Install anvi'o. Installation tutorials found on https://anvio.org/install/.

The processing of fasta files into usable contig databases can be scaled up using workflows (tutorial found on https://anvio.org/tutorials/scaling-up/). Anvi'o uses snakemake to automate common steps, such as the reformatting of fasta files, translating fasta files into contigs and running functional and taxonomy annotations.

```
#preparation for workflow (necessary on my device)
conda install pulp=2.7.0
pip install --upgrade snakemake

#set up annotation sources
anvi-setup-kegg-data
anvi-setup-ncbi-cogs
anvi-setup-scg-taxonomy

#create configuration file for our contigs
anvi-run-workflow -w contigs -c config-contigs-default.json

#create workflow graph to visualize each step
anvi-run-workflow -w contigs -c config-contigs-default.json --save-workflow-graph

#check stats
anvi-display-contigs-stats 02_CONTIGS/DC01-contigs.db
```

After this workflow, we have a contigs.db file for each fasta file. These contigs.db are input for the pangenomics workflow (tutorial found on https://merenlab.org/2016/11/08/pangenomics-v2/). These contigs.db get stored in a genomes storage object using a text file that lists the location of the contigs.db files.

```
#first, test system setup
anvi-self-test --suite pangenomics

#create genomes storage database
anvi-gen-genomes-storage -e external_genomes.txt -o strep_genomes.db

#run the pangenome analysis
anvi-pan-genome -g STREP-GENOMES.db -n Streptomyces -T 6 #using 6 threads
#this produces a -PAN.db file

#display the pangenome in the Anvi'o interactive environment
anvi-display-pan -p Streptomyces/StreptomycesPAN.db -g STREP-GENOMES.db
```

A gene cluster summary file can be downloaded from the interactive environment. This contains annotation information and the sequence of each gene cluster. This information was used to quantify the gene clusters for specific terms of interest, the analysis of which was done in R. These quantifications were then written into a data file that was imported into the Anvi'o pangenome to be displayed as metadata.

```
#import metadata file into pangenome
anvi-import-misc-data phenotype_table.txt -p Streptomyces/Streptomyces-PAN.db --target-data-table layers --contigs-mode

#display the pangenome in the Anvi'o interactive environment
anvi-display-pan -p Streptomyces/Streptomyces-PAN.db -g STREP-GENOMES.db
```