Overview of generic* amplicon workflow

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

When working with your own data you should never follow any pipeline blindly. There can be critical differences based on your data.

sequencing facility → fastq files → demultiplex (split samples by barodes) → quality filter/trim (remove adapters/primers) → fasta files

might be done by sequencing facility

@HISEQ2500:282:1:1101:1220:1944 1
ATCGGATCG...
+
<G.<G<AGGII...

Some tools:
- sabre
- fastx_demux (usearch/vsearch)
- idemp
- fastx barcode splitter (fastx-toolkit)

fastqc/multiqc
plotQuality (dada2)

Some tools:
- trimmomatic
- fastq_filter (usearch/vsearch)
- bbduk.sh (bbtools suite of tools)
- filterAndTrim (dada2)

>HISEQ2500:282:1:1101:1220:1944 1
ATCGGATCG...

dereplication → chimera removal → generate OTUs / resolve ASVs

Usually within a workflow, e.g.:
- dada2
- usearch/vsearch
- mothur

dada2 ●
qiime2 ▲

Standard outputs

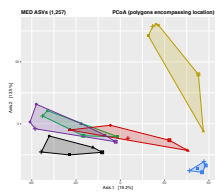
count table
fasta file
taxonomy

	Sample_A	Sample_B	. . .
Seq_1	0	428	. . .
Seq_2	306	323	. . .
Seq_3	217	1	. . .
. . .	. . .	. . .	. . .

Some tools:
- most workflows
- SEPP/TIPP
- RDP classifier

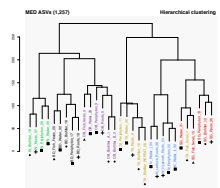
Analysis ▲

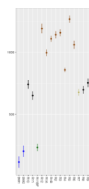
Some tools:
- phyloseq
- Breakaway
- DivNet
- CORNCOB
- SpiecEasi
- DESeq2



Beta diversity

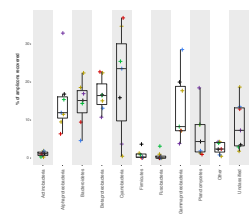
e.g. dissimilarity metrics
ordination
hierarchical clustering





Alpha diversity

e.g. richness
evenness
diversity



Taxonomic summaries

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