			
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Anaerococcus	no	2.74	0.81	4.63	2.72			
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Ezakiella	yes	1.09	0.11	1.86	1.47			
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Ezakiella	no	0.28	0.08	0.51	0.33			
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Finegoldia	yes	2.50	0.69	3.95	2.03			
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Finegoldia	no	3.65	0.58	5.73	3.52			
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Peptoniphilus	yes	2.40	1.56	3.03	3.27			
Firmicutes Clostridia Peptostreptococcales-Tissierellales Family XI Peptoniphilus	no	1.50	0.61	2.08	1.35			
Firmicutes Negativicutes Veillonellales-Selenomonadales Veillonellaceae Dialister	yes	1.99	0.07	3.06	4.13			
Firmicutes Negativicutes Veillonellales-Selenomonadales Veillonellaceae Dialister	no	0.55	0.36	0.59	0.84			
Fusobacteriota Fusobacteriia Fusobacteriales Fusobacteriaceae Fusobacterium	yes	1.13	0.06	2.23	0.11			
Fusobacteriota Fusobacteriia Fusobacteriales Fusobacteriaceae Fusobacterium	no	1.29	0.00	0.54	0.29			
Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Escherichia	yes	0.33	0.00	0.83	0.05			
Shigella								

Proteobacteria Gammaproteobacteria Enterobacterales Enterobacteriaceae Escherichia	5.49	0.00	16.61	1.95
Shigella				



genus_sexactive_post_top20



```
write_csv(Genus_Sexuallyactive_Post_top20, file = "results/Genus_SexuallyActive_Descriptiv
write_csv(Genus_Sexuallyactive_Trans_top20, file = "results/Genus_SexuallyActive_Descripti

# Analysis in Trans group with male sexual partner only
genus_Sexofpartner = psmelt_rel_genus_sexuallyactive_trans.top %>%
  filter(!is.na(SexOfPartner)) %>%
  mutate(SexOfPartner2 = ifelse(SexOfPartner %in% c("M", "M+F"), "M", "F only"))
genus_sexactive_trans_malepartner_top20 = ggplot(genus_Sexofpartner,
  aes(x = tax, y = Abundance, fill = SexuallyActive)) +
  geom_boxplot(position="dodge", alpha = 0.5) +
  geom_point(aes(color = SexuallyActive), position = position_dodge(width = 0.7), alpha =
  scale_fill_manual(values = c("black", "lightgrey")) +
  scale_color_manual(values = c("black", "lightgrey")) +
  coord_flip() +
  theme_bw() +
  facet_wrap(~SexOfPartner2) +
  lims(y = c(0,100)) +
  labs(title = "Trans-men group", y = "Abundance (%)", x = "", color = "Sexually Active",
ggsave(filename = "results/Genus_SexuallyActive_Descriptive_Trans_SexOfPartner.pdf",
  genus_sexactive_trans_malepartner_top20,
```

```
height = 8, width = 11)
```

6.0.7 Influence of the duration of amenorrhea on microbiome composition

Both postmenopausal women as well as trans men undergoing GAHT experience amenorrhea. This section explores a potential interaction between time since last menstruation and composition of the vaginal microbiome.

