

## 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.

Please note that these files will be made available for a period of **28-days** for users to retrieve before they are permanently deleted.

## 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<sup>1</sup>. 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<sup>2</sup>. 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<sup>3</sup>.

#### Beta-diversity analysis

Bray-Curtis distance scores are a quantitative measure of community dissimilarity<sup>4</sup>.

## Instructions on how to view data (qzv files) using Q2-Viewer

QIIME2 View is an *entirely* client-side interface for viewing data, and functions without a server. Importantly, this means that files provided (*drag-drop*) to Q2-View are not sent anywhere beyond your browser.

For more information, please visit: <https://view.qiime2.org/about/>

Please find instructions below to begin viewing your results interactively.

For best viewing of microbiota data, please:

Download all qzv files to a local computer (e.g., laptop, and/or desktop computer).

Visit the Q2-View webpage using the following link:

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

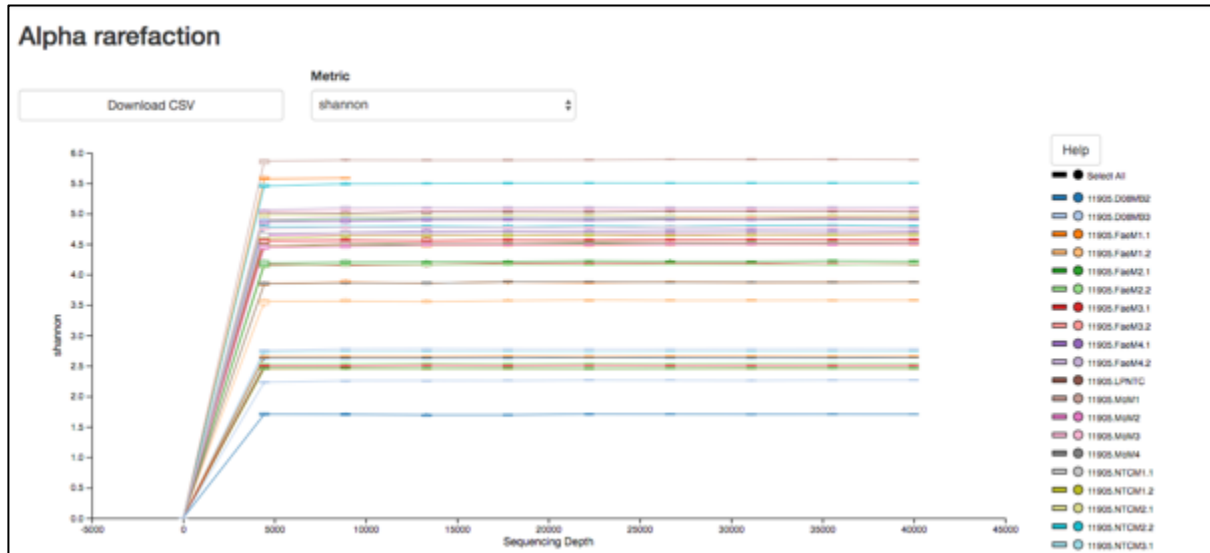
Please note that Google Chrome is the preferred web browser, and a new drag-drop home page is required for each new file to be viewed.



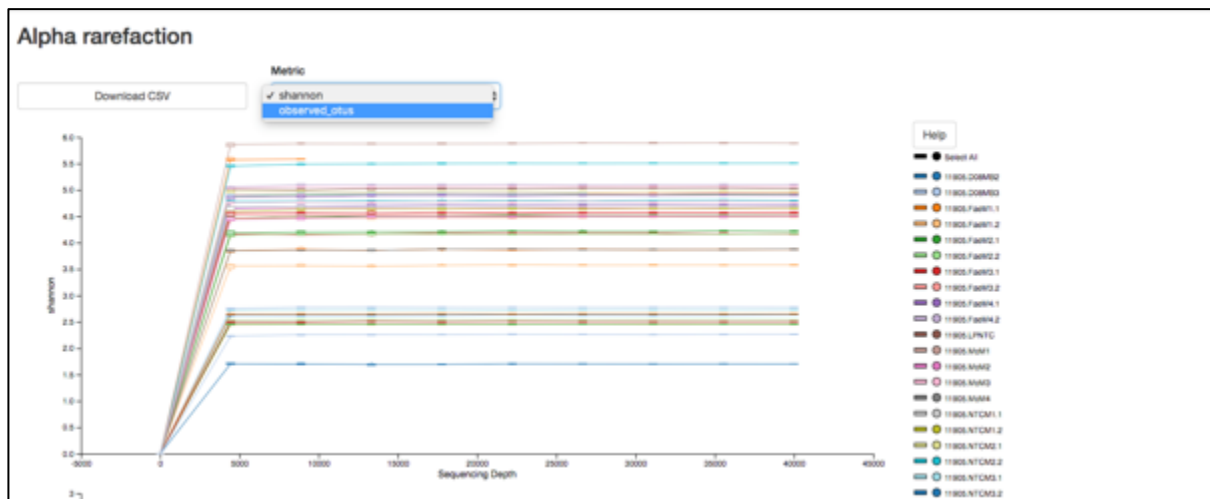
## Alpha-rarefaction curves:

Please drag-drop the *alpha-rarefaction curve* qzv file to the Q2-View home page.

