

Subject: PAPRICA update
From: "Claypool, Joshua Thomas" <joshua.t.claypool@exxonmobil.com>
Sent: 6/19/2018 6:45:35 AM
To: "Jeff Bowman (jsbowman@ucsd.edu)" <jsbowman@ucsd.edu>;
CC: "Summers, Zarath M" <zarath.m.summers@exxonmobil.com>;
Attachments: buildExternal.txt; internalDB.txt; keggPathways.txt; paprica-make_ref-dev.txt; paprica-place-it.txt; paprica-tally-pathways-dev.txt; paprica-taxa-runDB.txt

Hi Jeff,

These are the re-written scripts to help rebuild PAPRICA using KEGG. Just change the ending from '.txt' to '.py' except for keggPathways.

The build process is:

1. Make a directory with the previous models folder in it.
2. Place all these scripts inside the directory you made
3. paprica-make_ref-dev.py
4. buildExternal.py
5. internal.py

Then to run:

1. `paprica-taxa-runDB.py -i otu_rep.fasta -c otu_taxa_table.txt -o output_directory`
 - `otu_rep.fasta` is a fasta of unique sequences. QIIME usually output representative OTU sequences and that's what is supposed to go here
 - `otu_taxa_table.txt` is just a tab-delimited OTU table
 - output directory can be anything. Default name if nothing is specified is 'outFolder/'

I *think* I've caught most of the errors all around but let me know if you run into any. I avoided Eukaryotes for this build in terms of time, but minor modifications should allow it no problem. I have a working draft of a manuscript that will hopefully get released in a couple of weeks.

All the best and kind regards,
Josh