This analysis is performed within each group separately and results are compared.

```
# generate relevant phyloseq object
ps_menopause = subset_samples(ps_samples, grp != "02_premenopausal")

sample_data(ps_menopause)$DurationMenopause # this is the duration of menopause in postmen

[1] 11  2 10 22 12 22  4 28  7 26 11  2 18 20 17 24 23 19 23 27 13  1 13  4  3
[26] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA

sample_data(ps_menopause)$DurationAmenorrhea # duration of amenorrhea in trans group (mont

[1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA
[26] 19 42  2 33 17 35  6  2 37 14  0 60 46  4 26  9 16 27  0 74  9  1 35 16 18

sample_data(ps_menopause)$DurationMenopauseMonths = sample_data(ps_menopause)$DurationMenopause
sample_data(ps_menopause)$DurationCombined = ifelse(is.na(sample_data(ps_menopause)$DurationMenopauseMonths),
  sample_data(ps_menopause)$DurationAmenorrhea, sample_data(ps_menopause)$DurationMenopauseMonths)

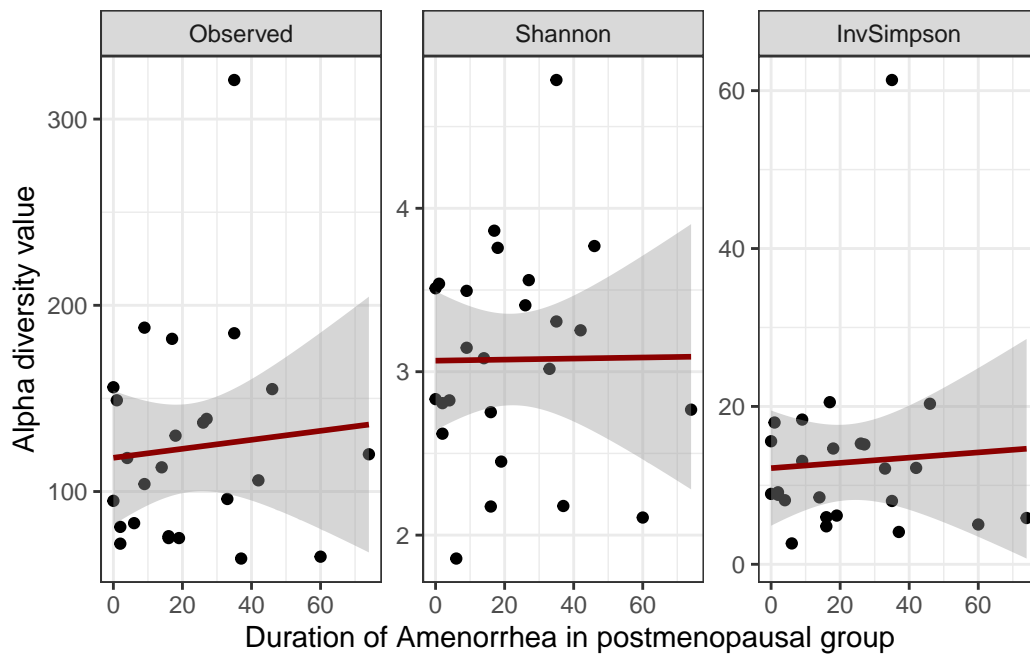
ps_menopause_rel = transform_sample_counts(ps_menopause, function(x) x/sum(x) * 100)
# Diversity analysis
adiv_menopause = plot_richness(ps_menopause, measures = c("Observed", "Shannon", "InvSimpson"))

adiv_menopause = adiv_menopause$data %>%
  select(samples, grp, DurationMenopause, DurationAmenorrhea, DurationCombined, variable, value)

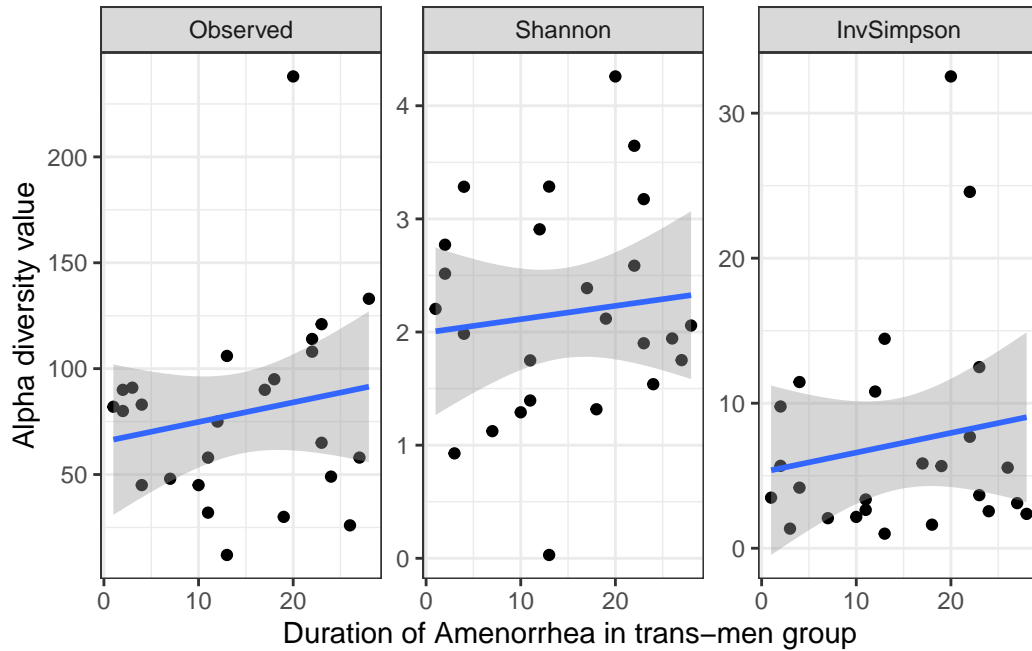
corr_trans = filter(adiv_menopause, grp == "03_trans-men") %>%
  group_by(variable) %>%
  rstatix::cor_test(value, DurationAmenorrhea)
kable(corr_trans)
```

variable	var1	var2	cor	statistic	p	conf.low	conf.high	method
Observed	value	DurationAmenorrhea	0.40840	0.4031351	0.691	-	0.4635524	Pearson
						0.3220252		
Shannon	value	DurationAmenorrhea	0.6095	0.0457909	0.964	-	0.4031576	Pearson
						0.3870434		
InvSimpson	value	DurationAmenorrhea	0.6570	0.2738670	0.787	-	0.4421820	Pearson
						0.3459110		

```
ggplot(filter(adiv_menopause, grp == "03_trans-men"), aes(x = DurationAmenorrhea, y = value)) +
  geom_point() +
  geom_smooth(method = "lm", color = "darkred") +
  facet_wrap(~variable, scales = "free") +
  labs(x = "Duration of Amenorrhea in postmenopausal group", y = "Alpha diversity value")
```



```
ggplot(filter(adiv_menopause, grp == "01_postmenopausal"), aes(x = DurationMenopause, y = value)) +
  geom_point() +
  geom_smooth(method = "lm") +
  facet_wrap(~variable, scales = "free") +
  labs(x = "Duration of Amenorrhea in trans-men group", y = "Alpha diversity value")
```



```
corr_postmeno = filter(adv_menopause, grp == "01_postmenopausal") %>%
  group_by(variable) %>%
  rstatix::cor_test(value, DurationMenopause, method = "pearson")
kable(corr_postmeno)
```

