2016 SISMID Module 14: Introduction to Metagenomic Data Analysis

Instructors:
Alexander V. Alekseyenko [Alex] (alexander.alekseyenko@gmail.com)
Paul Joseph McMurdie [Joey] (joey711@gmail.com)

Materials: http://sismid16m14.bot-tak.org

Monday, July 25

8:00 am – 8:30 am Coffee

8:30 am – 10:00 am Class Session:

8:30 am – 8:45 am Introductions

8:45 am – 9:00 am Introduction: Metagenomics in Biology and Medicine

9:00 am – 10:00 am Lecture 1: Metagenomics assays, overview of microbiome analysis

10:00 am – 10:30 am Break

10:30 am – 12:00 pm Class Session:

10:30 am – 12:00 pm Laboratory 1: Sequence denoising w/ DADA2 and Intro to phyloseq

12:00 pm - 1:30 pm Lunch Break

1:30 pm – 3:00pm Class Session:

3:00 pm – 3:30 pm Break

1:30 pm – 2:15 pm Lecture 2: Descriptive statistics, normalizations & testing

2:15 pm – 3:00 pm Laboratory 2: Quality control, transformations, filtering, univariate testing, multiple comparison

3:30 pm – 5:00 pm Class Session:

3:30 pm – 4:30 pm Lecture 3: Mixture models for microbiome data

4:30 pm – 5:00 pm Laboratory 3: Mixture models for differential abundance testing

5:00 pm – 6:00pm Networking Reception

Tuesday, July 26

8:00 am – 8:30 am Coffee

8:30 am – 10:00 am Class Session:

8:30 am – 8:45 am Questions and review

8:45 am – 9:15 am Lecture 4: Evolutionary trees in microbiome data analysis

9:15 am – 10:00 am Laboratory 4: Trees, network manipulation using phyloseq

10:00 am – 10:30 am Break

10:30 am – 12:00 pm Class Session:

10:30 am – 11:15 am Lecture 5: Ecological distance metrics; Principal Coordinates Analysis

11:15 am – 12:00 pm Laboratory 5: Computing distance matrices; PCoA

12:00 pm - 1:30 pm Lunch Break

1:30 pm – 3:00pm Class Session:

1:30 pm – 2:30 pm: Lecture 6: Generalized multivariate analysis of variance

2:30 pm – 3:00 pm: Laboratory 6: PERMANOVA with extensions

3:00 pm – 3:30 pm Break

3:30 pm – 5:00 pm Class Session:

3:30 pm – 4:15 pm Lecture 7: Machine learning with microbiome data

4:15 pm – 5:00 pm Laboratory 7: Clustering and classification.

Wednesday, July 27

8:00 am – 8:30 am Coffee

8:30 am – 10:00 am Class Session:

8:30 am – 8:45 am Questions and review

8:45 am – 9:15 am Lecture 8: Predicting metagenomic composition from 16S survey data

9:15 am – 10:00 am Laboratory 8: Working with functional data

10:00 am – 10:30 am Break

10:30 am – 12:00 pm Class Session:

