qiime usage

All stages of the analysis were done in qiime2-2021.2

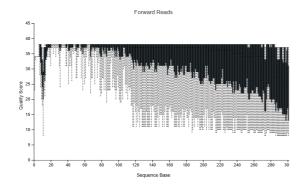
Preprocessing + taxonomical analysis

qiime demux summarize --i-data demux-paired-end.qza --o-visualization demux-summary-1.qzv

Demultiplexed sequence counts summary

<u>Aa</u> Property	# forward reads	# reverse reads
<u>Minimum</u>	23988	23988
Median	206576	206576
Mean	185457	185457
<u>Maximum</u>	346110	346110
<u>Total</u>	13167447	13167447

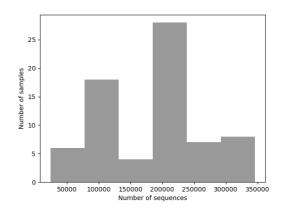
Forward Reads Frequency Histogram

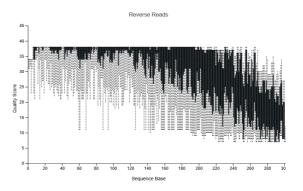


Parametric seven-number summary for position 68

Aa Box plot feature	■ Percentile	# Quality

Reverse Reads Frequency Histogram





Parametric seven-number summary for position 85

1

Aa Box plot feature	■ Percentile	# Quality
(Not shown in box plot)	2nd	11
Lower Whisker	9th	27
Bottom of Box	25th	36
Middle of Box	50th (Median)	38
Top of Box	75th	38
<u>Upper Whisker</u>	91st	38
(Not shown in box plot)	98th	38

Aa Box plot feature	■ Percentile	# Quality score
(Not shown in box plot)	2nd	11
Lower Whisker	9th	19
Bottom of Box	25th	32
Middle of Box	50th (Median)	38
Top of Box	75th	38
<u>Upper Whisker</u>	91st	38
(Not shown in box plot)	98th	38

Demultiplexed sequence length summary

Reverse Reads

Aa Total Sequences Sampled	10000
2%	300 nts
9%	300 nts
<u>25%</u>	300 nts
50% (Median)	300 nts
<u>75%</u>	300 nts
91%	301 nts
98%	301 nts

Forward Reads

Aa Total Sequences Sampled	≡ 10000
<u>2%</u>	297 nts
9%	300 nts
<u>25%</u>	301 nts
50% (Median)	301 nts
<u>75%</u>	301 nts
91%	301 nts
98%	301 nts

qiime dada2 denoise-single --i-demultiplexed-seqs demux-paired-end.qza --p-trim-left 15 --o-representative-sequences rep-seqs-dada2.qza --o-table table-dada2.qza --o-denoising-stats stats-dada2.qz --p-trunc-len 300

qiime metadata tabulate --m-input-file stats-dada2.qz.qza --o-visualization stats-dada2.qz

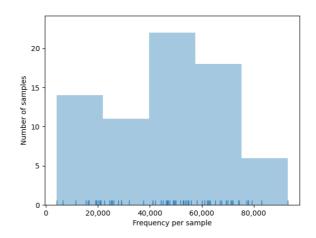
qiime feature-table summarize --i-table table-dada2.qza --o-visualization table-dada2.qzv

Table summary

<u>Aa</u> Metric	
Number of samples	71
Number of features	10,652
Total frequency	3,301,944

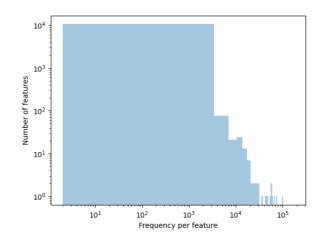
Frequency per sample

<u>Aa</u> Property	
Minimum frequency	4,056.0
1st quartile	25,501.0
Median frequency	49,051.0
3rd quartile	62,992.5
Maximum frequency	93,183.0
Mean frequency	46,506.25352112676



Frequency per feature

<u>Aa</u> Property	
Minimum frequency	2.0
1st quartile	6.0
Median frequency	19.0
3rd quartile	69.0
Maximum frequency	175,016.0
Mean frequency	309.98347728126174



