

1. Prerequisites -

Install miniconda, qiime2=2023.7 and usearch. For details to install qiime2=2023.7, check [qiime2 documentation](#),

Follow this tutorial for detailed guidance -

<https://currentprotocols.onlinelibrary.wiley.com/doi/full/10.1002/cpbi.100>

2. Paired-end fastq files merged using [usearch](#)

Example command -

```
usearch11.0.667_i86linux32 -fastq_mergepairs
raw_seq_dir/001AU46701-16S_V1-V3_L6RGF_GATCGTCGCG-CTGGATATGT_L001_R1.fastq
-reverse
raw_seq_dir/001AU46701-16S_V1-V3_L6RGF_GATCGTCGCG-CTGGATATGT_L001_R2.fastq
-fastqout 001AU46701-16S_V1-V3_L6RGF_GATCGTCGCG-CTGGATATGT_L001_R1R2.fq
```

3. Creating metadata file from Raw Reads

```
echo -e "sample-id\tabsolute-filepath" > manifest.tsv

for f in *R1R2.fq; do
  n=$(basename "$f")
  echo -e "${n}\t${PWD}/${f}"
done >> manifest.tsv
```

4. Using manifest file to import raw reads into QIIME 2 artifacts

```
qiime tools import \  
  --input-path manifest.tsv \  
  --type 'SampleData[SequencesWithQuality]' \  
  --input-format SingleEndFastqManifestPhred33V2 \  
  --output-path paired_end-demux.qza
```

5. Creating a summary of the demultiplexed artifact to interactively explore the metadata (Optional Commands)

Summarizing manifest.tsv

```
qiime metadata tabulate \  
  --m-input-file manifest.tsv \  
  --o-visualization metadata-summary.qzv
```

qiime2view

File: metadata-summary.qzv

VisualizationDetailsProvenance

Download metadata TSV file

This file won't necessarily reflect dynamic sorting or filtering options based on the interactive table below.

Search:

sample-id	absolute-filepath
seq-types	category
001AU27472-16S_V1-V3-L6RGF-GCACCACCAA-TGGTACTGAT_L001_R1R2.fq	/home/salmaneunus27/001AU27472-16S_V1-V3-L6RGF-GCACCACCAA-TGGTACTGAT_L001_R1R2.fq
001AU33237-16S_V1-V3-L6RGF-GTTCGGAGTT-CGCCTTCTGA_L001_R1R2.fq	/home/salmaneunus27/001AU33237-16S_V1-V3-L6RGF-GTTCGGAGTT-CGCCTTCTGA_L001_R1R2.fq
001AU46701-16S_V1-V3-L6RGF-GATCGTCGCG-CTGGATATGT_L001_R1R2.fq	/home/salmaneunus27/001AU46701-16S_V1-V3-L6RGF-GATCGTCGCG-CTGGATATGT_L001_R1R2.fq

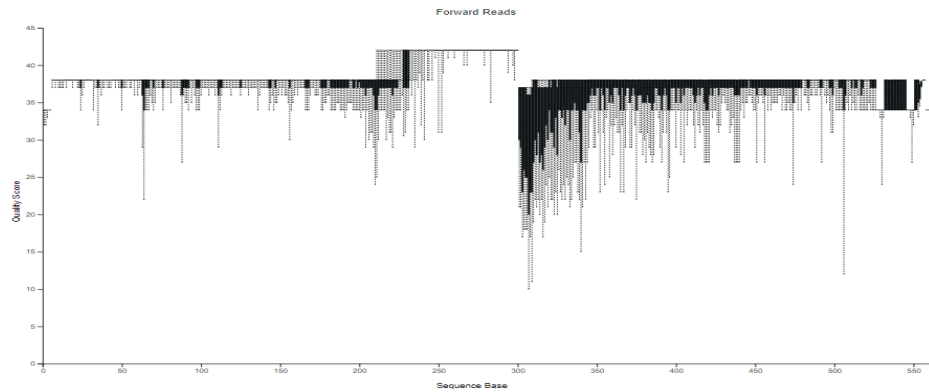
To get some basic info.

```
qiime tools inspect-metadata manifest.tsv
```

```
qiime2 2023.7.7 salmaneunus27@salman: ~  
=====  =====  
COLUMN NAME  TYPE  
=====  =====  
absolute-filepath  categorical  
=====  =====  
IDS: 3  
COLUMNS: 1
```

Summarizing paired-end reads

```
qiime demux summarize \  
--i-data paired_end-demux.qza \  
--o-visualization pe-demux.qzv
```



The plot at position 227 was generated using a random sampling of 10000 out of 119713 sequences without replacement. The minimum sequence length identified during subsampling was 35 bases. Outlier quality scores are not shown in box plots for clarity.

