BST281: Genomic Data Manipulation, Spring 2019

Wednesday 03: Metagenomics

This lecture continues our discussion from Monday on ‘omics analyses of microbial communities. Today, we’ll discuss methods based on shotgun sequencing of community DNA.

## Amplicon vs. shotgun sequencing

## Whereas amplicon sequencing surveyed a particular gene in (mostly) bacteria, shotgun sequencing profiles the complete DNA content of a community, including native bacteria, archaea, fungi, and viruses, as well as contaminant DNA. Shotgun sequencing of microbial community DNA is “metagenomics” in the strict sense, as it considers multiple genomes. It can profile the complete functional content of a community (e.g. gene families and pathways) in addition to taxonomy, often with better resolution than amplicon sequencing.

## Taxonomic profiling by mapping

Taxonomic profiling can be performed by mapping community reads to reference genomes. Requiring greater genome coverage breadth boosts specificity, while coverage depth is proportional to community abundance. Similar analyses can be performed using only clade-specific marker genes, which boosts specificity and computational performance. Binning approaches rapidly assign reads to taxa based on *k*-mers overlap.

*Strain profiling by mapping*

## Metagenomic strains can be profiled at the gene presence/absence or single-nucleotide variant level. The former is coarser grained but more feasible at lower coverage depths. Both approaches can be used to examine population structure within an individual microbial species.

## Functional profiling by mapping

Functional profiling is performed using more sensitive mapping techniques (e.g. translated search) to identify remote homology, but as a result is more computationally demanding. Raw gene family abundances are regrouped to broader categories (e.g. pathways). Across samples from a given environment type, community function tends to be more conserved than taxonomy.

## Metagenomic assembly

Metagenomic assembly experiences the technical challenges of single-genome assembly alongside issues from mixing genomes of variably-abundant species. Contigs must be “binned” into candidate genomes by supervised (e.g. taxonomic placement-based) and/or unsupervised (e.g. composition- and coverage-based) approaches. Candidate genomes are then annotated using approaches designed for isolate genomes. Despite its challenges, metagenomic assembly is powerful for recovering novel gene and genome sequences.

# Suggested textbook reading

* Lesk, p204-208 (metagenomics case study)

# Related literature

* Review article: [Quince, Christopher, et al. "Shotgun metagenomics, from sampling to analysis." Nature biotechnology 35.9 (2017): 833.](https://www.nature.com/articles/nbt.3935)
* Expanded Human Microbiome Project (HMP1-II): [Lloyd-Price, Jason, et al. "Strains, functions and dynamics in the expanded Human Microbiome Project." Nature 550.7674 (2017): 61.](https://www.nature.com/articles/nature23889)
* MAGs from TARA Oceans: [Delmont, Tom, et al. “Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes.” Nature Microbiology 3.7 (2018): 804.](https://www.nature.com/articles/s41564-018-0176-9)