Introductory Bioinformatic Course to Sequencing Data Processing

16S metagenomics

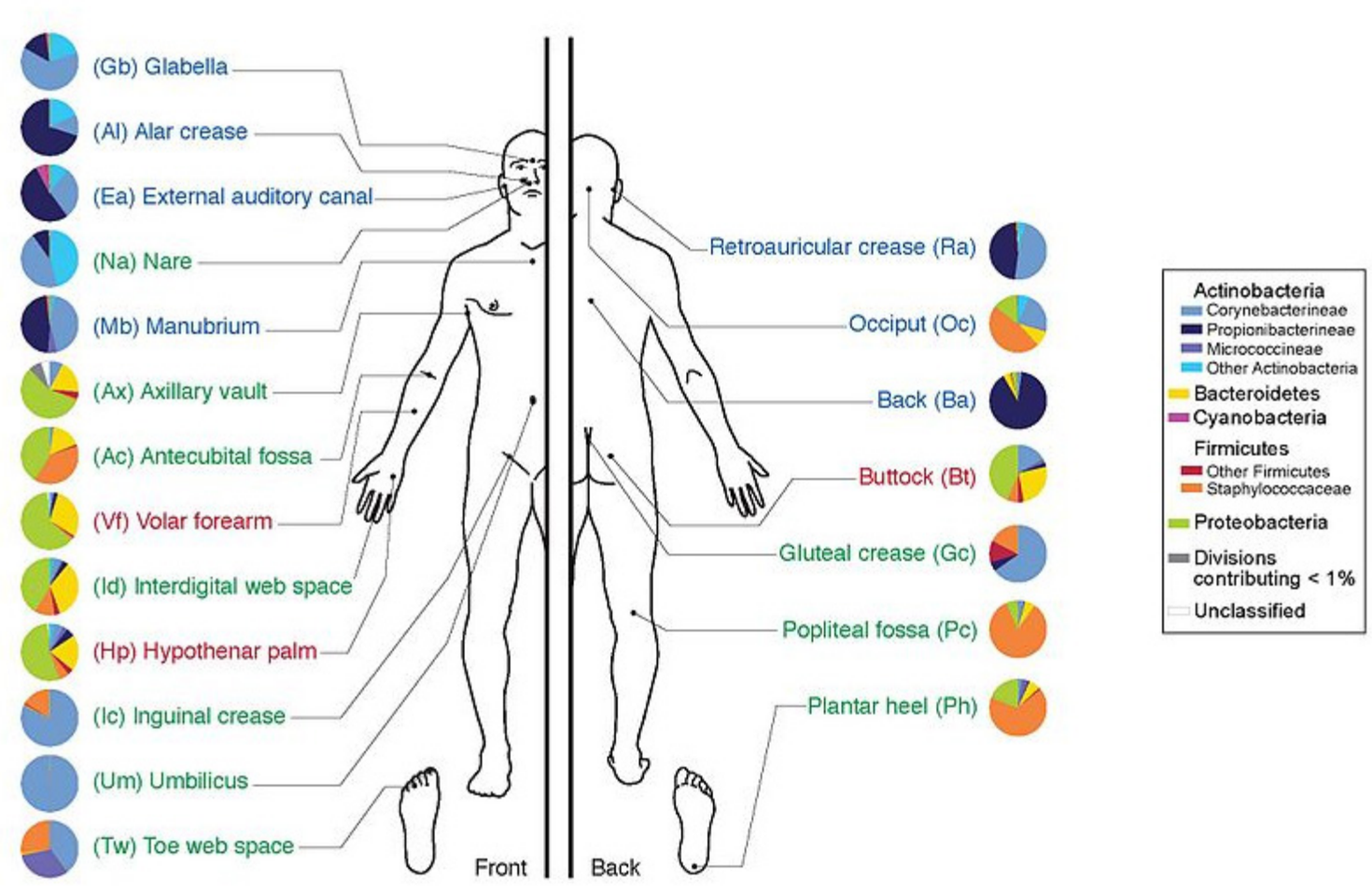
Overview

- Applications of metagenomics
- Amplicon (16S) Approach
- Shotgun Approach
  - Meta-genomics
  - Meta-transcriptomics
  - Meta-proteomics
  - ...

