Meta Tutorial

xyz

2021/9/3

Table of Contents

[Install metawrap 1](#_Toc81578845)

[Insall blast DB 2](#_Toc81578846)

[Install checKM DB 2](#_Toc81578847)

[Install Kraken DB 3](#_Toc81578848)

[Install salmon megahit 3](#_Toc81578849)

[Unzip 3](#_Toc81578850)

[Assemble 4](#_Toc81578851)

[Quantity contig 5](#_Toc81578852)

[Bining 5](#_Toc81578853)

[Species annotation 6](#_Toc81578854)

[Refine bin 6](#_Toc81578855)

[Visualize bins 7](#_Toc81578856)

[Species annotation of bins 7](#_Toc81578857)

[Function annotation of bins 7](#_Toc81578858)

# Install metawrap

# Express Installation  
conda create --name metawrap --channel ursky metawrap-mg=1.3.2  
  
# fix error  
# Can't locate Bio/Root/Version.pm in   
# @INC (you may need to install the Bio::Root::Version module)  
cd ~/miniconda3/envs/metawrap  
ln -s lib/perl5/site\_perl/5.22.0/ perl5  
which config-metawrap  
cp ~/config-metawrap ~/miniconda3/envs/metawrap/bin/config-metawrap  
  
# fix bowtie2-build-s: symbol lookup error, undefined symbol  
conda install tbb=2020.2  
  
conda activate metawrap

## Insall blast DB

[Aspera download link](https://www.ibm.com/aspera/connect/)

ftp  
open ftp.ncbi.nlm.nih.gov  
# user  
anonymous  
cd /blast/db/  
passive  
mls nt.\*.tar.gz download.list.txt  
cd /blast/db/v4/  
mls nt\_v4.\*.tar.gz downloadV4.list.txt  
bye  
  
nohup cat NCBI.nt.download.list.txt | \  
 xargs -n 1 -P 1 \  
 bash -c '~/.aspera/connect/bin/ascp -v -k 1 -T -l `  
 `1000m -i ~/asperaweb\_id\_dsa.openssh `  
 anonftp@ftp.ncbi.nlm.nih.gov:/blast/db/$0 ./' \  
 >> downloadlog.txt 2>&1 &  
   
nohup cat downloadV4.list.txt | \  
 xargs -n 1 -P 1 \  
 bash -c '~/.aspera/connect/bin/ascp -v -k 1 -T -l `  
 `1000m -i ~/asperaweb\_id\_dsa.openssh `  
 anonftp@ftp.ncbi.nlm.nih.gov:/blast/db/v4/$0 ./' \  
 >> downloadlog.txt 2>&1 &  
for a in nt\*.tar.gz; do tar xzf $a; done  
  
vim ~/miniconda3/envs/metawrap/bin/config-metawrap  
   
~/.aspera/connect/bin/ascp -v -k 1 -T -l 1000m \  
 -i ~/asperaweb\_id\_dsa.openssh \  
 anonftp@ftp.ncbi.nlm.nih.gov:/pub/taxonomy/taxdump.tar.gz ./  
tar -xvf taxdump.tar.gz

## Install checKM DB

# checkM database  
mkdir CHECKM\_DB  
cd CHECKM\_DB  
wget \  
https://data.ace.uq.edu.au/public/CheckM\_databases/checkm\_data\_2015\_01\_16.tar.gz  
tar -xvf \*.tar.gz  
rm \*.gz  
cd ..  
checkm data setRoot CHECKM\_DB

## Install Kraken DB

mkdir /home/xiongyi/database/KRAKEN\_DB  
cd /home/xiongyi/database/KRAKEN\_DB  
kraken2-build --download-library bacteria --db MY\_KRAKEN2\_DB  
kraken2-build --download-library archaea --db MY\_KRAKEN2\_DB  
kraken2-build --download-library fungi --db MY\_KRAKEN2\_DB  
kraken2-build --download-library viral --db MY\_KRAKEN2\_DB  
  
#!/bin/sh -login  
#PBS -o /home/xiongyi/database/KRAKEN\_DB  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N kraken2-build  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate metawrap  
cd /home/xiongyi/database/KRAKEN\_DB  
kraken2-build --build --db MY\_KRAKEN2\_DB --threads 24  
  
vim kraken2-build.sh  
qsub kraken2-build.sh

## Install salmon megahit

conda create -c bioconda -n salmon salmon  
conda create -c bioconda -n soil megahit

# Unzip

cd cai2  
mkdir cleanFastaq KRAKENreads \  
 INITIAL\_BINNING BIN\_REFINEMENT \  
 BIN\_REASSEMBLY BLOBOLOGY QUANT\_BINS \  
 BIN\_CLASSIFICATION FUNCT\_ANNOT QUANT\_CONTIG  
  