Parametric seven-number summary for position 227

Box plot feature	Percentile	Quality score
(Not shown in box plot)	2nd	26
Lower Whisker	9th	36
Bottom of Box	25th	38
Middle of Box	50th (Median)	38
Top of Box	75th	38
Upper Whisker	91st	42

6. Sequence Quality Control & Feature Table Construction

Initial filtering based on quality scores

```
qiime quality-filter q-score \  
--i-demux paired_end-demux.qza \  
--o-filtered-sequences pe-filtered.qza \  
--o-filter-stats pe-filter-stats.qza
```

Clustering sequences into OTUs Using q2-vsearch

Reference - <https://github.com/qiime2/docs/blob/master/source/tutorials/otu-clustering.rst>

Using the filtered sequences from the previous step, we will dereplicate the sequences to form a FeatureTable[Frequency] artifact and a FeatureData[Sequence] artifact

```
qiime vsearch dereplicate-sequences \
--i-sequences pe-filtered.qza \
--o-dereplicated-table feature-table.qza \
--o-dereplicated-sequences rep-seqs.qza
```

Closed-Reference Clustering of the OTU table

Clustering is performed at 85% identity against the Greengenes 13_8 85% OTUs reference database.

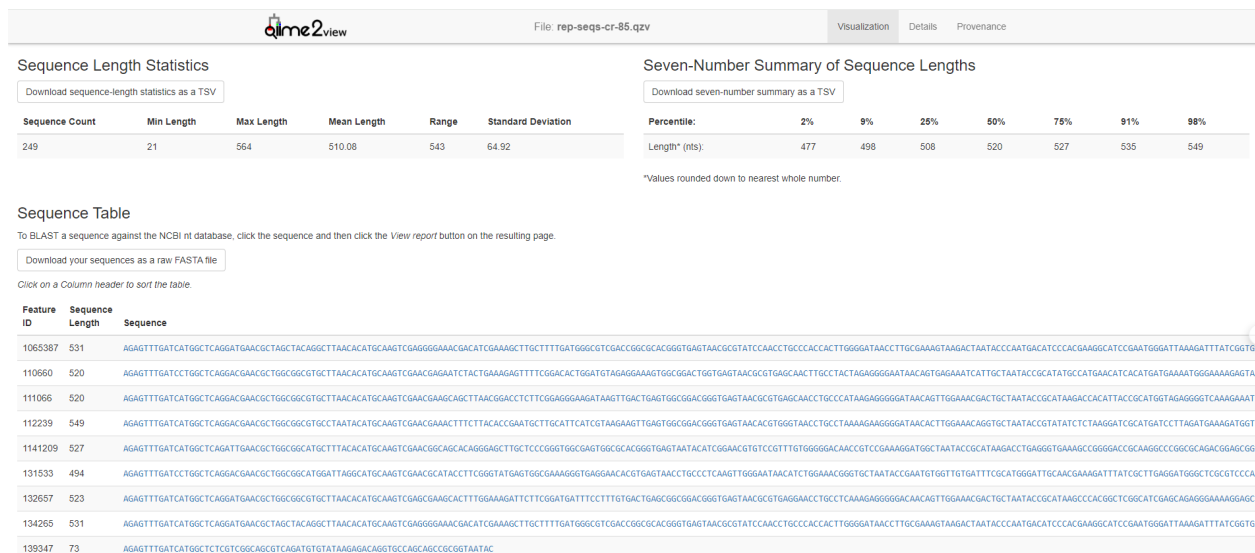
Downloading greengenes OTUs reference database

```
wget https://data.qiime2.org/2023.9/tutorials/otu-clustering/85_otus.qza
```

```
qiime vsearch cluster-features-closed-reference --i-table feature-table.qza
--i-sequences rep-seqs.qza --i-reference-sequences 85_otus.qza --p-perc-identity 0.85
--o-clustered-table table-cr-85.qza --o-clustered-sequences rep-seqs-cr-85.qza
--o-unmatched-sequences unmatched-cr-85.qza
```

