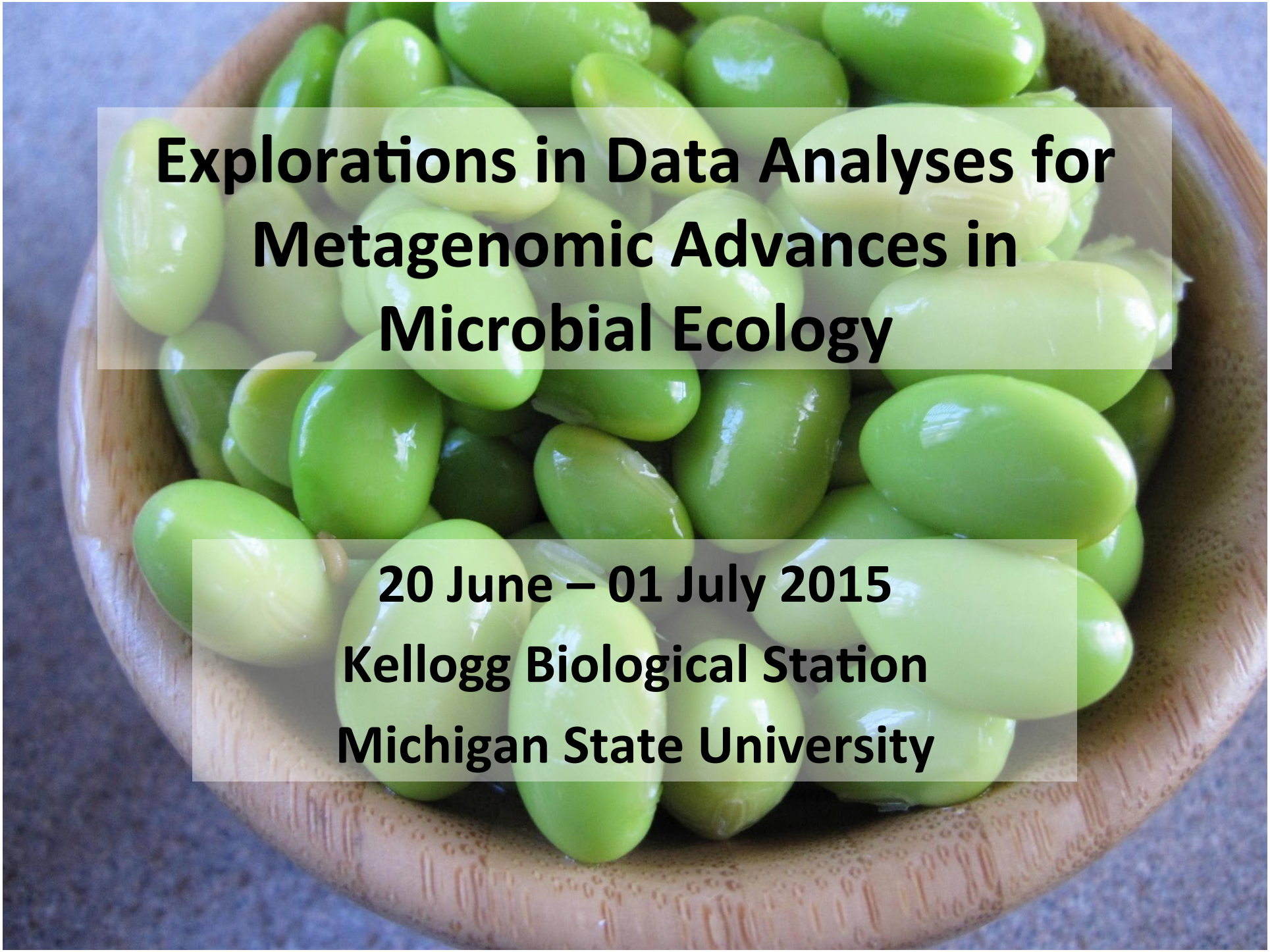
GOOD MORNING!

- Please make a nametag (back table)
- Take a linux/unix handout (back table)
- Take 1 **blue** and 1 **pink** stickie (back table)
- If you haven't already, we need you to fill out the course pre-assessment. Follow the emailed link to complete.
  - It is totally okay if you don't know the answers!



