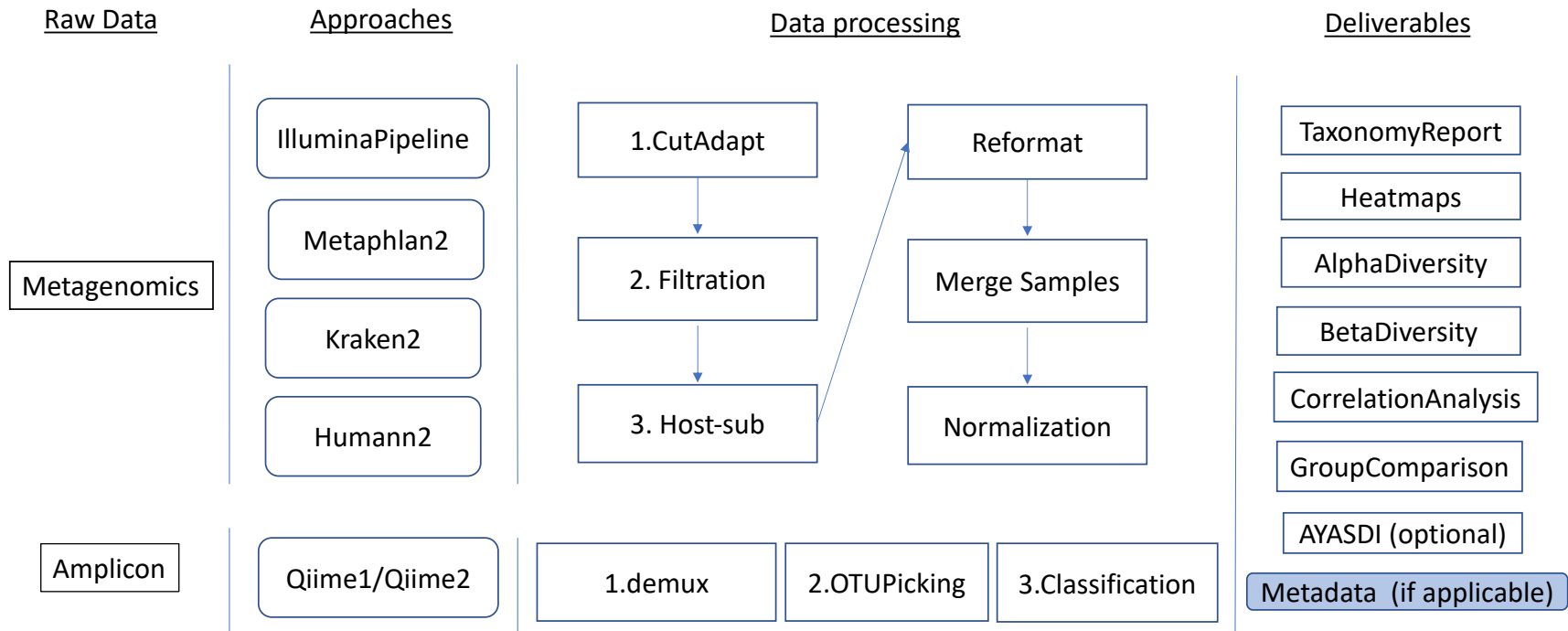


Metagenomics Pipeline



16S/ITS analysis pipeline

Installation of Qiime2 and tutorials:

<https://docs.qiime2.org/2019.7/>

Building/downloading databases:

<https://docs.qiime2.org/2019.7/tutorials/importing/>

<https://docs.qiime2.org/2019.7/data-resources/>

Forum for QA:

<https://forum.qiime2.org/>

Commands used for pipeline:

File preparation

```
mkdir emp-single-end-sequences
```

move R1 and index fastq file into the emp-single-end-sequences directory

```
qiime tools import --type EMPSingleEndSequences --input-path emp-single-end-sequences --output-path emp-single-end-sequences.qza
```

