MG_HW12: Anvi'o interactive

Bonnie Hurwitz and Murat Meren

Abstract

This protocol describes basic functions to view your contigs and profiles databases in Anvi'o.

Citation: Bonnie Hurwitz and Murat Meren MG_HW12: Anvi'o interactive. **protocols.io** dx.doi.org/10.17504/protocols.io.gg8btzw
Published: 22 Nov 2016

Protocol

Start anvi'o on your laptop

Step 1.

Using the previous protocol, fire up the anvio software and navigate to the directory with your contigs db and profiles directories.

cmd COMMAND

```
cd skinmicrobiome
```

Launch anvi'o interactive

Step 2.

Start up anvi'o interactive viewer. This command will start up your default browser and open a web page for viewing your data.

cmd COMMAND

```
anvi-interactive -p SAMPLES-MERGED/PROFILE.db -c contigs.db
```

ANNOTATIONS

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]

Read a detailed overview of anvi'o interactive viewer

Step 3.

For a detailed overview of the anvi'o interactive viewer and all of the things you can do with it. See the weblink below. We will go over major components of the viewer in class.

LINK: http://merenlab.org/2016/02/27/the-anvio-interactive-interface/

In the layers tab

Step 4.

Click on the draw button in the layers tab in anvi'o to draw a circle phylogram. Try zooming in and out of the viewer to see more or less detail. Try switching the colors for each of the body sites, and redraw the figure.

Through the layers tab you can,

- **Change general settings for the tree** (i.e., switching between circle or rectengular displays, changing tree radius or width), **and layers** (i.e., editing layer margins, or activating custom layer margins).
- **Load or save states** to store all visual settings, or load a previously saved state.
- **Customize individual layers** by switching between different **display modes**depending on the layer type (i.e., 'text' or 'color' mode for categorical layers, or 'bar' or 'intensity' mode for numerical layers), **set normalization** (i.e., 'square-root', or 'log' normalization), **minimum, and maximum** cutoff values for numerical layers, or set **layer height**, and **layer margin** (i.e., its distance from the previous layer).
- Use the **multi-selector** at the bottom to change settings for multiple layers at once.

In the bins tab

Step 5.

In the bins tab, click on the 'Load bin collection' button. Select 'CONCOCT' from the pop up window.

Questions:

How many bins did the CONCOCT algorithm find? These are meant to represent different 'genomes' in your metagenomic sample. Do you see bins that are comprised of reads from only certain skin sites?

Anvi'o allows you to create selections of items shown in the display (whether they are metagenomic contigs, 16S rRNA tags, or any other type of information). Bins tab allow you to maintain these selections. Any selection on the tree will be added to active bin in this tab (the state radio button next to a bin defines its activity). Through this tab you can,

- **Create or delete bins**, **set bin names**, **change the color of a given bin**, or sort bins based on their name, the number of units they carry, or completion and contamination estimates (completion / contamination estimates are only computed for genomic or metagenomic analyses).
- View **the number of selected units** in a given bin, and see the **list of names in the selection** by clicking the button that shows the number of units described in the bin.
- **Store a collection of bins**, or **load a previously stored collection**.

Can you create better bins?

In the mouse tab

Step 6.

In the mouse tab, mouse over different regions of the graph. You can mouse over different regions of the graph to see data about individual contigs. You can also mouse over different levels to see more or less of the data in each of the circles, going from center (with the most info) to the ends (with the least).

Right click to get detailed info on a contig

Step 7.

You can also right click on a contig and select "inspect" to see the detailed informaton about that contig, including genes and annotations (kegg and pfam) you loaded in the contigs db. Or, you can do a blast on the fly.

Check out the samples tab.

Step 8.

Samples tab is for the additional data you provide the interface through a samples database (see samples order and samples information sections above). Through this layer you can,

- **Change the order of layers** using automatically-generated or user-provided orders of layers using the Sample order combo box,
- **Customize individual samples information entries**.

Changes in this tab can be reflected to the current display without re-drawing the entire tree unless the sample order is changed.

Create a summary of the samples

Step 9.

The interactive viewer is fun to play with, but you might also want to create a summary of the data. You can do this from the anvi'o command line:

cmd COMMAND

```
anvi-summarize -p SAMPLES-MERGED/PROFILE.db -c contigs.db -o SAMPLES-SUMMARY -C CONCOCT
```

Look at the summary

Step 10.

Was Anvi'o able to determine the taxonomy for each of the bins? Click on one of the buttons that says "NONE". What taxa are present in that bin? Do you see a mixture?

ANNOTATIONS

Bethanie Rachele Edwards 07 Apr 2018

What criteria does CONCOCT use when determining the taxonomy of each bin? >50% of sequences ID'd as a particular OTU?

In the summary, check completeness

Step 11.

How complete are the bins from CONCOCT? Look at the "Compl." tab to get an estimate fomr Anvi'o on how complete your bins are based on the composition of core bacterial genes.

In the summary, check redundancy

Step 12.

When we try to estimate the completeness of a genome bin, we identify single-copy genes that appear more than once as "contamination".

Jed Fuhrman suggested that the use of "redundancy" would be more appropriate, since these are not always contaminations (sometimes there are hits due to not-very-specific HMM profiles, etc), *and* the word "contamination" is a bit scary.

Compare ad contrast your data in anvio with your partner's data

Step 13.

Do you have similar results across the two time points?

Refine bins

Step 14.

Can you use some of anvi'os interactive features to create better and more complete bins?