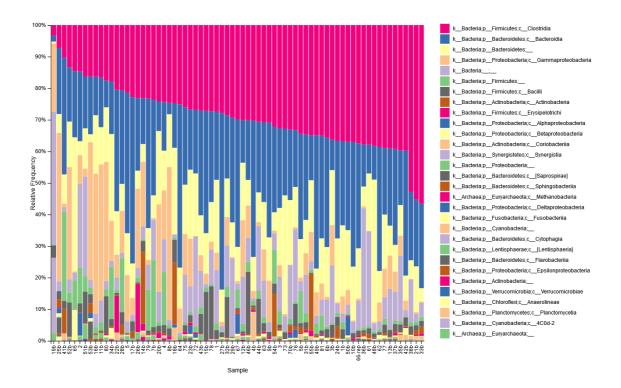
qiime feature-table tabulate-seqs --i-data rep-seqs-dada2.qza --o-visualization rep-seqs.qzv

Taxonomy classification

qiime metadata tabulate --m-input-file taxonomy.qza --o-visualization taxonomy.qzv

qiime taxa barplot --i-table table-dada2.qza --i-taxonomy taxonomy.qza --m-metadata-file metadata_old.tsv --o-visualization taxa-bar-plots.qzv

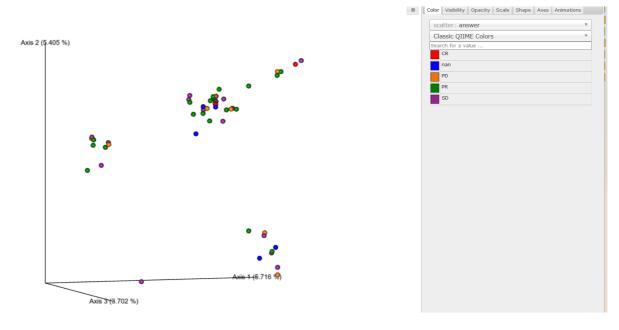
Example of barplot in level 3



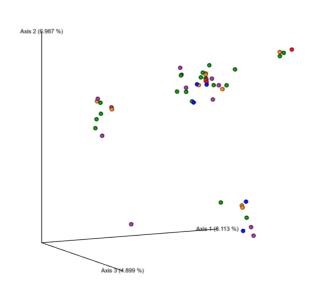
Diversity analysis

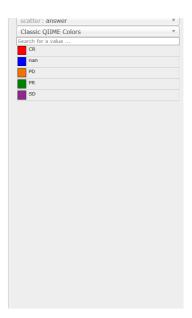
```
qiime diversity core-metrics-phylogenetic \
--i-phylogeny rooted-tree.qz.qza \
--i-table table-dada2.qza \
--p-sampling-depth 4 \
--m-metadata-file metadata.csv \
--output-dir ./phylo
```

Clusters by jaccard distances:

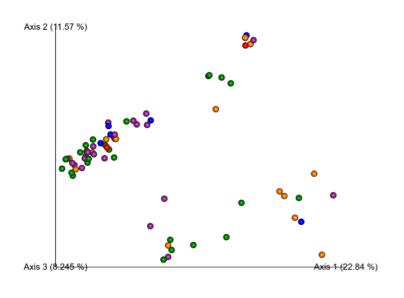


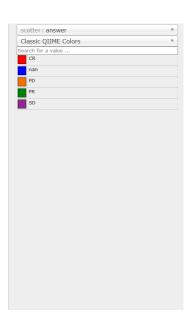
Clusters by bray-curtis distances:



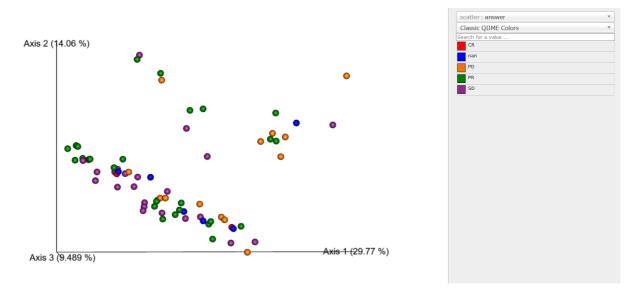


Clusters by unweighted unifrac distances:





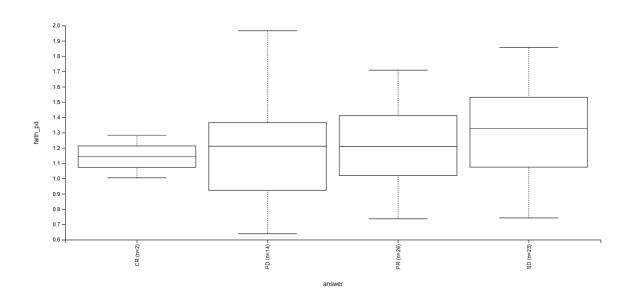
Clusters by weighted unifrac distances:



Alpha diversity analysis

qiime diversity alpha-group-significance $\$

- --i-alpha-diversity ./phylo3/faith_pd_vector.qza \
 --m-metadata-file metadata.tsv \
 --o-visualization ./phylo3/faith-pd-group-significance.qzv

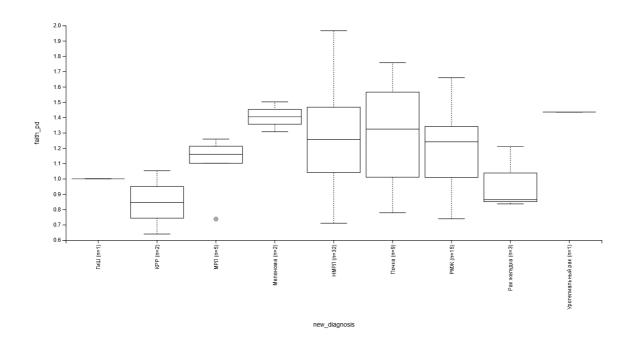


Kruskal-Wallis (all groups)

<u>Aa</u> Property	# Result	
<u>H</u>	1.598	
p-value	0.66	

Kruskal-Wallis (pairwise)

Aa Group 1		≡н		# q-value
<u>CR (n=2)</u>	PD (n=14)	0,025	0,874	0.874
<u>CR (n=2)</u>	PR (n=26)	0,286	0,592	0.874
<u>CR (n=2)</u>	SD (n=23)	0,642	0,423	0.846
<u>PD (n=14)</u>	PR (n=26)	0,065	0,799	0.874
PD (n=14)	SD (n=23)	0,715	0,398	0.846
PR (n=26)	SD (n=23)	1,044	0,307	0.846



Kruskal-Wallis (all groups)

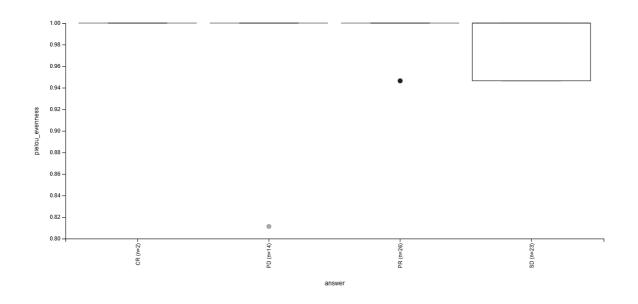
<u>Aa</u> Property	# Result
<u>H</u>	9.839793762575425
p-value	0.2764508938127179

Kruskal-Wallis (pairwise)

<u>Aa</u> Property	≡ 1	≡н	□ p-value	
Group 1	Group 2			
<u>ГиШ (n=1)</u>	KPP (n=2)	0	1	1
<u>ГиШ (n=1)</u>	МРЛ (n=5)	0,7714	0,38	0,5761
<u>ГиШ (n=1)</u>	Меланома (n=2)	1,5	0,221	0,5383
<u>ГиШ (n=1)</u>	НМРЛ (n=32)	1,1029	0,294	0,5383
<u>ГиШ (n=1)</u>	Почка (n=9)	0,7576	0,384	0,5761
<u>ГиШ (n=1)</u>	РМЖ (n=15)	0,9529	0,329	0,5383
<u>ГиШ (n=1)</u>	Рак желудка (n=3)	0,2	0,655	0,7366
<u>ГиШ (n=1)</u>	Уротелиальный рак (n=1)	1	0,317	0,5383
<u>KPP (n=2)</u>	МРЛ (n=5)	2,4	0,121	0,5383
<u>KPP (n=2)</u>	Меланома (n=2)	2,4	0,121	0,5383
<u>KPP (n=2)</u>	НМРЛ (n=32)	2,8339	0,092	0,5383
<u>KPP (n=2)</u>	Почка (n=9)	2	0,157	0,5383
<u>KPP (n=2)</u>	РМЖ (n=15)	2,2222	0,136	0,5383
<u>KPP (n=2)</u>	Рак желудка (n=3)	0,3333	0,564	0,6764
<u>KPP (n=2)</u>	Уротелиальный рак (n=1)	1,5	0,221	0,5383
<u>МРЛ (n=5)</u>	Меланома (n=2)	3,75	0,053	0,5383
<u>МРЛ (n=5)</u>	НМРЛ (n=32)	1,0441	0,307	0,5383
<u>МРЛ (n=5)</u>	Почка (n=9)	1,2844	0,257	0,5383
<u>МРЛ (n=5)</u>	РМЖ (n=15)	1,1905	0,275	0,5383
<u>МРЛ (n=5)</u>	Рак желудка (n=3)	0,5556	0,456	0,5968
<u>МРЛ (n=5)</u>	Уротелиальный рак (n=1)	2,1429	0,143	0,5383
<u>Меланома (n=2)</u>	НМРЛ (n=32)	0,5357	0,464	0,5968