variable	var1	var2	cor	statistic	p	conf.low	conf.high	method
Observed	value	DurationMenopause	0.618	0.8607750	0.398	-0.2348644	0.5344828	Pearson
Shannon	value	DurationMenopause	0.611	0.5281901	0.602	-0.2985727	0.4836819	Pearson
InvSimpson	value	DurationMenopause	0.616	0.7669201	0.451	-0.2530096	0.5205587	Pearson

```
# correlations are not significant for trans and postmenopausal group separately

# Next, both groups are combined: in months
corr_menopause = adv_menopause %>%
  group_by(variable) %>%
  rstatix::cor_test(value, DurationCombined, method = "pearson")
kable(corr_menopause)
```


variable	var1	var2	cor	statistic	p	conf.low	conf.high	method
Observed	value	DurationCombined	0.21	-	0.1410	-	0.0713220	Pearson
				1.497145		0.4623821		
Shannon	value	DurationCombined	0.30	-	0.0351	-	-	Pearson
				2.168592		0.5327673	0.0222174	
InvSimpson	value	DurationCombined	0.14	-	0.3200	-	0.1403448	Pearson
				1.005403		0.4057416		

```

comb_menopause_plot = ggplot(adv_menopause, aes(x = DurationCombined, y = value)) +
  geom_point(aes(color = grp)) +
  geom_smooth(method = "lm", color = "darkgrey") +
  facet_wrap(~variable, scales = "free") +
  labs(x = "Duration of Amenorrhea in postmenopausal and trans-men (months)", y = "Alpha diversity") +
  scale_color_manual(values = custom_pal[c(1,3)])

ggsave(plot = comb_menopause_plot, device = "pdf",
        filename = "MbVagTrans_Results_files/figure-html/duration-menopause_comb.pdf", height = 10)

## Maaslin2 analysis of abundance with duration of amenorrhea
ps_menopause_rel_genus = tax_glom(ps_menopause_rel, taxrank = "Genus")

genus_MenoPauseDurationMonths = Maaslin2(input_data = otu_table(ps_menopause_rel_genus),
      input_metadata = as(sample_data(ps_menopause_rel_genus), "data.frame"),
      output = "results/maaslin2/genus_MenopausedurationMonths",
      min_abundance = 5,
      min_prevalence = 0.2,
      max_significance = 0.2,
      fixed_effects = "DurationCombined")

```

```

2024-03-06 21:55:53.46621 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_MenopausedurationMonths/maaslin2_log.txt"
2024-03-06 21:55:53.640418 INFO::Writing function arguments to log file
2024-03-06 21:55:53.643104 INFO::Verifying options selected are valid
2024-03-06 21:55:53.643389 INFO::Determining format of input files
2024-03-06 21:55:53.643624 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:53.645472 INFO::Formula for fixed effects: expr ~ DurationCombined
2024-03-06 21:55:53.645752 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:53.64597 INFO::Total samples in data: 50
2024-03-06 21:55:53.646176 INFO::Min samples required with min abundance for a feature not to be dropped: 5

```

```

2024-03-06 21:55:53.647815 INFO::Total filtered features: 313
2024-03-06 21:55:53.648092 INFO::Filtered feature names from abundance and prevalence filter
2024-03-06 21:55:53.648547 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:53.648774 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:53.648976 INFO::Running selected normalization method: TSS
2024-03-06 21:55:53.649375 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:53.651754 INFO::Running selected transform method: LOG
2024-03-06 21:55:53.652133 INFO::Running selected analysis method: LM
2024-03-06 21:55:53.652388 INFO::Fitting model to feature number 1, ASV2
2024-03-06 21:55:53.653423 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:53.654362 INFO::Fitting model to feature number 3, ASV7
2024-03-06 21:55:53.655276 INFO::Fitting model to feature number 4, ASV15
2024-03-06 21:55:53.656248 INFO::Fitting model to feature number 5, ASV17
2024-03-06 21:55:53.657243 INFO::Fitting model to feature number 6, ASV31
2024-03-06 21:55:53.658128 INFO::Fitting model to feature number 7, ASV50
2024-03-06 21:55:53.660771 INFO::Counting total values for each feature
2024-03-06 21:55:53.661466 INFO::Writing filtered data to file results/maaslin2/genus_Menopar
2024-03-06 21:55:53.662351 INFO::Writing filtered, normalized data to file results/maaslin2/g
2024-03-06 21:55:53.663135 INFO::Writing filtered, normalized, transformed data to file resu
2024-03-06 21:55:53.663926 WARNING::Deleting existing residuals file: results/maaslin2/genus
2024-03-06 21:55:53.664322 INFO::Writing residuals to file results/maaslin2/genus_Menopausedr
2024-03-06 21:55:53.66475 WARNING::Deleting existing fitted file: results/maaslin2/genus_Menop
2024-03-06 21:55:53.665076 INFO::Writing fitted values to file results/maaslin2/genus_Menopar
2024-03-06 21:55:53.665456 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:53.665998 INFO::Writing the significant results (those which are less than c
2024-03-06 21:55:53.66638 INFO::Writing heatmap of significant results to file: results/maas
[1] "There is not enough metadata in the associations to create a heatmap plot. Please review
2024-03-06 21:55:53.66692 INFO::Writing association plots (one for each significant associat
2024-03-06 21:55:53.667478 INFO::Plotting associations from most to least significant, group
2024-03-06 21:55:53.667752 INFO::Plotting data for metadata number 1, DurationCombined
2024-03-06 21:55:53.668236 INFO::Creating scatter plot for continuous data, DurationCombined

2024-03-06 21:55:53.72933 INFO::Creating scatter plot for continuous data, DurationCombined v

Warning: Removed 2 rows containing missing values (`geom_point()`).

2024-03-06 21:55:53.793825 INFO::Creating scatter plot for continuous data, DurationCombined

2024-03-06 21:55:53.860334 INFO::Creating scatter plot for continuous data, DurationCombined

Warning: Removed 1 rows containing missing values (`geom_point()`).
Removed 2 rows containing missing values (`geom_point()`).

