https://github.com/picrust/picrust2/wiki/Installation

16s功能注释有很多种软件，主流大概是picrust2跟tax4fun2，两种方法我都跑了，tax4fun2相对简单，占的ram应该是相对较少，直接R就能跑了；不过跑出来的是KO跟KEGG pathway；picrust2功能相对更多，还可以跑代谢酶EC；听说是更适用于人体肠道，各有优势；picrust2建议用MetaCyc数据库

qiime2里面可以添加picrust2的插件，官网有教程，但功能比独立版少，我就不研究了

没科学上网就老老实实加载清华源阿里云源等等

##添加北外镜像，清华的好像不咋行  
conda config --add channels https://mirrors.bfsu.edu.cn/anaconda/cloud/bioconda/  
conda config --add channels https://mirrors.bfsu.edu.cn/anaconda/cloud/conda-forge/  
conda config --add channels https://mirrors.bfsu.edu.cn/anaconda/pkgs/free/  
conda config --add channels https://mirrors.bfsu.edu.cn/anaconda/pkgs/main/  
conda config --set show\_channel\_urls yes

#不建议conda直接装，直接下源代码

wget https://github.com/picrust/picrust2/archive/v2.5.1.tar.gz

tar xvzf v2.5.1.tar.gz

cd picrust2-2.5.1

conda env create -f  picrust2-env.yaml#创建picrust2独立环境

#有mamba的话，安装就conda install mamba -n -c conda-forge

#更新一下：

https://mamba.readthedocs.io/en/latest/mamba-installation.html#fresh-install-recommended

conda install -n base --override-channels -c conda-forge mamba 'python\_abi=\*=\*cp\*'

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

conda activate picrust2

mamba env create -f picrust2-env.yaml

conda activate picrust2

pip install --editable .#pip安装editable

#测试

pytest

#test报错的就去github看看，一般都是网络问题有些包没下好，重新下一遍就好了

<https://github.com/picrust/picrust2/issues>

#entire pipeline，直接一步跑完下面的，picrust2优点

picrust2\_pipeline.py -s rep-seqs.fasta -i feature-table.biom -o picrust2\_out\_pipeline -p 20

#不会生成pred\_metagenome\_contrib.tsv.gz文件

#以下是picrust2的工作流程

#建树，把ASV的序列跟参考16s序列进行比对

time place\_seqs.py -s rep-seqs.fasta -o placed\_seqs.tre -p 20 -t sepp --intermediate intermediate/place\_seqs --verbose

-s 输入序列

--min\_align 0.8：相似度低于0.8的序列会被移除

-p 线程

-t epa ng/sepp：将序列放入参考树的放置工具。必须输入“epa ng”或“sepp”之一（默认值：epa ng）

# 如果没有足够的RAM来运行上述命令，则可以指定应使用SEPP而不是EPA-ng，执行此步骤带有该-t sepp选项。

--intermediate：输出中间文件

# 预测16S拷贝数

time hsp.py -i 16S -t placed\_seqs.tre -o marker\_nsti\_predicted.tsv.gz -p 20 -n

#添加EC编号，EC号是根据其催化的化学反应定义的一种基因家族，用EC推断MetaCyc途径水平

time hsp.py -i EC -t placed\_seqs.tre -o EC\_predicted.tsv.gz -p 20 -n

#添加KO编号

time hsp.py -i KO -t placed\_seqs.tre -o KO\_predicted.tsv.gz -p 20 -n

-n 计算nsti值，OTU与其最相似的已测序基因组间系统进化距离的平均值，越小越可靠

-i TRAIT\_OPTION：指定预测类别，包括16S，COG，EC，KO，PAFM，TIGRFAM，PHENO

NSTI>2的ASV通常是噪声， NSTI值的分布对于确定群落的整体特征以及是否存在异常值可能很有用

#宏基因组预测

# 预测E.C.丰度

time metagenome\_pipeline.py -i feature-table.biom -m marker\_nsti\_predicted.tsv.gz -f EC\_predicted.tsv.gz -o EC\_metagenome\_out --strat\_out

-f : KO/EC功能注释表

-m: 拷贝数文件；

--min\_reads INT：丰度过滤 (default: 1)；

--min\_samples INT：频率过滤 (default: 1)；

--strat\_out：选项指示分层输出文件（这是一个长格式表，指示ASV如何贡献EC）也将被输出。

EC\_metagenome\_out目录中创建了4个输出文件：

EC\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz：每个样本的总EC数量丰度；unstratification没有分层

EC\_metagenome\_out/pred\_metagenome\_contrib.tsv.gz：“贡献”格式的分层表，细分了每个样本中ASV如何促进基因家族丰度；

EC\_metagenome\_out/seqtab\_norm.tsv.gz：通过预测的16S拷贝数标准化的ASV丰度表；

EC\_metagenome\_out/weighted\_nsti.tsv.gz：每个样本的平均NSTI值（考虑到ASV的相对丰度时）。该文件对于识别数据集中的异常样本很有用。在PICRUSt1中，建议的加权NSTI值分别为<0.06和> 0.15，分别为好和差。临界值对于弄清样本与其他数据集的比较方式很有用，但是加权NSTI得分> 0.15不一定意味着预测毫无意义。

