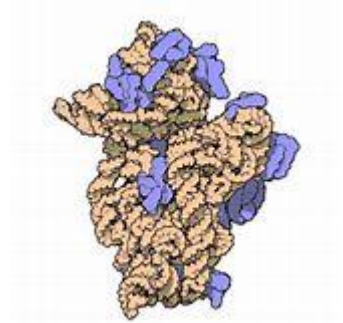


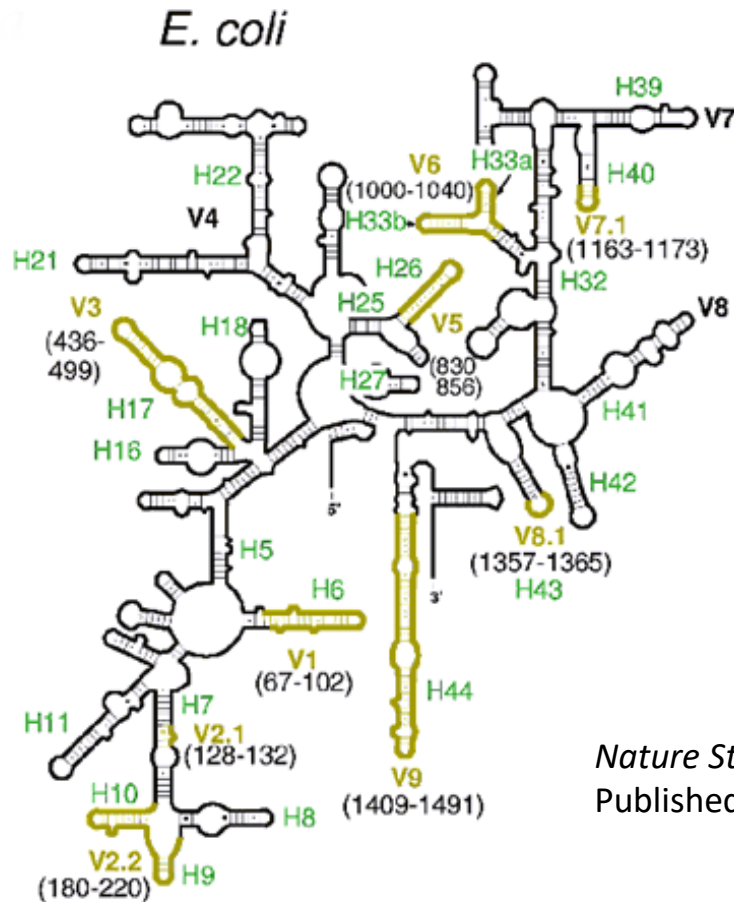
# Meta-genomics

Scope	Question	Answer
One gene – 16S rRNA	Who are there?	Genus, Species, OTUs
All genes – metagenomics	What can they do?	Functional potentials
All genes – metatranscriptomic	What are they doing?	Metabolic pathways Biological function

# What is 16S rRNA?



- **16S ribosomal RNA** - the component of the 30S small subunit of a prokaryotic ribosome that binds to the Shine-Dalgarno sequence during protein synthesis
- Approximately **1,500** nt in length



# Prokaryotic 16S rRNA

*Nature Structural Biology* **9**, 750 - 755 (2002)

