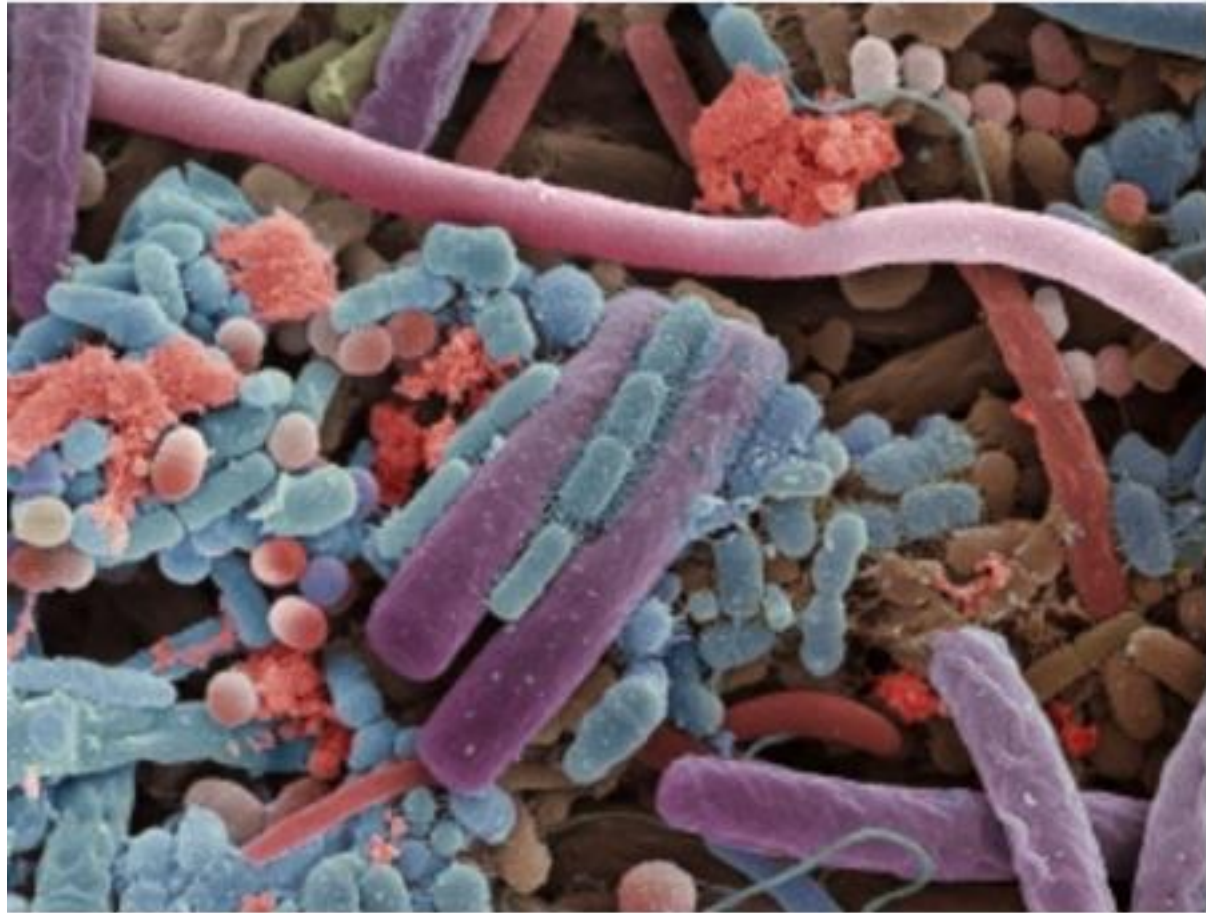


# **Metagenomics**

02-710 Computational Genomics

# Metagenomics

- Investigation of the microbes that inhabit oceans, soils, and the human body, etc. with sequencing technologies
- Cooperative interactions between microbes and their hosts
  - microbial participation in host functions such as defense, metabolism and reproduction

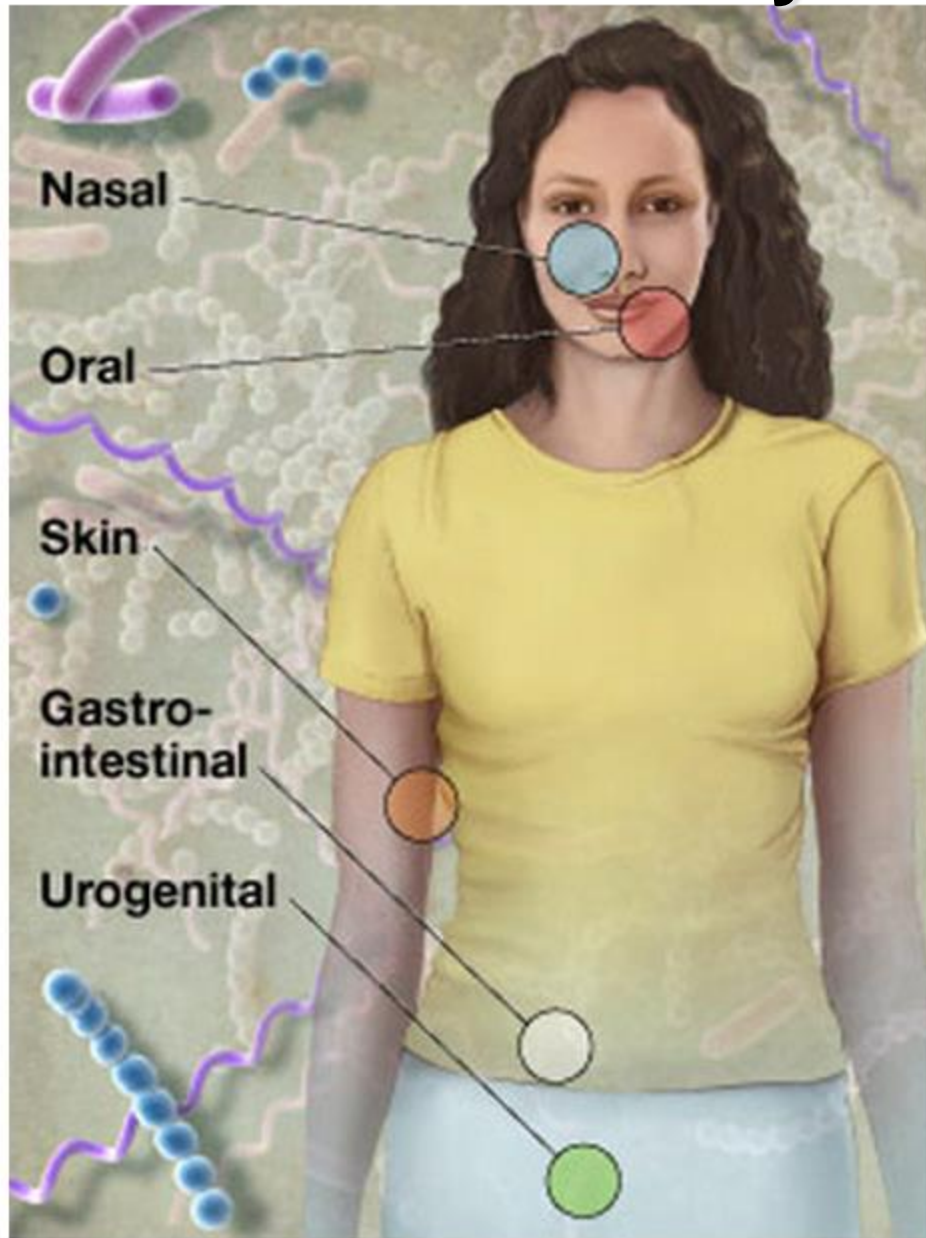


Tongue bacteria in human

# Metagenomics

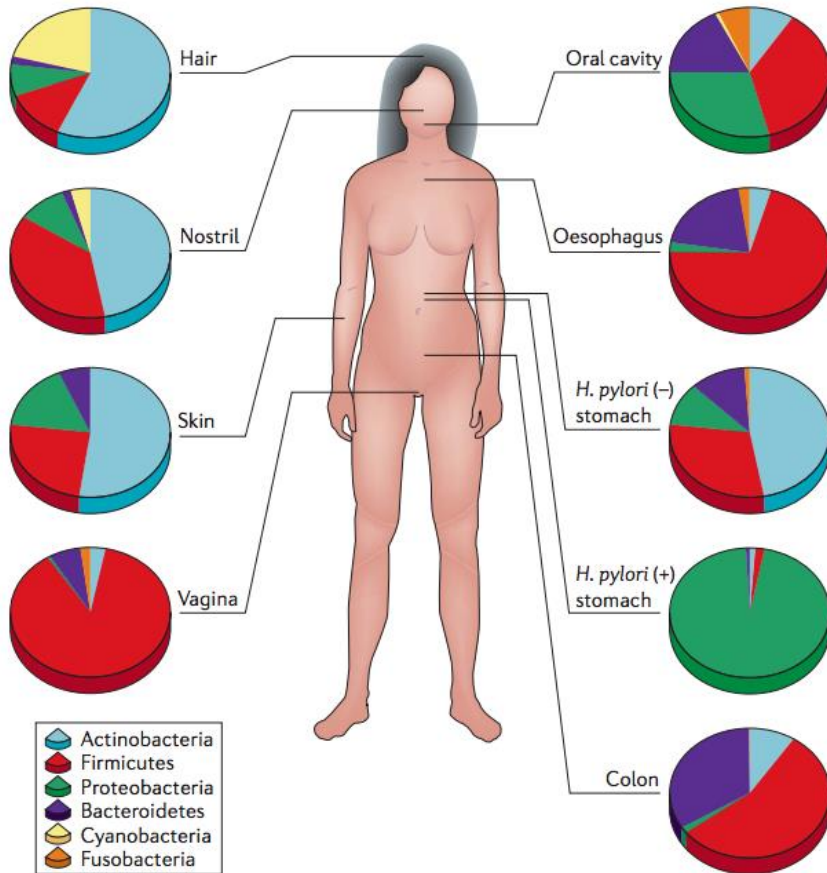
- Human microbiota: 100 trillion microbial cells, 10 times as many as human cells
- Human microbiome
  - Consists of archaea, bacteria, and viruses
  - What are the composition and gene content of human microbiome?
  - What are the differences of microbiome composition across individuals?
  - What are the differences of microbiome composition across body parts?