<u>Aa</u> Property	≡ 1	≣н	□ p-value	
<u>Меланома (n=2)</u>	Почка (n=9)	0,0556	0,814	0,8876
<u>Меланома (n=2)</u>	РМЖ (n=15)	1,4222	0,233	0,5383
<u>Меланома (n=2)</u>	Рак желудка (n=3)	3	0,083	0,5383
<u>Меланома (n=2)</u>	Уротелиальный рак (n=1)	0	1	1
<u>НМРЛ (n=32)</u>	Почка (n=9)	0,004	0,95	1
<u>НМРЛ (n=32)</u>	РМЖ (n=15)	0,638	0,424	0,5968
<u>НМРЛ (n=32)</u>	Рак желудка (n=3)	3,125	0,077	0,5383
<u>НМРЛ (n=32)</u>	Уротелиальный рак (n=1)	0,3971	0,529	0,6562
<u>Почка (n=9)</u>	РМЖ (n=15)	0,5556	0,456	0,5968
<u>Почка (n=9)</u>	Рак желудка (n=3)	1,4444	0,229	0,5383
<u>Почка (n=9)</u>	Уротелиальный рак (n=1)	0,2727	0,602	0,6985
<u>РМЖ (n=15)</u>	Рак желудка (n=3)	1,5474	0,214	0,5383
<u>РМЖ (n=15)</u>	Уротелиальный рак (n=1)	0,9529	0,329	0,5383
<u>Рак желудка (n=3)</u>	Уротелиальный рак (n=1)	1,8	0,18	0,5383

- qiime diversity alpha-group-significance \
 --i-alpha-diversity ./phylo3/evenness_vector.qza \
 --m-metadata-file metadata.tsv \
 --o-visualization ./phylo3/evenness-group-significance.qzv



Kruskal-Wallis (all groups)

<u>Aa</u> Property	≡ Result
<u>H</u>	4,163
n-value	0.244

Kruskal-Wallis (pairwise)

<u>Aa</u> Property	≣ 1	≡н	□ p-value	
Group 1	Group 2			
<u>CR (n=2)</u>	PD (n=14)	0,143	0,705	0,705
<u>CR (n=2)</u>	PR (n=26)	0,452	0,502	0,602
<u>CR (n=2)</u>	SD (n=23)	0,982	0,322	0,558
PD (n=14)	PR (n=26)	0,796	0,372	0,558
PD (n=14)	SD (n=23)	2,889	0,089	0,535
PR (n=26)	SD (n=23)	1,483	0,223	0,558

Beta diversity analysis

- qiime diversity beta-group-significance \
 --i-distance-matrix ./phylo3/weighted_unifrac_distance_matrix.qza \
 --m-metadata-file metadata.tsv \

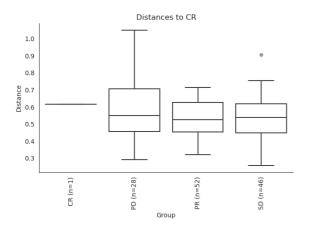
 - --m-metadata-column answer \
 --o-visualization ./phylo3/weighted-unifrac-answer-significance.qzv \
 - --p-pairwise