```

Warning: Removed 1 rows containing missing values (`geom_point()`).

```
# 4 ASVs
```

```
tax_table(ps_menopause_rel_genus) %>%  
  as.data.frame() %>%  
  rownames_to_column("ASV") %>%  
  filter(ASV %in% filter(genus_MenoPauseDurationMonths$results, qval < 0.2)$feature)
```

	ASV	Kingdom	Phylum	Class	Order
1	ASV2	Bacteria	Firmicutes	Bacilli	Lactobacillales
2	ASV3	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales
3	ASV17	Bacteria	Firmicutes	Negativicutes	Veillonellales-Selenomonadales
4	ASV31	Bacteria	Firmicutes	Clostridia	Peptostreptococcales-Tissierellales

	Family	Genus	Species
1	Lactobacillaceae	Lactobacillus	<NA>
2	Prevotellaceae	Prevotella	<NA>
3	Veillonellaceae	Dialister	<NA>
4	Family XI	Anaerococcus	<NA>

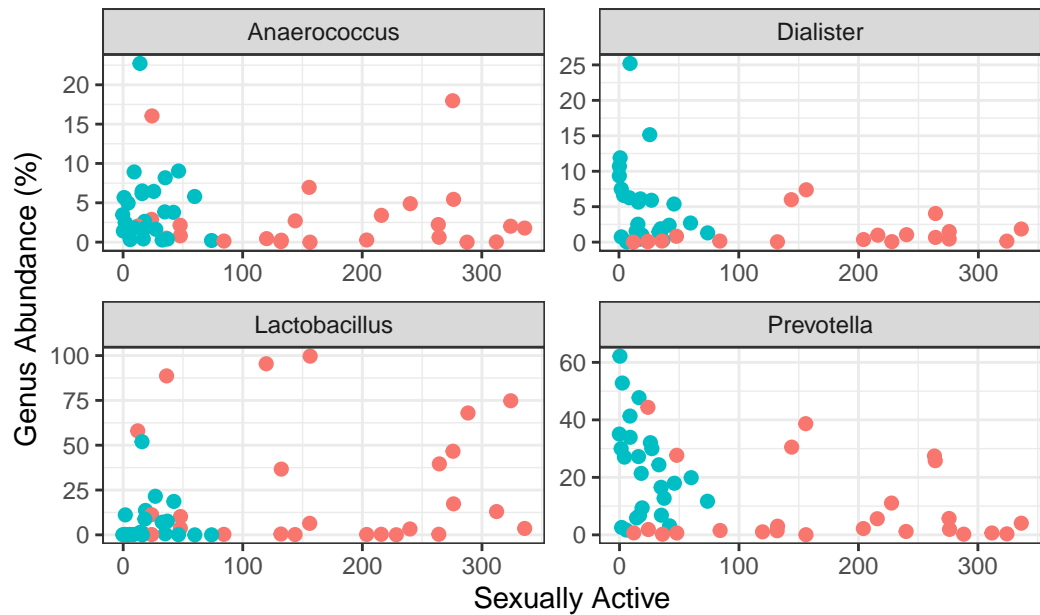
```
# Lactobacillus, dialister, Anaerococcus, Prevotella
```

```
# plot DurationCombined vs these 4 genera
```

```
psmelt_rel_genus_durationComb = psmelt(ps_menopause_rel_genus)  
maaslin_res_genus_durationComb = ggplot(filter(psmelt_rel_genus_durationComb, OTU %in% fil  
  aes(x = DurationCombined, y = Abundance, color = grp)) +  
  geom_jitter(size = 2) +  
  facet_wrap(~Genus, scales = "free") +  
  scale_y_continuous(limits = function(x){c(0, max(0.1, x))}) +  
  theme(legend.position = "none")+  
  labs(x = "Sexually Active",  
    y = "Genus Abundance (%)",  
    caption = "Maaslin2 results")
```

```
maaslin_res_genus_durationComb
```

Warning: Removed 11 rows containing missing values (`geom_point()`).



Maaslin2 results

```
ggsave(maaslin_res_genus_durationComb, filename = "results/maaslin2/genus_DurationMenopaus",
       height = 5, width = 6)
```

Warning: Removed 12 rows containing missing values (`geom_point()`).