Option1, use the wrapper script:

```
sh ~/pipelines/Github/MetagenomicsPipeline/16S_pipeline.V8_Mac.sh  
/Users/chengguo/Desktop/Project/BRA/BRA_merged/MK_BRA_AllWave3Samples_Mapping.txt 10000 1000 Wave  
~/Desktop/Hengchuang/16S_reference/gg-13-8-99-515-806-nb-classifier.qza ~/Desktop/Hengchuang/16S_reference/gg-13-5-97.rep_set.otu.qza N Dam
```

Option2, use individual commands:

```
qiime demux emp-single --i-seqs emp-single-end-sequences.qza --m-barcodes-file RISE-ITS-mapping.txt --m-barcodes-column  
BarcodeSequence --o-per-sample-sequences demux.qza --p-rev-comp-barcodes  
qiime demux summarize --i-data demux.qza --o-visualization demux.qzv
```

```
zcat <sequences.fastq.gz | head
```

```
qiime dada2 denoise-single --i-demultiplexed-seqs demux.qza --p-trim-left 10 --p-trunc-len 250 --o-representative-sequences rep-  
seqs-dada2.qza --o-table table-dada2.qza --p-n-threads 0 --o-denoising-stats stats-dada2.qza
```

```
qiime feature-classifier classify-sklearn --i-classifier ~/Desktop/Hengchuang/ITS_reference/UNITE/2017.12/unite-ver7-dynamic-  
classifier-01.12.2017.qza --i-reads rep-seqs-dada2.qza --o-classification taxonomy.qza
```

```
mv table-dada2.qza table.qza
```

```
mv rep-seqs-dada2.qza rep-seqs.qza
```

```
qiime feature-table summarize --i-table table.qza --o-visualization table.qzv --m-sample-metadata-file RISE-ITS-mapping.txt
```

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

```
qiime alignment mafft --i-sequences rep-seqs.qza --o-alignment aligned-rep-seqs.qza
```

```
qiime alignment mask --i-alignment aligned-rep-seqs.qza --o-masked-alignment masked-aligned-rep-seqs.qza
```

```
qiime phylogeny fasttree --i-alignment masked-aligned-rep-seqs.qza --o-tree unrooted-tree.qza
```

```
qiime phylogeny midpoint-root --i-tree unrooted-tree.qza --o-rooted-tree rooted-tree.qza
```

```
for f in rep-seqs.qza table.qza taxonomy.qza ; do echo $f; qiime tools export --input-path $f --output-path exported; done
```

```
qiime tools export --input-path rooted-tree.qza --output-path exported/
```

```
mv exported/tree.nwk exported/tree.rooted.nwk
```

```
qiime tools export --input-path unrooted-tree.qza --output-path exported/
```

```
mv exported/tree.nwk exported/tree.unrooted.nwk
```

```
biom add-metadata -i exported/feature-table.biom -o exported/feature-table.taxonomy.biom --observation-metadata-fp  
exported/taxonomy.tsv --observation-header OTUID,taxonomy,confidence
```

```
biom convert -i exported/feature-table.taxonomy.biom -o exported/feature-table.taxonomy.txt --to-tsv --header-key taxonomy
```

```
biom convert -i exported/feature-table.taxonomy.biom -o exported/feature-table.txt --to-tsvmv exported/tree.nwk  
exported/tree.rooted.nwk
```

```
biom convert -i exported/feature-table.taxonomy.biom -o exported/feature-table.txt --to-tsv
```

```
perl ~/pipelines/Github/MetagenomicsPipeline/stat_otu_tab.pl -unif min exported/feature-table.taxonomy.txt -prefix  
exported/Relative/otu_table --even exported/Relative/otu_table.even.txt -spestat exported/Relative/classified_stat_relative.xls
```

```
mv exported/Relative/otu_table.p.relative.mat exported/Relative/otu_table.Phylum.relative.txt  
mv exported/Relative/otu_table.c.relative.mat exported/Relative/otu_table.Class.relative.txt  
mv exported/Relative/otu_table.o.relative.mat exported/Relative/otu_table.Order.relative.txt  
mv exported/Relative/otu_table.f.relative.mat exported/Relative/otu_table.Family.relative.txt  
mv exported/Relative/otu_table.g.relative.mat exported/Relative/otu_table.Genus.relative.txt  
mv exported/Relative/otu_table.s.relative.mat exported/Relative/otu_table.Species.relative.txt
```

```
qiime taxa barplot --i-table table.Wave3.MS.qza --i-taxonomy taxonomy.qza --m-metadata-file  
MK_BRA_Wave3_Mother.Stool_MappingUpdate1.txt --o-visualization taxa-bar-plots.qzv
```

```
qiime diversity alpha --i-table table.Wave3.BM.qza --p-metric chao1 --o-alpha-diversity core-metrics-results/chao1_vector.qza
```

```
for c in faith_pd evenness shannon observed_otus chao1; do echo $c; qiime diversity alpha-group-significance --i-alpha-diversity core-metrics-results/${c}_vector.qza --m-metadata-file MK_BRA_Wave3_Breast.Milk_MappingUpdate.txt --o-visualization core-metrics-results/${c}-group-significance.qzv; done
```

```
qiime diversity core-metrics-phylogenetic --i-phylogeny rooted-tree.qza --i-table table.Wave3.IS.qza --p-sampling-depth 10000 --m-metadata-file InfantStool_Mapping_20190806.txt --output-dir core-metrics-results
```

```
qiime diversity alpha-rarefaction --i-table table.Wave3.IS.qza --p-max-depth 10000 --m-metadata-file InfantStool_Mapping_20190806.txt --i-phylogeny rooted-tree.qza --p-metrics chao1 --p-metrics shannon --p-metrics observed_otus --p-metrics faith_pd --o-visualization alpha-rarefaction_withchao1.qzv
```

