**Module 1: Read QC**

For read QC, I use BBTools. First adaptor trim w/ BBduk, then merge reads with BBMerge, then quality trim with BBduk.

Parameters in clean\_reads.sh; requires fasta of adaptors (adaptors\_illumina.fa)

Run for each sample:

./clean\_reads.sh FA\_SC

**Module 2: De novo assembly with SPAdes**

See assemble\_reads.sh

**Module 3 (optional): GC/coverage filtering**

GC/coverage based contaminant removal to improve computational efficiency prior to running redundans

1. Map to preliminary *de novo* assembly w/BWA-MEM
   1. ./aln.sh FA\_SC
2. Calculate GC/coverage for each contig
   1. ./cov\_gc\_len.sh FA\_SC

**Module 4: filter based on GC/cov (optional) and collapse haplotigs with *redundans***

See diploidify.sh

**Module 5: Further filter to remove highly heterozygous regions that did not collapse and likely assembled into alternate alleles**

1. Map to collapsed *de novo* assembly w/BWA-MEM
   1. ./aln2.sh FA\_SC
2. Calculate GC/coverage for each contig
   1. ./cov\_gc\_len\_blast.sh FA\_SC
3. Analysis in R
   1. See blob\_fig.R