Application

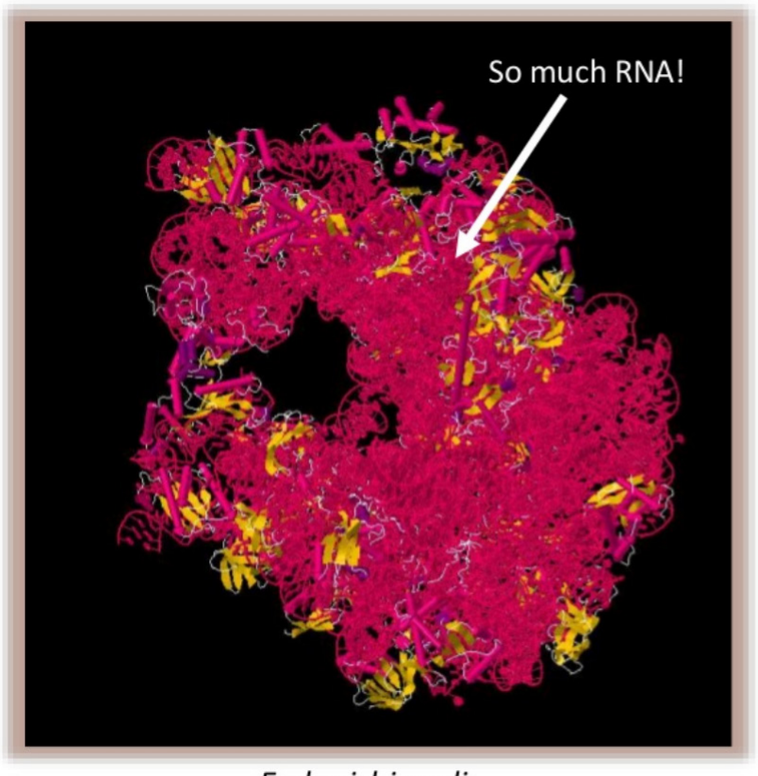
- Environmental Studies
- Medical / veterinary analysis

  Humans / Fish are full of microorganisms affecting health, drug efficacy, etc.

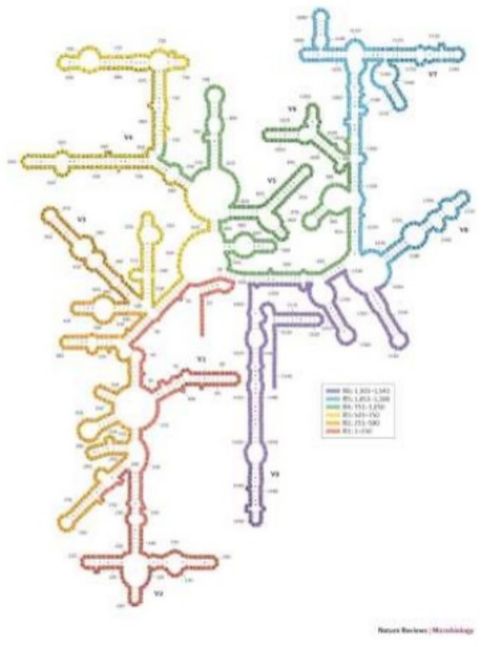


Amplicon

- Targetted approach (e.g. 16S/18S rRNA gene)
- Amplifies bacteria, not host or environmental fungi, plants, etc.



