Simon C-Microbial-MAP

This app seek to answer the question:

> "Given a 16S/18S sequence for my organism of interest, what information can I retrieve on its geographic distribution from the CMAP database?"

The input is a FASTA file of 16S sequences which are BLASTed to find hits which are then used to query the CMAP database using the Python "Opedia" to find the geographic information which is visualized using the R "Oce" module. The output of the application is set of PNG files.

Cf.:

- https://cmap.readthedocs.io/en/latest/catalog/catalog.html
- https://pypi.org/project/opedia/
- https://cran.r-project.org/web/packages/oce/vignettes/oce.html

Local Installation

To run this code locally on your computer, you will need to install:

- Python 3.x
- Microsoft ODBC drivers for you platform; for Mac, be sure to install http://www.freetds.org/
- https://pypi.org/project/opedia/
- BLAST+
- R, Oce

From the "scripts" directory, you can run `blast2cmap.py`:

```
$ ./blast2cmap.py -h
usage: blast2cmap.py [-h] -q str [-b str] [-p str] [-i float] [-c float]
                     [-o str]

BLAST hits to CMAP visualization

optional arguments:
  -h, --help            show this help message and exit
  -q str, --query str   Query file for BLAST (default: None)
  -b str, --blast_db str
                        BLAST db (default: blast)
  -p str, --blast_program str
                        BLAST program (default: blastn)
  -i float, --perc_identity float
                        BLAST percent identity (default: 97.0)
```

```
-c float, --qcov_hsp_perc float
                   BLAST percent query coverage per hsp (default: 100.0)
-o str, --outdir str  Output directory (default: out)
```

Singularity

A Singularity image with all dependencies is available on Stampede2 in the shared iMicrobe directory (/work/05066/imicrobe/singularity) that you can use directly from the command line, or you can download the image from ftp://ftp.imicrobe.us/singularity and run it locally on your own machine.

Run Online

If you have a CyVerse account, you can run the app on the Stampede2 cluster at TACC via iMicrobe (https://www.imicrobe.us/#/apps/83).

Run via CyVerse

If you wish to run the CyVerse/Stampede2 app directly, you must install the CyVerse SDK (https://github.com/cyverse/cyverse-sdk).

First you need to create a login token:

`$ auth-tokens-create -S -U <taccuser> -P <password>`

Or

`$ auth-tokens-refresh`

Then you can create a job template (JSON format)

`$ jobs-template -A r-oce-0.0.1 > job.json`

Edit the "job.json" file to indicate your desired parameters, and submit the job:

`$ jobs-submit -F job.json`

If your job is accepted, you will be shown a job ID which you can use to to monitor the job with the -W (watch) flag:

`$ jobs-status -W <job-id>`

When the job completes, you can use the CyVerse Discovery Environment to inspect the ouput from the job or use:

`$ jobs-output-list <job-id>`

To download the results to your local computer:

```
$ jobs-output-get -r <job-id>
```

To-Do

- Separate vizualizations by size fraction, data set

Authors

- Jesse McNichol mcnichol@alum.mit.edu
- Mohammad Ashkezari mdehghan@uw.edu
- Ken Youens-Clark kyclark@email.arizona.edu