The page should automatically refresh and load an alpha rarefaction curve.  
The default metric displayed is Shannon's diversity index.



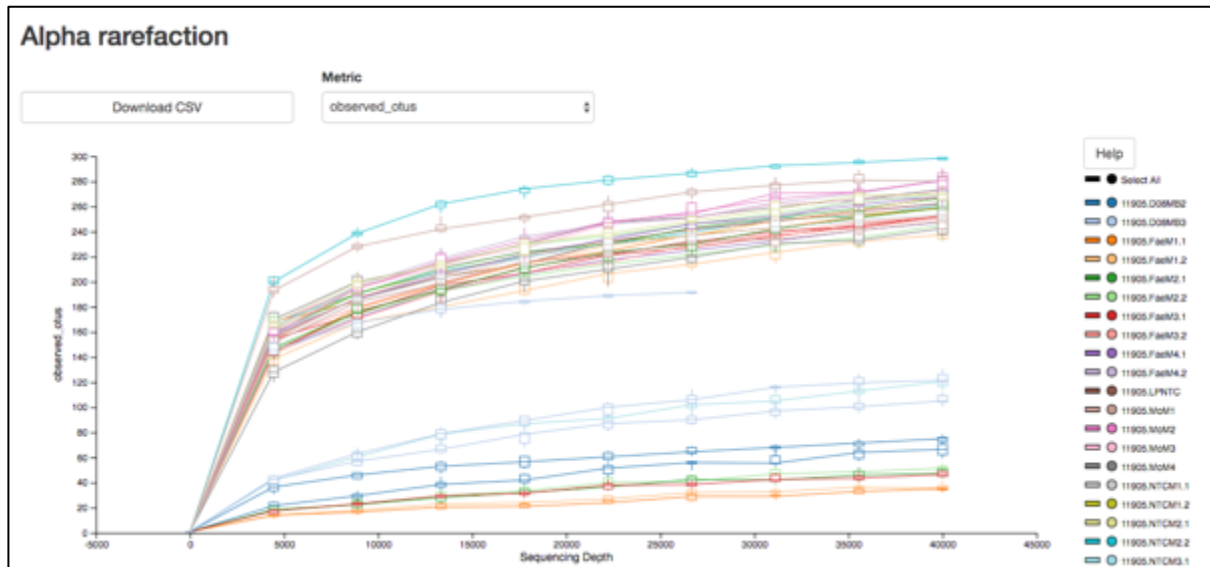
Please select "*observed\_otus*" from the **Metric** drop-down menu.



Each sample can be “toggled” on and off to facilitate greater breakdown and visualization of each sample on the graph.

The rectangular, and/or the circular symbols on the right-hand side panel can be “clicked” to toggle each sample on/off.

*e.g., click both the rectangular and circular symbols for “Select All”, and then select the samples of interest to draw the data points back on the graph. The rectangular symbol draws connecting lines, while the circular symbol draws the data points.*