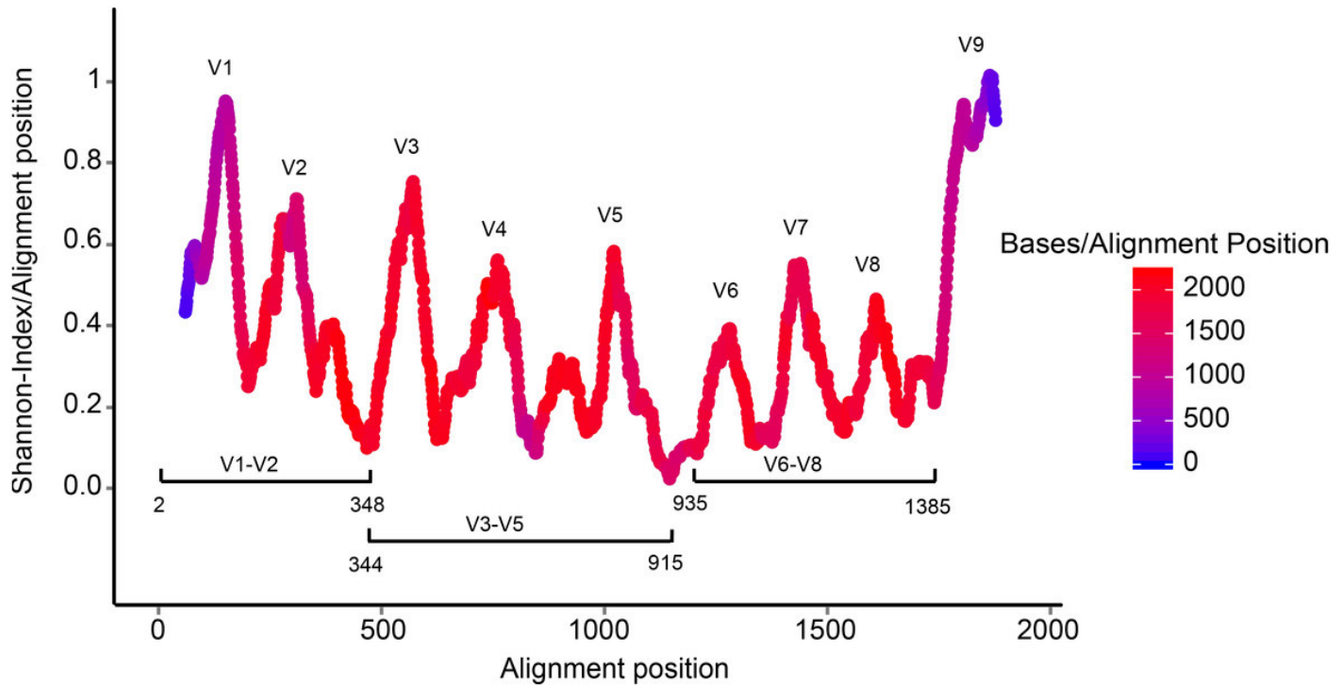
Escherichia coli
ribosome (PDB 4YBB)



Yarza et al. (2014)

Amplicon

- Highly conserved gene: easy to target across all bacteria
- With variable regions: distinguish between genus



Amplicon

Pros

- Well-established
- Inexpensive (€50/sample)

Cons

- V-region choice can bias results
- Is based on a very well conserved gene, making it hard to resolve species and strains
- Requires good reference 16S

Shotgun meta- genomic/transcriptomic

Aims to sequence the “whole” meta-genome/transcriptome

Pros:

- Not biased by amplicon primer set
- Not limited by conservation of the amplicon
- Can also provide functional information