# Sampling of a new environment: the human body



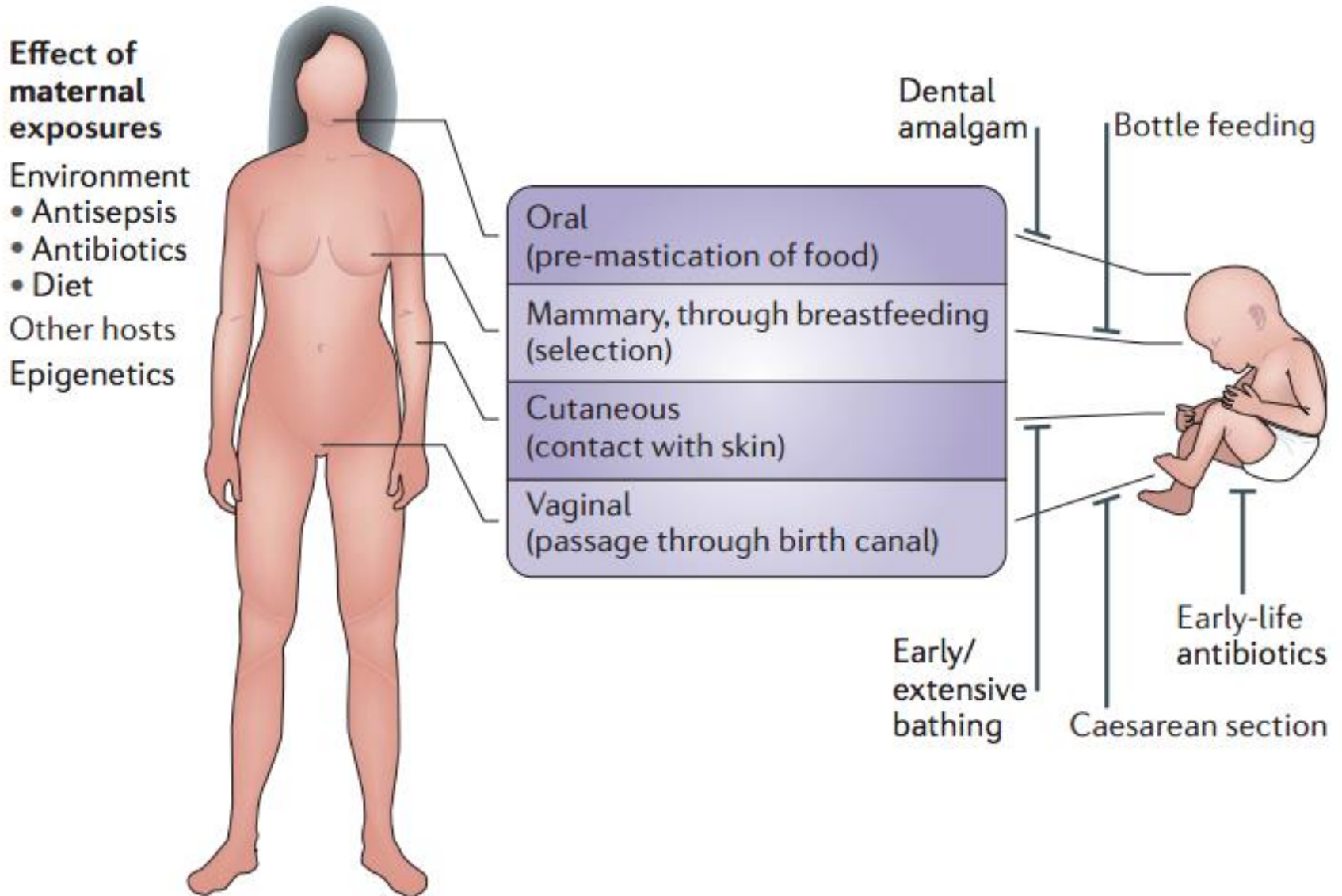
Human microbiome  
project consortium

# Microbiomes in Human Body

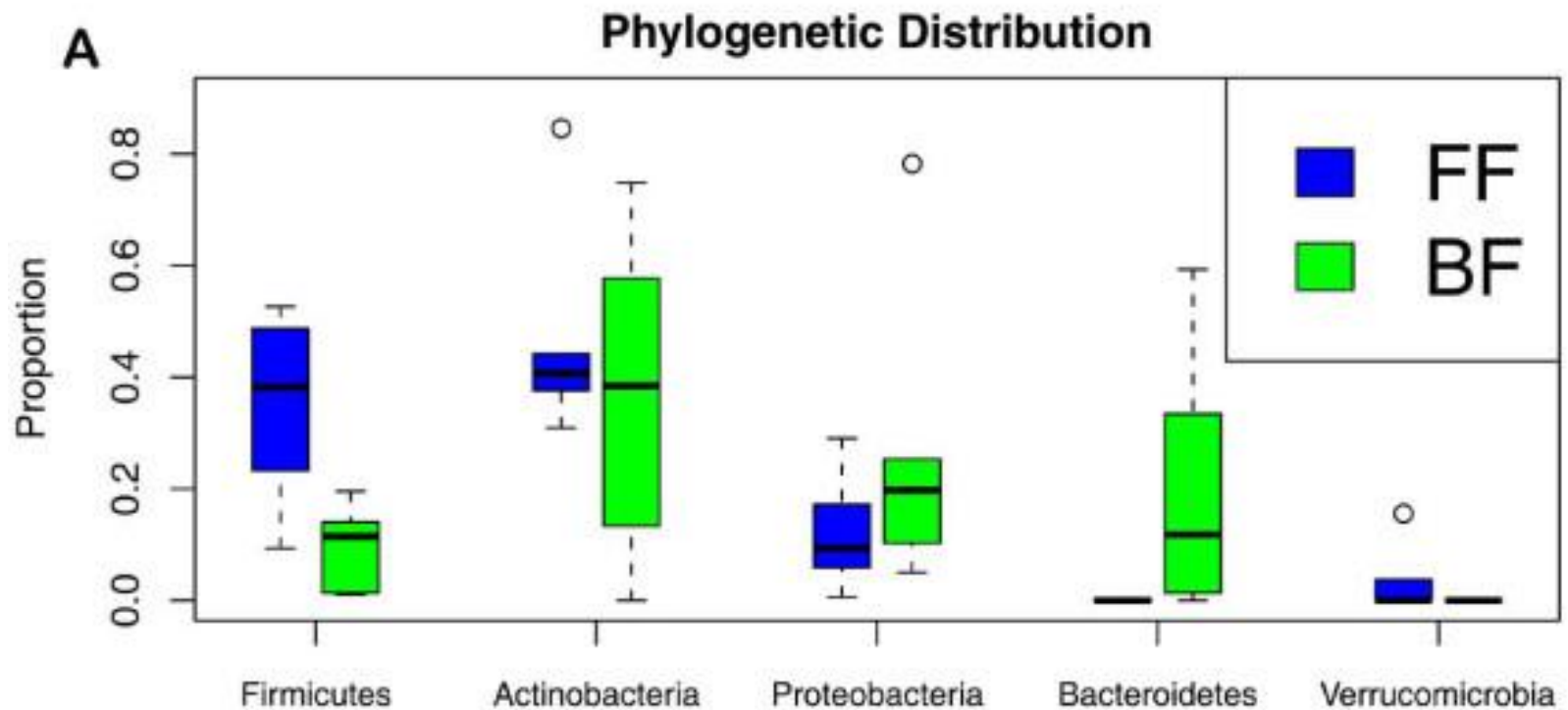


- The microbiome composition varies by anatomical sites.
- Substantial interpersonal variation
- Interpersonal variation is more substantial than temporal variability

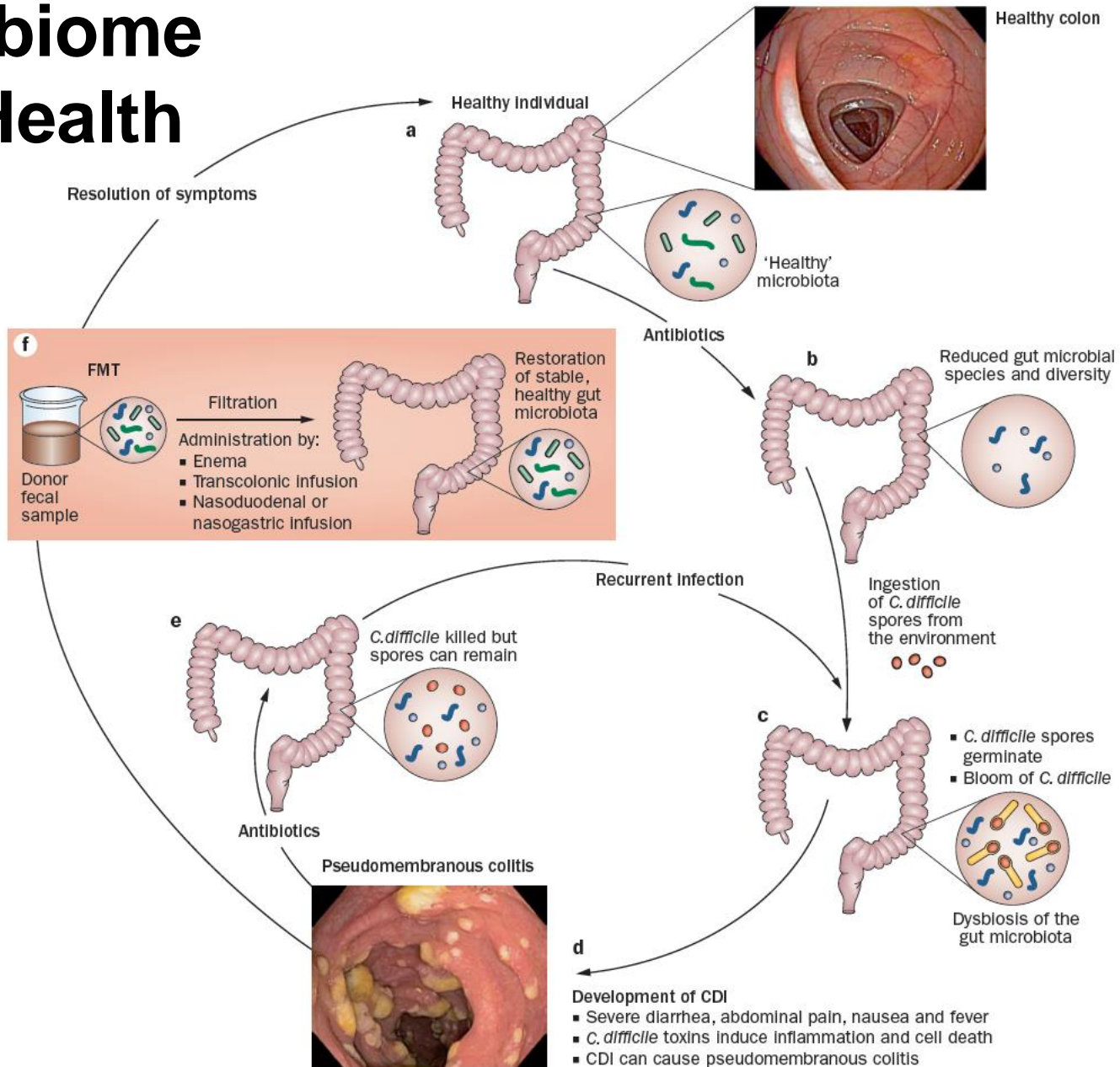
# Inheritance of the Microbiome from Mother to Baby



