BST281: Genomic Data Manipulation, Spring 2017

Monday 08: Metagenomics

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.

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.

Metagenomic and metatranscriptomic sequencing allow taxonomic and functional profiling.

 Can be assembly based (de novo) or reference based.

 Taxonomic profiling = figuring out which organisms are present and how much, like amplicon data.

 Functional profiling = which gene families and pathways are present, and how much.

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.

Human Microbiome Project one of many to translate host-associated microbial community studies to health.

# 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

# Literature

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

[A human gut microbial gene catalogue established by metagenomic sequencing. Qin, Nature 2010](https://www.ncbi.nlm.nih.gov/pubmed/20203603)

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

[Personalized Nutrition by Prediction of Glycemic Responses. Zeevi, Cell 2015](https://www.ncbi.nlm.nih.gov/pubmed/26590418)