###Procrustes analysis

```
qiime diversity procrustes-analysis --i-reference exported.NoS/core-metrics-results.9800/unweighted_unifrac_pcoa_results.qza --i-other exported.NoS/picrust2_out_pipeline/pathways_out/core-metrics-results/bray_curtis_pcoa_results.qza --output-dir procrustes-analysis
```

```
qiime emperor procrustes-plot --i-reference-pcoa exported.NoS/core-metrics-results.9800/unweighted_unifrac_pcoa_results.qza --i-other-pcoa exported.NoS/picrust2_out_pipeline/pathways_out/core-metrics-results/bray_curtis_pcoa_results.qza --m-metadata-file mapping_all_20200910.txt --o-visualization procrustes-analysis/procrustes-plot.qzv
```

```
qiime diversity mantel --i-dm1 exported.NoS/core-metrics-results.9800/unweighted_unifrac_distance_matrix.qza --i-dm2 exported.NoS/picrust2_out_pipeline/pathways_out/core-metrics-results/bray_curtis_distance_matrix.qza --o-visualization procrustes-analysis/mantel.qzv  
Saved Visualization to: procrustes-analysis/mantel.qzv
```

###

Picrust2 for functional prediction with 16S data

Official tutorial: <https://github.com/picrust/picrust2>

Commands used for pipeline:

```
picrust2_pipeline.py -s study_seqs.fna -i study_seqs.biom -o picrust2_out_pipeline -p 1
```

```
add_descriptions.py -i EC_metagenome_out/pred_metagenome_unstrat.tsv.gz -m EC \  
-o EC_metagenome_out/pred_metagenome_unstrat_descrip.tsv.gz
```

```
add_descriptions.py -i KO_metagenome_out/pred_metagenome_unstrat.tsv.gz -m KO \  
-o KO_metagenome_out/pred_metagenome_unstrat_descrip.tsv.gz
```

```
add_descriptions.py -i pathways_out/path_abun_unstrat.tsv.gz -m METACYC \  
-o pathways_out/path_abun_unstrat_descrip.tsv.gz
```