Explorations in Data Analyses for Metagenomic Advances in Microbial Ecology

20 June – 01 July 2015

Kellogg Biological Station

Michigan State University

Overview Lecture

- Our goals for EDAMAME
- Course logistics: Schedule and expectations
- Getting warmed up: What is a microbial community?
  - Traits of microbial communities
  - The ”OTU”
- Our Tutorial Dataset for the Week

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

Our goals

- Provide a safe & welcoming place to learn
- Lots of help from many people with different backgrounds – help each other out. Share your expertise and discuss challenges/troubleshoot together
- Many guests to provide insight into different tools and research areas
- Research specific help when possible

Our expectations

- Ask lots of questions, and try really hard to get all you need to execute analyses independently when you return to your group
- Don’t be afraid to ask for help when you need it! (we all have to do this sometime)
- Acceptance and patience (in both directions)

Our hopes

- Enthusiasm!
- Engagement!
- Fearlessness!
- Fun!

Our Learning Goals

- Overarching Goals are posted the wiki: https://github.com/edamame-course/2015-tutorials/wiki
- More specific objectives

A Snapshot of our action packed days

- 7-8: Breakfast. Head's up: They close promptly.
- 9:00 am-ish – Lecture
- 10:30 am – Morning Tutorial
- 12-1 pm – Lunch
- 1:15 pm – Afternoon Tutorial
- 4 pm – Break
- 5-6:30 – Dinner
- 8 pm – Guest lecture
- 9 pm- ? – Social time, fire pit

Introductions

- Prof. Ashley Shade - MSU
- Dr. Josh Herr – MSU… to Nebraska!
- Dr. Tracy Teal – Data Carpentry

- Dr. Jin Choi - Iowa State
- Siobhan Cusack - MSU
- Dr. Sang-Hoon Lee – MSU & Korea University
- Jackson Sorensen - MSU
- Paul Wilburn –KBS local – ask him for the insider's scoop, and tours!

Our Esteemed Guest Lecturers

• Pat Schloss Team	University of Michigan
• Jay Lennon	Indiana University
• Stuart Jones	University of Notre Dame
• Jim Tiedje	Michigan State University
• Jim Cole	Michigan State University
• Qiong Wang	Michigan State University
• Vince Young	University of Michigan
• Ariane Peralta	East Carolina University
• Vincent Denef	University of Michigan
• Sarah Evans	Kellogg Biological Station
• Matt Scholz	MSU HPCC

Food and drink

- Most meals will be at the KBS dining hall. Over the weekend they will be closed for a few meals. We will arrange for group meals; please do your part to chip in – we'll keep the cost as economical as possible.
- Snacks !

- We can also make group arrangements to head to "town" – check with one of the TAs. You might want to head to the market on your own. Kalamazoo is not too far away.

Recreational stuff

- Some options at KBS are volleyball, frisbee, bocci ball, swimming
- You may have to check with the KBS office for some of the options.
- There are good places to run, to swim, to hike, to bike, to fish, to boat
- There is also a few laundry rooms; we have a few pods of detergent

WIFI

- MSUnet Guest 3.0

Red/Green stickies...

- Red sticky means “I am in need of help...”; Green means “I’m doing ok”

- You don’t have to use them all the time, but we may ask some of you to put them up so we can get an assessment of where we are at as a group.

- Before Lunch and before afternoon break, we will collect “minute notes” for anonymous feedback about how you’re doing. Red for a question/complaint and Green for a complement

Web and social media

- All the tutorials are on our github wiki and will be updated as we go: https://github.com/edamame-course/2015-tutorials/wiki/Schedule
  - Bookmark it!
  - Course webpage not updated this week
- Etherpad for group note taking (link on wiki home page): https://edamame2015.etherpad.mozilla.org/1?
- Course email list – please let me know if you do NOT want your name and email distributed to the group
- Tweet #edamame2015
- Blog: MO BIO

Code of Conduct

Please read the course code of conduct:

- https://github.com/edamame-course/docs/blob/gh-pages/extra/edamame_code_of_conduct.md

- Bottom line:
- Let's all be nice to one another. Disrespectful conduct is grounds for immediate dismissal.