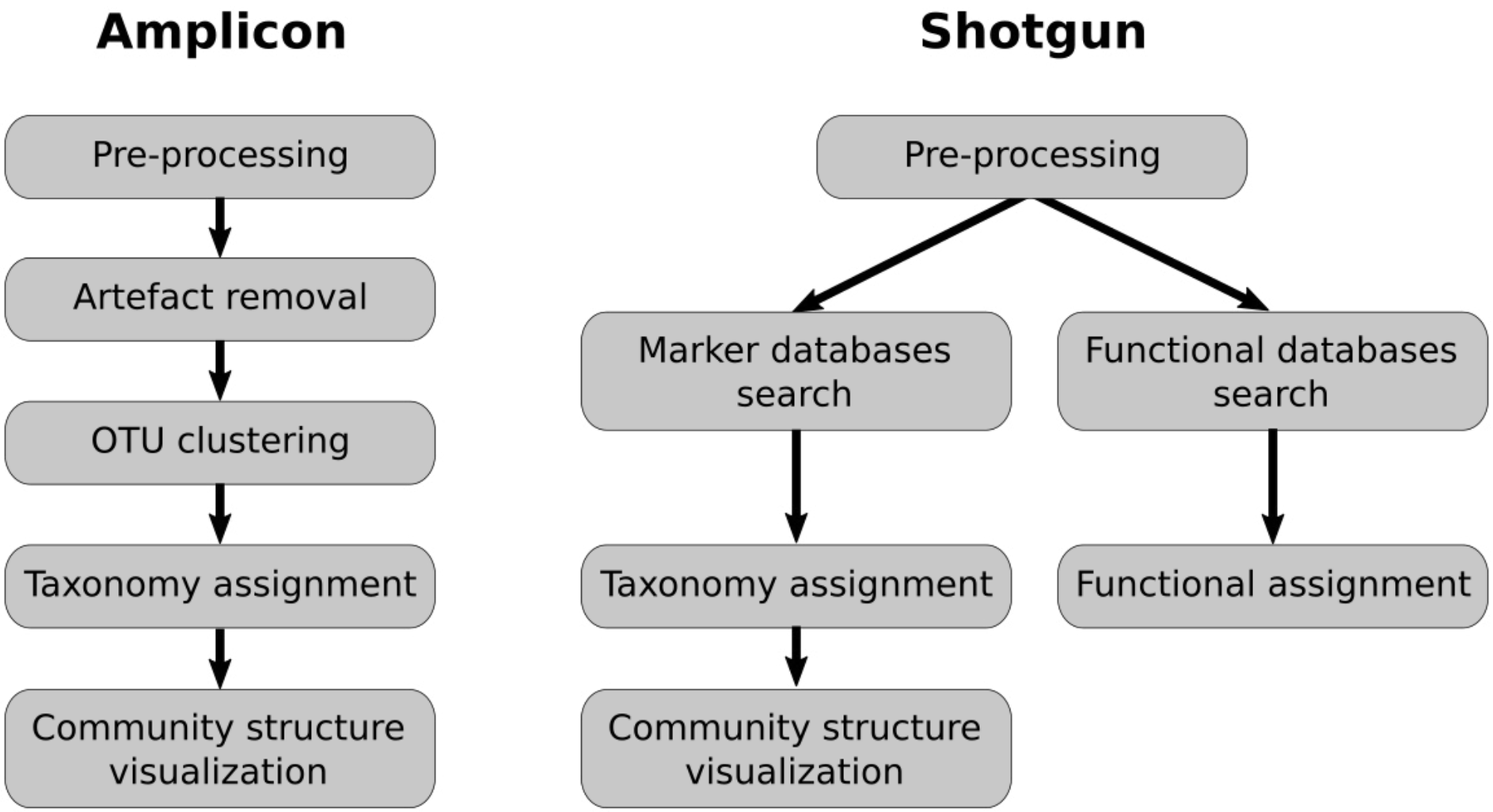
Cons:

- Environmental contamination, including host
- More expensive (€500/sample)
- Complex data analysis
- Requires high performance computing, high memory, high compute capacity
- Requires very good reference genomes

End-to-End

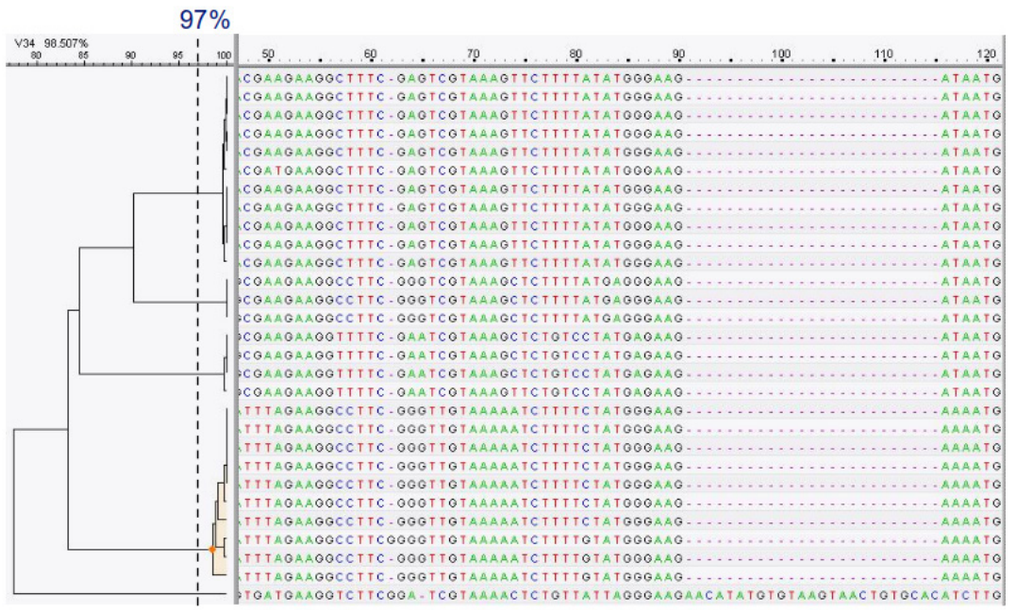
Every step in this process can have serious impact on the results