Metagenomics analysis pipeline

Background information:

<http://huttenhower.sph.harvard.edu/humann2>
<https://bitbucket.org/biobakery/biobakery/wiki/metaphlan2>
<https://ccb.jhu.edu/software/kraken2/index.shtml?t=manual>
<https://ccb.jhu.edu/software/bracken/>
<https://github.com/jiwoongbio/FMAP>
<https://bitbucket.org/biobakery/biobakery/wiki/lefse>
<https://github.com/gc26762524/MetagenomicsPipeline.git>

Commands used for pipeline:

##Raw fastq files should be analyzed with IlluminaPipelines to the step of hostsubtraction.

```
ls /share/data/IlluminaData/PCS/Host_subtracted_2019-01-12/bowtie/host_repeats_genomic/P*R1.host_repeats_genomic.unmapped.R1.fastq.gz > PCS.hostsubR1R1.list
```

#Metaphlan2

```
perl run_metaphlan2.pl RAO.hostsubR1R1.list PE 2
```

```
python ~/software/metaphlan2/utils/merge_metaphlan_tables.py LGY-0* > LGY.merged_metaphlan_table.20180301.txt
```

```
python ~/software/metaphlan2/utils/metaphlan_hclust_heatmap.py -c bbcry --top 25 --minv 0.1 -s log --in LGY.merged_metaphlan_table.20180301.txt --tax_lev g --out LGY.merged_abundance_table.genus.png
```

```
python ~/software/metaphlan2/utils/metaphlan_hclust_heatmap.py -c bbcry --top 25 --minv 0.1 -s log --in LGY.merged_metaphlan_table.20180301.txt --tax_lev p --out LGY.merged_abundance_table.phylum.png
```

```
python ~/software/metaphlan2/utils/metaphlan_hclust_heatmap.py -c bbcry --top 25 --minv 0.1 -s log --in LGY.merged_metaphlan_table.20180301.txt --tax_lev f --out LGY.merged_abundance_table.family.png
```

```
python ~/software/metaphlan2/utils/metaphlan_hclust_heatmap.py -c bbcry --top 100 --minv 0.1 -s log --in  
LGY.merged_metaphlan_table.20180301.txt --tax_lev s --out LGY.merged_abundance_table.species.png
```

#Humann2

```
perl run_humann.pl RAO.hostsubR1R1.list SE
```

```
for f in ../../Pool*/Metagenome/Humann/LGY*/*genefamilies.tsv ; do echo $f; ln -s $f ./;done
```

```
for f in LGY-0*.tsv; do echo $f; humann2 --input $f --pathways unipathway --output ./; done
```

```
humann2_join_tables -i ./ -o LGY.all.genefamilies.tsv --file_name genefamilies.tsv
```

```
humann2_join_tables -i ./ -o LGY.all.pathabundance.tsv --file_name pathabundance.tsv
```

```
humann2_join_tables -i ./ -o LGY.all.pathcoverage.tsv --file_name pathcoverage.tsv
```

```
humann2_renorm_table -i LGY.all.pathabundance.tsv -o LGY.all.pathabundance.cpm.tsv --units cpm
```

```
humann2_renorm_table -i LGY.all.genefamilies.tsv -o LGY.all.genefamilies.cpm.tsv --units cpm
```