Published online: 16 September 2002; | doi:10.1038/nsb841

0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 bp



**CONSERVED REGIONS:** unspecific applications

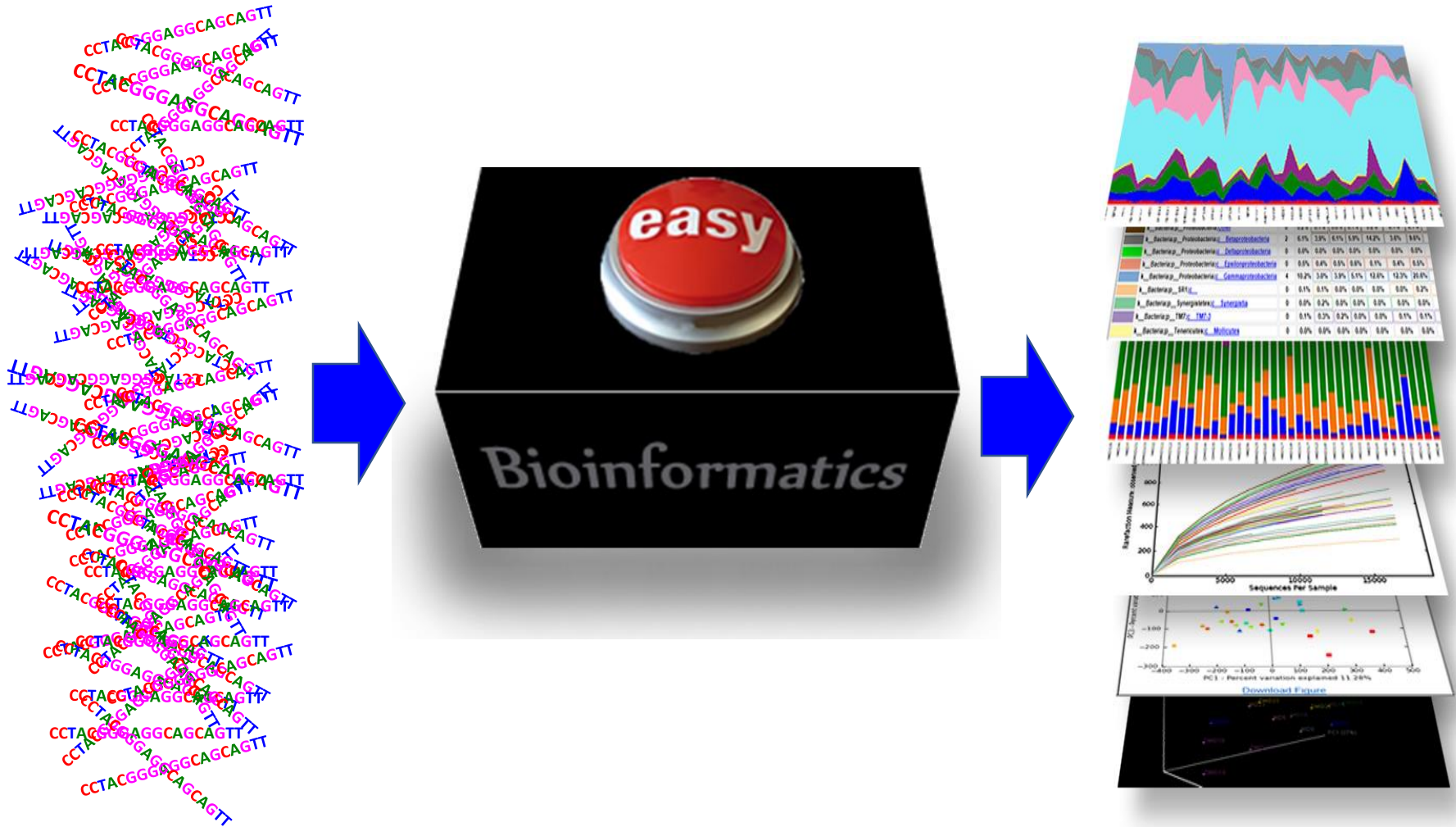
**VARIABLE REGIONS:** group or species-specific applications

**Alimetrics Ltd**  
Koskelontie 19 B  
02920 Espoo  
Finland

# Why 16S rRNA as the marker gene?

- **SIMILAR** enough – both Bacteria and Archaea have it
- **DIFFERENT** enough – to differential different species
- **SHORT** enough – easy (easier) to sequence
- **LONG** enough – to contain both conserve and variable domains