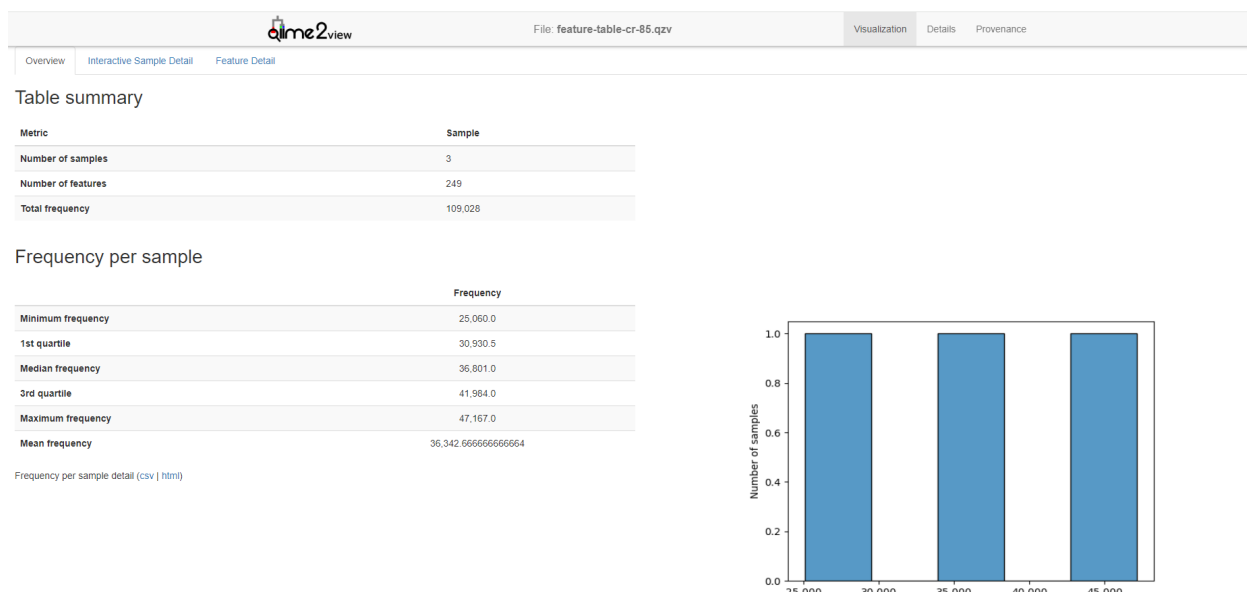
Visualizing the representative sequences after Closed-Reference Clustering

```
qiime feature-table tabulate-seqs --i-data rep-seqs-cr-85.qza --o-visualization
rep-seqs-cr-85.qzv
```



Visualizing feature table

```
qiime feature-table summarize --i-table table-cr-85.qza --m-sample-metadata-file manifest.tsv  
--o-visualization feature-table-cr-85.qzv
```



7. Generating Feature table in BIOM and CSV format

```
qiime tools export --input-path table-cr-85.qza --output-path exported-feature-table
```

	134265	1065387	551008	587572	215040	552456	185940	4308184	183870	530320
001AU27472-16S_V1-V3-L6RGF-GCACCACCAA-TGGTACTGAT_L001_R1R2.fq	35662.0	1400.0	4804.0	1.0	611.0	0.0	480.0	3.0	327.0	0.0
001AU33237-16S_V1-V3_L6RGF_GTTTCGGAGTT-CCGCTTCTGA_L001_R1R2.fq	15282.0	2273.0	2396.0	2.0	75.0	0.0	1181.0	2.0	17.0	6.0
001AU46701-16S_V1-V3_L6RGF_GATCGTCGCG-CTGGATATGT_L001_R1R2.fq	9094.0	1532.0	3943.0	440.0	804.0	346.0	2390.0	13.0	453.0	718.0

See details here to read .biom file and convert from .biom to .csv -

<https://colab.research.google.com/drive/1LcprGCrPnvrzujMZH37NTolMIUAvToWx?usp=sharing>

All OTU_ids matched with previous otu_taxonomy table -

https://colab.research.google.com/drive/1ivNYwkXj9f4EgGbRMzc-6w2U1frx59_k?usp=sharing