#Kraken2/Bracken

```
perl run_kraken2.pl RAO.hostsubR1R1.list 2 /share/data/software/kraken2/kraken2/Dec_2018_bacteria/ SE
```

```
for f in *report; do echo $f; python ~/pipelines/Github/Bracken/src/est_abundance.py -i $f -k  
/share/data/software/kraken2/kraken2/Dec_2018_bacteria/database140mers.kmer_distrib -l S -o ${f}.bracken;done
```

```
perl Braken2_to_OTUtable.pl /share/data/16sData/taxid2OTU_ranks.txt RSA.EuPathKraken2.list
```

```
bash ~/pipelines/Github/MetagenomicsPipeline/visualize_otu_table_by_qiime2.sh  
/Users/chengguo/Desktop/Project/LGY/Kraken2/All.merged.OTU.Renamed.txt  
/Users/chengguo/Desktop/Project/LGY/LGY_Mapping.Keemei.tsv SampleType,Family,MotherDaughter LGY
```



```
while read f; do echo $f; echo 'perl ~/pipelines/Github/FMAP/FMAP_mapping.pl' $f '>' ${f}.mapping.txt | qsub -V -N $f -cwd -l mem=3G,time=24:: -o $PWD -e $PWD -pe smp 4; done < x.pool1
```

```
for f in *mapping.txt; do echo $f; perl ~/pipelines/Github/FMAP/FMAP_quantification.pl $f > ${f}.abundance.txt; done
```

```
ls *abundance.txt | tr "\n" " "
```

```
perl ~/pipelines/Github/FMAP/FMAP_table.pl -n Case_P10_AB037.mapping.txt.abundance.txt  
Case_P10_AB047.mapping.txt.abundance.txt Case_P10_AB049.mapping.txt.abundance.txt  
Case_P10_AB052.mapping.txt.abundance.txt Control_P9_CD085.mapping.txt.abundance.txt  
Control_P9_EF027.mapping.txt.abundance.txt Control_P9_EF036.mapping.txt.abundance.txt  
Control_P9_GH061.mapping.txt.abundance.txt Control_P9_GH062.mapping.txt.abundance.txt  
Control_P9_GH063.mapping.txt.abundance.txt Control_P9_IJ029.mapping.txt.abundance.txt > all.merged.abundance.KeepID.KO.txt
```

```
perl ~/pipelines/Github/MetagenomicsPipeline/ConvergeKO2Module.pl all.merged.abundance.KeepID.KO.txt >  
all.merged.abundance.KeepID.Module.txt
```

```
perl ~/pipelines/Github/MetagenomicsPipeline/ConvergeKO2Pathway.pl all.merged.abundance.KeepID.KO.txt >  
all.merged.abundance.KeepID.Pathway.txt
```

#LEFSE

```
Rscript ~/pipelines/Github/MetagenomicsPipeline/make_lefse_input.R -i all.merged.abundance.KeepID.Pathway.txt -m  
RSA_mapping.txt -g Group1,Group2
```

```
for f in *.lefse.txt; do echo $f; perl -p -i.bak -e 's/\r/\n/g' $f; lefse-format_input.py $f "${f}.in" -c 1 -u 2 -o 1000000; run_lefse.py ${f}.in  
${f}.1vsALL.res; run_lefse.py -y 1 ${f}.in ${f}.ALLvsALL.res; done
```

```
for f in *res; do echo $f; cat $f | sort -k3,3 -k4,4 -r > ${f}.txt; done
```

```
for f in *res.txt; do echo $f; perl ~/pipelines/Github/MetagenomicsPipeline/ConvertID2Annotation_LEFSE_output.pl  
~/pipelines/Github/MetagenomicsPipeline/FMAP_data/KEGG_orthology_pathway_module.txt $f > ${f}.anno; done
```

```
perl ~/pipelines/rename.pl 's/KeepID/KeepAnnotation/' *txt.anno
```

```
perl ~/pipelines/rename.pl 's/txt\.anno/txt/' *txt.anno
```

```
for f in all.merged.abundance.KeepAnnotation*1vsALL.res.txt; do echo $f; base=$(basename $f .txt);echo $base; perl -p -i.bak -e  
's/\r/\n/g' $f; lefse-plot_res.py --left_space 0.3 --dpi 300 ${base}.txt ${base}.png; lefse-plot_res.py --max_feature_len 200 --  
orientation h --format pdf --left_space 0.3 --dpi 300 ${base}.txt ${base}.pdf; done
```