Microbiome sequence analysis SOP

Microbiome Analysis pipeline

Quality assessments and Chimera detection

1. Primer and barcodes trimming using Qimme
2. Ambigious base calls with homopolymer runs filtering using Qimme
3. De-noising sequences and Chimeras will be detected using Qimme USEARCH method on de-multiplexed sequences.

The Qimme analysis workflow

1. OTUs will be generated and chimeras will be removed using Qiime using USEARCH method on a set of de-multiplexed sequences.
2. Operational taxonomic units (OTUs) will be defined by clustering with 97% sequence similarity cutoffs (at 3% divergence).
3. Then the representative sequence for an OTU will be chosen as the most abundant sequence showing up in that OTU's by collapsing identical sequences, and choosing the one that was read the most abundant sequences.
4. Then representative sequences will be aligned against Greengenes database core set using PyNAST alignment method. The minimum sequence length of 150 and the minimum percent id of 75% will be used for the alignment.
5. The RDP Classifier program will used to assign the taxonomy to the representative set of sequences using a pre-built database of assigned sequence of reference set.
6. Sequence alignments will be filtered for sequence comprised of gaps and excessive variable locations before generating phylogenetic tree relating the sequences.
7. OTU tables (A table matrix of OTU abundance in each sample with taxonomic identifiers) will be generated using taxonomic assignments.

Downstream analysis

Community summarization by taxonomic composition

- OTUs clusters will be grouped by samples by different taxonomic levels (division, class, family, etc.) and area and bar plots will be generated to show taxonomic abundance.

Measuring Population Diversity

- For Alpha Diversity (within-sample taxonomic diversity):
  - Multiple rarefactions analysis will be performed using QIMME, PhyloSeq and other tools with default selection of 10 sequence/sample, and stepping up to 57601 sequence /sample in increments of 5759.

  - Alpha diversity within the samples will be computed using multiple OTU tables. The Chao1 metric (estimates the species richness.), the observed species metric (the count of unique OTUs found in the sample) and Phylogenetic Distance (PD_whole_tree) will be created. Using these three matrices rarefaction plots will be generated.

    - Beta Diversity (between sample taxanomic diversity i). Beta diversity distance matrix of the distances of all samples to all other samples that reflects the dissimilarity between those samples will be created. ii). This distance matrix will be visualized using Principal Coordinate Analysis (PCoA) to visualize distances between samples on an x-y-z plot iii). Using UPGMA (Unweighted Pair Group Method with Arithmetic mean) hierarchical clustering method.

OTU significance and co-occurence analysis

- ANOVA (ANOVA) method will be used to determine whether OTU relative abundance is different between categorie.That is to measure if OTUs are increased or decreased in relative abundance
- Following software will be used for the detection of differentially abundant taxonomic groups:
    - Metstat
    - Qimme
    - Vegan
    - phyloseq
    - Boruta

Class comparison and functional profiling will be performed using

- LEfSe & PICRUSt, MEGAN, FANTOM using following database KEGG, MetaCyc and COG

Visualization

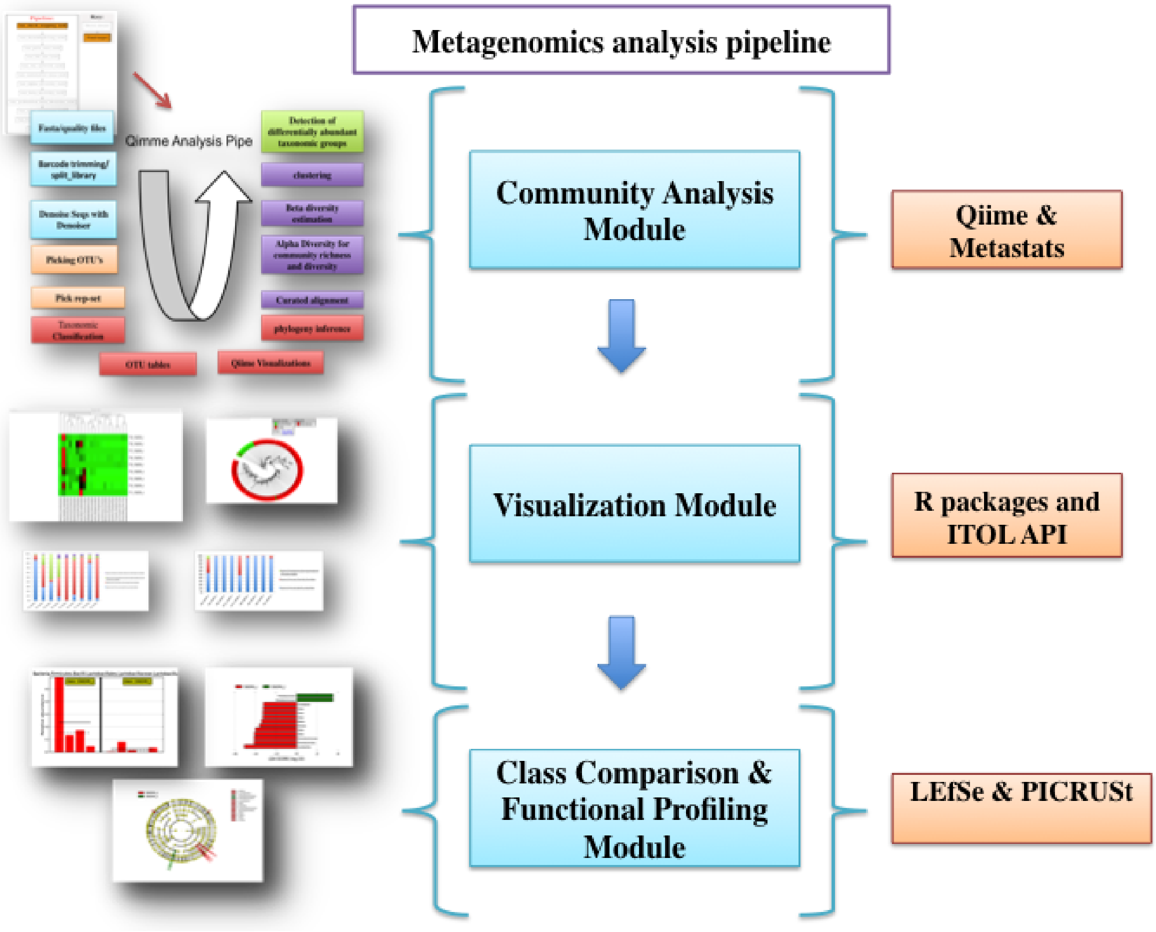
Heatmap Visualization of maximum relative abundance

