

Tools and Algorithms in Bioinformatics

GCBA815, Fall 2017

Week-13
Metagenomics Data Analysis

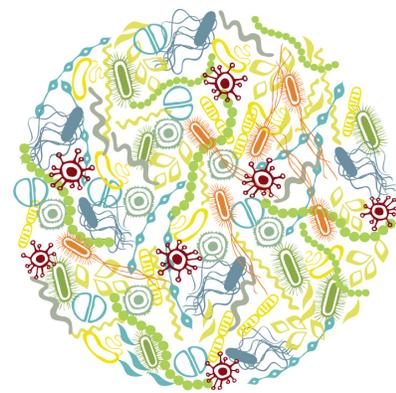
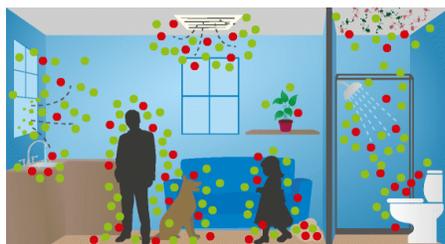
Sanjit Pandey, MS
Bioinformatics Analyst (Guda lab)

Department of Genetics, Cell Biology & Anatomy
University of Nebraska Medical Center



Introduction

- Microbes are found virtually everywhere in nature



<http://learn.genetics.utah.edu/content/microbiome/intro/>

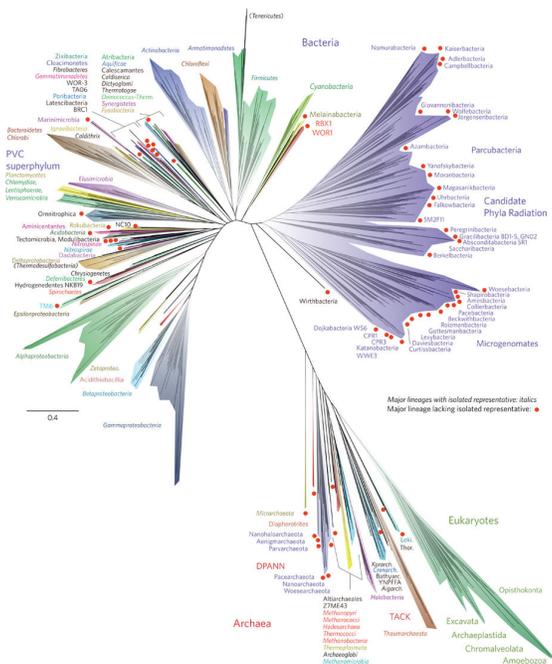


Environments



Tree of Life

A new view of the tree of life Laura A. Hug, Brett J. Baker, Karthik Anantharaman1, Christopher T. Brown, Alexander J. Probst, Cindy J. Castelle, Cristina N. Butterfield, Alex W. Hermsdorf3, Yuki Amano4, Kotaro Ise, Yohey Suzuki, Natasha Dudek, David A. Relman, Kari M. Finstad9, Ronald Amundson9, Brian C. Thomas and Jillian F. Banfield



Human Microbiome

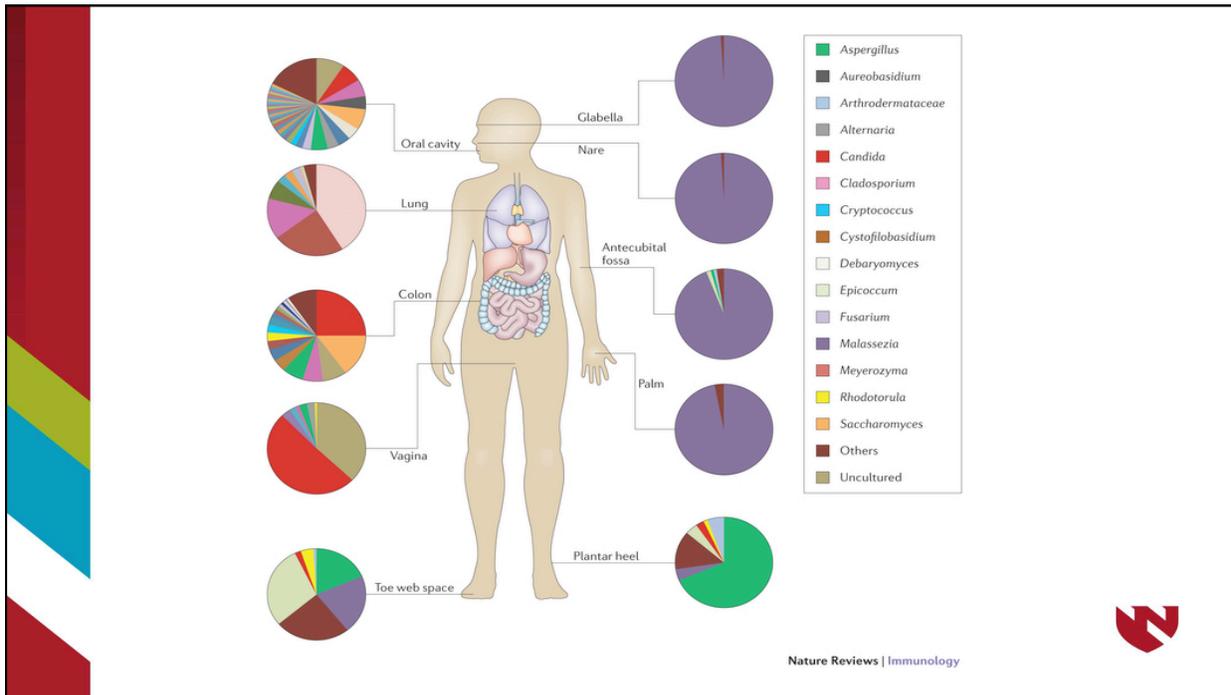
- The Human Microbiome
 - Microorganisms living in association with human body
- Recent studies suggest that there are around 39 trillion bacteria in a human body
- Human body has around 30 trillion human cells.