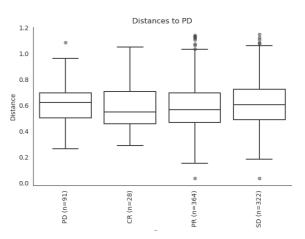
Overview

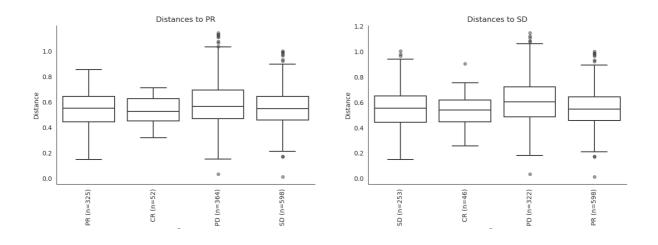
<u>Aa</u> Property	■ PERMANOVA results
method name	PERMANOVA
test statistic name	pseudo-F
sample size	65
number of groups	4
test statistic	1.4614
<u>p-value</u>	0.082
number of permutations	999

Pairwise permanova results

<u>Aa</u> Property	≡ 1	# Sample size	# Permutations	# pseudo-F	# p-value	# q-value
Group 1	Group 2	0	0	0	0	0
<u>CR</u>	PD	16	999	0.789842	0.625	0.883
CR	PR	28	999	0.599877	0.775	0.883
CR	SD	25	999	0.524779	0.883	0.883
<u>PD</u>	PR	40	999	2.11232	0.058	0.174
<u>PD</u>	SD	37	999	2.991727	0.011	0.066
<u>PR</u>	SD	49	999	0.855005	0.512	0.883



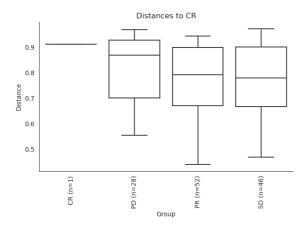


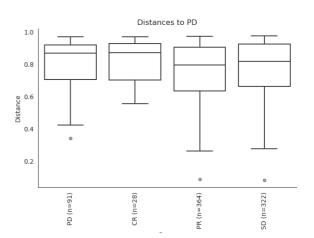


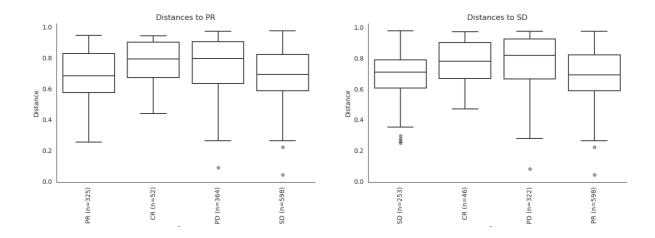
qiime diversity beta-group-significance --i-distance-matrix ./phylo/unweighted_unifrac_distance_matrix.qza --m-metadata-file me tadata.tsv --m-metadata-column answer --o-visualization ./phylo/unweighted-unifrac-answer-significance.qzv --p-pairwise

Overview

<u>Aa</u> Property	■ PERMANOVA results
method name	PERMANOVA
test statistic name	pseudo-F
sample size	65
number of groups	4
test statistic	1.46927
<u>p-value</u>	0.032
number of permutations	999







Pairwise permanova results

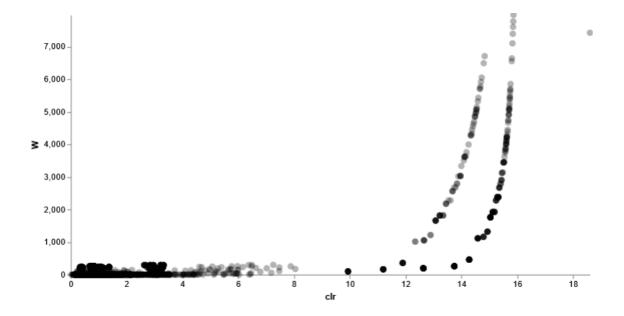
<u>Aa</u> Property	≡ 1	# Sample size	# Permutations	# pseudo-F	# p-value	# q-value
Group 1	Group 2	0	0	0	0	0
CR	PD	16	999	0.957013	0.503	0.503
<u>CR</u>	PR	28	999	1.212632	0.199	0.3165
CR	SD	25	999	1.146111	0.211	0.3165
<u>PD</u>	PR	40	999	1.723926	0.06	0.18
<u>PD</u>	SD	37	999	2.28916	0.015	0.09
<u>PR</u>	SD	49	999	1.08105	0.322	0.3864

ANCOM diversity analysis

```
qime feature-table filter-samples --i-table table-dada2.qza --m-metadata-file metadata.tsv --o-filtered-table nmrl-table.qza

qime composition add-pseudocount --i-table nmrl-table.qza --o-composition-table comp-nmrl-table.qza

qime composition ancom --i-table comp-nmrl-table.qza --m-metadata-file metadata.tsv --m-metadata-column answer --o-visuali zation ancom-answer.qzv
```