8. Taxonomic Classification

Downloading 3 required files from the inventory

```
wget
https://github.com/BenKaehler/readytowear/raw/master/data/gg_13_8/515f-806r/human-stool.qza

wget
https://github.com/BenKaehler/readytowear/raw/master/data/gg_13_8/515f-806r/ref-seqs.qza

wget https://github.com/BenKaehler/readytowear/raw/master/data/gg_13_8/515f-806r/ref-tax.qza
```

Training classifiers using these files

```
qiime feature-classifier fit-classifier-naive-bayes --i-reference-reads ref-seqs.qza
--i-reference-taxonomy ref-tax.qza --i-class-weight human-stool.qza --o-classifier
human_gut_classifier.qza
```

Assign taxonomy to our representative sequences using our newly trained classifier:

```
qiime feature-classifier classify-sklearn \
--i-reads rep-seqs-cr-85.qza \
--i-classifier human_gut_classifier.qza \
--o-classification agrf-taxonomy.qza
```

Now, we will visualizing the taxonomy:

```
qiime metadata tabulate \
--m-input-file agrf-taxonomy.qza \
--m-input-file rep-seqs-cr-85.qza \
--o-visualization agrf-taxonomy.qzv
```

qiime2view				File: agrf-taxonomy.qzv	Visualization	Details	Provenance
ID	Taxon	Confidence	Sequence				
197469	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__Alisipipes	0.7351869852415325	AGAGTTTGTATCGGCTCAGGATGAACGCTAGCTACAGGCTTAACACATGCAAGTCGAACGGAAGCGCTTGGAAAGATTCTTCGGATGATTTCTCTTGCTGACTGAGCGGCGAGCGGTGAGTAACGCGTGGTAACCTGCTCATACAGGGGGATAACAGTTAGAAAGTACTGCTAATACCGCATAGACCACGGTACCGG				
290002	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__Alisipipes	0.7733976935317363	AGAGTTTGTATCGGCTCAGGATGAACGCTGCGCGGCTGCCTAACACATGCAAGTCGAACGGAAGCACTTCTGAATGAGGCTTCGTCGAAGTTCAGATTTGACTTAAGGCGGACGCGGTGAGTAATGATGAGCAACCTGCCCTGCAGTGGGGGACAACAGTTGGAAGACGACTGCTAATACCGCATGATGATCTGGAAGG				
297390	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__Alisipipes	0.741151606996709	AGAGTTTGTATCGGCTCAGGATGAACGCTGCGCGGCTGCCTAACACATGCAAGTCGAACGGAAGCACTTCTTCGGAAGAACGCGCGAGCGGTGAGTAACGCGTGGTAACCTGCGCTGTACACAGCGATAACATACCBAAGGTATGCTAATACGAGATAATATGCTTTTATCGCATGTAAGATGATCAAAGTTTTTGGCGT				
298149	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__Alisipipes	0.7302399176854026	AGAGTTTGTATCGGCTCAGGATGAACGCTGCGCGGCTGCCTAACACATGCAAGTCGAACGGAAGCACTTCTTCGGAAGAACGCGCGAGCGGTGAGTAACGCGTGGTAACCTGCGCTGTACACAGCGATAACATACCBAAGGTATGCTAATACGAGATAATATGCTTTTATCGCATGTAAGATGATCAAAGTTTTTGGCGT				
298728	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__Alisipipes	0.7309745520222917	AGAGTTTGTATCGGCTCAGGATGAACGCTGCGCGGCTGCCTAACACATGCAAGTCGAACGGAAGCACTTCTTCGGAAGAACGCGCGAGCGGTGAGTAACGCGTGGTAACCTGCGCTGTACACAGCGATAACATACCBAAGGTATGCTAATACGAGATAATATGCTTTTATCGCATGTAAGATGATCAAAGTTTTTGGCGT				
300457	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__Alisipipes	0.7084241446130658	AGAGTTTGTATCGGCTCAGGATGAACGCTGCGCGGCTGCCTAACACATGCAAGTCGAACGGAAGCACTTCTTCGGAAGAACGCGCGAGCGGTGAGTAACGCGTGGTAACCTGCGCTGTACACAGCGATAACATACCBAAGGTATGCTAATACGAGATAATATGCTTTTATCGCATGTAAGATGATCAAAGTTTTTGGCGT				
355868	k__Bacteria; p__Bacteroidetes; c__Bacteroidia; o__Bacteroidales; f__Rikenellaceae; g__Alisipipes	0.7800781122149469	AGAGTTTGTATCGGCTCAGGATGAACGCTGCGCGGCTGCCTAACACATGCAAGTCGAACGGAAGCACTTCTTCGGAAGAACGCGCGAGCGGTGAGTAACGCGTGGTAACCTGCGCTGTACACAGCGATAACATACCBAAGGTATGCTAATACGAGATAATATGCTTTTATCGCATGTAAGATGATCAAAGTTTTTGGCGT				
524633	k__Bacteria; p__Bacteroidetes;	0.751354597668002	AGAGTTTGTATCGGCTCAGGATGAACGCTGCGCGGCTGCCTAACACATGCAAGTCGAACGGAAGCACTTCTTCGGAAGAACGCGCGAGCGGTGAGTAACGCGTGGTAACCTGCGCTGTACACAGCGATAACATACCBAAGGTATGCTAATACGAGATAATATGCTTTTATCGCATGTAAGATGATCAAAGTTTTTGGCGT				

