MG_HW9: Getting started with Anvi'o

Murat Eren and Bonnie Hurwitz

Abstract

Adapted from http://merenlab.org/2016/06/22/anvio-tutorial-v2/

Citation: Murat Eren and Bonnie Hurwitz MG_HW9: Getting started with Anvi'o. **protocols.io** dx.doi.org/10.17504/protocols.io.gc3bsyn
Published: 08 Nov 2016

Before start

The goal of this tutorial is to provide a brief overview of the anvi'o workflow for the analysis of assembly-based shotgun metagenomic data. Throughout this tutorial you will primarily learn about the following topics:

- Process your contigs

Protocol

Open an Anvi'o terminal
Step 1.

If you are here, you must have already installed the platform (hopefully without much trouble), and have run the infamous "mini test" successfully.

It is always a good idea to stick with stable versions of the platform, as the snapshots from the codebase can be very unstable and/or broken. However we also need people who like to live at the edge, and who would follow the development, test new features, join discussions, and push us to do better.

Download contigs.fa from the HPC to create the contigs.db for Anvi'o.
Step 2.

We are going to download the contigs.fa file from the HPC

To run the anvi'o metagenomic workflow, you will need these files:

- **A FASTA file of your contigs**. We shall call it `contigs.fa` throughout this manual. We will assume that `contigs.fa` contains contigs from a co-assembly. However, it may also have been a reference genome from NCBI, a metagenome-assembled genome (MAG), or a bunch of genes you are interested in profiling. Regardless of what it contains, following steps will not change too much.

- **Your FASTA file must have simple deflines, and if it doesn't have simple deflines, you must fix your FASTA file prior to mapping. We created this in a past protocol, be sure you download fixed-contigs.fa**

cmd COMMAND
```
scp netid@sftp.hpc.arizona.edu:/rsgrps/bh_class/username/assembly/megahit-out/fixed-contigs.fa contigs.fa
```
copying to a file called contig.fa into the anvi'o directory.

ANNOTATIONS
Emma Skidmore 10 Nov 2016

Is this meant to be ran in anvi'o? I'm getting an error message (bash: scp: command not found) when I do.

Creating a contigs database in Anvi'o

Step 3.

An anvi'o contigs database will keep all the information related to your contigs: positions of open reading frames, k-mer frequencies for each contigs, where splits start and end, functional and taxonomic annotation of genes, etc. The contigs database is an essential component of everything related to anvi'o metagenomic workflow.

Create the initial contigs.db from the contigs.fa file using the command below.

When you run this command, `anvi-gen-contigs-database` will,

- **Compute k-mer frequencies** for each contig (the default is 4, but you can change it using `--kmer-size` parameter if you feel adventurous).

- **Soft-split contigs** longer than 20,000 kbp into smaller ones (you can change the split size using the `--split-length`). When gene calling step is not skipped, the process splitting contigs will consider where genes are and avoid cutting genes in the middle. For very very large assemblies this process can take a while, and you can skip it with `--skip-mindful-splitting` flag.

- **Identify open reading frames** using Prodigal, the bacterial and archaeal gene finding

program developed at Oak Ridge National Laboratory and the University of Tennessee. If you don't want gene calling to be done, you can use the flag `--skip-gene-calling` to skip it. If you have your *own* gene calls, you can provide them to be used to identify where genes are in your contigs. All you need to do is to use the parameter `--external-gene-calls` (see down below for the format).

cmd COMMAND
```
anvi-gen-contigs-database -f contigs.fa -o contigs.db
```

ANNOTATIONS
Emily Wall 22 Nov 2016

Where should we be doing this? I keep getting this error whether I run it in on cygwin or through anvio: -bash: anvi-gen-contigs-database: command not found

James Thornton Jr 29 Nov 2016

PC users

When you scp your files using Cygwin, move those files to a new folder in Documents. Then in docker quickstart terminal navigate to that folder and do pwd to get the full path. Then to launch Anvio:

docker run --rm -v /path/to/files:/my_data -p 8080:8080 -it meren/anvio:latest

Additional troubleshooting- if having issues do docker ps and see if there are existing sessions. If so do docker kill [session id]

Run HMMs

Step 4.

Although this is absolutely optional, you shouldn't skip this step. Anvi'o can do a lot with hidden Markov models (HMMs provide statistical means to model complex data in probabilistic terms that can be used to search for patterns, which works beautifully in bioinformatics where we create models from known sequences, and then search for those patterns rapidly in a pool of unknown sequences to recover hits). To decorate your contigs database with hits from HMM models that ship with the platform (which, at this point, constitute multiple published bacterial single-copy gene collections), run this command:

cmd COMMAND
```
anvi-run-hmms -c contigs.db
```

NOTES
Bonnie Hurwitz 08 Nov 2016

When you run this command (without any other parameters),

- It will utilize multiple default bacterial single-copy core gene collections and identify hits among your genes to those collections using HMMER. If you have already run this once, and now would like to add an HMM profile of your own, that is easy. You can use `--hmm-profile-dir` parameter to declare where should anvi'o look for it.

- Note that the program will use only one CPU by default, especially if you have multiple of them available to you, you should use the `--num-threads` parameter. It significantly improves the runtime, since HMMER is truly an awesome software.

Read the Anvi'o paper

Step 5.

Read the Anvi'o paper. Why are we using hidden markov models to profile our contigs? How many raw hits were detected on your contigs from Campbell_et_al and Creevey_et_al? What are these papers about? What did they do to create these profiles? Add this into your google report (as a few lines of text) when you describe your steps in the methods for importing contigs into Anvi'o.