Bioinformatics module: Community ecology using 16S amplicon sequencing

By: Jacob Heil

**Objectives**

Day 1

1. Introduction to community ecology and how amplicon sequencing can be used to quantify whole microbial communities.
2. Use R script to translate raw reads into ASVs and produce a taxonomic tree
3. Complete worksheet

Day 2

1. Introduction to community statistics
2. Use R script to analyze a microbial community
3. Group project
4. Complete worksheet

**Preparation**

1. Data: For day 1 you will be provided with a set of raw 16S amplicon sequencing reads. For day 2 you will be provided with a set of whole community samples and metadata.
2. All files can be found in the module folder provided for you
3. Read the paper provided in the folder prior to class and answer relevant worksheet questions
4. Open R scripts and install all packages prior to class

A close up of a flower

Description automatically generated

**Background**

Study system

Microbial community ecology reveals the hidden world microbes and has important applications for understanding microbes, their hosts, and by extension whole ecosystems. In this module we will be looking at microbial communities living in pitcher plants. Pitcher plants have modified leaves and house liquid to dissolve prey and acquire nutrients. The microbes living in this liquid have important functions for the plant and constitute an ecological community. To better understand this community, we can use whole community amplicon sequencing and ecological community analyses. This analysis will provide insight into the factors that drive the composition and diversity of these communities.

Workflow

The workflow begins with raw amplicon sequencing reads and produces quantified microbial communities, measurements of diversity, measurements of similarity between communities, models of community change in response to variables, and more. First we extract DNA from whole pitcher communities in the lab. Next the extracted DNA is sequenced using next-generation amplicon sequencing. When the sequences come back they are demultiplexed, sorting sequences by their original sample, and resolved into amplicon sequence variants (ASVs) using software, for example we will be using the dada2 package in R. ASVs are an approximation of species. Each ASV is a unique DNA sequence found in our data after the process of DADA2. There are other ways to approximate species from sequence data, for example operational taxonomic units (OTUs), which are groups of sequences that share a certain percent similarity. We will use ASVs because they are more precise and represent all unique sequences in the data.. When the ASVs have been identified we can use the sequences and their quantities to perform statistical community analyses. See the flowchart below for an example of the bioinformatics steps for analyzing amplicon sequencing data.

A screenshot of a cell phone

Description automatically generated