16S rRNA Gene Sequencing Data Analysis

Demultiplexed raw amplicon sequences will be processed using an open-source software package Quantitative Insights Into Microbial Ecology, QIIME 2 2018.8 (REF 1). Denoising, and dereplication, of paired-end sequences including chimera removal and trimming of reads based on positional quality scores will be performed using the Divisive Amplicon Denoising Algorithm 2 (DADA2), an amplicon-specific error-correction method that models and corrects Illumina-sequenced amplicon errors (REF 2). Briefly, a feature table containing counts of each unique sequence variant in the samples will be constructed using DADA2. A feature is essentially any unit of observation, e.g., an operational taxonomic unit (OTU), an amplicon sequence variant (ASV), a gene or a metabolite. In QIIME2, most features are OTUs or ASVs; OTUs are identified via clustering method VSEARCH (REF 3) or, ASVs are identified via DADA2 or DEBLUR (REF 4). An OTU is a cluster of sequences that differ by less than a fixed dissimilarity threshold (typically 3%). ASVs delineate sequences even if they vary by only one base pair, giving us distinct units that otherwise would be lost with any form of OTU clustering. However, there can be instances where greater resolution isn't helpful (it depends on the dataset and on what you are trying to find out).

DIVERSITY METRICS: In order to calculate alpha diversity metrics including observed feature counts (or observed OTUs), and Shannon and Simpson diversity indices, the feature table containing ASVs will be rarefied. We further calculate an estimate of microbial beta-diversity using weighted or unweighted UniFrac distances as implemented in QIIME2. A number of diversity metrics like unifrac distance require the construction of a phylogenetic tree. A rooted phylogenetic tree will be constructed with a set of sequences representative of the ASVs using FastTree or RAXML method (REF 5). A summary of beta diversity relationships will be visualized using principal coordinate analysis (PCoA) plots.

TAXONOMIC ASSIGNMENT: Taxonomic composition analysis will be performed to know what kinds of organisms are present in each sample. The taxonomy of each ASV is established, by matching to the GreenGenes (v13_8, 97% clustered OTUs), Silva, RDP or Human Oral Microbiome Database (HOMD) database, based on a naive Bayesian classifier with default parameters (REF 6,7,8,9). For differential abundance test, ANCOM (Analysis of Composition of Microbiomes) will be used to identify features that are differentially abundant across groups and, at a specific taxonomic level (REF 10).

MOCK MICROBIAL COMMUNITIES: The inclusion of negative, reagent-only controls i.e. DNA extraction kit and PCR controls, and a mock community is strongly encouraged. Mock microbial communities comprising genomic DNA or whole cells. The DNA mock community will assess the efficiency of PCR, sequencing and analysis steps whereas the whole cell mock community is useful for establishing the efficiency of the DNA extraction steps. Mock communities are available from the ATCC (American Type Culture Collection) and Zymo Research (REF 11).

Updated on 11/05/2018

REQUIREMENTS:

a. Raw data (eg. FASTQ format)
b. Forward and reverse primers used to amplify 16S rRNA gene region (For example primers for V3-V4 region; 515f PCR Primer (GTGCCAGCMGCCGCGGTAA) and 806r PCR Primer (GGACTACHVGGGTWTCTAAT).
c. Metadata spreadsheet (use this example, https://data.qiime2.org/2018.6/tutorials/moving-pictures/sample_metadata.tsv)
d. Select one of the following databases for taxonomic reference data
   I. expanded Human Oral Microbiome Database (eHOMD)
   II. Silva 132 99% OTUs (for V3-V4 region) and
   III. Greengenes 13_8 97% OTUs (any region of 16S rRNA eg., V3, V4 or V3-V4 etc.)
   IV. Ribosomal Database Project's Training Set 16 (11.5 release of the RDP database)

DELIVERABLES:

a. Raw data QC report, which includes stats about read trimming and filtering, and chimera filtering
b. Classified ASVs (or OTUs) table with abundance information
c. Phylogenetic Tree
d. Alpha and beta rarefaction plots
e. Alpha and beta diversity reports that includes: Distance matrices; PCoA plots
f. Taxonomy classification report and representative sequences file
g. Finally, analysis/methodology description for manuscript

REFERENCES:

1. https://qiime2.org
2. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. (2016a). DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13: 581–583.
3. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4:e2584. doi:10.7717/peerj.2584.
4. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ et al. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2: e00191–16.
5. Price MN, Dehal PS, Arkin AP. Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 2009; 26(7):1641–50.
6. http://greengenes.secondgenome.com
7. https://www.arb-silva.de
8. http://rdp.cme.msu.edu
9. http://www.homd.org/index.php
10. Xia, Y.; Sun, J. Hypothesis testing and statistical analysis of microbiome. Genes & Diseases. V.1, p.1-11, 2017.
11. Pollock J., Glendinning L., Wisedchanwet T., Watson M. (2018). The madness of microbiome: attempting to find consensus 'Best practice' for 16S microbiome studies. Appl. Environ.

Microbiol. 84.

Questions? Comments?

Please email us at EICC@emory.edu