6.1 Influence of sex of partner

```
## Analysis in complete cohort
### alpha diversity
rstatix::shapiro_test(filter(adiv$data, variable=="Observed"), value)
```

```
# A tibble: 1 x 3
  variable statistic      p
  <chr>      <dbl>    <dbl>
1 value      0.875 0.00000226
```

```
rstatix::shapiro_test(filter(adiv$data, variable=="Shannon"), value)
```

```
# A tibble: 1 x 3
  variable statistic      p
  <chr>      <dbl> <dbl>
1 value      0.984 0.490

rstatix::shapiro_test(filter(adiv$data, variable=="InvSimpson"), value)
```

```
# A tibble: 1 x 3
  variable statistic      p
  <chr>      <dbl> <dbl>
1 value      0.659 6.43e-12
```

```
adiv$data %>%
  filter(variable %in% c("Observed", "Shannon", "InvSimpson")) %>%
  group_by(variable) %>%
  rstatix::kruskal_test(value~SexOfPartner)
```

```
# A tibble: 3 x 7
  variable .y.      n statistic    df      p method
* <fct>    <chr> <int>    <dbl> <int> <dbl> <chr>
1 Observed value    75    -41.0     2      1 Kruskal-Wallis
2 Shannon  value    75    -33.9     2      1 Kruskal-Wallis
3 InvSimpson value    75    -31.4     2      1 Kruskal-Wallis
```

```
adiv$data %>%
  filter(variable %in% c("Observed", "Shannon", "InvSimpson")) %>%
  group_by(variable) %>%
  rstatix::anova_test(value~SexOfPartner)
```

```
# A tibble: 3 x 8
  variable Effect      DFn  DFd      F      p `p<.05` ges
* <fct>    <chr>    <dbl> <dbl> <dbl> <dbl> <chr> <dbl>
1 Observed SexOfPartner    3    71  3.62 0.017  "*"  0.133
2 Shannon  SexOfPartner    3    71  6.96 0.000358  "*"  0.227
3 InvSimpson SexOfPartner    3    71  1.34 0.267  ""   0.054
```

Observed is not distributed normally, and no significant effect of sex of partner is detected using Kruskal Wallis test. However, Shannon and InvSimpson indices show a significant effect.

```

adonis_bc_partnersex = vegan::adonis2(D_BC ~ phyloseq::sample_data(ps_filtered_rel)$SexOfP
adonis_pw_bc_partnersex = pairwise.adonis(D_BC, phyloseq::sample_data(ps_filtered_rel)$Sex

adonis_uf_partnersex = vegan::adonis2(D_UF ~ phyloseq::sample_data(ps_filtered_rel)$SexOfP
adonis_pw_uf_partnersex = pairwise.adonis(D_UF, phyloseq::sample_data(ps_filtered_rel)$Sex

adonis_wUF_partnersex = vegan::adonis2(D_wUF ~ phyloseq::sample_data(ps_filtered_rel)$SexO
adonis_pw_wUF_partnersex = pairwise.adonis(D_wUF, phyloseq::sample_data(ps_filtered_rel)$S

kable(adonis_pw_bc, digits = 3, caption = "Bray-Curtis")

```

Table 22: Bray-Curtis

pairs	Df	SumsOfSqsF	Model	R2	p.value	p.adjusted	sig
01_postmenopausal vs 02_premenopausal	1	1.839	4.972	0.094	0.001	0.003	*
01_postmenopausal vs 03_trans-men	1	1.255	3.474	0.067	0.001	0.003	*
02_premenopausal vs 03_trans-men	1	3.351	10.303	0.177	0.001	0.003	*

```

kable(adonis_pw_uf, digits = 3, caption = "Unweigthed Unifrac")

```

Table 23: Unweigthed Unifrac

pairs	Df	SumsOfSqsF	Model	R2	p.value	p.adjusted	sig
01_postmenopausal vs 02_premenopausal	1	1.104	5.145	0.097	0.001	0.003	*
01_postmenopausal vs 03_trans-men	1	0.704	4.133	0.079	0.002	0.006	*
02_premenopausal vs 03_trans-men	1	2.263	12.762	0.210	0.001	0.003	*

```

kable(adonis_pw_wUF, digits = 3, caption = "Weighted Unifrac")

```

Table 24: Weighted Unifrac

pairs	Df	SumsOfSqsF	Model	R2	p.value	p.adjusted	sig
01_postmenopausal vs 02_premenopausal	1	0.559	8.149	0.145	0.001	0.003	*
01_postmenopausal vs 03_trans-men	1	0.413	5.341	0.100	0.001	0.003	*
02_premenopausal vs 03_trans-men	1	1.369	23.593	0.330	0.001	0.003	*