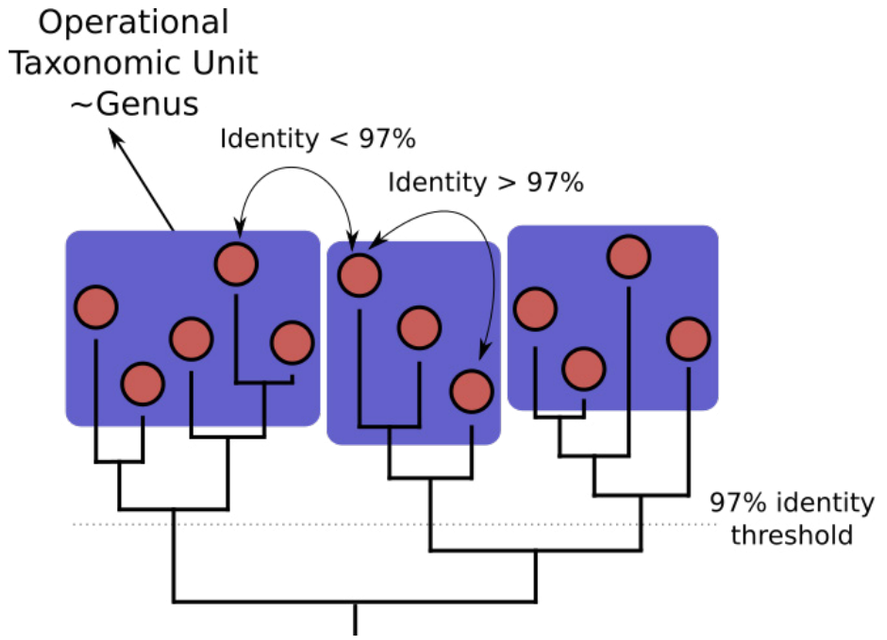
Analysis pipelines



OTU Clustering

Cluster on 97% sequence similarity for genus-level differentiation





Search marker database and taxonomy assignment

Homology with reference databases

- Amplicon

- Berkeley lab
- August 2013
- 202,421 entries

green genes
16S rRNA gene database and workbench compatible with ARB
greengenes.lbl.gov

- Max Planck Institute
- July 2015
- 172,418 entries

silva
high quality ribosomal RNA databases

- Shotgun: MetaPhlAn2 database

  ~1M unique clade-specific marker genes

  ~17,000 reference genomes (bacterial and archaeal, viral and eukaryotic)

Accuracy depends on quality and completeness of database used

Databases are inevitable incomplete

Results: OTU table

	A	B	C	D	E	F	G	H
1	OTU	Reads	Taxonomy					
2	Otu0001	342	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
3	Otu0002	265	Bacteria	Firmicutes	Bacilli	Bacillales	Listeriaceae	Listeria
4	Otu0003	222	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
5	Otu0004	191	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
6	Otu0005	184	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus
7	Otu0006	170	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	Clostridium
8	Otu0007	157	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	unclassified
9	Otu0008	152	Bacteria	Actinobacteria	Actinobacteria	Propionibacteriales	Propionibacteriaceae	Propionibacterium
10	Otu0009	144	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
11	Otu0010	143	Bacteria	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	Neisseria
12	Otu0011	139	Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia-Shigella
13	Otu0012	125	Bacteria	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus
14	Otu0013	112	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus
15	Otu0014	94	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter
16	Otu0015	77	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	Rhodobacter
17	Otu0016	67	Bacteria	Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	Actinomyces
18	Otu0017	44	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
19	Otu0018	41	Bacteria	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	Helicobacter
20	Otu0019	7	Bacteria	Deinococcus-Thermus	Deinococci	Deinococcales	Deinococcaceae	Deinococcus
21	Otu0020	1	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
22	Otu0021	1	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	unclassified
23	Otu0022	1	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
24	Otu0023	1	Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides
25	Otu0024	1	Bacteria	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
26	Otu0025	1	Bacteria	Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	unclassified
27								

