Learning Objectives for EDAMAME 2015

Lead Instructor: Ashley Shade, shadeash@msu.edu
Michigan State University

1. **Increase computing literacy**
    1. Determine in what situations a remote session is useful
    2. Start, detatch, and attach to a tmux session
    3. Understand what the shell is, how to access it from your computer, and how to use it.
    4. Navigate in around in a Unix file system to view and manipulate files
    5. Find help and other resources for using the Shell and a Linux/Unix file system

2. **Develop proficiency in cloud computing**
    1. Select a community AMI image to initiate an Amazon EC2 instance
    2. Make and use a "key" for secure EC2 connection
    3. Use ssh to connect to a running Amazon EC2 instance
    4. Use scp to transfer files between a personal computer and an EC2 instance
    5. Use wget to download data from external storage to an EC2 instance
    6. Install axillary software on an Amazon EC2 instance
    7. *Understand what a high performance computing cluster is, and what resources it can provide for sequence analysis*
    8. Find help and other resources for Amazon EC2 cloud computing

3. **Analyze microbial amplicon sequences**
    1. Understand the information provided in Illumina "raw" fastq files
    2. Install axillary software on a QIIME EC2 instance
    3. Use FastQC to assess the overall quality of raw sequencing data, and determine the parameters that are important specifically to metagenomes
    4. Install auxillary software on the QIIME Amazon EC2 image
    5. Subsample a large amplicon dataset for workflow development and tutorial testing
    6. Assemble paired-end reads
    7. Execute a shell script to automate a process
    8. Explore input and output files for QIIME workflows and scripts
    9. Understand the structure and components of a good mapping file
    10. Move sequences into the QIIME environment from an outside tool using "add_qiime_labels.py"
    11. Obtain summary information about sequence files (fasta, fna, fastq)
    12. Define operational taxaonomic units (OTUs)
    13. Execute a QIIME workflow script, and understand the separate steps in the workflow
    14. Align sequences, assign taxonomy, and build a tree with representative sequences from OTU definitions
    15. Extract summary information from a biom OTU table
    16. Subsample a dataset to an even sequencing coverage across all ccommunity observations
    17. Calculate and visualize within-sample (alpha) diversity
    18. Calculate resemblance matrices from an OTU table
    19. Visualize comparative diversity across a priori categorical groups

20. Convert .biom formatted OTU tables to text files for use outside of QIIME
21. *Develop a proficiency with mothur for amplicon analysis*
22. Find help and other resources for microbial amplicon analysis

4. Analyze microbial shotgun metagenome sequences

1. Recognize that there is not one tool for an entire metagenome analysis
2. Assess the quality of "raw" metagenome data
3. Trim raw reads to meet quality standards using fastx-toolkit
4. Use digital normalization to remove redundant data
5. Trim/Filter out errors from sequences by identifying low coverage kmers in high coverage areas
6. Understand the limitations and strengths of metagenome assembly
7. Assemble a metagenome with MEGAHit, cite alternative assemblers available
8. Summarize and assess the assembly quality
9. *Perform targeted gene assembly from a microbial metagenome dataset using Xander*
10. Find help and other resources for microbial metagenome analysis

5. Apply ecological statistics to analyze and interpret microbial sequencing data

1. Install R and R studio on a personal computer
2. Understand what an R package is and how to install it
3. Import, manipulate, and export data tables using R
4. Differentiate data modes in R, including numeric, complex, logical, character modes
5. Differentiate data structures in R, including vector, matrix/array, list, and data frame structures
6. Transform and standardize community structure data
7. Calculate pairwise resemblance matrices in R to summarize relationships between community structure or taxa occurrences
8. Visualize community composition using ordination and clustering
9. Test hypotheses about differences in community structure across samples or sites
10. Identify correlations between community composition, environmental conditions, and spatial configuration of sites
11. Find help and other resources for working with R

6. Access resources provided by public sequence databases

1. *Understand what the RDP is, how to access it on the web, and what analysis tools it offers*
2. Make a local BLAST database and execute a query of it
3. List at least two sequence databases or repositories that are accessible to you
4. Identify features of public sequence databases that are useful to your research