

# Meta Tutorial

xyz

2021/9/3

## Contents

<b>Install metawrap</b>	<b>1</b>
Install blast DB . . . . .	2
Install checkKM DB . . . . .	3
Install Kraken DB . . . . .	3
Install salmon megahit . . . . .	3
<b>Unzip</b>	<b>3</b>
<b>Assemble</b>	<b>4</b>
<b>Quantity contig</b>	<b>5</b>
<b>Bining</b>	<b>5</b>
<b>Species annotation</b>	<b>6</b>
<b>Refine bin</b>	<b>6</b>
<b>Visualize bins</b>	<b>7</b>
<b>Species annotation of bins</b>	<b>7</b>
<b>Function annotation of bins</b>	<b>7</b>

## 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
```

## Install blast DB

Aspera download link

```
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 classification_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
```