10:30 am – 11:00 am Lecture 9: Networks

11:00 am – 11:30 am Laboratory 9: SpiecEasi networks tutorial

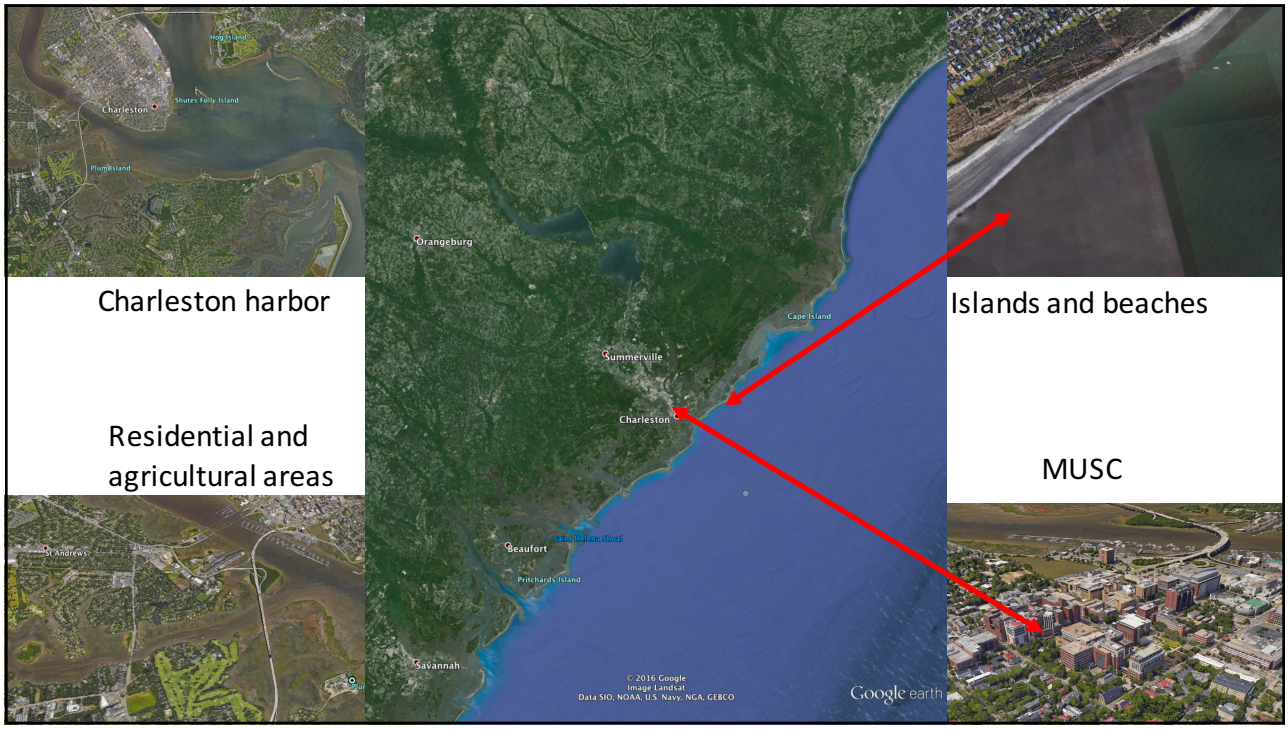
11:30 am – 12:00 pm Questions, feedback, references, resources

Introduction: Metagenomics in Biology and Medicine

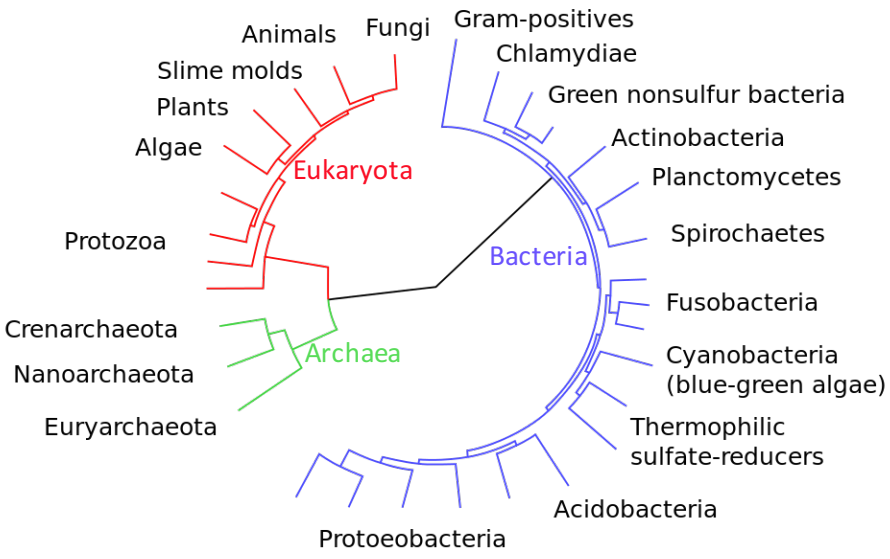
What is a microbiome?

- The totality of microbes in a defined environment, especially their genomes and interactions with each other and surrounding environment.
  - A population of a single species/strain is a culture, extremely rare outside of lab, some infections
  - A microbiome is a mixed population of different microbial species (microbial ecosystem)
- Joshua Lederberg (1925 – 2008): "the ecological community of commensal, symbiotic, and pathogenic microorganisms that literally share our body space" (Lederberg and McCray Scientist. 2001;15:8).