	id	Taxon	Confidence	Sequenc
0	#q2:types	categorical	categorical	categori
1	1065387	k__Bacteria; p__Firmicutes; c__Clostridia; o__...	0.7176739941419537	AGAGTTTGATCATGGCTCAGGATGAACGCTAGCTACAGGCTTAAC/
2	110660	k__Bacteria	0.998410293516579	AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCTTAAC/
3	111066	k__Bacteria; p__Firmicutes; c__Clostridia; o__...	0.7497422366344827	AGAGTTTGATCATGGCTCAGGACGAACGCTGGCGGCGTGCTTAAC/
4	112239	k__Bacteria	0.9980215995275771	AGAGTTTGATCATGGCTCAGGACGAACGCTGGCGGCGTGCCTAAT/
5	1141209	k__Bacteria; p__Firmicutes	0.712636651358992	AGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCATGCTTTAC/
6	131533	k__Bacteria; p__Firmicutes; c__...	0.8163357852277051	AGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCATGGATTAGC/

Taxonomy Based Filtering of Data

Excluding sequences that are unexpected such as those from chloroplasts or mitochondria

```
qiime taxa filter-table --i-table table-cr-85.qza --i-taxonomy agrf-taxonomy.qza --p-mode
contains --p-include p__ --p-exclude 'p__;Chloroplast,Mitochondria' --o-filtered-table
filtered-table-3.qza
```

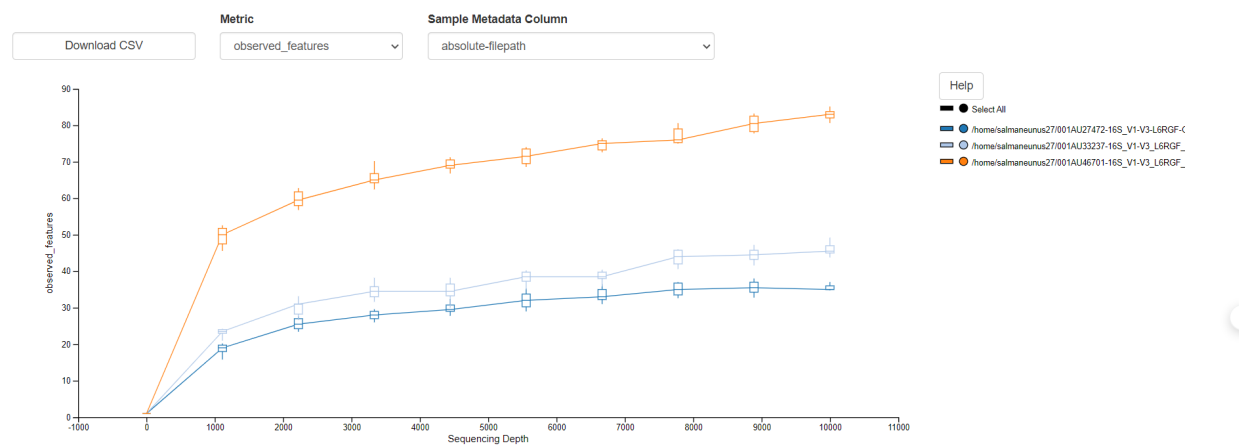
In this step we are removing those features to save computational time

```
qiime feature-table filter-seqs --i-data rep-seqs-cr-85.qza --i-table filtered-table-3.qza
--o-filtered-data filtered-sequences-2.qza
```