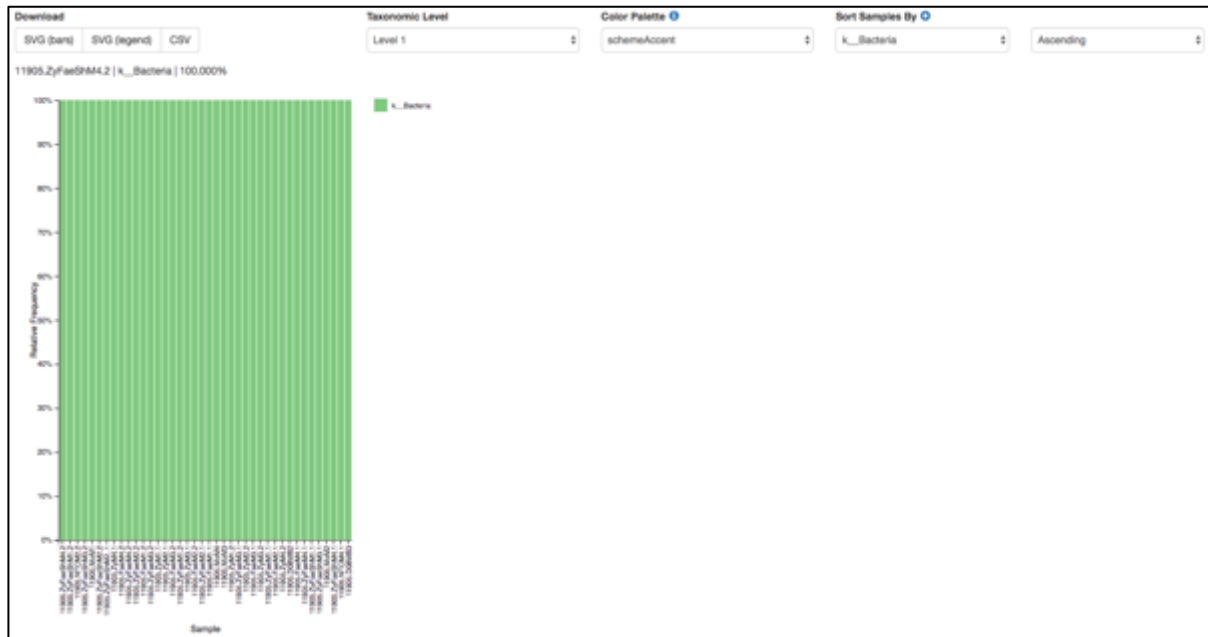


## Biodiversity plots

Please drag-drop the *biodiversity plot* qzv file to the Q2-View home page.

The page should automatically refresh and load a biodiversity plot.

The default metric displayed is the relative frequency of features (i.e., sub-OTUs) per sample at the Domain taxonomic level (i.e., Level 1).



The level of taxonomy visualized on the bar plot can be edited by selecting one of the following in the **Taxonomic Level** drop-down menu:

Level 1	Domain	e.g., Bacteria
Level 2	Phyla	e.g., Firmicutes
Level 3	Class	e.g., Clostridia
Level 4	Order	e.g., Clostridiales
Level 5	Family	e.g., Lachnospiraceae
Level 6	Genus	e.g., <i>Faecalibacterium</i>
Level 7	Species	e.g., <i>prausnitzii</i>

Despite the availability of selecting Level 7, and some features being assigned to the species level, we would advise against drawing conclusions higher than genus level of taxonomy without, for example, (i) culture-dependent, (ii) whole genome sequencing of isolates, and/or (ii) shotgun metagenomic data.

Samples can be sorted by multiple **Sample Metadata** values in an effort to reveal first-order trends in biodiversity.

*e.g., First select the level of taxonomy, then sort samples by Sample ID, and then sort by Cohort descriptions (e.g., pre- and post-antibiotic treatment).*

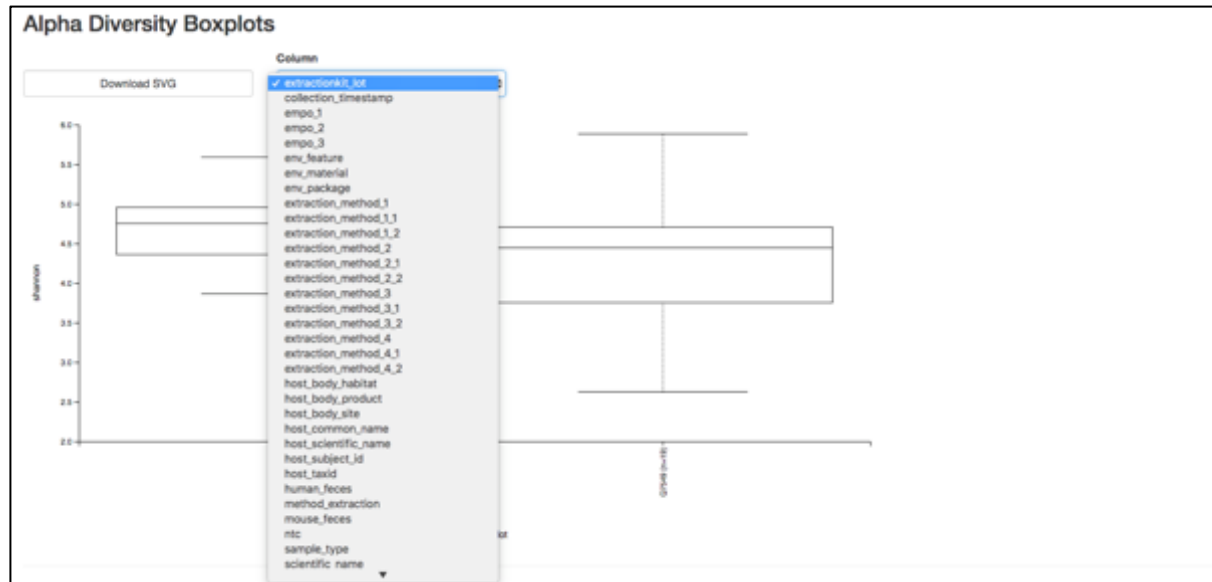
Select the **Sort Samples By** drop-down menu, and select the first value to sort upon. Then select whether this value should be sorted in ascending or descending order. Click the **+** symbol in the **Sort Samples By +** and select the next value to sort upon.