- The heat map will be generated using RDP classified OTU table by removing genera with relative read abundance less than 1% of at least 1 sample. This leaves us with filtered genera.
- The dendogram for the samples in the heat map will be based on the distance matrices calculated by vegan R package using Bray-Curtis dissimilarity matrix and clustering using average linkage hierarchical clustering method.

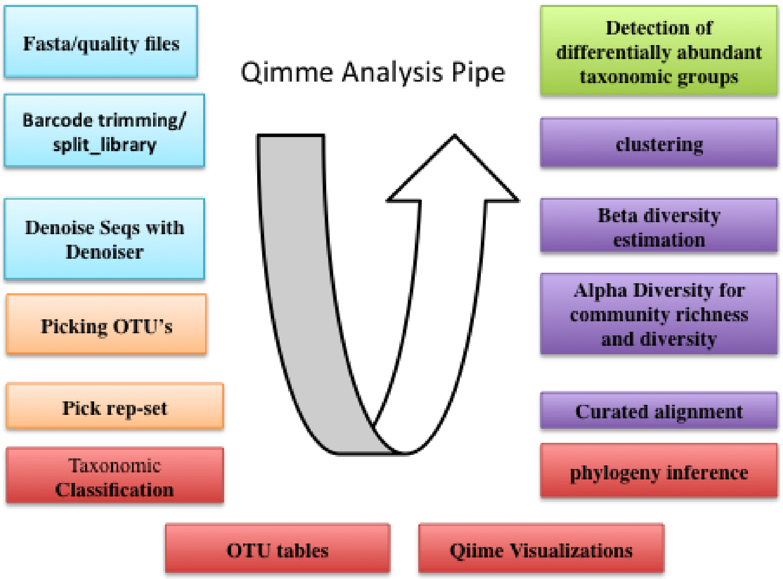
Other Visulizations

- Other visualization will be performed using ITOL API, MEGAN, QIMME, PhyloSeq, R scripts/Python scripts, FigTree, CytoScape, Krona, MetaSee and Sysbiocube
- Phylogenetic tree will be visualized using ITOL API, MEGAN, FigTree
- PCA and taxonomic Bar charts will be visualized using Krona, QIMME, Sysbiocube, R scripts/Python scripts, PhyloSeq
- OTUs network visualization will be performed using CytoScape
- Functional profiling based visualization will be performed using PICRUST and LEfSE and MEGAN

Microbiome Sequence analysis pipeline


Metagenomics analysis pipeline
Fasta/quality files
Barcode trimming/ split_library
Denoise Seqs with Denoiser
Picking OTU's
Pick rep-set
Taxonomic Classification
OTU tables
Qiime Visualizations
Qimme Analysis Pipe
Detection of differentially abundant taxonomic groups
clustering
Beta diversity estimation
Alpha Diversity for community richness and diversity
Curated alignment
phylogeny inference
Community Analysis Module
Visualization Module
Class Comparison & Functional Profiling Module
Qiime & Metastats
R packages and ITOL API
LEfSe & PICRUSt

Qimme analysis Pipeline in detail



Qimme Analysis Pipe

Fasta/quality files

Barcode trimming/ split_library

Denoise Seqs with Denoiser

Picking OTU’s

Pick rep-set

Taxonomic Classification

OTU tables

Qiime Visualizations

phylogeny inference

Curated alignment

Alpha Diversity for community richness and diversity

Beta diversity estimation

clustering

Detection of differentially abundant taxonomic groups