Human Microbiome Project

- Human microbiome project: About 3,000 were isolated and sequenced from an estimated 10,000 species
- Human gut alone is estimated to harbor ~1000 species
- Abnormal gut function is considered a risk factor for a series of autoimmune diseases
- A healthy gut is the key to a healthy body





Symbiotic & Pathogenic

Good and Bad Bacterial Flora



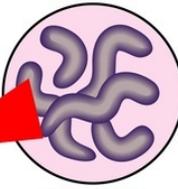
BIFIDOBACTERIA
The various strains help to regulate levels of other bacteria in the gut, modulate immune responses to invading pathogens, prevent tumour formation and produce vitamins.



ESCHERICHIA COLI
Several types inhabit the human gut. They are involved in the production of vitamin K2 (essential for blood clotting) and help to keep bad bacteria in check. But some strains can lead to illness.



LACTOBACILLI
Beneficial varieties produce vitamins and nutrients, boost immunity and protect against carcinogens.



CAMPYLOBACTER
C. jejuni and C. coli are the strains most commonly associated with human disease. Infection usually occurs through the ingestion of contaminated food.

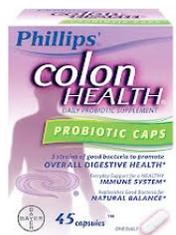


ENTEROCOCCUS FAECALIS
A common cause of post-surgical infections.



CLOSTRIDIUM DIFFICILE
Most harmful following a course of antibiotics when it is able to proliferate.

Probiotics and Prebiotics



Potential Health Benefits

- Useful for the treatment of
 - Gastroenteritis
 - Antibiotic-associated diarrhea
 - Lactose intolerance
 - Colon cancer
- Lowers serum cholesterol
- Anti-aging effects
- Improve immune functions and fight allergies and infections
- Prevent reoccurrences of inflammatory bowel disease (IBD)
- Improve some symptoms of irritable bowel syndrome



Study of Microbes

- Traditional approaches depended on cultivated clonal cultures
 - Involved culture of individual organisms and 16s rRNA gene
 - Difficult to study microbes that cannot be cultured in lab (less than 1% of microbes can be grown/cultured in lab)
- Modern approaches involves studying genetic material recovered directly from environmental samples
 - Allows us to identify almost all the species in the sample at once



Study of Microbes

What is there ?

How many of them are there ?

What are they doing ?