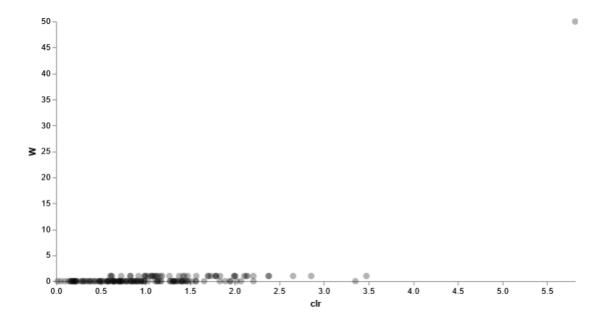


ANCOM statistical results

<u>Aa</u> Property	# W
8cfe521992ddbd4e0f6135d565177430	7977
31edfa54e028429a52fdc8493a10b7c1	7778
cd348c9798781ea702c36ca56b53981f	7609
2caff55d1333461efc2ba3148d9be5c4	7427
26c8d6ad6e8d7a24c495a82b4fd20867	7395
fb669f41e001ec53df6a80d55696757e	7103

qiime taxa collapse --i-table ./metadata/nmrl-table.qza --i-taxonomy taxonomy.qza --p-level 6 --o-collapsed-table ./metadat a/nmrl-table-l6.qza

qiime composition add-pseudocount --i-table ./metadata/nmrl-table-l6.qza --o-composition-table ./metadata/comp-nmrl-table-l6.qz



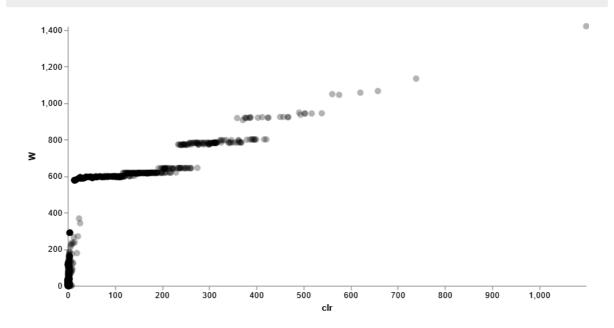
ANCOM statistical results

Aa Property	# W
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Veillonellaceae;g_Anaerosinus	50

qiime feature-table filter-samples --i-table table-dada2.qza --m-metadata-file metadata.tsv --p-where "answer='PD'" --o-fil tered-table gut-table.qza

qiime composition add-pseudocount --i-table gut-table.qza --o-composition-table comp-gut-table.qza

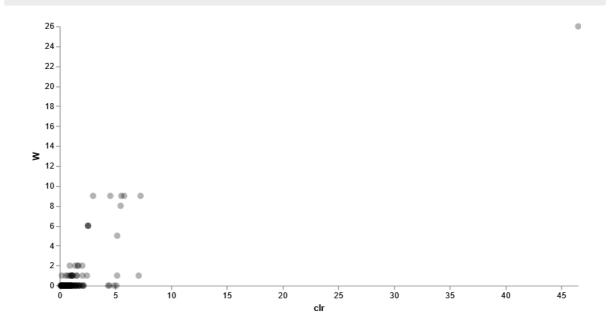
qiime composition ancom --i-table comp-gut-table.qza --m-metadata-file metadata.tsv --m-metadata-column new_diagnosis --o-v isualization ancom-Subject.qzv



qiime taxa collapse --i-table ./metadata/gut-table.qza --i-taxonomy taxonomy.qza --p-level 6 --o-collapsed-table ./metadat a/gut-table-l6.qza

qiime composition add-pseudocount --i-table gut-table-l6.qza --o-composition-table comp-gut-table-l6.qza

qiime composition ancom --i-table ./metadata/comp-gut-table-l6.qza --m-metadata-file ./metadata/metadata.tsv --m-metadata-column new_diagnosis --o-visualization ./metadata/l6-ancom-Subject.qzv