Adonis identifies significant influence of sex of partner oall indices of beta diversity. Pairwise comparisons remain significant for all groups ($p = 0.03$)

```
genus_SexOfPartner = Maaslin2(input_data = otu_table(ps_rel_genus_sexuallyactive),
  input_metadata = as(sample_data(ps_rel_genus_sexuallyactive), "data.frame"),
  output = "results/maaslin2/genus_SexOfPartner",
  min_abundance = 5,
  min_prevalence = 0.2,
  max_significance = 0.2,
  fixed_effects = "SexOfPartner")
```

```
2024-03-06 21:55:55.368404 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_SexOfPartner/maaslin2.log"
2024-03-06 21:55:55.547277 INFO::Writing function arguments to log file
2024-03-06 21:55:55.549565 INFO::Verifying options selected are valid
2024-03-06 21:55:55.549849 INFO::Determining format of input files
2024-03-06 21:55:55.550098 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:55.55239 INFO::Formula for fixed effects: expr ~ SexOfPartner
2024-03-06 21:55:55.552809 INFO::Factor detected for categorical metadata 'SexOfPartner'. Proceeding with factor
2024-03-06 21:55:55.553132 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:55.553595 INFO::Total samples in data: 72
2024-03-06 21:55:55.553864 INFO::Min samples required with min abundance for a feature not to be filtered: 5
2024-03-06 21:55:55.555608 INFO::Total filtered features: 316
2024-03-06 21:55:55.555987 INFO::Filtered feature names from abundance and prevalence filtering: 316
2024-03-06 21:55:55.556489 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:55.556736 INFO::Filtered feature names from variance filtering: 0
2024-03-06 21:55:55.556956 INFO::Running selected normalization method: TSS
2024-03-06 21:55:55.557379 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:55.559677 INFO::Running selected transform method: LOG
2024-03-06 21:55:55.560116 INFO::Running selected analysis method: LM
2024-03-06 21:55:55.5604 INFO::Fitting model to feature number 1, ASV1
```

```

2024-03-06 21:55:55.561616 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:55.562654 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:55.56365 INFO::Fitting model to feature number 4, ASV6
2024-03-06 21:55:55.566661 INFO::Counting total values for each feature
2024-03-06 21:55:55.567477 INFO::Writing filtered data to file results/maaslin2/genus_SexOfPartner
2024-03-06 21:55:55.568916 INFO::Writing filtered, normalized data to file results/maaslin2/genus_SexOfPartner
2024-03-06 21:55:55.569885 INFO::Writing filtered, normalized, transformed data to file results/maaslin2/genus_SexOfPartner
2024-03-06 21:55:55.570741 WARNING::Deleting existing residuals file: results/maaslin2/genus_SexOfPartner
2024-03-06 21:55:55.571191 INFO::Writing residuals to file results/maaslin2/genus_SexOfPartner
2024-03-06 21:55:55.571792 WARNING::Deleting existing fitted file: results/maaslin2/genus_SexOfPartner
2024-03-06 21:55:55.572133 INFO::Writing fitted values to file results/maaslin2/genus_SexOfPartner
2024-03-06 21:55:55.572521 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:55.57311 INFO::Writing the significant results (those which are less than 0.05)
2024-03-06 21:55:55.573477 INFO::Writing heatmap of significant results to file: results/maaslin2/genus_SexOfPartner
[1] "There are no associations to plot!"
2024-03-06 21:55:55.573958 INFO::Writing association plots (one for each significant association)
2024-03-06 21:55:55.57452 INFO::Plotting associations from most to least significant, grouped by metadata
2024-03-06 21:55:55.57479 INFO::Plotting data for metadata number 1, SexOfPartner
2024-03-06 21:55:55.575309 INFO::Creating boxplot for categorical data, SexOfPartner vs ASV3

```

```

genus_SexOfPartner2 = Maaslin2(input_data = otu_table(ps_rel_genus_sexuallyactive),
    input_metadata = as(sample_data(ps_rel_genus_sexuallyactive), "data.frame"),
    output = "results/maaslin2/genus_SexOfPartnergrp",
    min_abundance = 5,
    min_prevalence = 0.2,
    max_significance = 0.2,
    fixed_effects = c("SexOfPartner", "grp", "SexOfPartner*grp"))

```

```

2024-03-06 21:55:55.718531 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_SexOfPartnergrp/maaslin2.log"
2024-03-06 21:55:55.907498 INFO::Writing function arguments to log file
2024-03-06 21:55:55.909546 INFO::Verifying options selected are valid
2024-03-06 21:55:55.909794 INFO::Determining format of input files
2024-03-06 21:55:55.910027 INFO::Input format is data samples as rows and metadata samples as columns
2024-03-06 21:55:55.954827 WARNING::Feature name not found in metadata so not applied to formula
2024-03-06 21:55:55.955507 INFO::Formula for fixed effects: expr ~ SexOfPartner + grp
2024-03-06 21:55:55.95584 INFO::Factor detected for categorical metadata 'SexOfPartner'. Provide a reference level.
2024-03-06 21:55:55.956096 INFO::Factor detected for categorical metadata 'grp'. Provide a reference level.
2024-03-06 21:55:55.956324 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:55.956563 INFO::Total samples in data: 72
2024-03-06 21:55:55.956797 INFO::Min samples required with min abundance for a feature not to be filtered: 5