# Infant Diet Changes the Microbiome

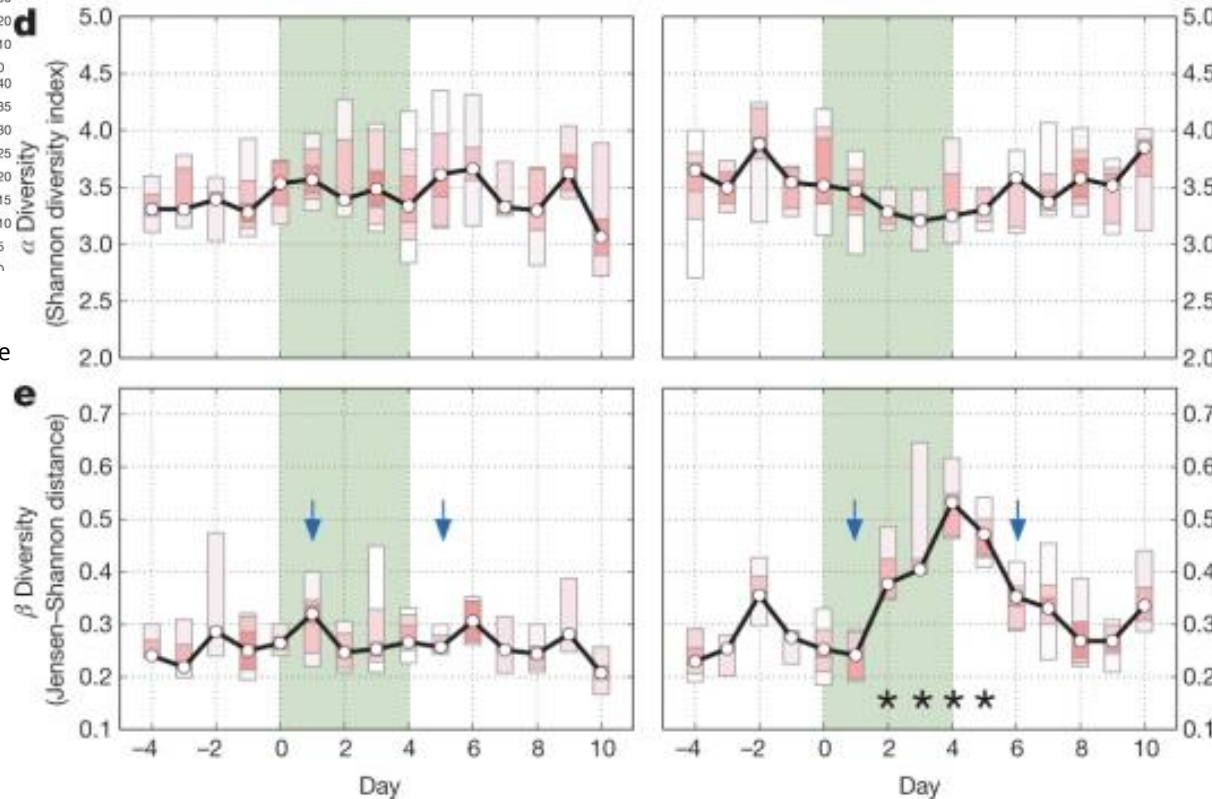
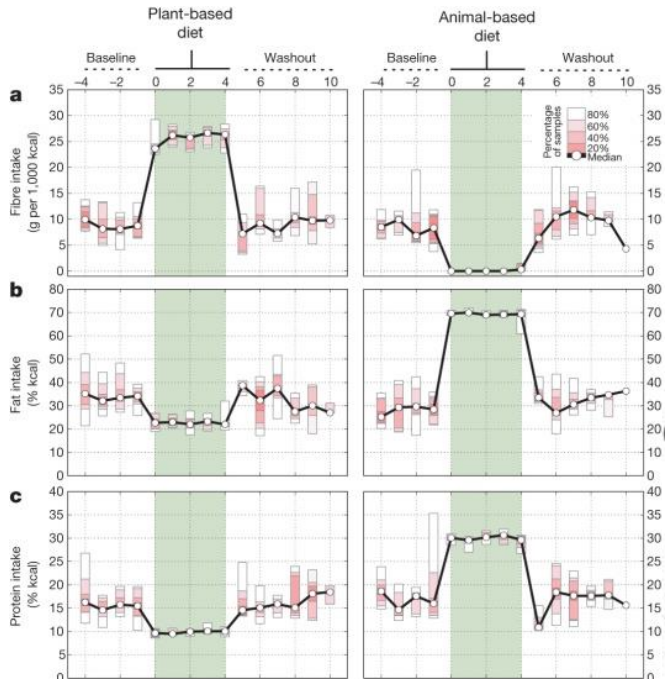


# Microbiome and Health





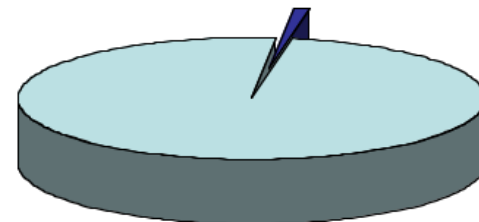
# Short-term diet alters the gut microbiota.



LA David *et al. Nature* **000**, 1-5 (2013) doi:10.1038/nature

# Methods for Characterizing Microbiome

- Use 16S rRNA as a culture-independent phylogenetic marker.
  - Disadvantage: 16S rRNA sequences rarely reveal the physiology of the cells.
- Sequence whole environmental DNA
  - Can show both the **physiology** and **taxonomy** of the cells