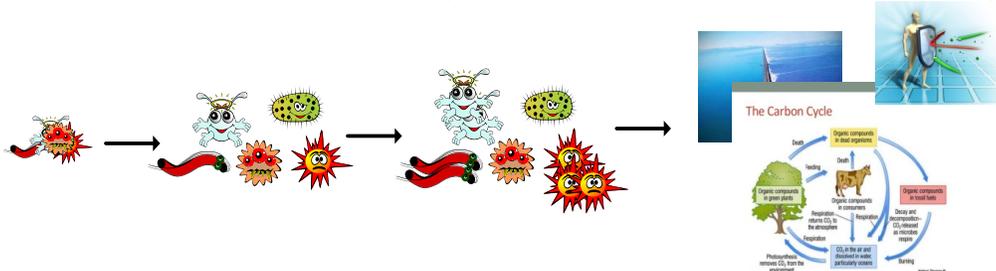


Image source: openclipart.org, <http://learn.genetics.utah.edu>, google images



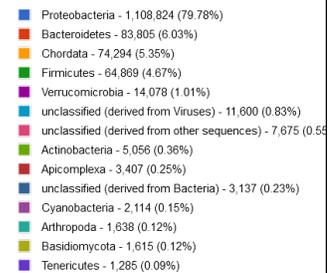
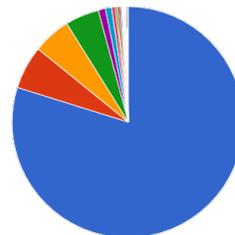
Case Studies

- Study of Fecal Microbiota Transplant (FMT) on *C. difficile* population using whole genome sequencing.
- Study of development of infant gut using whole genome sequencing
- Study of microbial content of air in pig farm using targeted sequencing of 16s region
- Study of effect of Betaine in preventing alcohol-induced gut dysbiosis using whole genome sequencing
- Study of change in community structure over time during decomposition using targeted sequencing of 16s region.



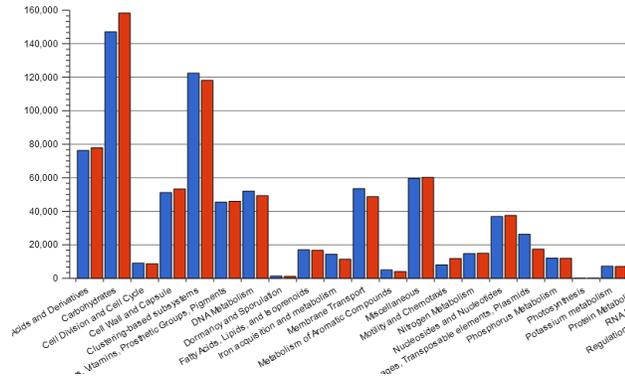
Case Studies

- Study of Fecal Microbiota Transplant (FMT) on *C. difficile* population using whole genome sequencing.
 - Fecal sample collected from recipient before and after transplant
 - Fecal sample collected from donor
 - Comparison of community change before transplant and over multiple days after transplant



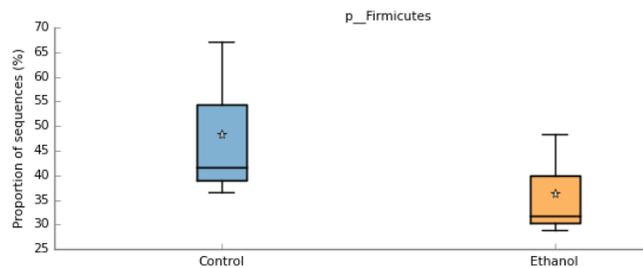
Case Studies

- Study of development of infant gut using whole genome sequencing
 - Samples collected from stool of infant at different time interval ranging from day three to 2 months.
 - Comparison of community change as the infant's gut microbiota develops over time.



Case Studies

- Study of effect of Betaine in preventing alcohol-induced gut dysbiosis using whole genome sequencing
 - Samples collected from rats before and after alcohol/betaine treatment
 - Comparison of community change in normal vs alcohol induced vs betaine treated samples



Case Studies

- Study of change in community structure over time during decomposition using targeted sequencing of 16s region.
 - Samples collected from decomposing mouse over 24 days in different moisture condition.
 - How does the composition of the bacteria change at different moisture condition and at different time interval



Metagenomics vs Genomics

- Genomics focuses on genetic study of single organism
- Metagenomics focuses on multiple organisms at the same time
 - Study of effect of specific bacteria in the environment
 - Study of collective effect of all bacteria involved



Metagenomics

- **Metagenomics** is the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species. (Chen, K.; Pachter, L. (2005))



Metagenomics - Sequencing

- **Metagenomic sequencing**

- **Targeted sequencing**

- Sequencing of specific region of the genome such as 16s rDNA.
- Cheaper to sequence but, does not capture viruses and eukaryotes
- Suffers from DNA extraction and amplification biases



