BST281: Genomic Data Manipulation, Spring 2019

Monday 03: Amplicon meta'omics

Metagenomics: culture-independent study of microbes and microbial ecology.

Sequencing is one of several cellular and molecular tools now employed:

 Started with in situ dyes and stains that eventually developed to target specific DNA sequences.

 This led to 16S rRNA gene amplicon sequencing as a "universal" marker gene.

 Conserved regions that can be targeted for PCR etc.

 Variable regions that can be used as a molecular "name tag."

 Good evolutionary properties for short- and long-term molecular clocking.

Taxonomic profiles from 16S data can be generated and analyzed in several ways.

 Binning to Operational Taxonomic Units (OTUs): assignment of reads to taxa defined by % identity threshold.

 Open reference = clustering, closed reference = classification.

 Exact sequence variants (ESVs) or similar (oligotypes etc.) use probabilistic models of frequency, errors.

Amplicon sequencing also includes 18S for eukaryotes, Internal Transcribed Spacers (ITS) particularly for fungi.

Many potential sources of unwanted technical bias in amplicon sequencing.

Extraction, primers, amplification, contamination.

Ecology studies overall community structure and interactions.

 Abundance = how much of an organism is present, prevalence = how many samples it's present in.

 Diversity = types and distribution of taxa in a community, richness = simple number of taxa.

 Qualitative = focus on which organisms are present, quantitative = focus on how abundant they are.

 Taxonomic = how many different organisms (by some definition), phylogenetic = how evolutionarily related.

 Alpha = within-sample (like an absolute value), beta = between-sample (like a distance or correlation score).

Ordination is one of several popular analysis methods.

 Projects high-dimensional community structure into two-dimensional scatter plot for summary visualization.

Features (taxonomic or functional) can be statistically analyzed much like RNA-seq or GWAS: association testing.

 Keeping in mind that microbiome data are sparse, compositional, and noisy.

# Textbooks

Meta'omics: Pevsner, Chapter 15 p700-711, stop before Brief Chronology

 p720-737, stop before GENOME ANALYSIS PROJECTS: ANNOTATION

 Chapter 17, p797-830, stop before COMPARISON OF BACTERIAL GENOMES

Prokaryotic 'omics: Lesk, Chapter 6 p191-212

# Literature

[Environmental genome shotgun sequencing of the Sargasso Sea. Venter, Science 2004](https://www.ncbi.nlm.nih.gov/pubmed/15001713)

[Structure, function, and diversity of the healthy human microbiome. Huttenhower, Nature 2012](https://www.ncbi.nlm.nih.gov/pubmed/22699609)