Visualizing the filtered taxonomy

```
qiime taxa barplot --i-table filtered-table-3.qza --i-taxonomy agrf-taxonomy.qza
--m-metadata-file manifest.tsv --o-visualization taxa-bar-plots-1.qzv
```


Alpha rarefaction



shannon.csv

Open with

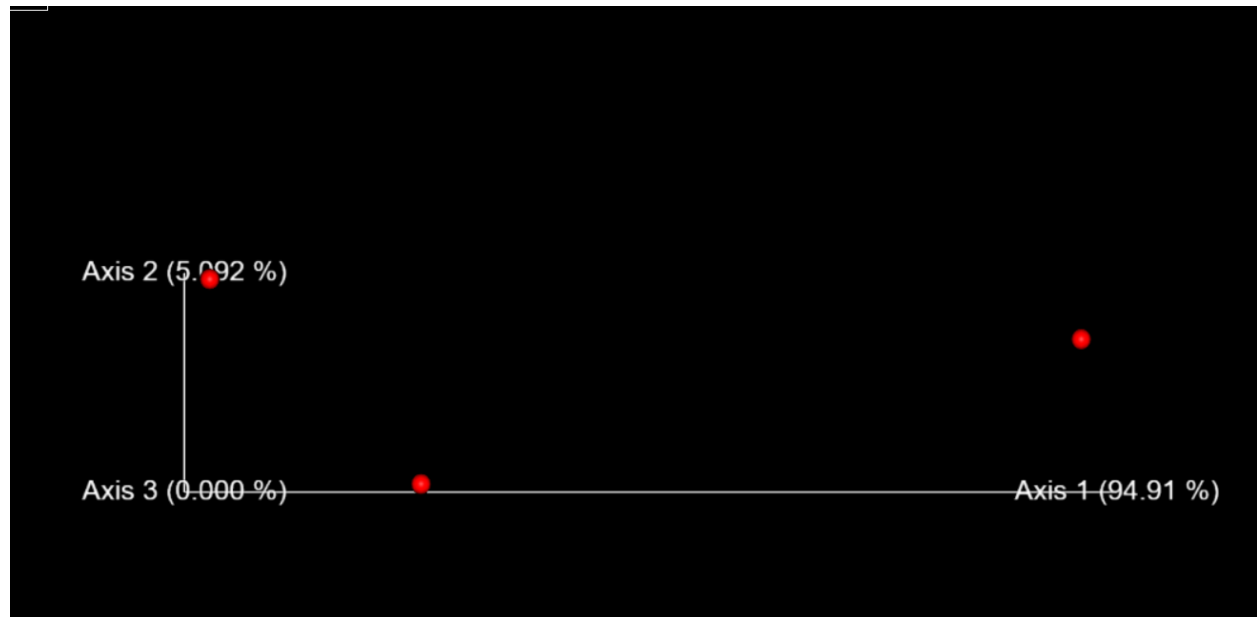
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	sample-id	depth-1_iter=1	depth-1_iter=2	depth-1_iter=3	depth-1_iter=4	depth-1_iter=5	depth-1_iter=6	depth-1_iter=7	depth-1_iter=8	depth-1_iter=9	depth-1_iter=10	depth-1112_iter=1	depth-1112_iter=2	depth-1112_iter=3	depth-1112_iter=4	depth-1112_iter=5	depth-1112_iter=6	depth-1112_iter=7	depth-1112
2	001AU27472-16S_V	0	0	0	0	0	0	0	0	0	0	1.118217064	1.115815523	1.185584482	1.172720022	1.157327444	1.002555005	1.247897477	1.09
3	001AU33237-16S_V	0	0	0	0	0	0	0	0	0	0	1.940891076	1.890068525	1.885324882	1.953839708	1.887352536	1.882679182	1.893771847	1.93
4	001AU46701-16S_V	0	0	0	0	0	0	0	0	0	0	3.308693718	3.341789498	3.213833001	3.19258002	3.348255072	3.255820044	3.117383891	3.27

More csv output files can be found here - [drive link](#)

Alpha & Beta diversity

```
qiime diversity core-metrics-phylogenetic --i-table filtered-table-3.qza --i-phylogeny phylogeny-align-to-tree-mafft-fasttree/rooted_tree.qza --p-sampling-depth 3400 --m-metadata-file manifest.tsv --p-n-jobs-or-threads 1 --output-dir child-norep-core-metrics-results
```

Weighted Unifrac distance



Statistical tests on diversity