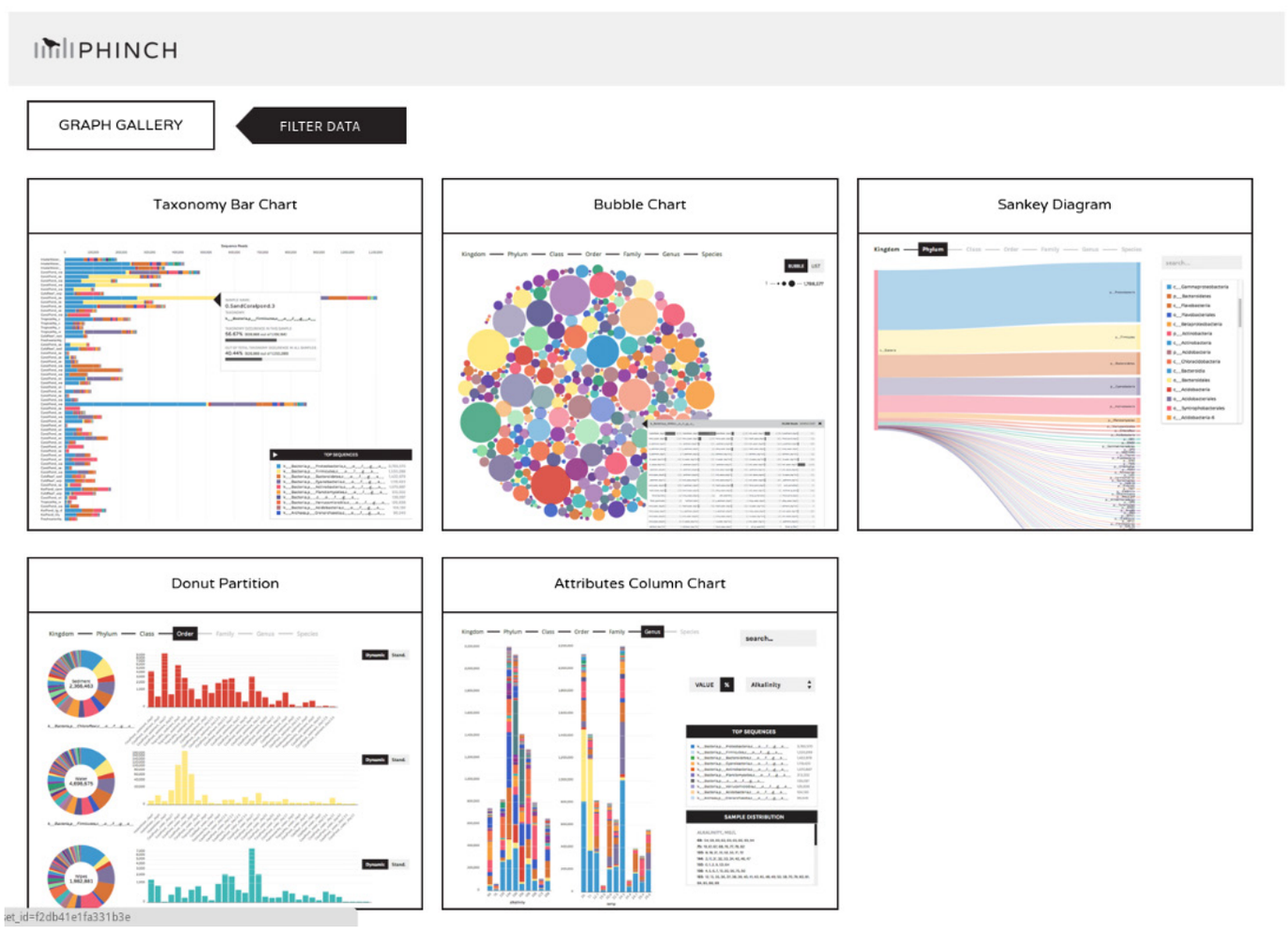
Results: Visualisations

- Krona

  interactive exploration of sample taxonomy

Results: Visualisations

- Phinch
  - BIOM file input
  - various visualizations
  - multi-sample data



Tutorial