Our Support comes from

- The BEACON Center for the study of evolution in action (MSU)
- MO-BIO (t-shirts!)
- Amazon Web Services
- MSU Office of Vice President for Research and Graduate Studies
- Kay Gross, director of KBS
- … Thank our sponsors! tweet, blog, etc

Mo BIO

Blogging opportunity!

Emelia DeForce is REALLY excited to have EDAMAME guest post!

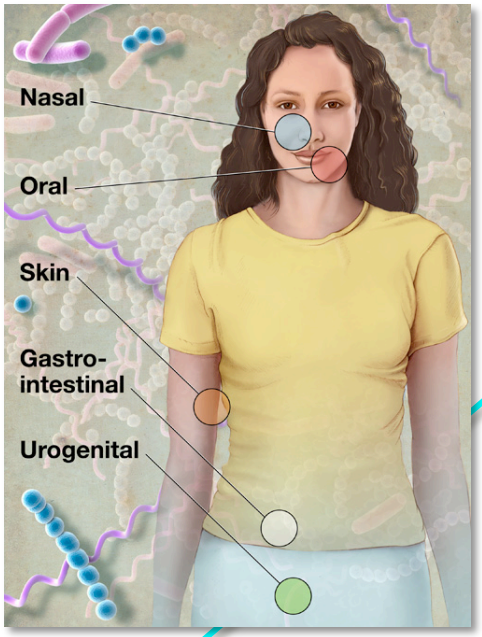
Any questions or comments?

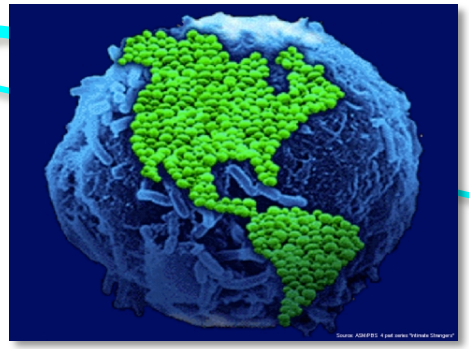
Let’s do it.

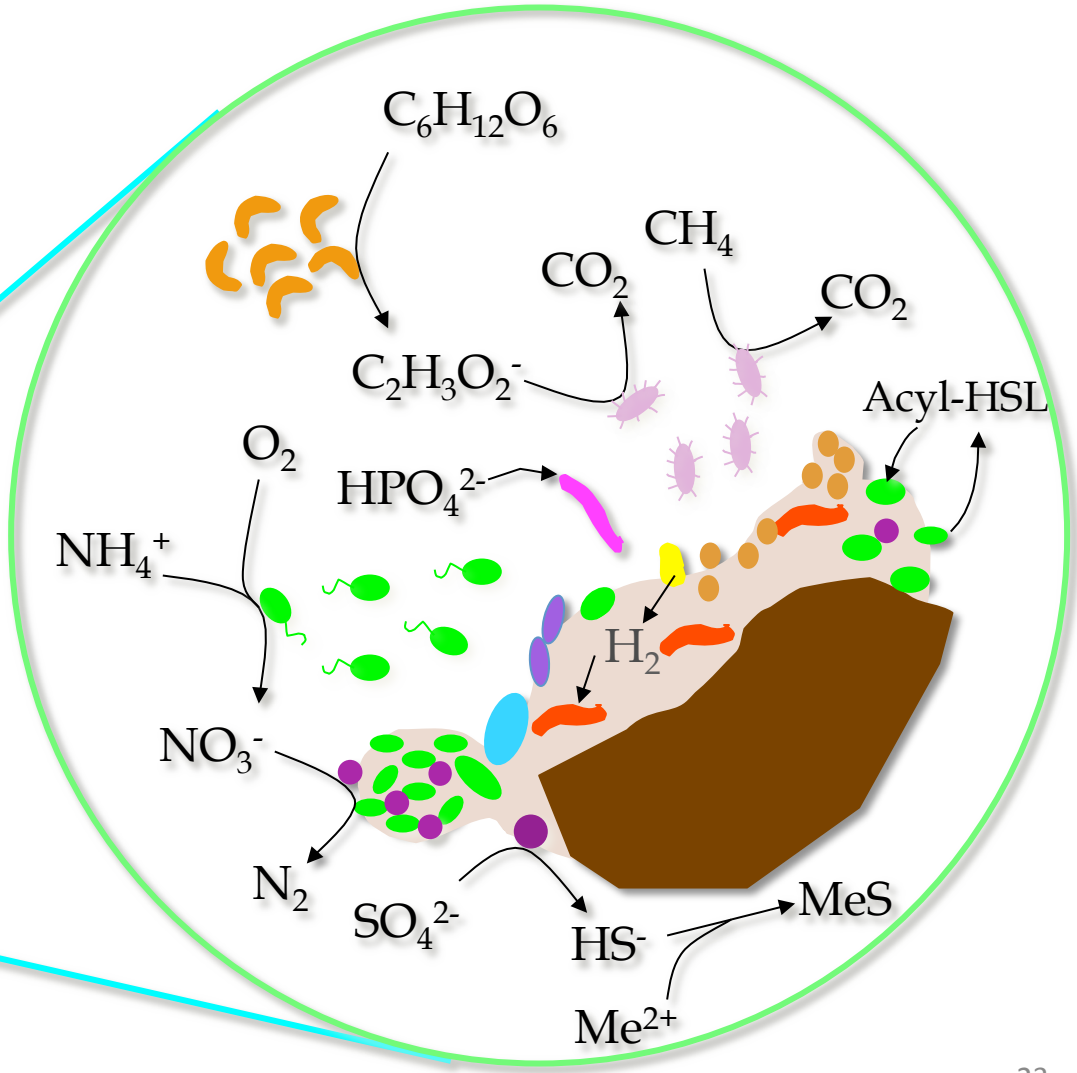
What are the Burning Questions in microbial ecology?