```



```

2024-03-06 21:55:55.958659 INFO::Total filtered features: 316
2024-03-06 21:55:55.959213 INFO::Filtered feature names from abundance and prevalence filter
2024-03-06 21:55:55.959732 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:55.960003 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:55.960249 INFO::Running selected normalization method: TSS
2024-03-06 21:55:55.960705 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:55.963347 INFO::Running selected transform method: LOG
2024-03-06 21:55:55.964005 INFO::Running selected analysis method: LM
2024-03-06 21:55:55.964435 INFO::Fitting model to feature number 1, ASV1
2024-03-06 21:55:55.966037 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:55.967677 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:55.968854 INFO::Fitting model to feature number 4, ASV6
2024-03-06 21:55:55.971885 INFO::Counting total values for each feature
2024-03-06 21:55:55.972732 INFO::Writing filtered data to file results/maaslin2/genus_SexOfPa
2024-03-06 21:55:55.973636 INFO::Writing filtered, normalized data to file results/maaslin2/g
2024-03-06 21:55:55.974455 INFO::Writing filtered, normalized, transformed data to file resu
2024-03-06 21:55:55.975313 WARNING::Deleting existing residuals file: results/maaslin2/genus
2024-03-06 21:55:55.975793 INFO::Writing residuals to file results/maaslin2/genus_SexOfPartn
2024-03-06 21:55:55.976268 WARNING::Deleting existing fitted file: results/maaslin2/genus_Sex
2024-03-06 21:55:55.976636 INFO::Writing fitted values to file results/maaslin2/genus_SexOfPa
2024-03-06 21:55:55.977022 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:55.977727 INFO::Writing the significant results (those which are less than
2024-03-06 21:55:55.978145 INFO::Writing heatmap of significant results to file: results/maas
2024-03-06 21:55:56.00937 INFO::Writing association plots (one for each significant associat
2024-03-06 21:55:56.01023 INFO::Plotting associations from most to least significant, groupe
2024-03-06 21:55:56.010529 INFO::Plotting data for metadata number 1, grp
2024-03-06 21:55:56.011082 INFO::Creating boxplot for categorical data, grp vs ASV3

2024-03-06 21:55:56.070066 INFO::Creating boxplot for categorical data, grp vs ASV6

2024-03-06 21:55:56.136451 INFO::Creating boxplot for categorical data, grp vs ASV1

2024-03-06 21:55:56.196974 INFO::Creating boxplot for categorical data, grp vs ASV1

2024-03-06 21:55:56.260916 INFO::Creating boxplot for categorical data, grp vs ASV3

2024-03-06 21:55:56.652596 INFO::Plotting data for metadata number 2, SexOfPartner
2024-03-06 21:55:56.653504 INFO::Creating boxplot for categorical data, SexOfPartner vs ASV6

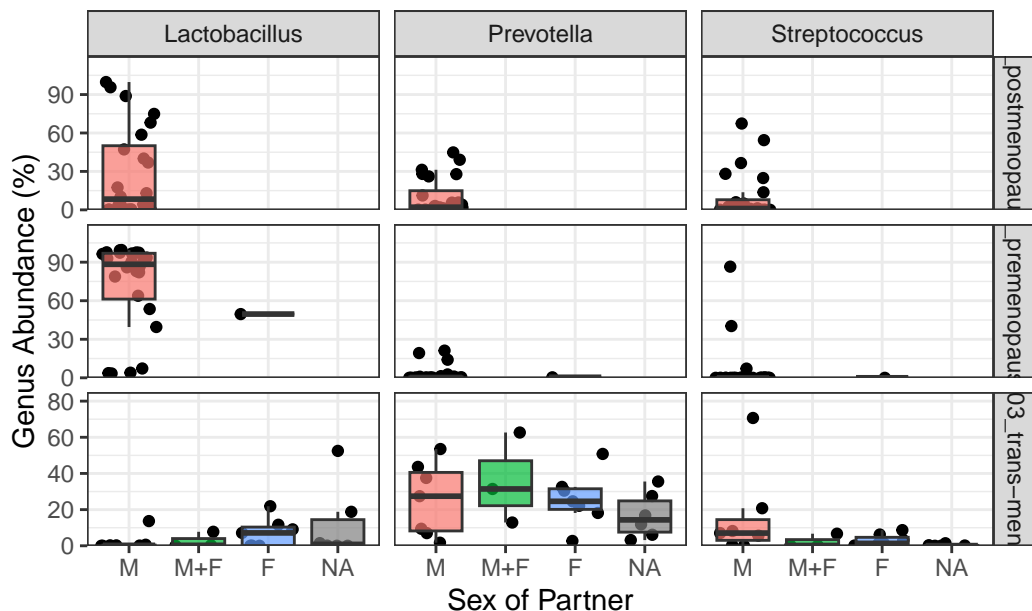
```

```
tax_table(ps_rel_genus_sexuallyactive) %>%
  as.data.frame() %>%
  rownames_to_column("ASV") %>%
  filter(ASV %in% filter(genus_SexOfPartner2$results, qval < 0.2)$feature)
```

	ASV	Kingdom	Phylum	Class	Order	Family
1	ASV1	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae
2	ASV3	Bacteria	Bacteroidota	Bacteroidia	Bacteroidales	Prevotellaceae
3	ASV6	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae
		Genus	Species			
1	Lactobacillus	<NA>				
2	Prevotella	<NA>				
3	Streptococcus	<NA>				

```
maaslin_res_genus_sexpartner = ggplot(filter(psmelt_rel_genus_sexuallyactive, OTU %in% fil
  aes(x = SexOfPartner, y = Abundance, fill = SexOfPartner))
  geom_jitter() +
  geom_boxplot(alpha = 0.7, outlier.shape = NA) +
  facet_grid(grp~Genus, scales = "free") +
  scale_y_continuous(limits = function(x){c(0, max(0.1, x))}) +
  theme(legend.position = "none")+
  labs(x = "Sex of Partner",
    y = "Genus Abundance (%)",
    caption = "Maaslin2 results") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.2)))
```

```
maaslin_res_genus_sexpartner
```