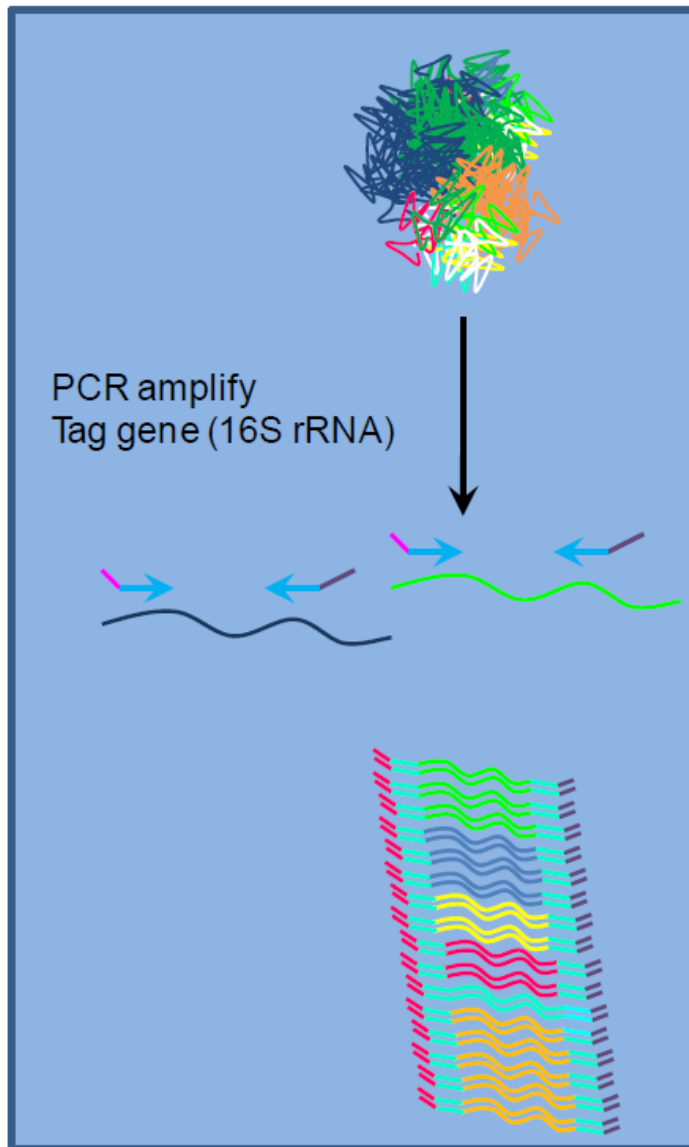
■ Uncultured ■ Cultured

Biosphere has  $10^{30}$  -  $10^{31}$  microbial genomes

**Target Gene**

**vs**

**Metagenomics**

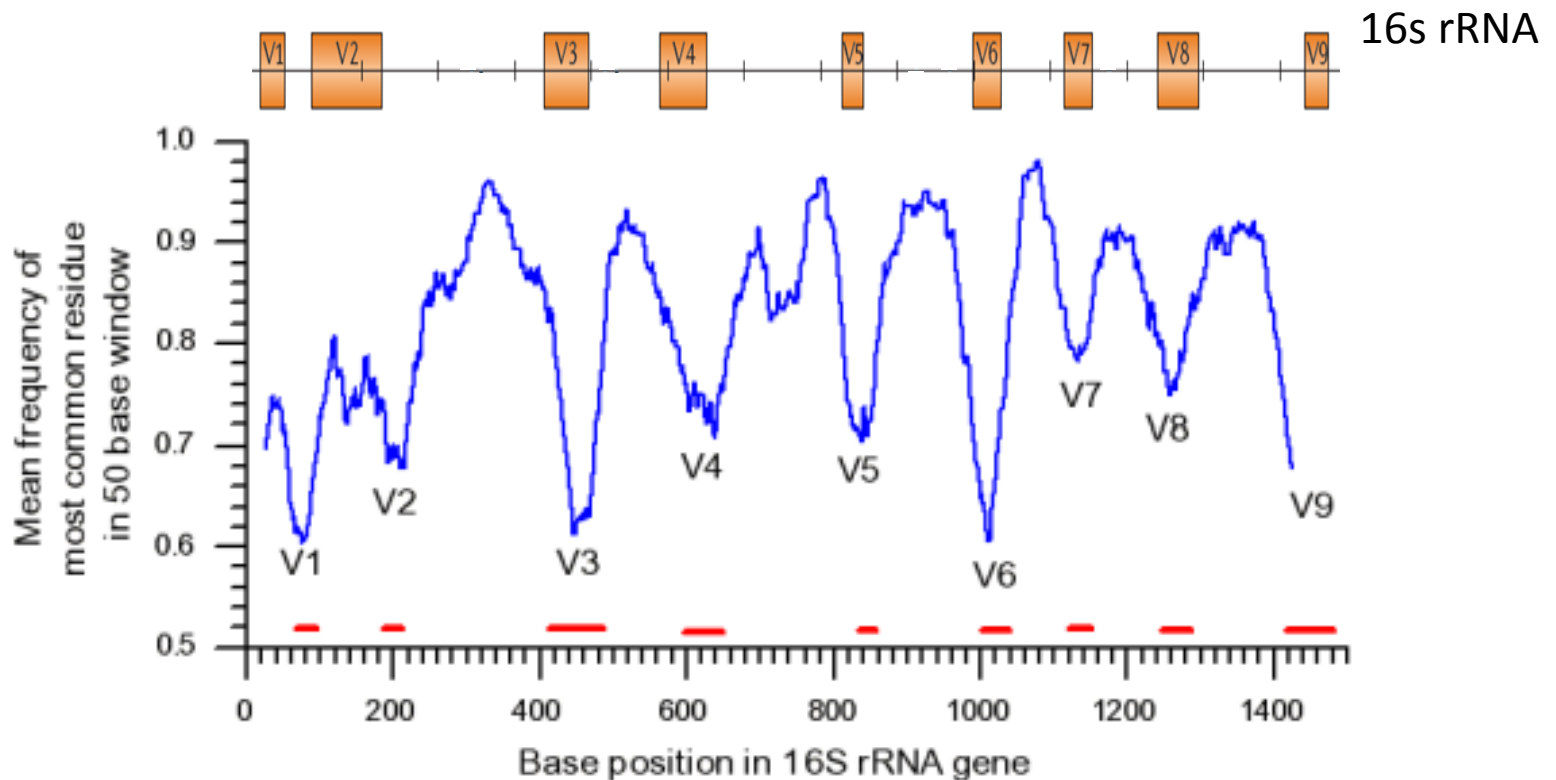


Shotgun  
Library



# 16s rRNA gene as a proxy for taxonomic assignment

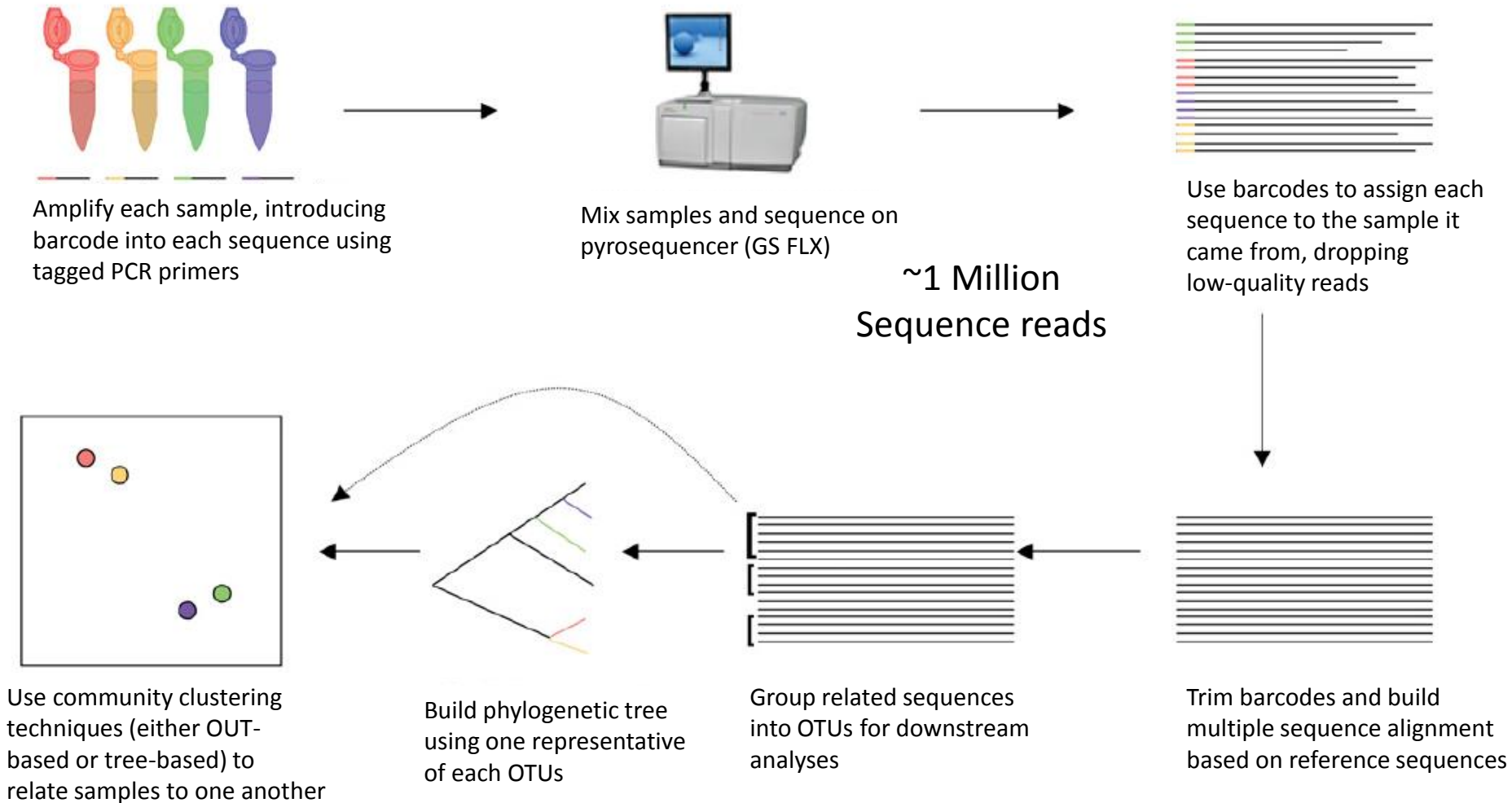
- Most widely used marker gene
- Has both highly conserved regions and highly variable regions



# Operational Taxonomic Unit (OTU)

- OTUs
  - cluster of similar sequence variants of the 16S rDNA marker gene sequence
  - used to categorize bacteria based on sequence similarity
  - OTU cluster is defined by a 97% identity threshold of the 16S gene sequence variants

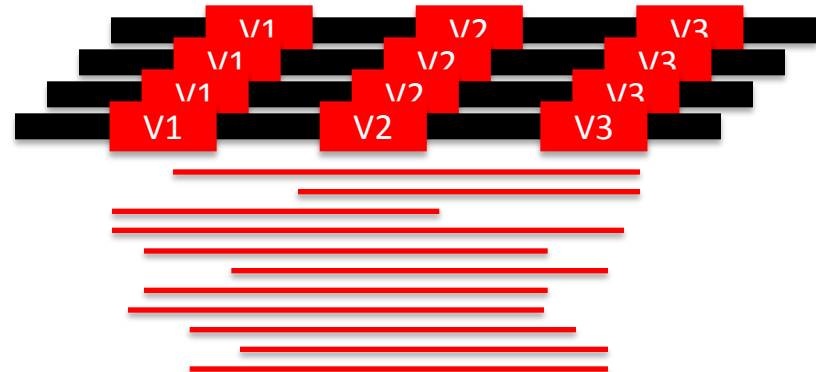
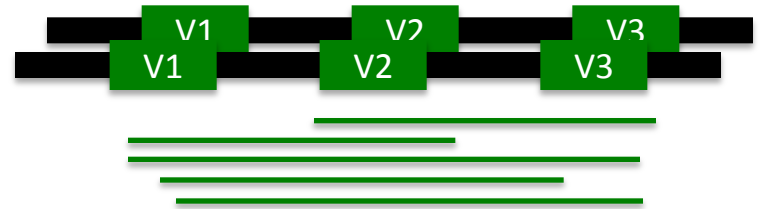
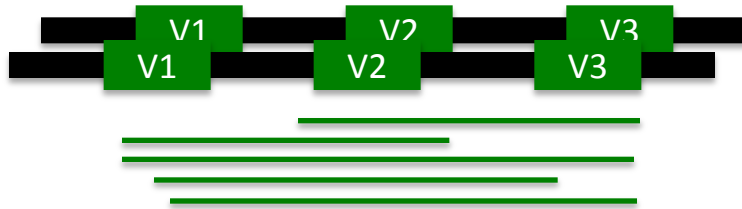
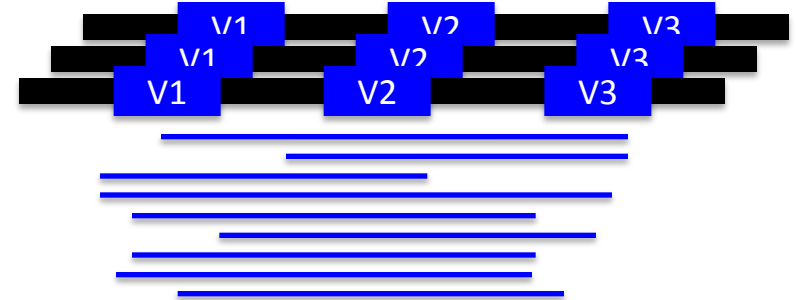
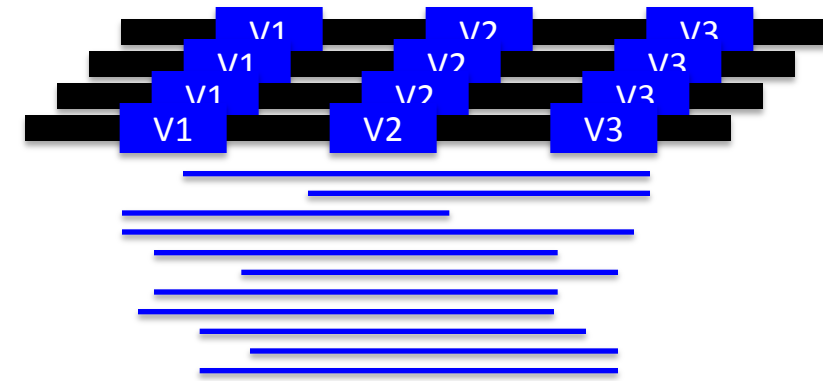
# Overview of workflow with NextGen Sequencing



