SourceTracker (example data)

Back to

All of the code in this page is meant to be run in on the command line unless otherwise specified.

Install SourceTracker

First, get SourceTracker from github.com
These commands must be run on the command line

```
# clone the repo
git clone https://github.com/danknights/sourcetracker.git

# Enter the repository folder
cd sourcetracker

# Create a system variable called SOURCETRACKER_PATH pointing to this directory
# We will add it to a .Renviron file in your home directory.
echo "" >> $HOME/.Renviron
echo "SOURCETRACKER_PATH=$PWD" >> $HOME/.Renviron
```
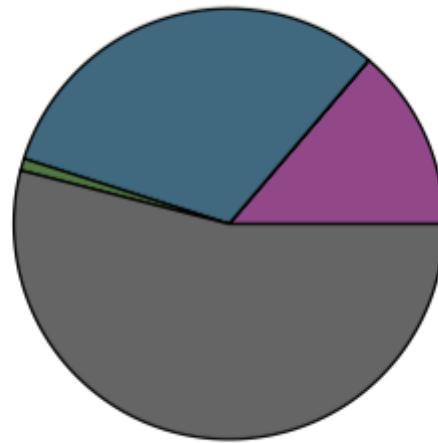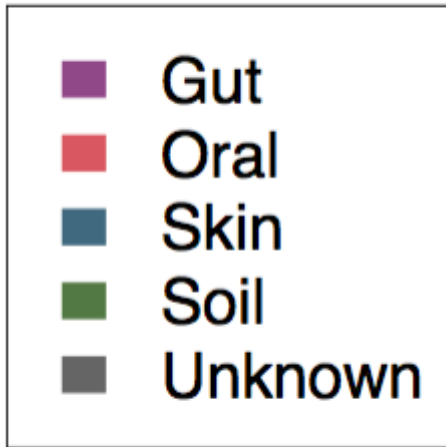
Run SourceTracker on example data

This will run SourceTracker with only 2 restarts. Note: on MSI, run this first: `module load R`

```
Rscript sourcetracker_for_qiime.r -i data/otus.txt -m data/metadata-subset.txt -o outdir -v -n 2
```

This will produce pie charts, bar charts, and distribution charts (see Figure 2C in SourceTracker paper). For example, we can see a mixture of gut and oral sources on the NICU surface in `sink_predictions_pie_NICU.pdf`:

NICU_UDB19

Run SourceTracker on other data

To run SourceTracker on other data, you need to do the following things:

1. Use a text-based OTU or Taxon table. You can get a text-based OTU table by running `biom convert -i otutable.biom -o otutable.txt --to-tsv` on your biom-formatted OTU table. You can get a text-based taxonomy summary by running the QIIME command `summarize_taxa.py -i otutable.biom -o taxa` and choosing one of the output files ending with `.txt`. If you are using an older version of QIIME

2. Include an "Env" column in your mapping file. You can open the mapping file in Excel or Google Docs and create a column indicating the known or expected source environment for each sample.

3. Include a "SourceSink" column in your mapping file. This column should say "source" for source samples and "sink" for sink samples.