- 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!

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**

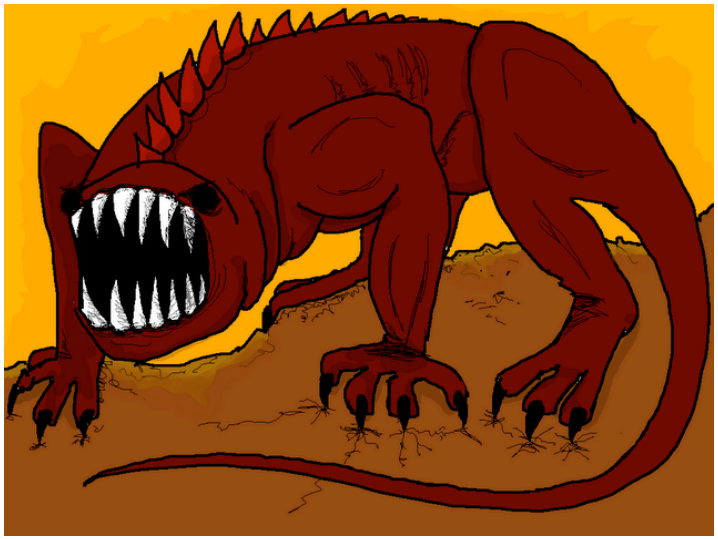
The "OTU"
operational taxonomic unit

- 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)

What are our options for sequencing and analysis?

- What sequencer?
- If amplicon, which gene? Which variable region?
- What quality control options?
- Defining OTUs
- Describing communities
- Testing hypotheses
- Visualizing results

Introduction to our Tutorial Dataset

- Motivation : get an idea of a complete analysis from start to finish
- Everyone working on the same dataset helps the instructors maximize their time when helping students to troubleshoot
- There is time dedicated to apply what you learned on the tutorial dataset to your own dataset

