

Explorations in Data Analyses for Metagenomic Advances in Microbial Ecology

10 July– 20 July 2016
Kellogg Biological Station
Michigan State University

GOOD MORNING!

- Please make a nametag (back table)
- Take a linux/unix handout (back table)
- Take 1 **blue** and 1 **pink** stickie (back table)

Overview Lecture

- Our goals for EDAMAME
- Course logistics: Schedule and expectations
- Getting warmed up: Why are we here? What is a microbial community?
- 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/2016-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/Tutorial
- 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 / activity
- 9 pm- ? – Social time, fire pit

Introductions – Your Instructors

- Ashley Shade – Michigan State
- Adina Howe – Iowa State
- Tracy Teal – Data Carpentry

Your FEARLESS TAs

Shade Lab

- John Chodkowski
- Siobhan Cusack
- Taylor Dunivin
- Jackson Sorensen

Howe Lab

- Jinylung Choi
- Phillip Colgan
- Jared Flater
- Fan Yang

Our Esteemed Guest Lecturers

- Pat Schloss University of Michigan
- Stuart Jones University of Notre Dame
- Jim Tiedje Michigan State University
- Jim Cole Michigan State University
- Rich Lenski Michigan State University
- Pat Bills MSU HPCC

Food and drink

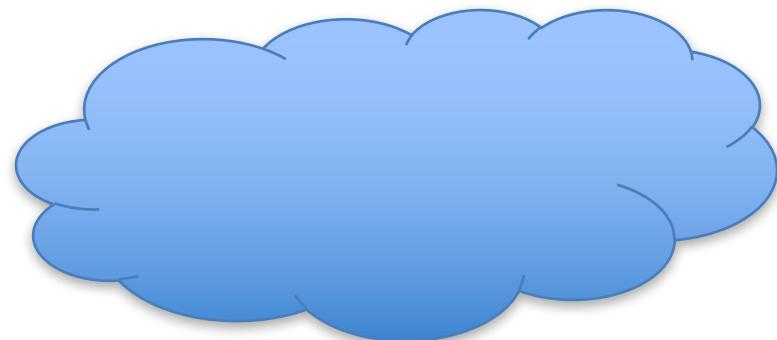
- Most meals will be at the KBS dining hall. Over the weekend they will be closed for a few meals. We will arrange for group meals; please do your part to chip in – we'll keep the cost as economical as possible.
- Snacks !
- We can also make group arrangements to head to “town” – check with one of the TAs. You might want to head to the market on your own. Kalamazoo is not too far away.
- A note on adult beverages

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 & Cloud

- MSUnet Guest 3.0
- We'll be working on the CLOUD! Amazon EC2 for many tutorials.



Learning Resources : Etherpad

- <https://public.etherpad-mozilla.org/p/EDAMAME-2016>
- Etherpad for shared note taking – there is a designated SCRIBE for each session, but everyone can add.

Learning Resources: Stickies!

- Put them on your lap top
- 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/Blue for a complement

For every tutorial, there will be

- Lead instructor
- Co-instructor
- Scribe – takes etherpad notes, updates tutorials immediately
- Several TAs to assist in the room and respond to red stickies

Learning Resources: Mendeley Group!

- <https://www.mendeley.com/groups/4688421/edamame/>
- Ask to be invited to drop papers in
- Also, Scribes will be updating LIVE

Web and social media

- All the tutorials are on our github wiki and will be updated as we go:
<https://github.com/edamame-course/2016-tutorials/wiki/Schedule>
- Webpage : edamamecourse.org
- Course email list – please let me know if you do NOT want your name and email distributed to the group
- Tweet #edamame2016 !! (Unofficial annual microbial haiku contest on Twitter)
- Blog: MO BIO Culture Dish
- Our friends at Notre Dame!

Code of Conduct

Please read the course code of conduct:

- https://github.com/edamame-course/docs/blob/gh-pages/extr/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 National Institutes of Health (R25)
- Institute for Cyber-Enabled Research (Snacks!)
- MO-BIO (t-shirts!)
- Amazon Web Services (computing!)
 - ... Thank our sponsors! Tweet! blog! Show them some love!

Mo BIO

Blogging opportunity!

Emelia DeForce is REALLY excited to have EDAMAME guest post!

Any questions or comments?

Let's do it.

Why are we here?

OUR MICROBIAL PLANET

MICROBES—life forms too tiny to see—play a surprisingly large role in life on Earth. Microbes are everywhere, and they do a lot of good for human health and our planet. In fact, disease-causing microbes make up only a very tiny fraction of the millions of types of microbes. Microbes... .

Think microbes are bad guys? Think again.



Visit www.nationalacademies.org/microbes to learn more!

This poster was derived from the National Research Council report *The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet* (2007). Reports from the National Academies are available from the National Academies Press, 500 Fifth St., NW, Washington, DC 20001; 800-624-6242; <http://www.nap.edu>. Reports are available online in a fully searchable format.

Support for this publication generously provided by the President's Circle Communications Initiative of the National Academies. Artwork by Nicole Ringer Fuller (www.says-art.com). Top photo at left by Tina Cavallo, University of Hawaii at Manoa. Bottom photo by Jeff Miller, University of Wisconsin-Madison.

Copyright 2008 by the National Academy of Sciences. All rights reserved.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The nation turns to the National Academies—National Academy of Sciences, National Academy of Engineering, Institute of Medicine, and National Research Council—for independent, objective advice on issues that affect people's lives.

www.national-academies.org

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?