<u>Aa</u> Property	# W	Reject null hypothe
k Bacteria;p Actinobacteria;c Coriobacteriia;o Coriobacteriales;f Coriobacteriaceae;g Eggerthella	26	~
k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Bifidobacteriales;f_Bifidobacteriaceae;g_Alloscardovia	9	~
k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Leuconostocaceae;g_Weissella	9	~
k Bacteria;p Fusobacteria;c Fusobacteriia;o Fusobacteriales;f Fusobacteriaceae;g	9	~
k_Bacteria;p_Firmicutes;c_Bacilli;o_Lactobacillales;f_Leuconostocaceae;g_Oenococcus	9	~
k_Bacteria;p_Fusobacteria;c_Fusobacteriia;o_Fusobacteriales;f_Leptotrichiaceae;_	9	~
k Bacteria;p Bacteroidetes;c Cytophagia;o Cytophagales;f Cyclobacteriaceae;	8	~
k_Bacteria;p_Verrucomicrobia;c_Verrucomicrobiae;o_Verrucomicrobiales;f_Verrucomicrobiaceae;_	6	~
k_Bacteria;p_Proteobacteria;c_Alphaproteobacteria;o_Rickettsiales;f_mitochondria;_	6	~
k_Bacteria;p_Actinobacteria;c_Actinobacteria;o_Actinomycetales;f_Corynebacteriaceae;g_Corynebacterium	6	~
k_Bacteria;p_Bacteroidetes;c_Bacteroidia;o_Bacteroidales;f_Porphyromonadaceae;_	5	~
k_Bacteria;p_Chloroflexi;c_Anaerolineae;o_Anaerolineales;f_Anaerolinaceae;_	2	~
k_Bacteria;p_Proteobacteria;c_Gammaproteobacteria;o_Pseudomonadales;f_Pseudomonadaceae;g_Pseudomonas	2	~
k Bacteria;p Proteobacteria;c Alphaproteobacteria;o Rhizobiales; ;	2	~
k_Bacteria;p_Bacteroidetes;c_Flavobacteriia;o_Flavobacteriales;_;_	2	~
k Bacteria;p Planctomycetes;c Planctomycetia;o Pirellulales;f Pirellulaceae;	2	~
k Bacteria;p Chloroflexi;c Anaerolineae;o Anaerolineales;f Anaerolinaceae;g Leptolinea	1	~
k Bacteria;p Proteobacteria;c Alphaproteobacteria;o Rhizobiales;f Beijerinckiaceae;	1	~
k_Bacteria;p_Bacteroidetes;c_Cytophagia;o_Cytophagales;f_Cyclobacteriaceae;g_Echinicola	1	~

Aa Property	# W	Rejection
k Bacteria;p Firmicutes;c Erysipelotrichi;o Erysipelotrichales;f Erysipelotrichaceae;g cc 115	1	~
k Bacteria;p Firmicutes;c Erysipelotrichi;o Erysipelotrichales;f Erysipelotrichaceae;g Clostridium	1	~
k Bacteria;p Bacteroidetes;c Bacteroidia;o Bacteroidales;f Rikenellaceae;g Alistipes	1	~
k Bacteria;p Bacteroidetes;c Bacteroidia;o Bacteroidales;f Porphyromonadaceae;g Petrimonas	1	~
k Bacteria;p Firmicutes;c Bacilli;o Lactobacillales;f Carnobacteriaceae;g Isobaculum	1	~
k Bacteria;p Bacteroidetes;c Bacteroidia;o Bacteroidales;f Marinilabiaceae;g Cytophaga	1	~
k Bacteria;p Firmicutes;c Clostridia;o Clostridiales;f ;g	1	~
k Bacteria;p Actinobacteria; ; ; ;	1	~
k Bacteria;p Firmicutes;c Clostridia;o Clostridiales;f Ruminococcaceae;g Anaerotruncus	1	~
k_Bacteria;p_Proteobacteria;c_Betaproteobacteria;o_Burkholderiales;f_Oxalobacteraceae;_	1	~
k Bacteria;p Proteobacteria;c Gammaproteobacteria;o Pasteurellales;f Pasteurellaceae;g Mannheimia	1	~
k_Bacteria;p_Firmicutes;c_Clostridia;o_Clostridiales;f_Ruminococcaceae;g_Oscillospira	1	~
k Bacteria;p Firmicutes;c Clostridia;o Clostridiales;f Lachnospiraceae;g Robinsoniella	1	~

Diversity analysis for dada2 pipeline

```
biom convert -i dada_normalized.tsv -o dada.biom --to-hdf5
```

```
qiime tools import \
    --input-path dada.biom \
    --type 'FeatureTable[Frequency]' \
    --input-format BIOMV210Format \
    --output-path dada.qza
```

```
qiime feature-table summarize \
--i-table dada.qza \
--o-visualization dada_feature.qzv \
--m-sample-metadata-file metadata.tsv
```