# Metagenomic Sequencing

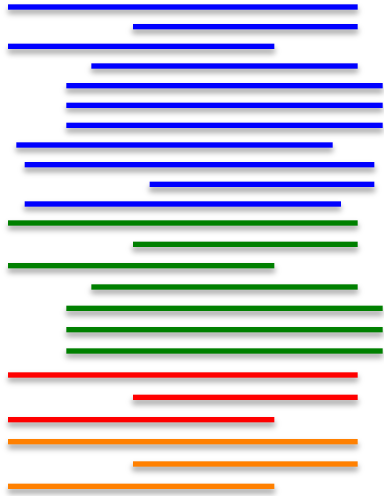
Sample 1

Sample 2

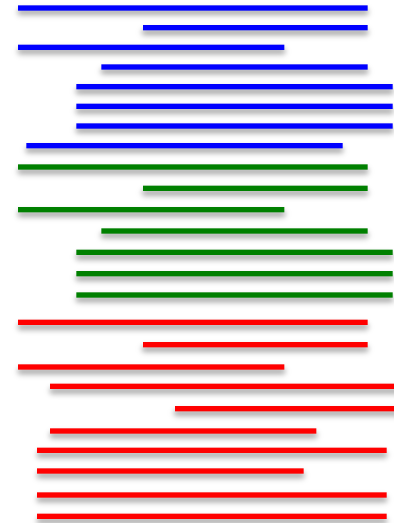


# Metagenomic Sequencing

Sample 1



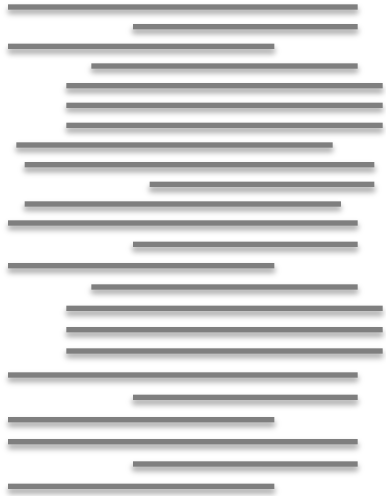
Sample 2





# Metagenomic Sequencing

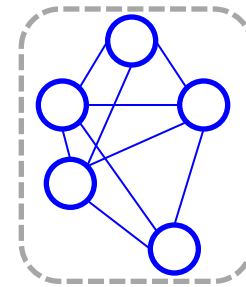
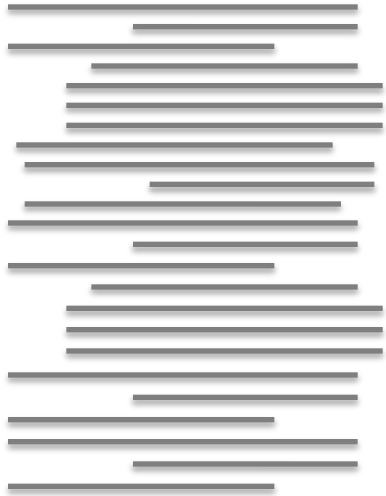
## Sample 1



## Sample 2

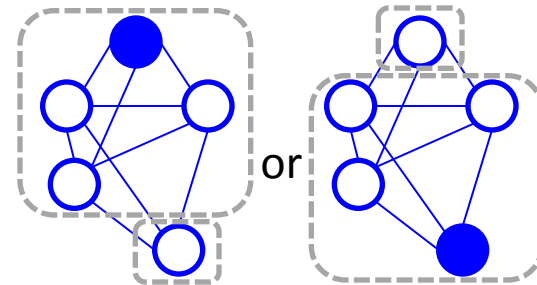


# Clustering of OTUs



= 1 OTU

Each sequence has a neighbor with at least 97% relatedness (Nearest Neighbor)

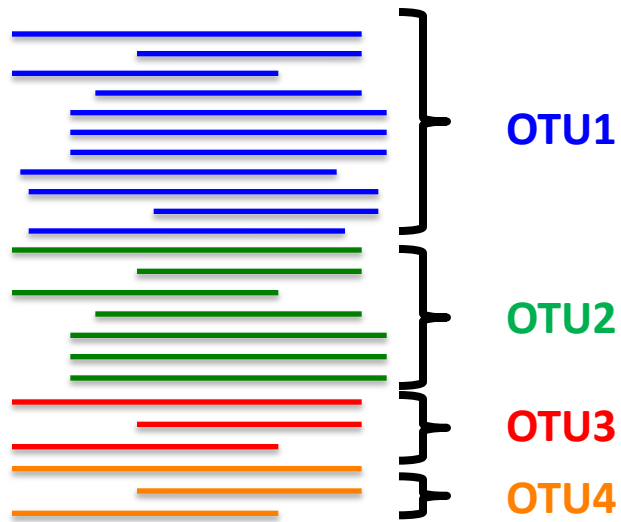


= 2 OTUs

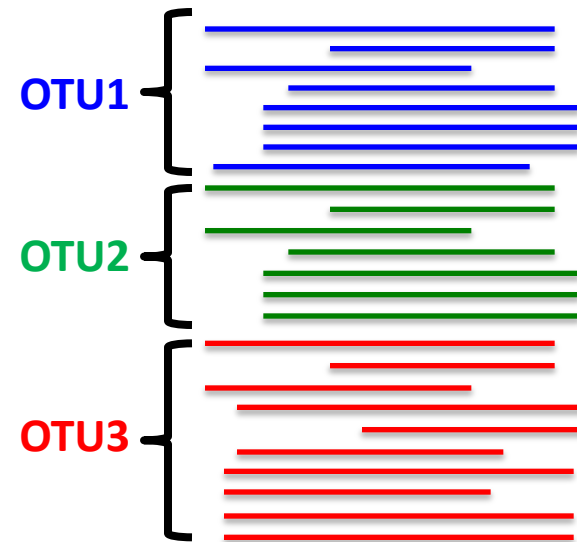
All sequences within a cluster have at least 97% relatedness (Furthest Neighbor)

