



# **Explorations in Data Analyses for Metagenomic Advances in Microbial Ecology**

**13-20 August 2014  
Kellogg Biological Station  
Michigan State University**

# Overview Lecture

- What are our loftiest goals for EDAMAME?
- What is a microbial community?
- Traits of microbial communities
- The "OTU"
- How does Illumina sequencing work?

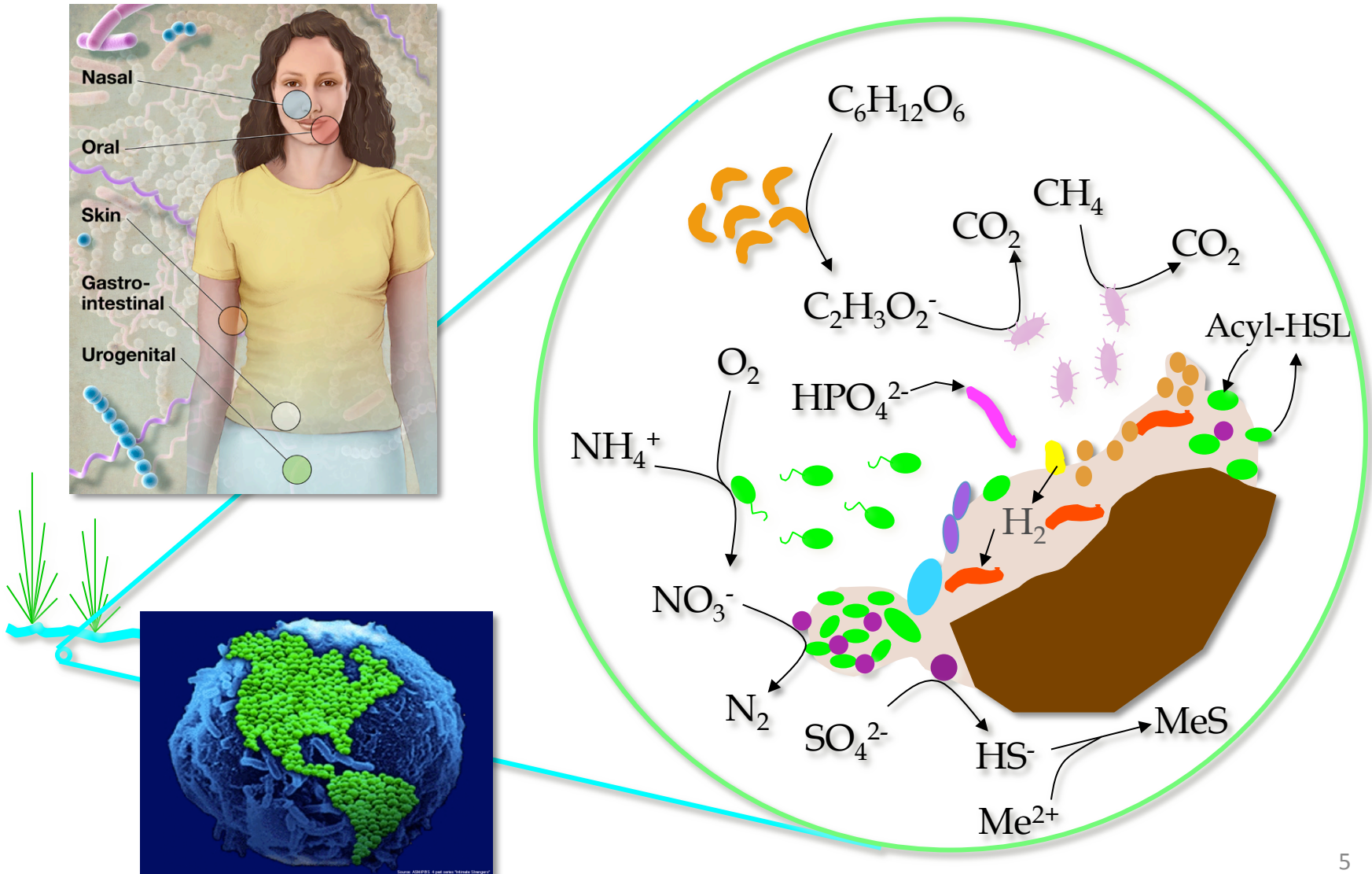
# What are the Burning Questions in microbial ecology?

- Exploration: describing patterns, understanding diversity, discovery (e.g., rare biosphere, dark matter)
- Community structure – function relationships
- Sequencing SOP – active area of research.  
How can we use this technology to answer our burning questions?
- Host – microbe relationships
- ...many more!

# Our goals for YOU

- Be audacious in the face of analyses!
  - Analysis is hard. Have no fear. It is completely normal to struggle.
  - Understand the problem in the pipeline /where the workflow was breaking down
  - Be able to find resources to fix problems
  - Where to find help and how to ask for help optimally
  - Learn how to critique and test others' analyses pipelines

# What is a microbial community?



# What is a microbial community?

- Many taxa (species;  $>2$ )
- Exist in the same locality
- **Interact with each other and/or with the environment**



# The “OTU” operational taxonomic unit

- Species = basic unit of classification
- Defined somewhat arbitrarily
- Typical = 97% sequence identity
  - Originally, identity based on *full length* 16S rRNA gene
  - roughly equivalent to genus level
  - Does not well-distinguish “taxa” for all bacteria (*e.g., Streptomyces*)