2

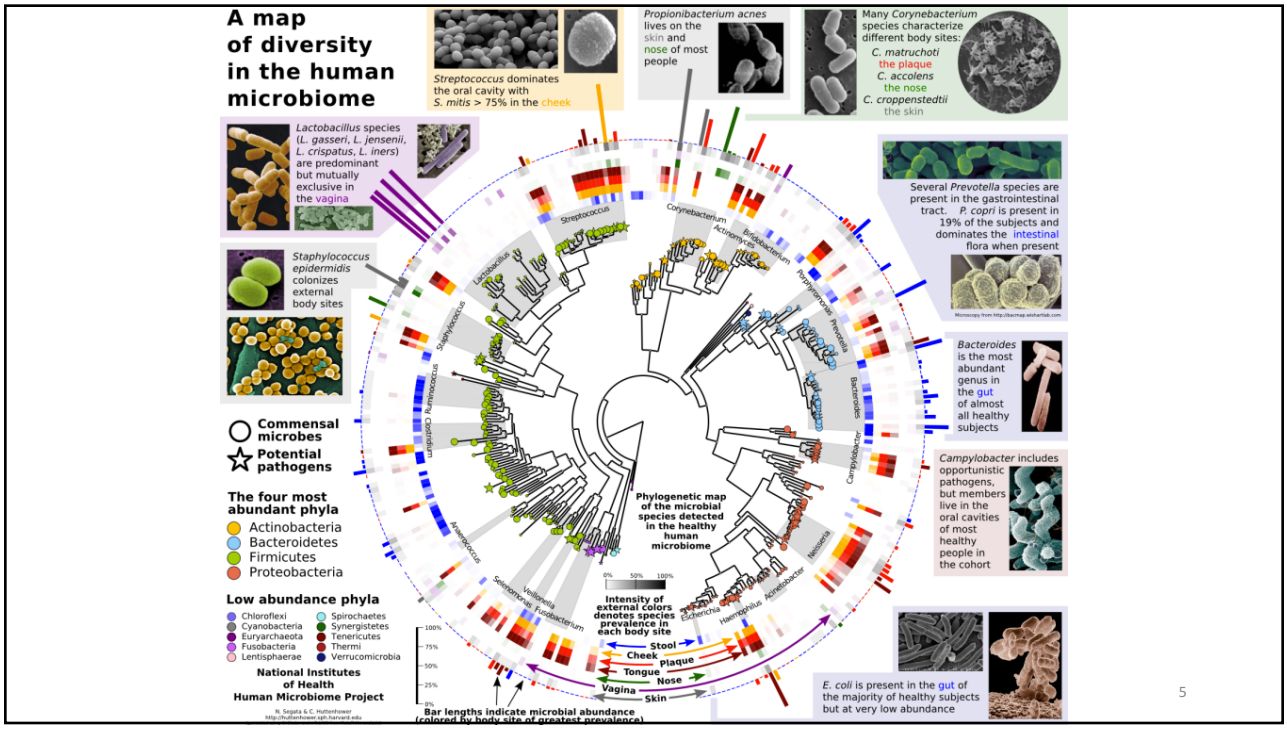


What are microbes?



http://en.wikipedia.org/wiki/Tree_of_life_(biology)

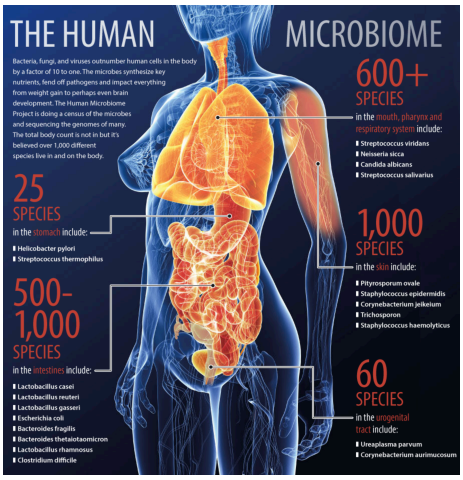
4



5

We are more microbes than we are humans?

- Human shelter 10 trillion microbes (10^{13}) in their gut alone, (we are made of 10 trillion cells).
- ~~Only 1 in 10 cells in your body carries 'your' DNA~~. Recent evidence suggests as many bacterial cells as human.
- It is estimated that there are 1000 species of bacteria living in the human gut.
- Compare also the number of human genes (~25,000) to the number of genes and variants that bacterial communities may carry (~4,000,000, see e.g. doi:10.1038/ncomms3151).