#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/cleanFastaq  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N unzip-fastaq  
#PBS -q cpu   
cd /home/xiongyi/cai2/cleanFastaq  
gunzip -c /home/xiongyi/caiyuanfeng/data2/E3\_clean\_R1.fq.gz > E3\_1.fastq  
gunzip -c /home/xiongyi/caiyuanfeng/data2/E3\_clean\_R2.fq.gz > E3\_2.fastq  
gunzip -c /home/xiongyi/caiyuanfeng/data2/E10\_clean\_R1.fq.gz > E10\_1.fastq  
gunzip -c /home/xiongyi/caiyuanfeng/data2/E10\_clean\_R2.fq.gz > E10\_2.fastq  
gunzip -c /home/xiongyi/caiyuanfeng/data2/E13\_clean\_R1.fq.gz > E13\_1.fastq  
gunzip -c /home/xiongyi/caiyuanfeng/data2/E13\_clean\_R2.fq.gz > E13\_2.fastq  
gunzip -c /home/xiongyi/caiyuanfeng/data2/E17\_clean\_R1.fq.gz > E17\_1.fastq  
gunzip -c /home/xiongyi/caiyuanfeng/data2/E17\_clean\_R2.fq.gz > E17\_2.fastq  
  
vim unzip-fastaq.sh  
qsub unzip-fastaq.sh

# Assemble

#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/cleanFastaq  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N merge  
#PBS -q cpu  
cd /home/xiongyi/cai2/  
cat cleanFastaq/\*\_1.fastq > cleanFastaq/All1.fastq  
cat cleanFastaq/\*\_2.fastq > cleanFastaq/All2.fastq  
  
vim merge.sh  
  
#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=124gb  
#PBS -N megahit  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate soil  
cd /home/xiongyi/cai2/  
megahit -1 cleanFastaq/All1.fastq \  
 -2 cleanFastaq/All2.fastq \  
 -o ASSEMBLY \  
 --min-count 2 \  
 -m 123 \  
 --k-list 27,37,47,57,67,77,87 \  
 -t 24  
   
cd /home/xiongyi/cai2/  
vim megahit.sh

# Quantity contig

#!/bin/sh -login  
#PBS -o /home/xiongyi/sunhongyang/MbPL202011862/QUANT\_CONTIG  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N QUANT\_CONTIG  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate salmon  
cd /home/xiongyi/sunhongyang/MbPL202011862/  
salmon index -t ASSEMBLY/final.contigs.fa \  
 -i QUANT\_CONTIG/contig\_index \  
 --threads 24  
# get fastq name  
files=(cleanFastaq/\*\_\*.fastq)  
# get sample name  
files=($(echo "${files[@]}" | tr ' ' '\n' | cut -d '\_' -f1 | tr '\n' ' '))  
# unique sample name  
files=($(echo "${files[@]}" | tr ' ' '\n' | sort -u | tr '\n' ' '))  
for file in ${files[@]}; do   
 salmon quant -i QUANT\_CONTIG/contig\_index \  
 -l A \  
 -1 ${file}\_1.fastq \  
 -2 ${file}\_2.fastq \  
 --validateMappings \  
 -o QUANT\_CONTIG/${file#\*/} \  
 --threads 24  
done  
  
vim QUANT\_CONTIG.sh  
qsub QUANT\_CONTIG.sh

# Bining

no maxbin because maxbin is too slow

#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/INITIAL\_BINNING  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N metawrap-binning  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate metawrap  
cd /home/xiongyi/cai2/  
metawrap binning -o INITIAL\_BINNING -t 24 \  
 -a ASSEMBLY/final.contigs.fa \  
 --metabat2 \  
 --concoct cleanFastaq/\*\_\*.fastq \  
 --run-checkm  
   
vim binning.sh   
qsub binning.sh

# Species annotation

#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/KRAKENreads  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N kraken-reads-run  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate metawrap  
cd /home/xiongyi/cai2/  
metawrap kraken2 -o KRAKENreads \  
 -t 24 \  
 cleanFastaq/\*\_\*.fastq ASSEMBLY/final.contigs.fa  
  
vim kraken-reads-run.sh   
qsub kraken-reads-run.sh

# Refine bin

#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/BIN\_REFINEMENT  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N bin\_refinement  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate metawrap  
cd /home/xiongyi/cai2/  
metawrap bin\_refinement -o BIN\_REFINEMENT \  
 -t 24 \  
 -A INITIAL\_BINNING/metabat2\_bins \  
 -B INITIAL\_BINNING/concoct\_bins \  
 -c 50 -x 10  
  
vim bin\_refinement.sh   
qsub bin\_refinement.sh

# Visualize bins

#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/BLOBOLOGY  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N BLOBOLOGY  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate metawrap  
cd /home/xiongyi/cai2/  
metawrap blobology -a ASSEMBLY/final.contigs.fa \  
 -t 24 \  
 -o BLOBOLOGY \  
 --bins BIN\_REFINEMENT/metawrap\_50\_10\_bins \  
 cleanFastaq/\*\_\*.fastq  
  
vim blobology.sh   
qsub blobology.sh

# Species annotation of bins

#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/BIN\_CLASSIFICATION  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N BIN\_CLASSIFICATION  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate metawrap  
cd /home/xiongyi/cai2/  
metawrap classify\_bins -b BIN\_REFINEMENT/metawrap\_50\_10\_bins \  
 -o BIN\_CLASSIFICATION \  
 -t 24  
   
vim classfication\_bins.sh

# Function annotation of bins

#!/bin/sh -login  
#PBS -o /home/xiongyi/cai2/FUNCT\_ANNOT  
#PBS -j oe  
#PBS -l nodes=1:ppn=24,walltime=999:00:00,mem=120gb  
#PBS -N BIN\_FUNCT\_ANNOT  
#PBS -q cpu  
source ~/miniconda3/etc/profile.d/conda.sh  
conda activate metawrap  
cd /home/xiongyi/cai2/  
metaWRAP annotate\_bins -o FUNCT\_ANNOT -t 24 -b BIN\_REFINEMENT/metawrap\_50\_10\_bins  
  
vim annotate\_bins.sh