# Ecological traits of microbial communities

*Understand the Nature of the Beast. Microbial community data are:*

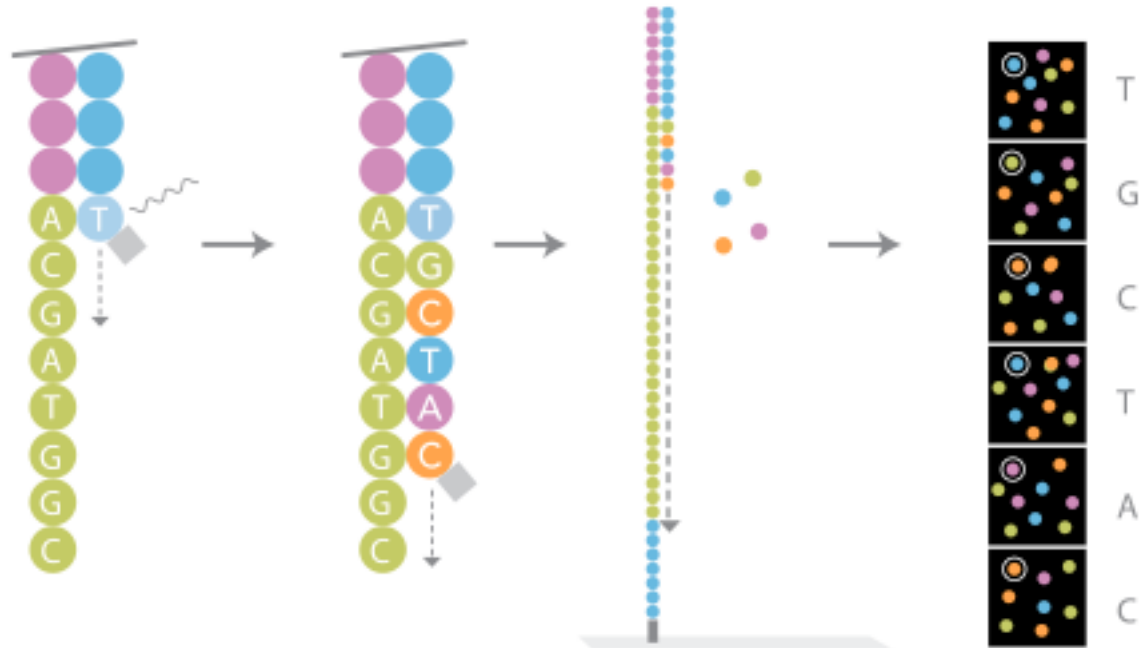
- “Species” rich
- Depend on operational taxonomic unit (OTU) definitions
- Dynamic : sensitive to environmental changes
- Distinctive: even very similar habitats “house” distinct microbial communities (e.g., every human has her own gut community)
- Influenced by dispersal?
- Influenced by gene-swapping (phage, HGT)
- Large proportion of dormant members
- Large proportion of rare members



(A beast, [hyperboleandahalf.blogspot.com](http://hyperboleandahalf.blogspot.com))



# How does Illumina sequencing work?



Illumina: sequencing by synthesis

Short reads : ~125 bp

Lots of data

Informative Video!

[www.youtube.com/watch?v=tuD-ST5B3QA&noredirect=1](http://www.youtube.com/watch?v=tuD-ST5B3QA&noredirect=1)

# Naming Samples Other Stories

Examples of a a good name

20\_A\_T1\_R1 (translation : subject 20, treatment A, timepoint 1, rep1 )

Example : A bad name

Ashley's sample/A

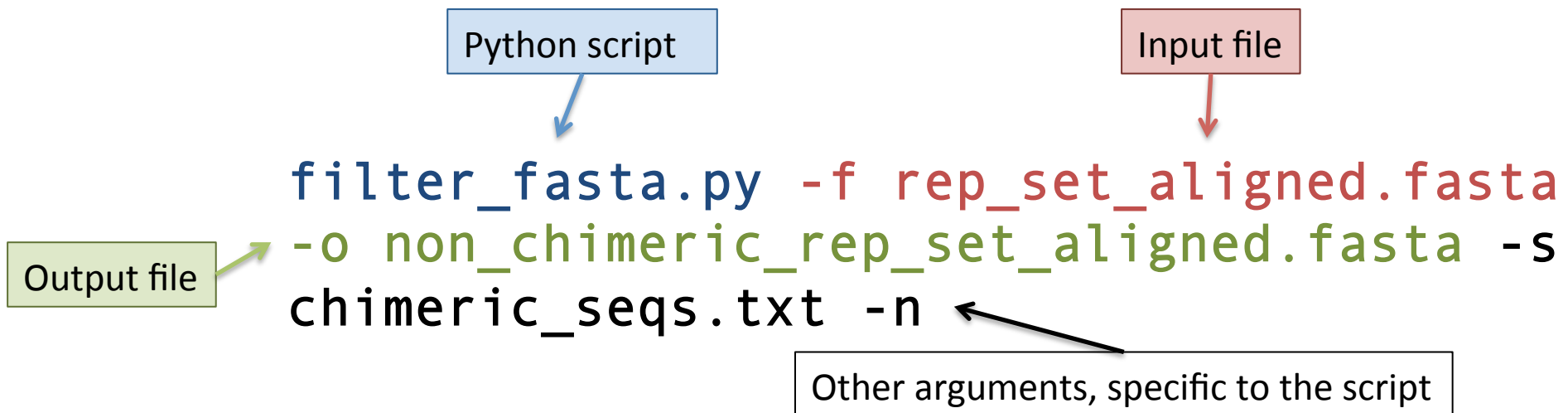
Example : kind of bad names

ALS1, ALS2, ALS3....ALS10, ALS11

Improved:

ALS01, ALS02, ALS03...ALS10, ALS11

# A look at python syntax & common arguments in QIIME



## Other common QIIME arguments

- m** analysis method, metric (sometimes map file)
- t** tree file
- a** alignment template file
- v** verbose = good for troubleshooting
- h** help
- f** force overwrite of an existing directory