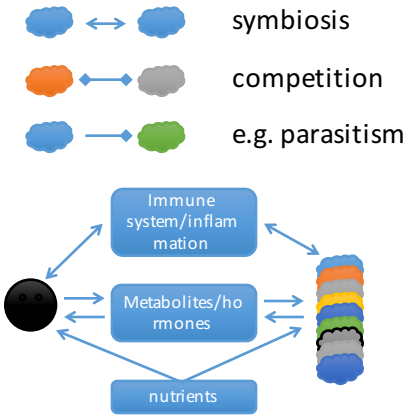


6

Mechanisms for host-microbe interactions

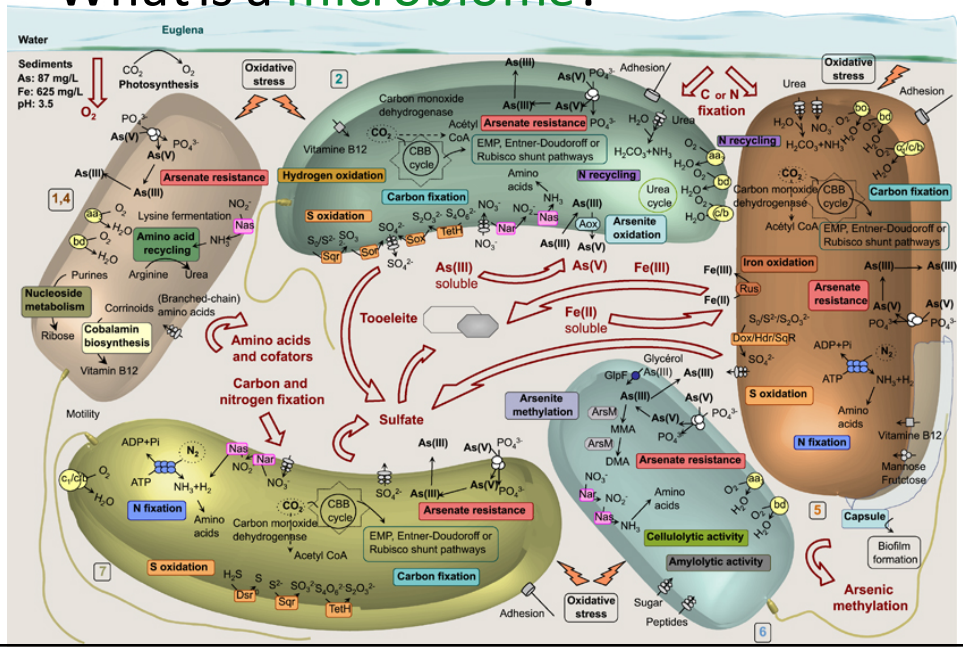
Interact:

- With each other
  - Via regular ecological mechanisms (competition)
- With the host
  - Produce and metabolize hormones and common nutrients
  - Host immune system
- With the environment



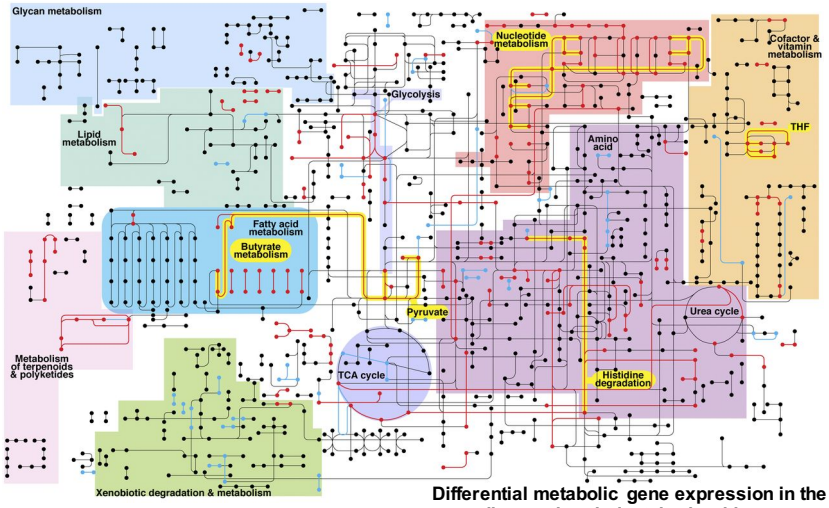
7

What is a microbiome?



8

Mechanisms for host-microbe interactions



Differential metabolic gene expression in the diseased periodontal microbiome.

Peter Jorth et al. mBio 2014; doi:10.1128/mBio.01012-14

9

Understanding the role of the microbiome in human disease, through Koch's postulates.

- Microbiomics effectively generalizes over the Koch's postulates:
  1. The microorganism must be found in abundance in all organisms suffering from the disease, *but should not be found in healthy organisms*.
  2. The microorganism must be isolated from a diseased organism and grown in pure culture*.
  3. The cultured microorganism should cause disease when introduced into a healthy organism.
  4. The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.
- Substitute *microbial community* for *microorganism*
- *How do we culture microbiomes if "it is estimated that as much as 20% to 60% of the human-associated microbiome, depending on body site, is uncultivable" (Genome Res. 2009 Dec; 19(12): 2317–2323)?