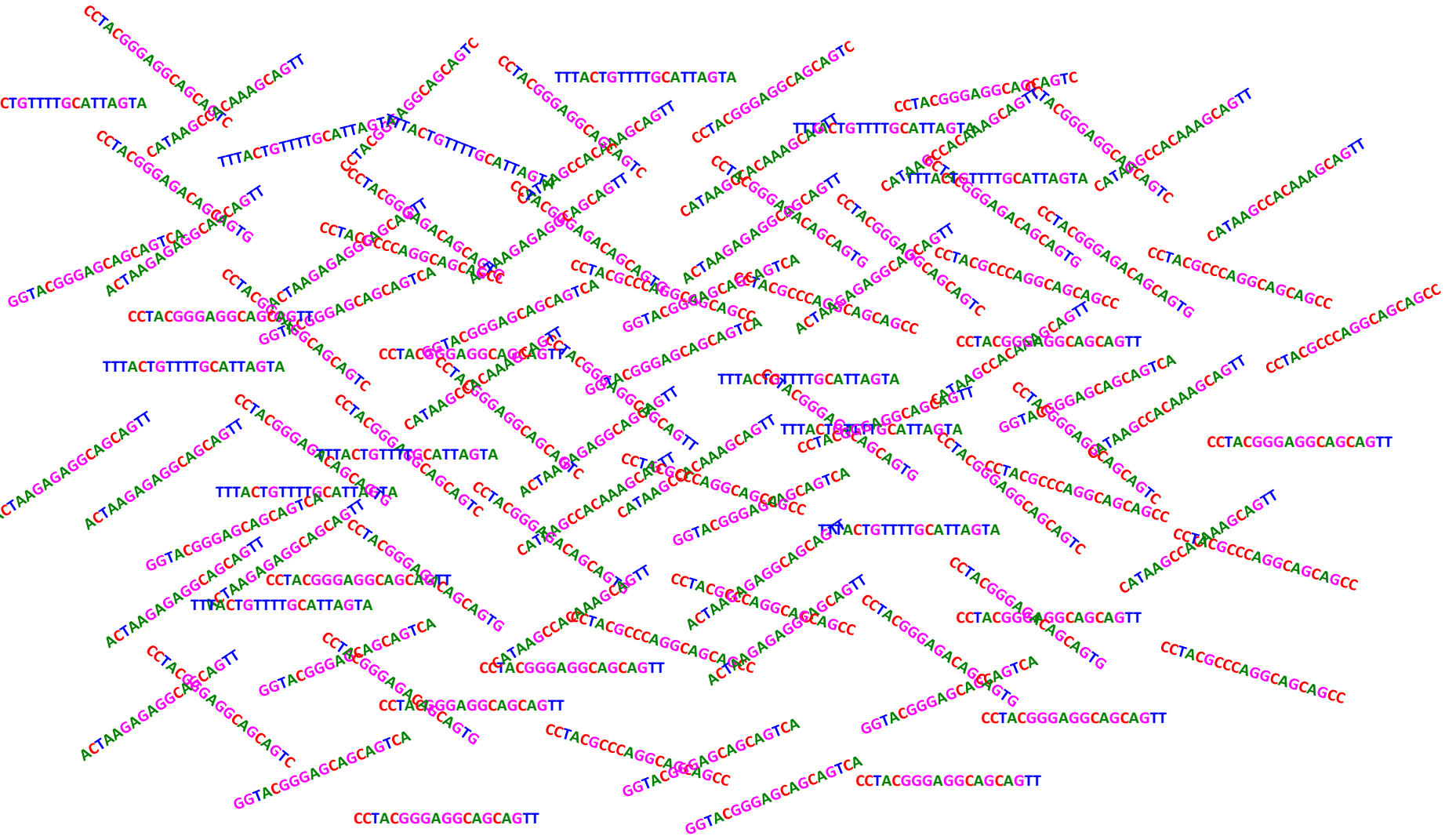
# The Bioinformatics Black Box



# Inside the blackbox

1. **Upstream analyses** – quality control, separate samples (de-barcoding, de-multiplexing)
2. **Middle tier analyses** – organize sequences into biological groups (OTUs, ASVs, taxa)
3. **Downstream analyses** – compare microbial compositions of the samples – diversity analysis