# Clustering of OTUs

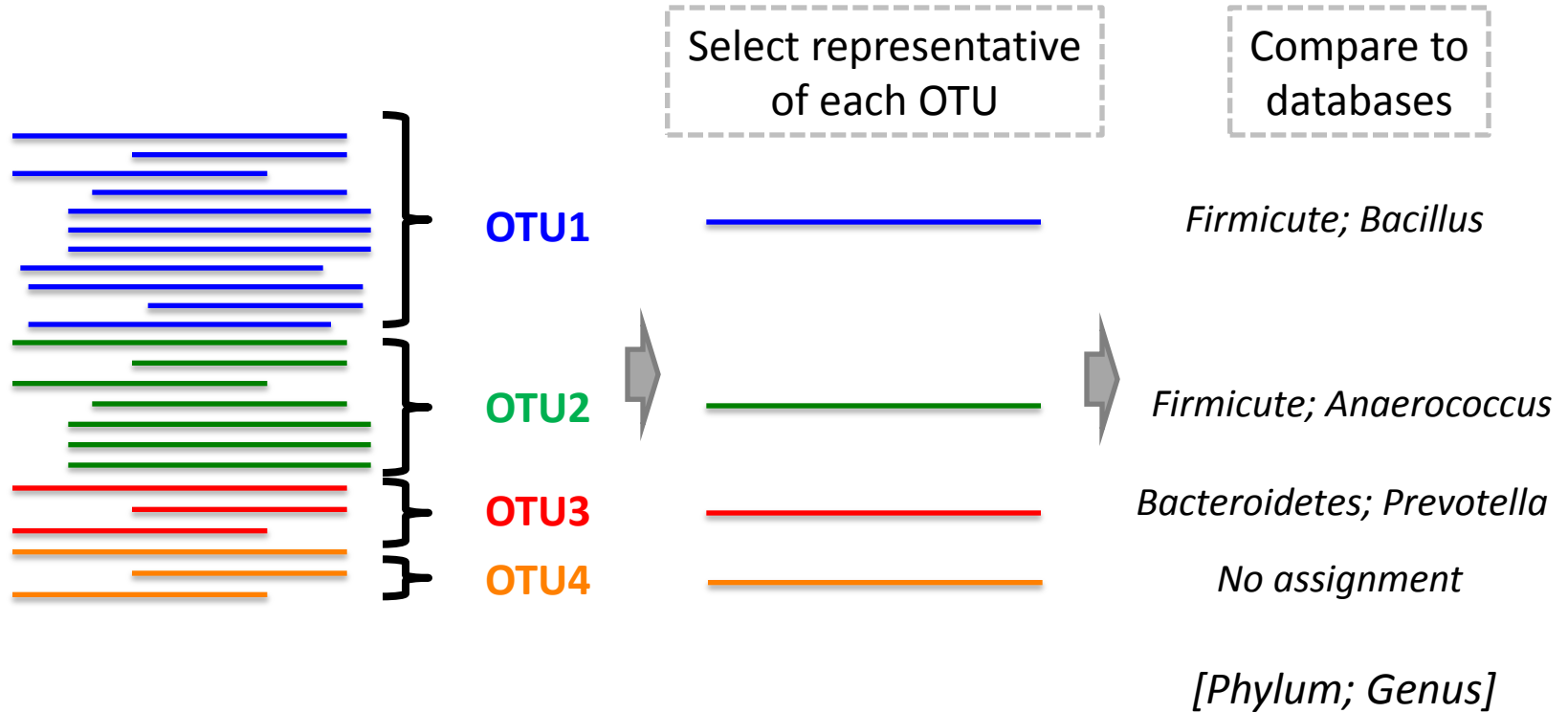
Sample 1



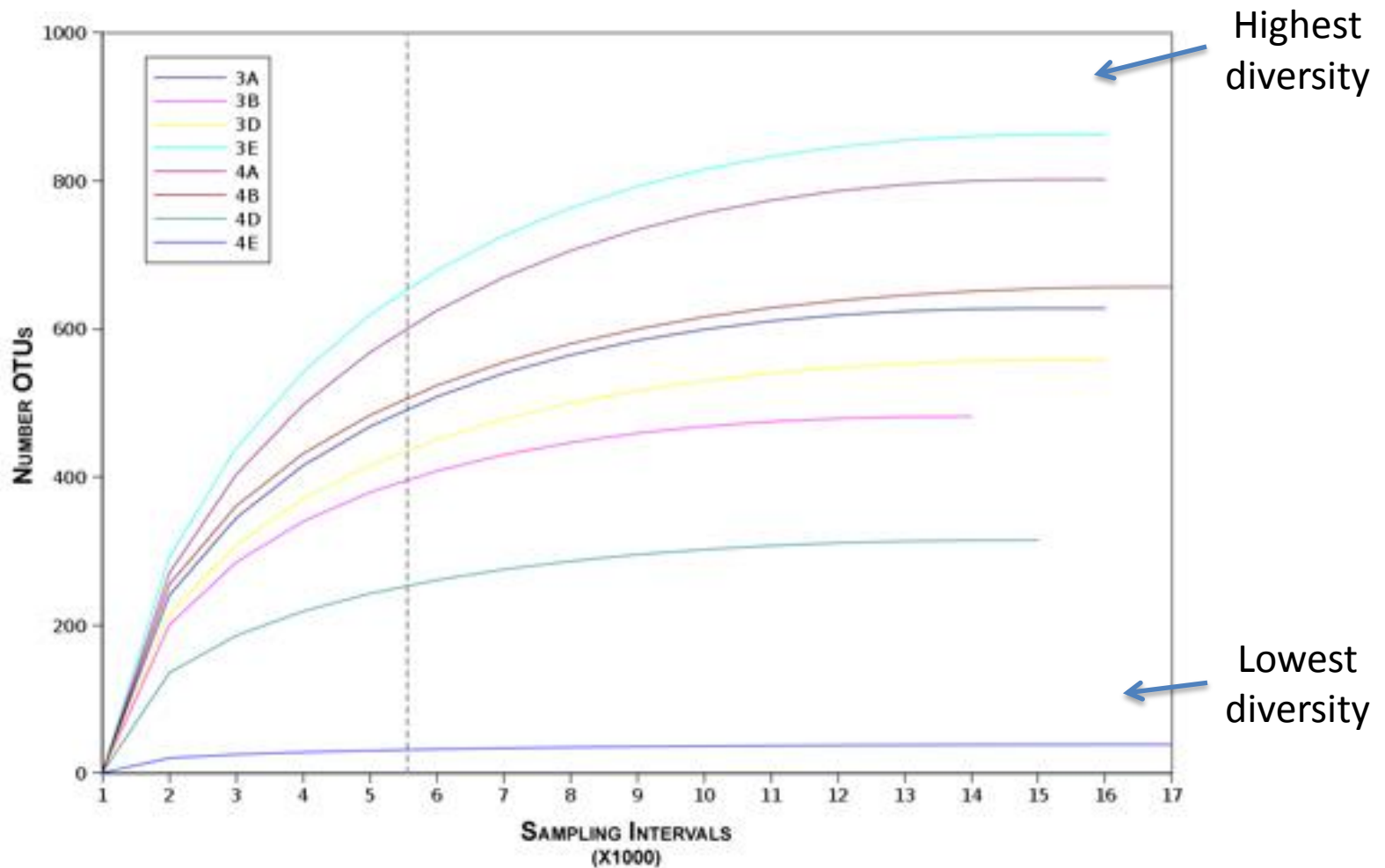
Sample 2



# Taxonomic assignments of OTUs



# Rarefaction curves to determine if sequencing is exhaustive



# Shotgun Sequencing for Human Microbiomes

- Reference based assembly
- **Denovo assembly**: assembling reads into contigs without reference genomes
  - a substantial fraction of the metagenome (~33%) is not well-represented by reference genomes
- Assembling reads from non-homogeneous sample with multiple species

# Shotgun Sequencing for Human Microbiomes

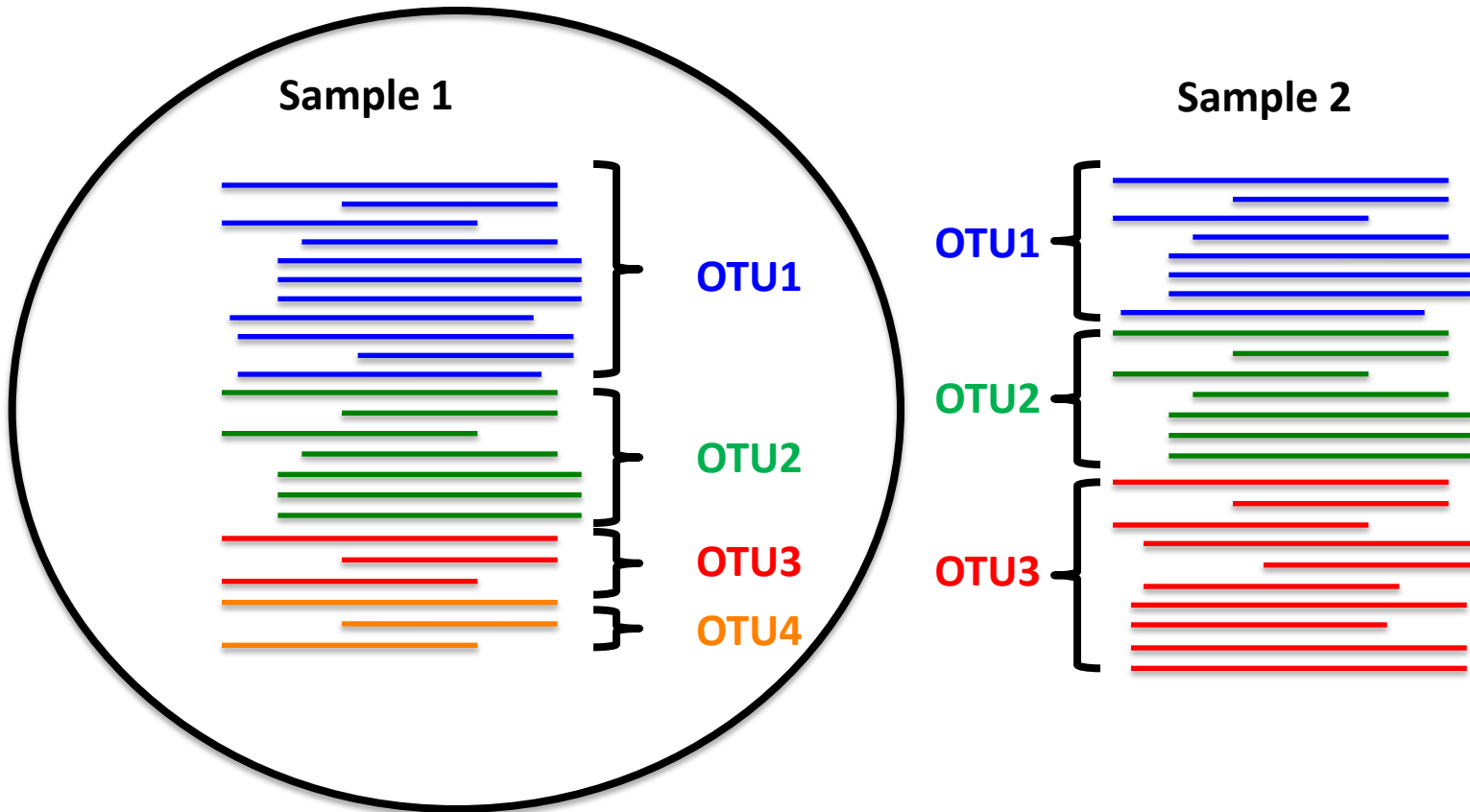
- Assigning unassembled sequences generated by shotgun high-throughput sequencing to the known gene sequences.
  - the assessment of interactions that occur
    - within the microbiome
    - between a microbiome and its host
- However, a substantial fraction of the metagenome (~33%) is not well-represented by reference genomes

# Analyses of Sequenced Microbiome Genome Data

- **Alpha-diversity vs beta-diversity**
  - How many taxa in a sample vs how many are shared across samples
- **Clustering individuals according to microbiome composition**

