**Overview of generic* amplicon workflow**

*This is generic; specific workflows can vary on the order of steps here and how they are done.

1. **sequencing facility**
   - ATCGGATCG...
   - HISEQ2500:282:1:1101:1220:1944 1
   - ATCGGATCG...

2. **fastq files**
   - demultiplex (split samples by barcodes)
   - Some tools:
     - sabre
     - fastx_demux (usearch/vsearch)
     - idemp
     - fastx barcode splitter (fastx-toolkit)

3. **quality filter/trim (remove adapters/primers)**
   - fastqc/multiqc
   - plotQuality (dada2)
   - Some tools:
     - trimmomatic
     - fastq_filter (usearch/vsearch)
     - bbduk.sh (bbtools suite of tools)
     - filterAndTrim (dada2)

4. **fasta files**

5. **dereplication**

6. **chimera removal**

   - might be done by sequencing facility
   - >HISEQ2500:282:1:1101:1220:1944 1
   - ATCGGATCG...

**Fastq files vs. fasta files**

- **fastq files**
- **fasta files**

**generate OTUs**

- **dada2**
- **qiime2**

**count table**

<table>
<thead>
<tr>
<th></th>
<th>Sample_A</th>
<th>Sample_B</th>
<th>. . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq_1</td>
<td>0</td>
<td>428</td>
<td></td>
</tr>
<tr>
<td>Seq_2</td>
<td>306</td>
<td>323</td>
<td></td>
</tr>
<tr>
<td>Seq_3</td>
<td>217</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>. . .</td>
<td>. . .</td>
<td>. . .</td>
<td>. . .</td>
</tr>
</tbody>
</table>

**encode files**

- **taxonomy**
- **resolve ASVs**

**Analysis**

- **dissimilarity metrics**
- **ordination**
- **hierarchical clustering**

- **Alpha diversity**
  - e.g. dissimilarity metrics
  - e.g. richness
  - e.g. evenness
  - e.g. diversity

- **Beta diversity**

**Some tools that provide whole workflows:**

- **dada2** runs within R (ASVs)
- **usearch/vsearch** runs at the command line (ASVs and OTUs)
- **mothur** runs at the command line (OTUs only currently)

**qiime2** provides a multi-interface environment that employs processing tools like those above, infrastructure for easily documenting all processing performed, and interactive visualizations.

*astrobiomike.github.io*