# Chaos organizing

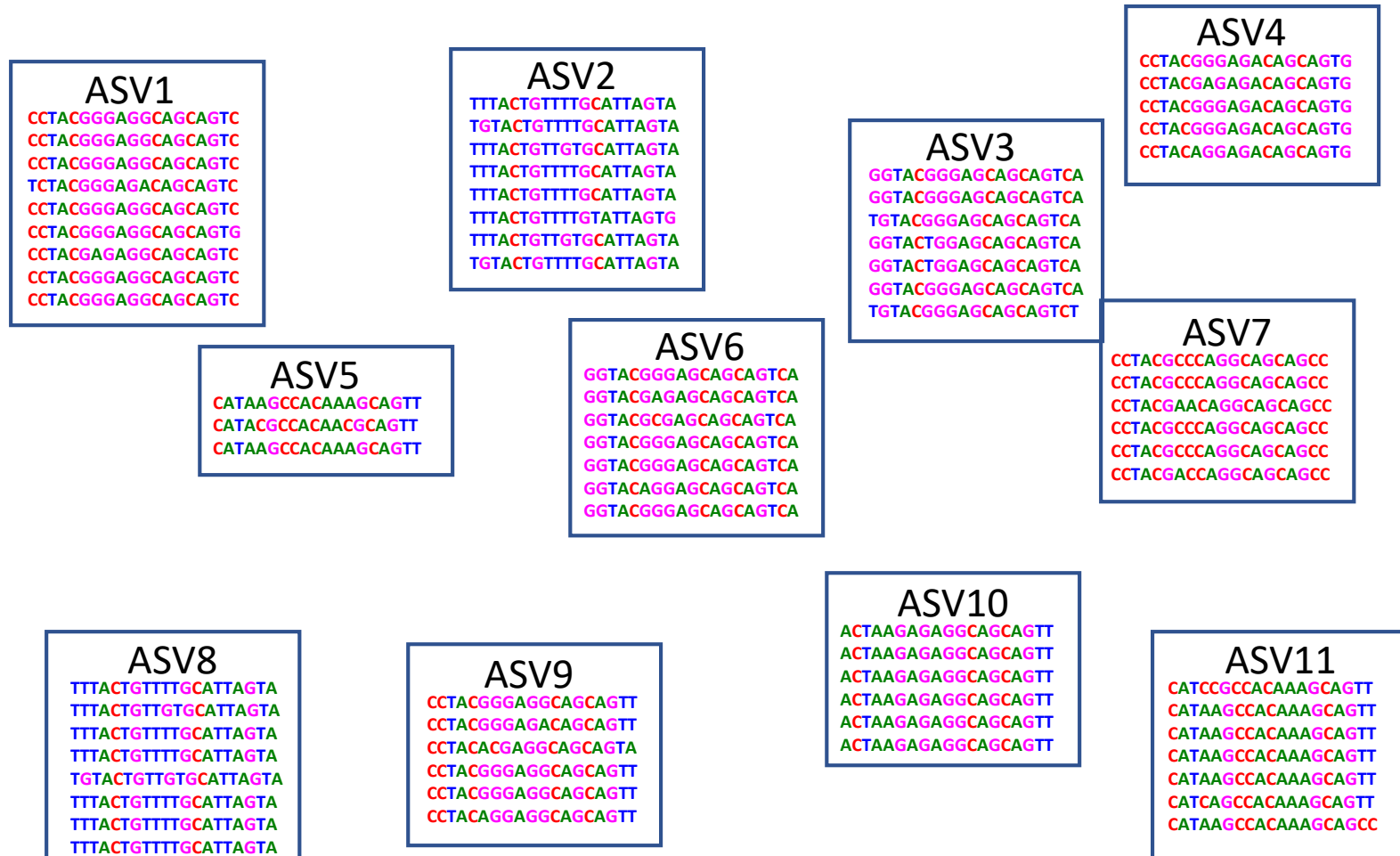


# OTU (operational taxonomic unit) picking – based on % sequence identity





# ASV (amplicon sequence variants) clustering – denoise sequences based on sequence quality error model



DADA2 Pipeline

# Identify representative sequences

OTU1

CCTACGGGAGGCAGCAGTC  
CCTACGGGAGGCAGCAGTC  
CCTACGGGAGGCAGCAGTC  
TCTACGGGAGACAGCAGTC  
CCTACGGGAGGCAGCAGTC  
CCTACGGGAGGCAGCAGTG  
CCTACGAGAGGCAGCAGTC  
CCTACGGGAGGCAGCAGTC  
CCTACGGGAGGCAGCAGTC

OTU2

TTTACTGTTTTGCATTAGTA  
TGTACTGTTTTGCATTAGTA  
TTTACTGTTTGCATTAGTA  
TTTACTGTTTTGCATTAGTA  
TTTACTGTTTTGCATTAGTA  
TTTACTGTTTTGTATTAGTG  
TTTACTGTTTGCATTAGTA  
TGTACTGTTTTGCATTAGTA

OTU4

CCTACGGGAGACAGCAGTG  
CCTACGAGAGACAGCAGTG  
CCTACGGGAGACAGCAGTG  
CCTACGGGAGACAGCAGTG  
CCTACGAGAGACAGCAGTG

OTU3

GGTACGGGAGCAGCAGTCA  
GGTACGGGAGCAGCAGTCA  
TGTACGGGAGCAGCAGTCA  
GGTACTGGAGCAGCAGTCA  
GGTACTGGAGCAGCAGTCA  
GGTACGGGAGCAGCAGTCA  
TGTACGGGAGCAGCAGTCT