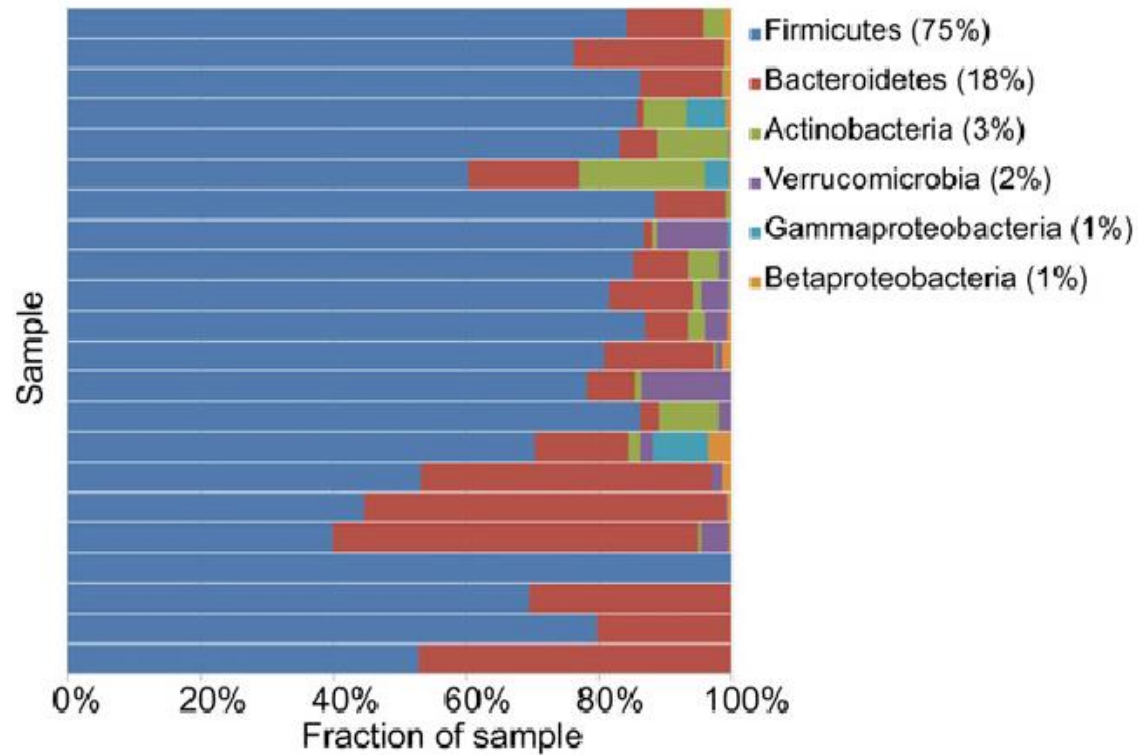


# $\alpha$ Diversity



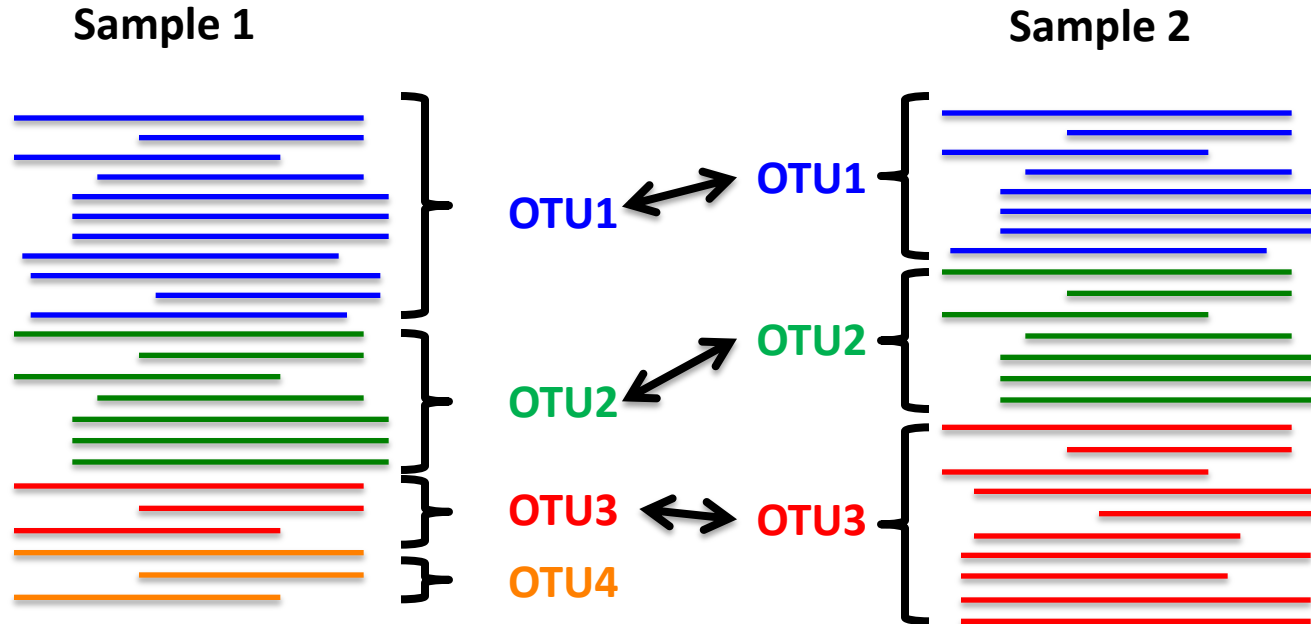
**Alpha diversity:** Diversity within each sample, i.e., how many taxa in a sample

# $\alpha$ Diversity



22 human gut samples with depth of coverage of at least 350 sequences per individual

# $\beta$ Diversity



**Beta-diversity**: Differences in species composition across samples, i.e., how many taxa are shared across samples

# $\beta$ Diversity

- Yue & Clayton's Theta measure

$$\theta_{YC} = \frac{\sum_{i=1}^{S_{12}} \frac{X_i}{n_{total}} \frac{Y_i}{m_{total}}}{\sum_{i=1}^{S_1} \left( \frac{X_i}{n_{total}} \right)^2 + \sum_{i=1}^{S_2} \left( \frac{Y_i}{m_{total}} \right)^2 - \sum_{i=1}^{S_{12}} \frac{X_i}{n_{total}} \frac{Y_i}{m_{total}}}$$

- $S_{12}$ : number of shared OTUs in A and B
- $S_1$  and  $S_2$ : observed number of OTUs in each community
- $X_i, Y_i$ : abundance of the  $i$ th shared OTU in A and B
- $n_{total}, m_{total}$ : total number of sequences sampled in A and B

# Enterotypes of Human Gut Microbiome

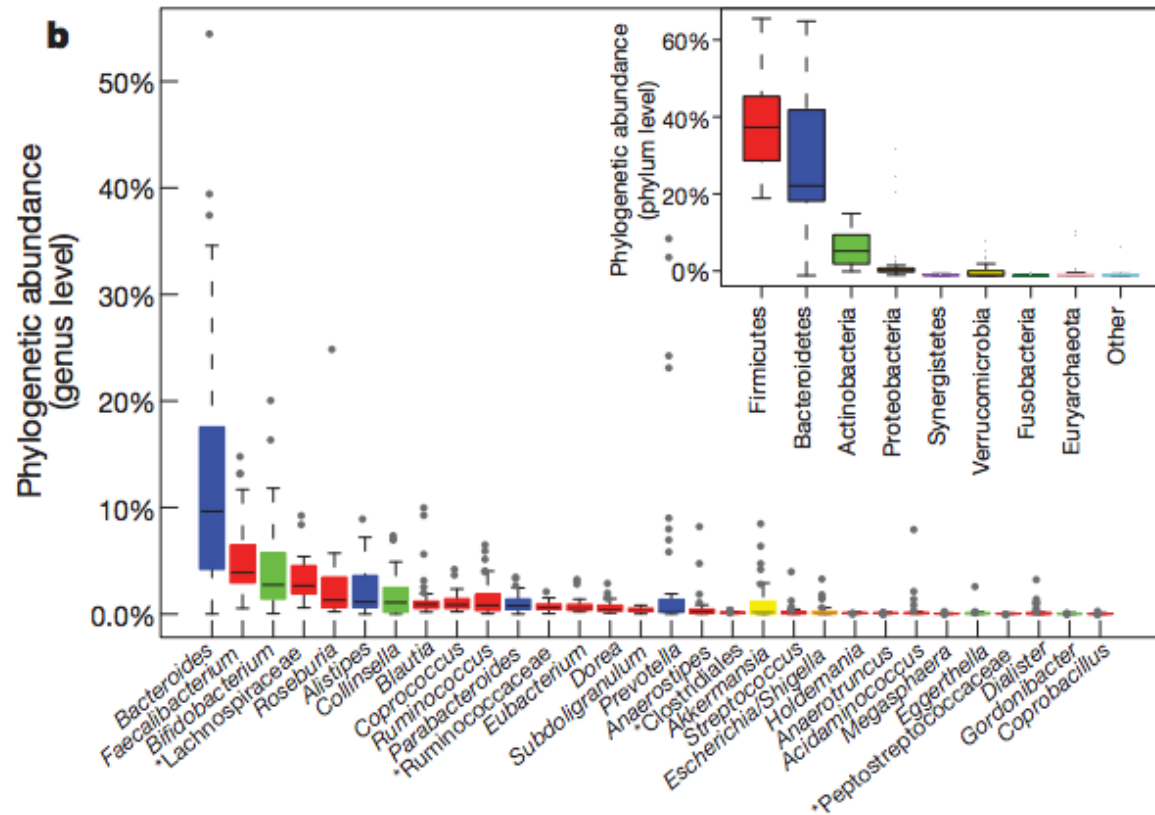
(Arumugan et al., Nature, 2011)

- Analysis of gut microbiome sequence data of 39 individuals including French, Spanish, Danish, Italian, American, and Japanese
- Enterotype: classification of individuals based on their microbiome ecosystem

# Enterotypes of Human Gut Microbiome

(Arumugan et al., Nature, 2011)