Table summary

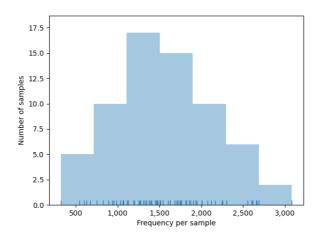
Aa Metric	
Number of samples	65
Number of features	89
Total frequency	102,120

Frequency per sample

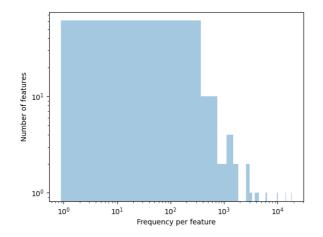
<u>Aa</u> Property	
Minimum frequency	318.3618967111167
1st quartile	1,127.1998461509056
Median frequency	1,506.7266086264692
3rd quartile	1,932.1853385551474
Maximum frequency	3,084.4565670690517
Mean frequency	1,571.0828567184328

Frequency per feature

<u>Aa</u> Property	

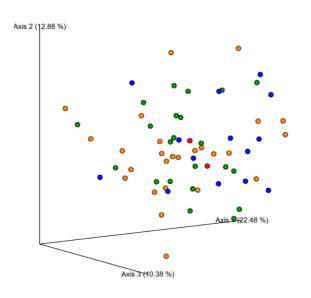


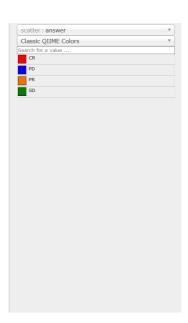
<u>Aa</u> Property	
Minimum frequency	0.879084641372472
1st quartile	9.897050753357469
Median frequency	51.59282314049115
3rd quartile	536.4770776069847
Maximum frequency	18,620.901263924585
Mean frequency	1,147.4200638954844



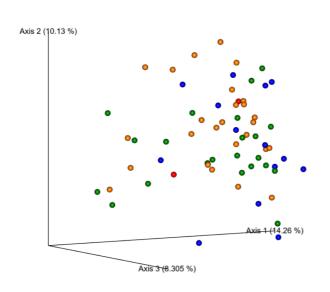
```
qiime diversity core-metrics \
--i-table dada.qza \
--m-metadata-file metadata.tsv \
--output-dir phylo \
--p-sampling-depth 500
```

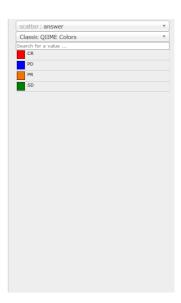
Clusters by bray-curtis distances:



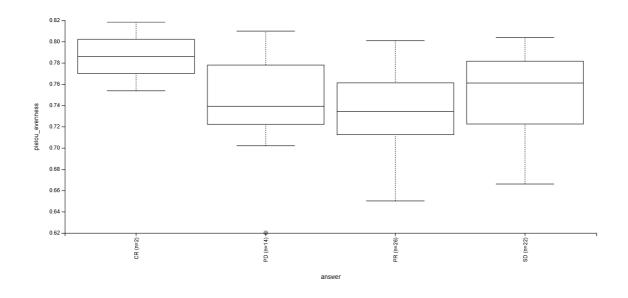


Clusters by jacard distances:





qiime diversity alpha-group-significance \
--i-alpha-diversity ./phylo/evenness_vector.qza \
--m-metadata-file metadata.tsv \
--o-visualization ./phylo/evenness-group-significance.qzv



Kruskal-Wallis (all groups)

<u>Aa</u> Property	# Result
<u>H</u>	4.98217455621301
p-value	0.17310707394219432

Kruskal-Wallis (pairwise)

<u>Aa</u> Property	≡ 1	≡н		
Group 1	Group 2			
<u>CR (n=2)</u>	PD (n=14)	1,613445	0,204008	0,408016
<u>CR (n=2)</u>	PR (n=26)	2,578249	0,108341	0,325024
<u>CR (n=2)</u>	SD (n=22)	1,090909	0,29627	0,444405
PD (n=14)	PR (n=26)	0,463147	0,496157	0,496157
PD (n=14)	SD (n=22)	0,557038	0,455456	0,496157
PR (n=26)	SD (n=22)	3,166548	0,075161	0,325024