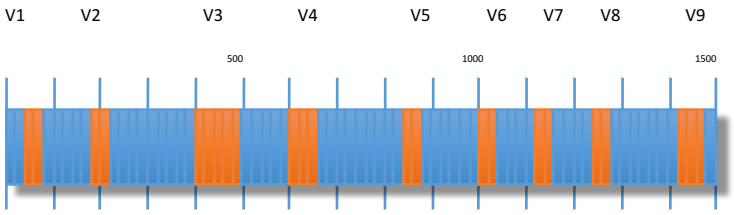
10

Discovery of *Culture Independent* Techniques

- 1980 – rRNA as evolutionary explained - C.Woese *Science*
- 1985 – Polymerase Chain Reaction (PCR) - K. Mullis *Science*
- 1985 – "Universal" Primers for rRNA sequencing - N. Pace *PNAS*
- 1989 – PCR amplification of 16S rRNA gene - Böttger *FEMS Microbiol.*
- 1996 – Large, curated rRNA database (RDP) - Maidak *Nuc.Acids Res*
- 2001 – term "microbiome" coined by Joshua Lederberg

11

Microbial community identification using targeted sequencing regions of 16S rRNA gene



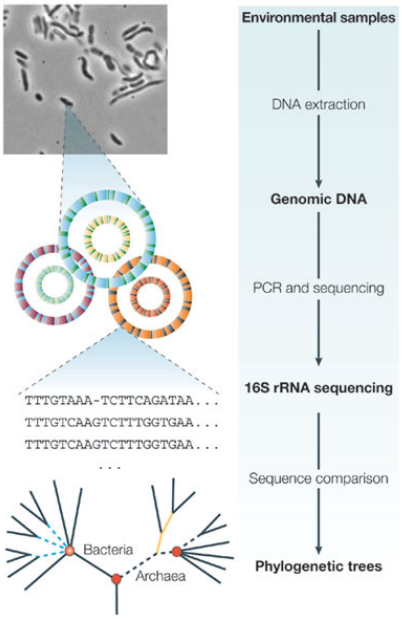
- Location of the hyper-variable regions of the 16S rRNA.
- Current technology does not allow for high-throughput sequencing of the entire 16S gene, only fragments can be sequenced.

12

How do we query a new microbiome?

- Single microbiome:

1. Break all cells, extract all DNA (gDNA)
2. PCR-amplify a universal gene from gDNA
3. DNA sequencing from pool of amplified genes
4. Cluster sequences according to species
5. Count each species and make a tree



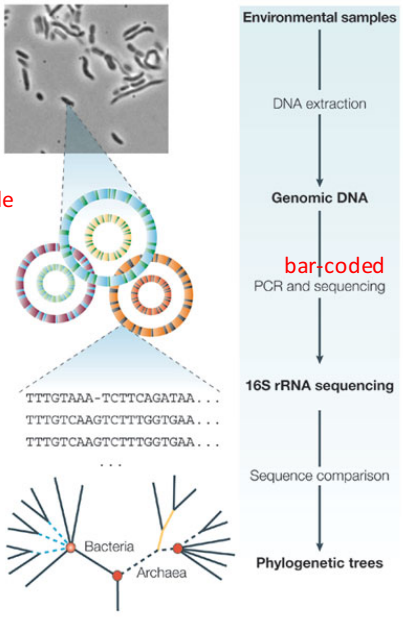
Tringe, S. G., & Rubin, E. M. (2005). Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics*, *6*(11), 805–814.

13

How do we query from many microbiomes??

- Many microbiomes in parallel:

1. Break all cells, extract all DNA (gDNA)
2. PCR-amplify a universal gene from gDNA
   using bar-coded primers, diff code for each sample
3. DNA sequencing from pool of amplified genes
   4a. "De-multiplex" barcode, ID source sample
4. Cluster sequences according to species
5. Count each species and make a tree



14

Functional identification

- Sequencing based
  - Whole metagenome sequencing
  - Whole meta-transcriptome sequencing
  - Custom: e.g. IgA-Seq
- Mass spectrometry based
  - Metaproteomics
  - Metabolomics:
    - Small molecule
    - Glycomics
    - Lipidomics
- Microbial gene expression
  - NanoString codesets

15

Some highlights of human microbiome research

- Pretreatment gut microbiome predicts chemotherapy-related bloodstream infection, *Genome Medicine,* **8**:49 (2016);
- Gut microbiome–derived metabolites modulate intestinal epithelial cell damage and mitigate graft-versus-host disease, *Nature Immunology* **17**, 505–513 (2016);
- Cigarette smoking and the oral microbiome in a large study of American adults, ISME J. 2016;
- The maternal interleukin-17a pathway in mice promotes autism-like phenotypes in offspring, Science 2016, 351(6276):933-9;
- The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes, Cell Host Microbe 2015.

16