## Alpha-diversity boxplots

Please drag-drop the *alpha-diversity* qzv file to the Q2-View home page.

The page should automatically refresh and load alpha diversity boxplots.  
Select the sample metadata value of interest from the **Column** drop-down menu.



Scroll down the page to inspect statistical analyses of the groups  
*e.g., Kruskal-Wallis analysis for all groups, or pairwise between the different groups*

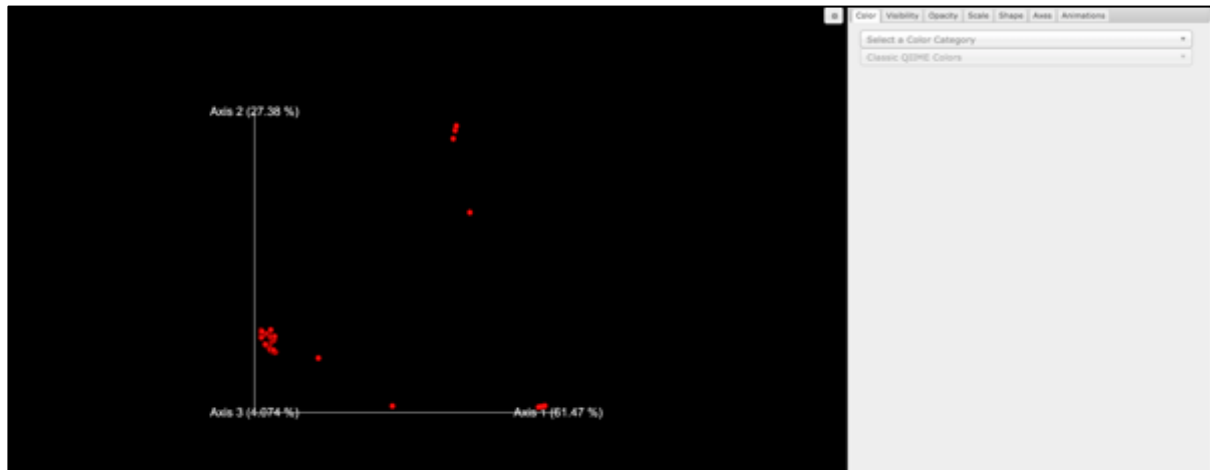


## Beta-diversity analysis

Please drag-drop the *Emperor plot* qzv file to the Q2-View home page.

The page should automatically refresh and load an Emperor plot.

The default settings will colour each data point in the sample space red.



Depending on the specifics of each study design, and the level of detail included in the metadata sheets, you may be able to colour code the data points accordingly (e.g., differentially colour pre- and post-infection samples). Additionally, if the study involves temporal sampling, the **Animation** tab on the right-hand panel provides the capability to highlight temporal trajectories of each data point on the sample space.

## 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 see below for contact details of key microbiome researchers from Doherty Applied Microbial Genomics:

**Prof. Timothy P. Stinear**

Scientific Director

E: [tstinear@unimelb.edu.au](mailto:tstinear@unimelb.edu.au)

**Prof. Benjamin P. Howden**

Medical Director

E: [bhowden@unimelb.edu.au](mailto:bhowden@unimelb.edu.au)

**A/Prof. Torsten Seemann**

Principal Bioinformatician

E: [t.seemann@unimelb.edu.au](mailto:t.seemann@unimelb.edu.au)

**Dr. Susan A. Ballard**

Principal Scientist

E: [susan.ballard@unimelb.edu.au](mailto:susan.ballard@unimelb.edu.au)

**Miss Nicole S. Isles**

Professional Research Assistant

E: [isles.n@unimelb.edu.au](mailto:isles.n@unimelb.edu.au)

**Dr. Jason Kwong**

NHMRC Early Career Fellow, and

Infectious Diseases & General Medicine physician (Austin Hospital)

E: [jason.kwong@unimelb.edu.au](mailto:jason.kwong@unimelb.edu.au)

**Dr. Andre Mu**

Research Fellow, and

Endeavour Fellow (Knight Lab, UCSD)

E: [andre.mu@unimelb.edu.au](mailto:andre.mu@unimelb.edu.au)

**Doherty Applied Microbial Genomics**

The Peter Doherty Institute for Infection and Immunity

792 Elizabeth Street | Melbourne | Victoria | Australia | 3000

[doherty.edu.au](http://doherty.edu.au)

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