Centralia, PA: burning since 1962

Nowiknow.com

Key Questions

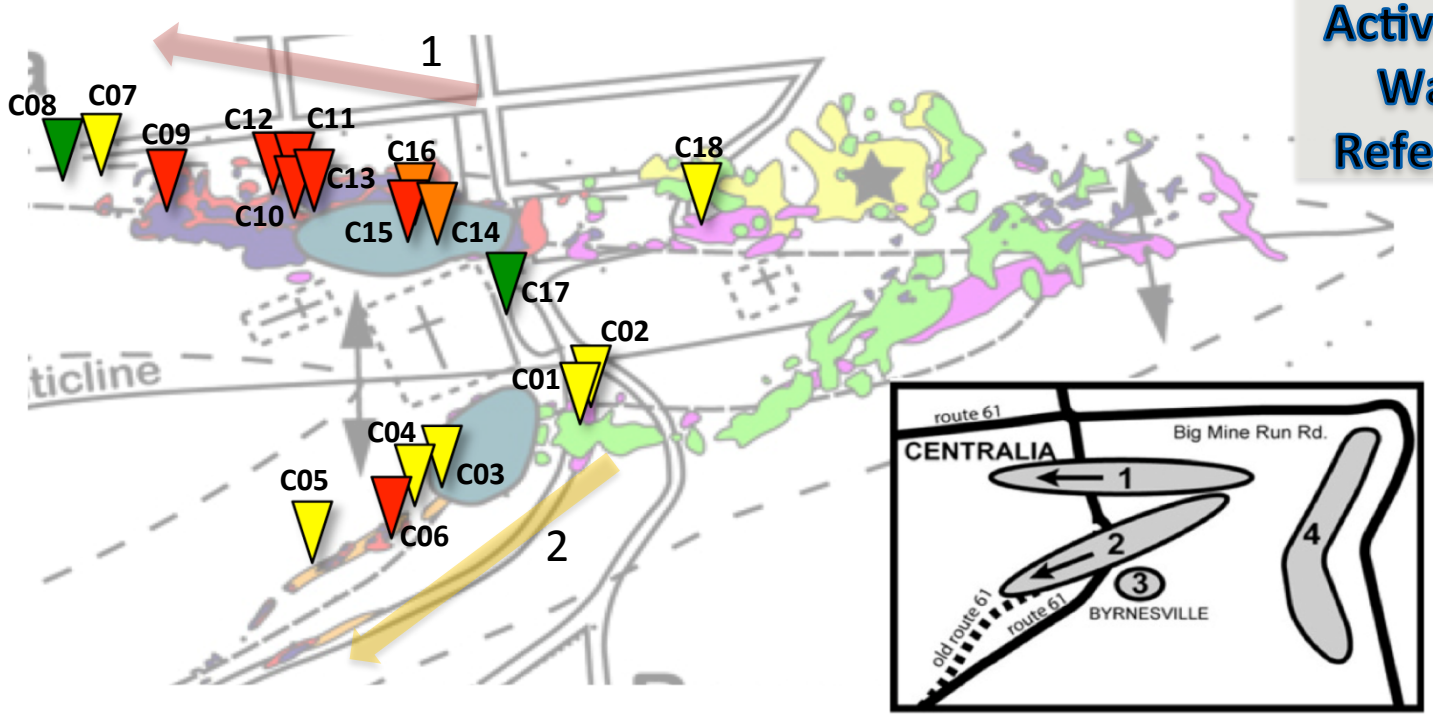
- What is the diversity and structure of microbial communities in Centralia soils?
- Do temperature/geochemical gradients or historical fire activity explain differences in community structure?
  - Stability: resistance and resilience
- What are the lifestyles of organisms in fire-affected sites?
  - Thermophiles: dormancy strategies, cellulose degradation, thermal tolerance and stress responses
  - Bioremediative organisms: heavy metals and other coal combustion products
  - Novel antibiotic producers?
- Is Centralia a site of novel microbial diversity?

Sample collection

Sieved (4 mm pore)

Underground coal mine fire

Recovered : 7
ActiveVent : 7
Warm : 2
Reference : 2



October 2014 sampling sites

Reference
Recovered
Warm
Active vent

Historical fire activity

1962, origin
1972
1980
1982
1983
1999
2002
2007
2009

Figure 1. Sampling sites

Sampling Period: Oct. 5-6, 2014

Datasets for the week

- 18 soils (0-20 cm cores) along active fire fronts 1 and 2
  1. Illumina paired-end V4 16S rRNA amplicon sequencing on each of 3 replicate DNA extractions. 54 total amplicon samples
  2. Soil chemistry and contextual data on each core – 18 total soils with measurements
  3. Metagenome sequencing on the DNA extracted from the soil of an active vent sample, Cen13 (temperature = 57 C) – 1 sample

Other things you should know about these datasets/ analyses

- Sequenced VERY deeply
  - We will be working with small datasets subsampled randomly from the full datasets
  - Subsampling is key for developing a workflow/ troubleshooting scripts
- We will be working entirely on "the cloud" using Amazon
- Please attribute the EDAMAME tutorials if you use or share them. We have a CC-BY license.