OTU5

CATAAGCCACAAAGCAGTT  
CATAAGCCACAAAGCAGTT  
CATAAGCCACAAAGCAGTT

OTU6

GGTACGGGAGCAGCAGTCA  
GGTACGAGAGCAGCAGTCA  
GGTACGCGAGCAGCAGTCA  
GGTACGGGAGCAGCAGTCA  
GGTACGGGAGCAGCAGTCA  
GGTACAGGAGCAGCAGTCA  
GGTACGGGAGCAGCAGTCA

OTU7

CCTACGCCCCAGGCAGAGCC  
CCTACGCCCCAGGCAGAGCC  
CCTACGAAAGGCAGAGCC  
CCTACGCCCCAGGCAGAGCC  
CCTACGCCCCAGGCAGAGCC  
CCTACGACCAAGGCAGAGCC

OTU8

TTTACTGTTTTGCATTAGTA  
TTTACTGTTTGCATTAGTA  
TTTACTGTTTTGCATTAGTA  
TTTACTGTTTTGCATTAGTA  
TGTACTGTTTGCATTAGTA  
TTTACTGTTTTGCATTAGTA  
TTTACTGTTTTGCATTAGTA  
TTTACTGTTTTGCATTAGTA

OTU9

CCTACGGGAGGCAGCAGTT  
CCTACGGGAGACAGCAGTT  
CCTACGAGAGGCAGCAGTA  
CCTACGGGAGGCAGCAGTT  
CCTACGGGAGGCAGCAGTT  
CCTACGAGAGGCAGCAGTT

OTU10

ACTAAGAGAGGCAGCAGTT  
ACTAAGAGAGGCAGCAGTT  
ACTAAGAGAGGCAGCAGTT  
ACTAAGAGAGGCAGCAGTT  
ACTAAGAGAGGCAGCAGTT  
ACTAAGAGAGGCAGCAGTT

OTU11

CATCCGCCACAAAGCAGTT  
CATAAGCCACAAAGCAGTT  
CATAAGCCACAAAGCAGTT  
CATAAGCCACAAAGCAGTT  
CATCAGCCACAAAGCAGTT  
CATAAGCCACAAAGCAGCC

