PICRUST2 Workflow

Contents

PICRUST2 wiki

1. Installing PICRUST

This was already done and does not need to be done again.

```
mamba create -n picrust2 -c bioconda -c conda-forge picrust2=2.5.2
```

2. Required input files

Two required files

- FASTA of amplicon sequences variants (ASVs; i.e. your representative sequences, not your raw reads)
- Tab-delimited table with ASV ids as the first column and sample abundances as all subsequent columns

Example of both files :

```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed,
```

3. Running the full pipeline

Running picrust throught command line (not R)

a. Activate the environnement

```
mamba activate picrust2
```

b. Run the full default pipeline

```
picrust2_pipeline.py -s fasta_file.fna -i asv_file.tsv -o picrust2_out -p 12
```

Main options :

- `-s` PATH - FASTA of unaligned study sequences
- `-i` PATH - Input table of sequence abundances (BIOM, TSV, or mothur shared file format)
- `-o` PATH - Output folder
- `-p` INT: Number of processes to run in parallel.
- `--skip_norm` - Skip normalizing sequence abundances by predicted marker gene copy numbers (typically 16S rRNA genes). This step will be performed automatically unless this option is specified (added in v2.2.0-b).
- `--remove_intermediate` - Remove the intermediate outfiles of the sequence placement and pathway inference steps.
- `--verbose` - If specified, print out wrapped commands to screen.

See wiki for further options.

4. Key output files

- `EC_metagenome_out` - Folder containing unstratified EC number metagenome predictions
- `pred_metagenome_unstrat.tsv.gz` sequence table normalized by predicted 16S copy number abundances (seqtab_norm.tsv.gz), and the per-sample NSTI values weighted by the abundance of each ASV (weighted_nsti.tsv.gz).
- `KO_metagenome_out` - As EC_metagenome_out above, but for KO metagenomes.
- `pathways_out` - Folder containing predicted pathway abundances and coverages per-sample, based on predicted EC number abundances.