IB516 Analytical Workflows Project Presentation

Jared Freedman

Background

- Some experience in R
  - Mostly using scripts given to me by an advisor, or using R as a calculator
- I am in the beginning stages of establishing an eDNA metabarcoding pipeline, and am in this class to learn how to create a manageable and reproducible project structure

Project Summary

- eDNA sampling of tributary and main-stem sites in the Grand Canyon
  - Assess the genetic connectivity of these communities
  - Metabarcoding of COI gene fragment using a degenerate primer targeting aquatic macroinvertebrates
- Are organisms dispersing into the mainstem from the tributaries? Dispersal between tributaries? Are there latitudinal trends in community composition?

Project Goals

- Create an eDNA sequence processing pipeline that is effective, modular, and reproducible for other projects
  - Including different primers or fieldwork sampling strategies
- Sequencing reads from Illumina MiSeq platform will need to be:
  1. Demultiplexed
  2. Primer sequences removed
  3. Remove singleton reads or reads with high expected error rates
  4. Group remaining reads into Operational Taxonomic Units (OTUs)
  5. *Align OTUs with published invertebrate COI fragments*
- Utilize published R scripts for completing these tasks

Anticipated Challenges

Anticipated Challenge	Possible Solution
Not receiving sequencing data in time	Find and use a different dataset (where to find this?)
Creating a new pipeline from scratch	Understand what each section of the pipeline is doing, don’t just plug and play
Not bogged down with “practice” scripts and output files	Intentional and organized data management
Creating something that is intelligible to other collaborators/advisors/mentees	Take the time to comment and deliberately structure the code/pipeline