Amplicon 16S microbiota profiling services

The Department of Microbiology and Immunology at The University of Melbourne has established Doherty Applied Microbial Genomics, located within the Doherty Institute for Infection and Immunity. Doherty Applied Microbial Genomics is closely aligned with the Microbiological Diagnostic Unit Public Health Laboratory, which provides a microbiological investigation, reference characterization, and advisory services for the public health community.

Doherty Applied Microbial Genomics has extended expertise in microbial ecology and microbiome sciences, with research academics actively involved in The University of Melbourne's Environmental Microbiology Research Initiative, and strong collaborations with world-leading microbiome experts.

Our amplicon-based 16S profiling of microbial communities conform to international standards by way of the Earth Microbiome Project, and offers high resolution in identifying exact sequence variants within an ecosystem. Our services take advantage of paired-end 16S sequencing on the Illumina platforms using universal prokaryotic 515F and 806R primers targeting the V4 region of the 16S small subunit rRNA gene. Through our strong collaborations with world-leading experts, we offer bioinformatic analyses that are at the vanguard of microbiome science.

Our standard reporting includes:

- **Quality control metrics**
  e.g., Number of sequences per sample
- **Alpha rarefaction curve analysis**
  i.e., An assessment of sequencing the microbial community to saturation
- **Biodiversity bar plots**
  i.e., Illustrating community composition as relative frequencies
- **Alpha-diversity analysis**
  e.g., Visually and statistically comparing, for example, Shannon's diversity index across different groupings
- **Beta-diversity analysis**
  e.g., An Emperor plot of, for example, Bray-Curtis distances representing a quantitative measure of community dissimilarity

We provide the primary sequence files (FASTQ), and post-QC files, including the Feature Table (BIOM), and Feature Sequences (FASTA). The Feature Table and Sequence files are provided in an effort to enable, and facilitate, targeted downstream bioinformatic analyses by our end users. Alternatively, please contact us for the potential of collaborative investigations of your dataset; for example, identifying key microbes that may be associated with a metadata variable (e.g., infection phase) using differential abundance analyses, and computing phylogenetically-aware beta-diversity analyses.

Overview of the workflow

Our standard workflow includes three main phases, and allows for analyses that conform to international standards by way of the Earth Microbiome Project.

(1) Sample processing and sequencing
Genomic DNA are extracted using the PowerSoil HTP kit, and amplified using universal prokaryotic 515F and 806R primers targeting the V4 region of the 16S small subunit rRNA gene[1]. Amplicon 16S rRNA gene sequences are generated using paired-end 150 bp sequencing on the Illumina MiSeq.

(2) Data processing including quality control
Raw FASTQ files are processed using Qiita for (i) quality control, (ii) demultiplexing sequences, (iii) trimming, and (iv) resolving exact sequence variants. Exact sequence variants are determined using Deblur[2]. Data processed through Qiita is, by default, sandboxed. This means your data is *secure and private*. The key advantage of Qiita is the possibility of computing meta-analyses with the >10,000 *public* studies available through Qiita. Please contact us if you have any concerns, and alternative arrangements can potentially be made to process the sequence data.

(3) Downstream statistical analyses
Majority of the downstream analyses are computed using the latest statistical algorithms available through QIIME2, and analyses visualized using QIIME2 View (Q2-View). Optional further analyses through collaborative efforts will be computed using various sources of bioinformatic tools (e.g., a suite of python scripts, and R packages). Please contact us for possible collaborative arrangements.

Quality control metrics
Sequences are filtered and those without a corresponding barcode, and without the correct primer sequence will be excluded from downstream analyses.

Alpha rarefaction curve analysis
Alpha rarefaction curves explore alpha diversity indices (e.g., Shannon's diversity index) as a function of sampling depth. This provides an assessment of whether if the microbial community has been sequenced to saturation.

Biodiversity bar plots
Sequences are resolved to high quality sequence variant data (i.e., sub-operational taxonomic units) using Deblur, and classified using a pre-trained Naïve Bayes classifier. The classifier was trained on the Greengenes 13.8 99% OTU database. Relative frequencies of each sequence variant are calculated and plotted as a bar plot.

Alpha-diversity analysis
Shannon's diversity index is a quantitative measure of community richness[3].

Beta-diversity analysis
Bray-Curtis distance scores are a quantitative measure of community dissimilarity[4].

References

1. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the national Academy of Sciences* 108, 4516–4522 (2011).
2. Amir, A. *et al.* Deblur rapidly resolves single-nucleotide community sequence patterns. *mSystems* 2, e00191–16 (2017).
3. Shannon, C. E. A Mathematical Theory of Communication. *The Bell System Technical Journal* 27, 379–423– 623 – 656 (1948).
4. Bray, R. & Curtis, J. T. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs* 27, 326–349 (1957).

Contact

For further enquiry please send an email to the address below:

damg-general@unimelb.edu.au

We acknowledge the Traditional Owners of the land on which we work, and pay our respects to the Elders, past and present.