EC预测表不是相对丰度

#预测KO丰度

time metagenome\_pipeline.py -i feature-table.biom -m marker\_nsti\_predicted.tsv.gz -f KO\_predicted.tsv.gz -o KO\_metagenome\_out --strat\_out

#计算通路丰度

#MetaCyc

#pred\_metagenome\_contrib

time pathway\_pipeline.py -i EC\_metagenome\_out/pred\_metagenome\_contrib.tsv.gz -o pathways\_out -p 20

# pred\_metagenome\_unstrat

time pathway\_pipeline.py -i EC\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz -o MetaCyc\_pathways\_out -m ./picrust2-2.5.1/picrust2/default\_files/pathway\_mapfiles/metacyc\_path2rxn\_struc\_filt\_pro.txt --intermediate MetaCyc\_pathways\_out/minpath\_working -p 20

-coverage：计算覆盖度

默认的分层途径丰度表表示每个ASV对整个群落途径的丰度有多少贡献，而不是仅在该ASV的预测基因组内预测的途径丰度是多少，所以用-per\_sequence\_contrib输出每个ASV的路径丰度

# MetaCyc通路丰度，基于EC结果汇总

# MetaCyc通路丰度

time pathway\_pipeline.py -i EC\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz -o MetaCyc\_pathways\_out\_per\_seq --per\_sequence\_contrib --per\_sequence\_abun EC\_metagenome\_out/seqtab\_norm.tsv.gz --per\_sequence\_function EC\_predicted.tsv.gz -p 20

-per\_sequence\_contrib：可计算每个序列的贡献度，需要指定-per\_sequence\_abun和-per\_sequence\_function特征表和具体序列的功能

# KEGG通路丰度，基于KO结果汇总

#KEGG通路丰度

time pathway\_pipeline.py -i KO\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz -o KEGG\_pathways\_out --no\_regroup -m ~ /picrust2-2.5.1/picrust2/default\_files/pathway\_mapfiles/KEGG\_pathways\_to\_KO.tsv-p 20

--no\_regroup：不会重新分组

time pathway\_pipeline.py -i KO\_metagenome\_out/pred\_metagenome\_contrib.tsv.gz -o KEGG\_pathways\_out --no\_regroup -m /home/qiime2/miniconda/envs/picrust2/lib/python3.8/site-packages/picrust2/default\_files/pathway\_mapfiles/KEGG\_pathways\_to\_KO.tsv -p 20

#注释

# 添加EC、KO注释

add\_descriptions.py -i EC\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz -m EC -o EC\_metagenome\_out/pred\_metagenome\_unstrat\_descrip.tsv.gz

add\_descriptions.py -i KO\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz -m KO -o KO\_metagenome\_out/pred\_metagenome\_unstrat\_descrip.tsv.gz

add\_descriptions.py -i MetaCyc\_pathways\_out/path\_abun\_unstrat.tsv.gz -m METACYC -o MetaCyc\_pathways\_out/path\_abun\_unstrat\_descrip.tsv.gz

#KEGG\_pathway，旧版KEGG数据库，建议用MetaCyc数据库

add\_descriptions.py -i KEGG\_pathways\_out/path\_abun\_unstrat.tsv.gz --custom\_map\_table /home/qiime2/miniconda/envs/picrust2/lib/python3.8/site-packages/picrust2/default\_files/description\_mapfiles/KEGG\_pathways\_info.tsv.gz -o KEGG\_pathways\_out/path\_abun\_unstrat\_descrip.tsv.gz

picrust2\_pipeline.py -s /media/sf\_qiime2share/MHZ/2a/rep-seqs.fasta -i /media/sf\_qiime2share/MHZ/2a/b/feature-table.biom -o /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline --verbose --in\_traits COG,EC,KO,PFAM,TIGRFAM -p 20

但是输出的结果没有注释也没有keggpathway

# 添加EC、KO、metacyc注释

add\_descriptions.py -i /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/EC\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz -m EC -o /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/EC\_metagenome\_out/pred\_metagenome\_unstrat\_descrip.tsv.gz

add\_descriptions.py -i /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/KO\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz -m KO -o /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/KO\_metagenome\_out/pred\_metagenome\_unstrat\_descrip.tsv.gz

add\_descriptions.py -i /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/pathways\_out/path\_abun\_unstrat.tsv.gz -m METACYC -o /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/MetaCyc\_pathways\_out/path\_abun\_unstrat\_descrip.tsv.gz

#KEGG\_pathway得自己跑出来

time pathway\_pipeline.py -i /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/KO\_metagenome\_out/pred\_metagenome\_unstrat.tsv.gz -o /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/KEGG\_pathways\_out --no\_regroup -m /home/qiime2/picrust2-2.5.1/picrust2/default\_files/pathway\_mapfiles/KEGG\_pathways\_to\_KO.tsv -p 20

#注释

add\_descriptions.py -i /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/KEGG\_pathways\_out/path\_abun\_unstrat.tsv.gz --custom\_map\_table /home/qiime2/picrust2-2.5.1/picrust2/default\_files/description\_mapfiles/KEGG\_pathways\_info.tsv.gz -o /media/sf\_qiime2share/MHZ/2a/b/picrust2\_out\_pipeline/KEGG\_pathways\_out/path\_abun\_unstrat\_descrip.tsv.gz