Maaslin2 results

```
ggsave(maaslin_res_genus_sexpartner, filename = "results/maaslin2/genus_SexOfPartner_panel",
       height = 5, width = 6)
```

```
# M/F distinction only in trans - filter for this group
ps_rel_genus_sexuallyactive_trans = subset_samples(ps_rel_genus_sexuallyactive, grp == "03_
genus_SexOfPartnerTrans = Maaslin2(input_data = otu_table(ps_rel_genus_sexuallyactive_tran
    input_metadata = as(sample_data(ps_rel_genus_sexuallyactive_trans), "data.frame")
    output = "results/maaslin2/genus_SexOfPartnerTrans",
    min_abundance = 5,
    min_prevalence = 0.2,
    max_significance = 0.2,
    fixed_effects = "SexOfPartner")
```

```
2024-03-06 21:55:57.27439 WARNING::Input is a matrix, passing through as.data.frame() .
[1] "Warning: Deleting existing log file: results/maaslin2/genus_SexOfPartnerTrans/maaslin2.
2024-03-06 21:55:57.461799 INFO::Writing function arguments to log file
2024-03-06 21:55:57.465255 INFO::Verifying options selected are valid
2024-03-06 21:55:57.465629 INFO::Determining format of input files
2024-03-06 21:55:57.465908 INFO::Input format is data samples as rows and metadata samples as
2024-03-06 21:55:57.467834 INFO::Formula for fixed effects: expr ~ SexOfPartner
2024-03-06 21:55:57.468134 INFO::Factor detected for categorical metadata 'SexOfPartner'. Pro
```

```

2024-03-06 21:55:57.468348 INFO::Filter data based on min abundance and min prevalence
2024-03-06 21:55:57.468556 INFO::Total samples in data: 23
2024-03-06 21:55:57.468758 INFO::Min samples required with min abundance for a feature not to
2024-03-06 21:55:57.470259 INFO::Total filtered features: 311
2024-03-06 21:55:57.470561 INFO::Filtered feature names from abundance and prevalence filter
2024-03-06 21:55:57.471021 INFO::Total filtered features with variance filtering: 0
2024-03-06 21:55:57.47127 INFO::Filtered feature names from variance filtering:
2024-03-06 21:55:57.471497 INFO::Running selected normalization method: TSS
2024-03-06 21:55:57.471912 INFO::Applying z-score to standardize continuous metadata
2024-03-06 21:55:57.474248 INFO::Running selected transform method: LOG
2024-03-06 21:55:57.474753 INFO::Running selected analysis method: LM
2024-03-06 21:55:57.475037 INFO::Fitting model to feature number 1, ASV1
2024-03-06 21:55:57.476279 INFO::Fitting model to feature number 2, ASV3
2024-03-06 21:55:57.47739 INFO::Fitting model to feature number 3, ASV4
2024-03-06 21:55:57.478505 INFO::Fitting model to feature number 4, ASV6
2024-03-06 21:55:57.479585 INFO::Fitting model to feature number 5, ASV17
2024-03-06 21:55:57.480952 INFO::Fitting model to feature number 6, ASV20
2024-03-06 21:55:57.482374 INFO::Fitting model to feature number 7, ASV24
2024-03-06 21:55:57.483619 INFO::Fitting model to feature number 8, ASV31
2024-03-06 21:55:57.484588 INFO::Fitting model to feature number 9, ASV50
2024-03-06 21:55:57.487917 INFO::Counting total values for each feature
2024-03-06 21:55:57.489018 INFO::Writing filtered data to file results/maaslin2/genus_SexOfPa
2024-03-06 21:55:57.490072 INFO::Writing filtered, normalized data to file results/maaslin2/g
2024-03-06 21:55:57.490765 INFO::Writing filtered, normalized, transformed data to file resu
2024-03-06 21:55:57.491422 WARNING::Deleting existing residuals file: results/maaslin2/genus
2024-03-06 21:55:57.49188 INFO::Writing residuals to file results/maaslin2/genus_SexOfPartne
2024-03-06 21:55:57.492338 WARNING::Deleting existing fitted file: results/maaslin2/genus_Se
2024-03-06 21:55:57.492721 INFO::Writing fitted values to file results/maaslin2/genus_SexOfPa
2024-03-06 21:55:57.493113 INFO::Writing all results to file (ordered by increasing q-values)
2024-03-06 21:55:57.493798 INFO::Writing the significant results (those which are less than
2024-03-06 21:55:57.494174 INFO::Writing heatmap of significant results to file: results/maas
[1] "There are no associations to plot!"
2024-03-06 21:55:57.494662 INFO::Writing association plots (one for each significant associat
[1] "There are no associations to plot!"

```

```

# no significant results in trans group only

```