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

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)

Introduction to our Tutorial Dataset

- Motivation : get an idea of a:
 - complete analysis from start to finish
 - the different types of complementary data that can be generated and how they may inform one another
- Everyone working on the same dataset helps the instructors maximize their time when helping students to troubleshoot
- There is Independent Study dedicated to apply what you learned on the tutorial dataset to your own dataset

Centralia, PA: burning since 1962





Centralia, Pennsylvania: An opportunity to investigate eco-evolutionary microbial dynamics in response to a press disturbance

Severe
disturbance

>80 C surface
temps., coal
combustion
pollutants

Ongoing

Active since 1962, will
continue 100+ years

Multi-generational
impact

Relevant for microbial
community turnover AND
evolutionary processes

Multi-
generational
recovery

Cools to temperate
conditions, pollutants
may remain

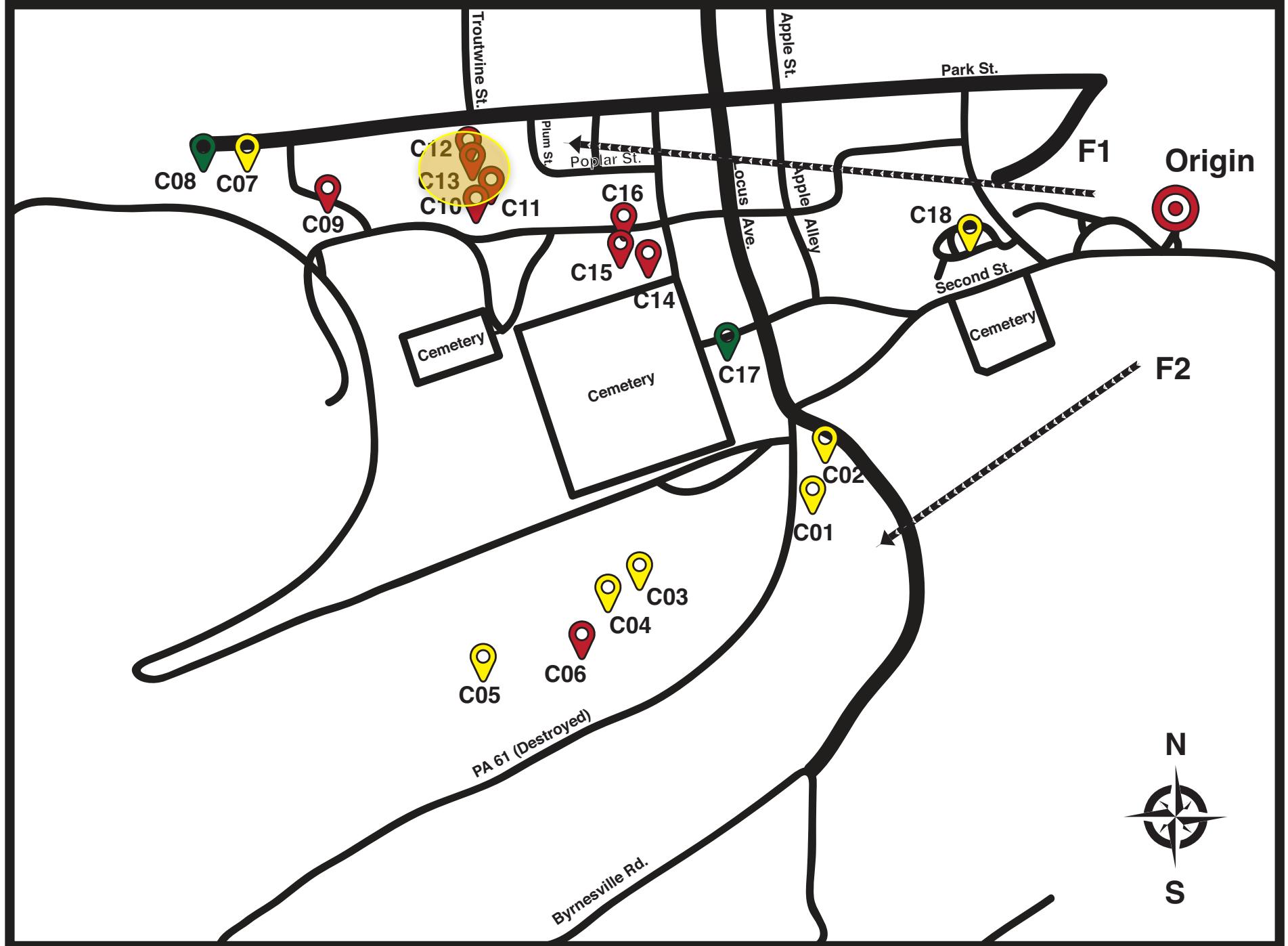
Some sites have been
cool already for decades

Centralia Research Themes 1

- **Understanding eco-evolutionary dynamics**
 - What role does **dormancy** have for **adaptation and diversification**?
 - What are the implications of temporary (but inter-generational) **extreme environmental filtering** for diversification?
 - What are the **origins of extremophiles** within non-extreme environments?
- **Discovering functional and phylogenetic diversity**
 - How do microorganisms **adapt to and cope with pollutants**?
 - Does the diversity in Centralia expand the breadth of knowledge about **terrestrial thermophiles**?

Centralia Research Themes 2

- Predicting community stability (resistance and resilience) to press disturbances
 - How does community structure, especially rare taxa and their seed bank, influence robustness?
 - How does press disturbance determine community assembly and succession?
 - What are the costs and benefits of general versus specialized stress responses to disturbances?
- Deciphering microbial interactions during disturbance transitions
 - Do member interactions change during disturbances, and how?
 - Do changed interactions recover?



Sample collection



Sieved (4 mm pore)

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.

Icing

- Paper draft on amplicon analysis – comments welcome!
- Reproduce the statistical analysis in R from the paper (advanced topics)
- Check out the amplicon analysis pipeline on the ShadeLab GitHub group
- Basically, get an comprehensive, open look at how we approach an analysis