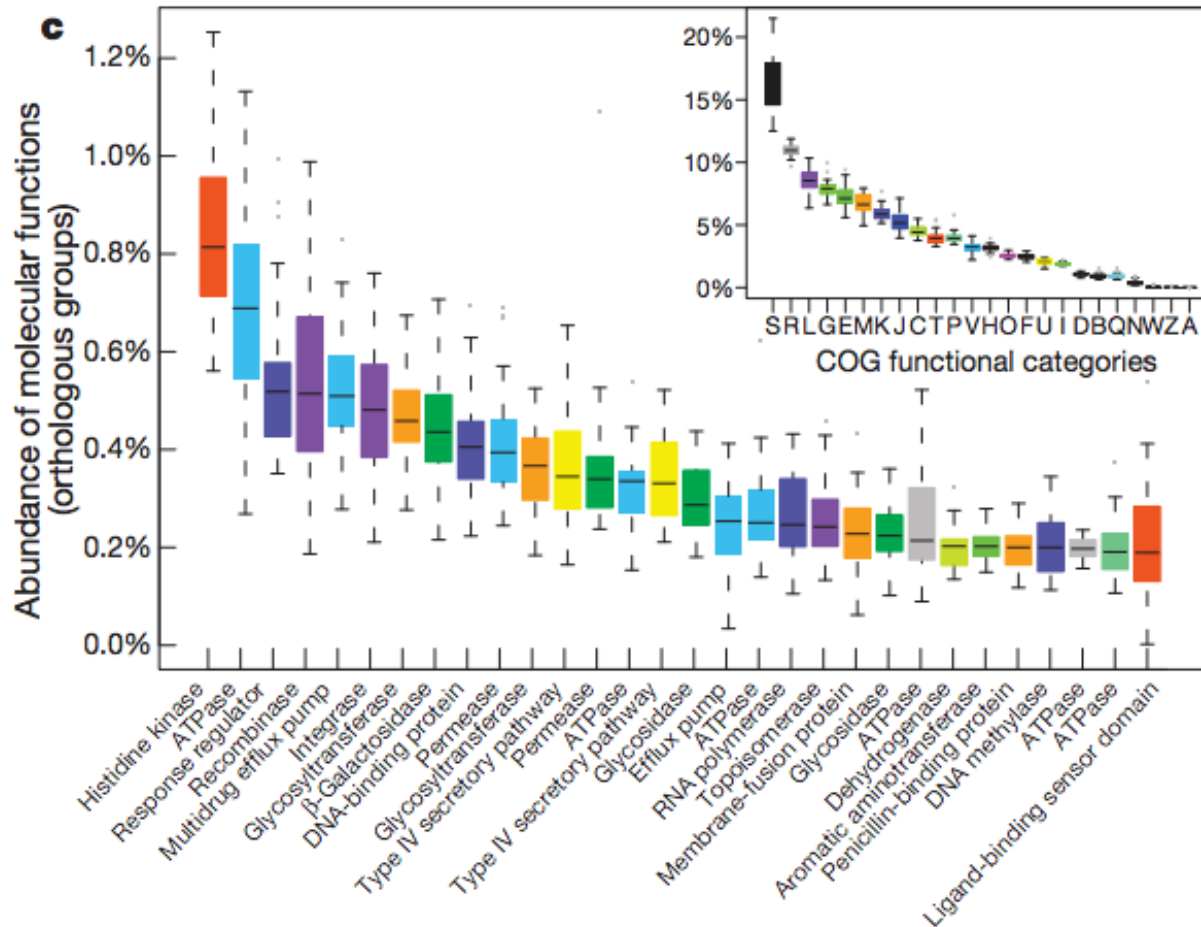
- Profiles of human gut microbiome (Boxplots represent individual variation)



# Human Gut Microbiome

(Arumugan et al., Nature, 2011)

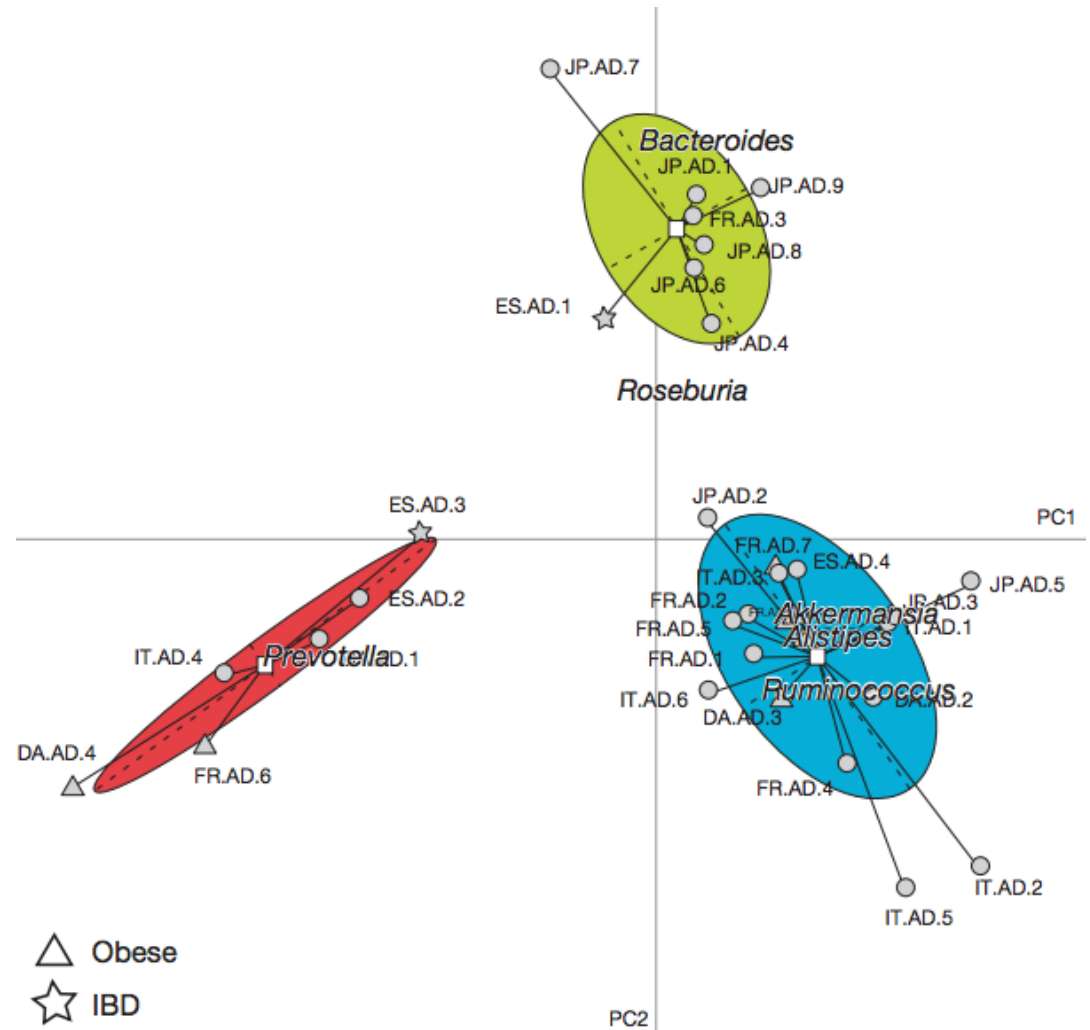
- Functional categories of the orthologous groups of genes



# Human Gut Microbiome

(Arumugan et al., Nature, 2011)

- The 39 individuals can be assigned to one of three Enterotypes
  - Bacteroides
  - Prevotella
  - Ruminococcus

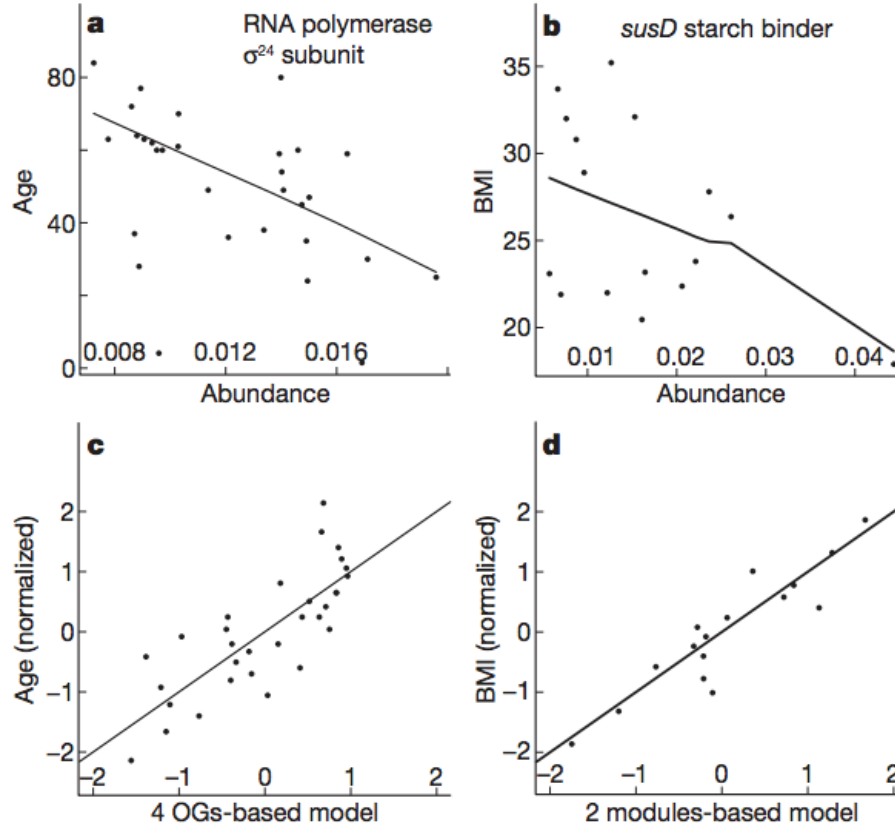




# Human Gut Microbiome

(Arumugan et al., Nature, 2011)

- Correlations between host properties and abundant microbiome species



# Microbiome and Human Health

- Causal link between microbiome variation and disease
  - Variant microbe populations that occur in specific disease states
  - Temporal microbial changes over the course of a disease
- Associations between human conditions and microbiota characteristics

Disease	Relevant finding
Psoriasis	Increased ratio of Firmicutes to Actinobacteria
Reflux oesophagitis	Oesophageal microbiota dominated by gram-negative anaerobes; gastric microbiota with low or absent <i>Helicobacter pylori</i>
Obesity	Reduced ratio of Bacteroidetes to Firmicutes
Childhood-onset asthma	Absent gastric <i>H. pylori</i> (especially the cytotoxin-associated gene A ( <i>cagA</i> ) genotype)
Inflammatory bowel disease (colitis)	Larger populations of Enterobacteriaceae
Functional bowel diseases	Larger populations of <i>Veillonella</i> and <i>Lactobacillus</i>
Colorectal carcinoma	Larger populations of <i>Fusobacterium spp.</i>
Cardiovascular disease	Gut-microbiota-dependent metabolism of phosphatidylcholine

# Microbiome and Human Health

- Gut microbiota and obesity
  - In mice
    - Genetically obese mice have decreased Bacteroidetes/Firmicutes ratios compared with their lean siblings
    - Transplantation of gut microbiota from the obese (ob/ob) to germ-free mice conferred an obese phenotype – the transferred microbiomes had increased capacities for energy harvest
  - In humans
    - the relative proportions of members of the Bacteroidete phylum increase with weight loss
    - Antibiotic use in human infancy (before the age of 6 months) was significantly associated with obesity development

# American Gut Project

<http://humanfoodproject.com/americangut/>

Project to sequence thousands of human microbiomes.

For \$99 you can send in your sample - this supports the scientists who are doing the work

Trying to connect it to diet - need a 7 day diet journal

# Summary

- Human microbiomes are being investigated for gene content and function using sequencing technology
- Classifying individuals into enterotypes based on gut microbiomes
- Characteristics of human microbiomes are often related to health issues