# Taxonomy assignment

*Fusobacterium nucleatum*

```
CCTACGGGAGCAGCAGTG
CCTACGAGAGCAGCAGTG
CCTACGGGAGCAGCAGTG
CCTACGGGAGCAGCAGTG
CCTACGAGAGCAGCAGTG
```

*Streptococcus mutans*

```
GGTACGGGAGCAGCAGTCA
GGTACGGGAGCAGCAGTCA
TGTACGGGAGCAGCAGTCA
GGTACTGGAGCAGCAGTCA
GGTACTGGAGCAGCAGTCA
GGTACGGGAGCAGCAGTCA
TGTACGGGAGCAGCAGTCT
```

*Veillonella parvula*

```
CCTACGCCCCAGCAGGCC
CCTACGCCCCAGCAGGCC
CCTACGAAAGCAGCAGGCC
CCTACGCCCCAGCAGGCC
CCTACGCCCCAGCAGGCC
CCTACGACCAAGCAGGCC
```

*Streptococcus mitis*

```
TTTACTGTTTGCATTAGTA
TGTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
TTTACTGTTTGTATTAGTG
TTTACTGTTTGCATTAGTA
TGTACTGTTTGCATTAGTA
```

*Streptococcus parasanguinis*

```
GGTACGGGAGCAGCAGTCA
GGTACGAGAGCAGCAGTCA
GGTACGCGAGCAGCAGTCA
GGTACGGGAGCAGCAGTCA
GGTACGGGAGCAGCAGTCA
GGTACAGGAGCAGCAGTCA
GGTACGGGAGCAGCAGTCA
```

*Selenomonas artemidis*

```
CATAAGCCACAAAGCAGTT
CATAAGCCACAAAGCAGTT
CATAAGCCACAAAGCAGTT
```

*Aggregatibacter actinomycetemcomitans*

```
ACTAAGAGAGGCAGCAGTT
ACTAAGAGAGGCAGCAGTT
ACTAAGAGAGGCAGCAGTT
ACTAAGAGAGGCAGCAGTT
ACTAAGAGAGGCAGCAGTT
ACTAAGAGAGGCAGCAGTT
```

*Lactobacillus plantarum*

```
CATCCGCCACAAAGCAGTT
CATAAGCCACAAAGCAGTT
CATAAGCCACAAAGCAGTT
CATAAGCCACAAAGCAGTT
CATAAGCCACAAAGCAGTT
CATCAGCCACAAAGCAGTT
CATAAGCCACAAAGCAGCC
```

*Lactobacillus reuteri*

```
CCTACGGGAGGCAGCAGTC
CCTACGGGAGGCAGCAGTC
CCTACGGGAGGCAGCAGTC
TCTACGGGAGCAGCAGTC
CCTACGGGAGGCAGCAGTC
CCTACGGGAGGCAGCAGTG
CCTACGAGAGGCAGCAGTC
CCTACGGGAGGCAGCAGTC
CCTACGGGAGGCAGCAGTC
```

*Streptococcus parasanguinis*

```
TTTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
TGTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
TTTACTGTTTGCATTAGTA
```

*Porphyromonas gingivalis*

```
CCTACGGGAGGCAGCAGTT
CCTACGGGAGCAGCAGTT
CCTACGAGAGGCAGCAGTA
CCTACGGGAGGCAGCAGTT
CCTACGGGAGGCAGCAGTT
CCTACGAGAGGCAGCAGTT
```

# Downstream analyses

- **Alpha diversity** – analyze the complexity of the microbial community in **each sample**
- **Beta diversity** – compare the complexity of different **groups of samples** (e.g., health vs disease)
- **Differential analysis** – identify **organisms** with different relative abundance in different groups
- **Functional inferring** – associate functions with species identified

# Workshop Objectives

- I know R
- I have used R
- I have used R to analyze NGS data
- I like R

# Methodology

- Mostly copy and paste
- Provide a comprehensive online tutorial
- There are more tutorials out there (Web, Youtube)
- Google is your good friend