Bacterial 16S rDNA

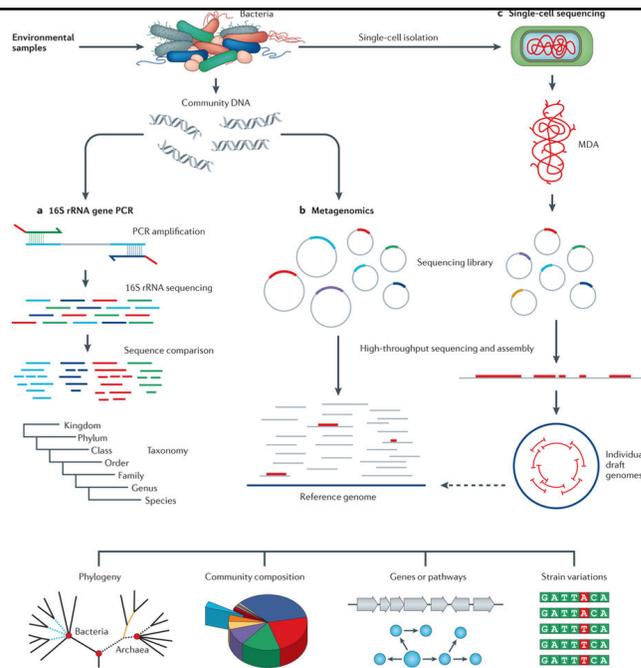
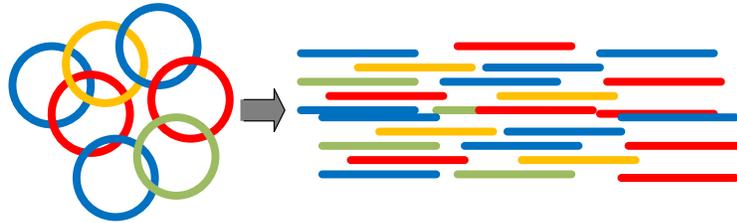


Metagenomics - Sequencing

- **Metagenomic sequencing**

- **Shotgun sequencing**

- Complete sequencing of entire metagenome in the sample
 - Relatively expensive and difficult to get enough coverage
 - Captures absolute measurement
 - Captures viruses and eukaryotes.



Lasken RS, McLean JS. Recent advances in genomic DNA sequencing of microbial species from single cells. *Nat Rev Genet.* 2014;15:577–584.



Metagenomics - Analysis

- Identification and Quantification



Metagenomics - Analysis

- Alignment based methods
 - Reads are aligned directly to the reference genomes
 - Sensitive to gene transfers, sequence lengths
 - Computes all possible pairwise comparisons; compute intensive
 - Well established algorithms and tools available

Target Sequence	5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
Query Sequence	5' TACTCACGGATGAGGTACTTTAGAGGC 3'



Metagenomics - Analysis

- Alignment free methods
 - Based on occurrences of subsequences (n-grams)
 - Memory intensive while computationally inexpensive
 - Less dependent on substitutions or genetic transfers



Analysis tools

Popular tools and methods

Tool	Input Data	Methodology	Highest identification level	Web based
MEGAN	WGS/16s	Alignment-based	Species	No
MetaPhlan	WGS	Alignment-based	Species	No
MG-RAST	WGS/16s	Alignment-based	Genus	<u>Yes</u>
QIIME	16s	Alignment-based	Species	No
MetaPhyler	WGS/16s	Alignment-based	Species	No
GOTTCHA	WGS	Alignment-based	Species	No
Kraken	WGS	LCA - based	Genus	No
MetaID	WGS	<u>N-gram-based</u>	<u>Strain</u>	No



Resources

- IMG/M: Integrated Microbial Genomes & Microbiomes - <https://img.jgi.doe.gov/cgi-bin/m/main.cgi>
- EBI Metagenomics - <https://www.ebi.ac.uk/metagenomics/>
- Human Microbiome Project - www.hmpdacc.org
- MG-RAST - <http://metagenomics.anl.gov/>
- Sequence Read Archive - www.ncbi.nlm.nih.gov/Traces/sra
- MEGAN - <http://ab.inf.uni-tuebingen.de/software/megan6/>
- STAMP - <http://kiwi.cs.dal.ca/Software/STAMP>
- Phyloseq - <http://bioconductor.org/packages/release/bioc/html/phyloseq.html>



Resources

- MG-RAST - <http://metagenomics.anl.gov/>
 - Widely used portal to store and analyze metagenomics data.
 - The pipeline performs qc, protein prediction, clustering, similarity based annotation.
 - Supports shotgun metagenomics data, 16s, 18s



Resources

- **STAMP** - <http://kiwi.cs.dal.ca/Software/STAMP>
 - Software package for analyzing taxonomic and metabolic profiles



Demonstration

- **MG-RAST** - <http://metagenomics.anl.gov/>
- **STAMP** - <http://kiwi.cs.dal.ca/Software/STAMP>
 - Installation
 - Some screen shots
 - Prepare sample data sets for demo
 - demo

