

基于组装的软件和数据库下载

环境搭建

```
1 conda create -n mgs python=3.7
2 conda activate mgs
```

软件安装

```
1 conda install bowtie2 ##比对
2 conda install fastqc ##质控
3 conda install kneaddata ##质控过滤去宿主
4 conda install metaphlan ##物种分类
5 conda install megahit ##组装
6 conda install parallel ##软件并行
7 conda install prodigal ##基因预测
8 sudo apt install cd-hit ##去冗余
9 conda install salmon ##基因定量
10 conda install seqkit ##数据整理 (类似excel)
```

如没有管理员权限，cd-hit可以这样下载

```
1 cd /home/ubuntu/Software/
2 wget https://github.com/weizhongli/cdhit/archive/refs/tags/V4.8.1.tar.gz
3
4 #解压文件
5 tar xvf V4.8.1.tar.gz
6
7 #进入文件夹，然后编辑
8 cd cdhit-4.8.1/
9 make
```

Trimmomatic下载

进入官网：

[USADELLAB.org – Trimmomatic](https://www.usadellab.org/~brent/trimmomatic/): 用于Illumina NGS数据的灵活读取修整工具

Trimmomatic: A flexible read trimming tool for Illumina NGS data

Citations

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

Downloading Trimmomatic

starting on version 0.40 we also offer a [github page](#) (as well as older versions)

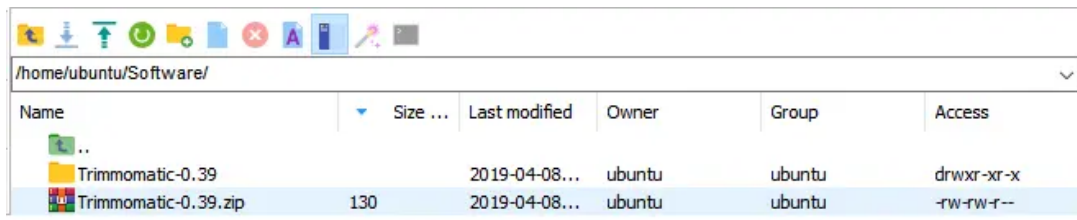
Version 0.39: [binary](#), [source](#) and [manual](#)

Version 0.36: [binary](#) and [source](#)

下载安装包上传到服务器：

建议在根目录下新建一个Software文件夹，把不通过conda等安装的软件统一放置于此

```
1 wget http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic/Trimmomatic-0.39.zip
2 unzip Trimmomatic-0.39.zip
```



Name	Size ...	Last modified	Owner	Group	Access
..					
Trimmomatic-0.39		2019-04-08...	ubuntu	ubuntu	drwxr-xr-x
Trimmomatic-0.39.zip	130	2019-04-08...	ubuntu	ubuntu	-rw-rw-r--

kneaddata、metaphlan和humann都有对应的数据库，因此需要同步进行数据库的下载，数据库存储在根目录下的database下

```
1 mkdir database
```

数据库下载

EasyMetagenome流程Pipeline下载

网页 <https://github.com/YongxinLiu/EasyMetagenome> 中Code – Download ZIP下载压缩包，上传至服务器/home/ubuntu/project/test_mgs/ 文件夹下（没有文件夹可以自己新建）

```
1 unzip EasyMetagenome-master.zip
2 # 改名
3 mv EasyMetagenome-master EasyMetagenome
4 ##修改权限
5 chmod -R 777 /home/ubuntu/project/test_mgs/EasyMetagenome
```

EasyMicrobiome流程Pipeline下载

网页 <https://github.com/YongxinLiu/EasyMicrobiome/tree/master> 中Code – Download ZIP下载压缩包，上传至服务器/home/ubuntu/project/test_mgs/ 文件夹下

```

1 unzip EasyMicrobiome-master.zip
2 # 改名
3 mv EasyMicrobiome-master EasyMicrobiome
4 ##添加linux命令可执行权限
5 chmod +x EasyMicrobiome/linux/*
6 ##添加软件至环境变量，否则需要指定软件的完整路径使用
7 # 将变量写入.bashrc，永久添加环境变量
8 echo "PATH=$PATH:`pwd`/EasyMicrobiome/linux:`pwd`/EasyMicrobiome/script" >
  > ~/.bashrc
9 ##重新激活
10 source ~/.bashrc

```

kneaddata数据库

```

1 cd database
2 mkdir kneaddata
3 cd kneaddata/
4 ##新建文件夹
5 mkdir human_genome
6 mkdir mouse_C57BL
7 mkdir human_transcriptome
8 mkdir ribosomal_RNA
9 ##数据库下载
10 kneaddata_database --download human_genome bowtie2 ./human_genome
11 kneaddata_database --download mouse_C57BL bowtie2 ./mouse_C57BL
12 kneaddata_database --download human_transcriptome bowtie2 ./human_transcri
  ptome
13 kneaddata_database --download ribosomal_RNA bowtie2 ./ribosomal_RNA
14
15 ##解压文件：tar -zxvf

```

一定要记得解压数据库中的tar.gz文件到对应文件夹

如若因为网络环境下载不了，可wget下载

```

1 kneaddata_database --available

```

```

(metagenome) [huodong3@cloud human_genome]$ kneaddata_database --available
KneadData Databases ( database : build = location )
human_genome : bowtie2 = http://huttenhower.sph.harvard.edu/kneadData_databases/Homo_sapiens_hg37_and_human_contamination_Bowtie2_v0.1.tar.gz
human_genome : bmtagger = http://huttenhower.sph.harvard.edu/kneadData_databases/Homo_sapiens_BMTagger_v0.1.tar.gz
human_transcriptome : bowtie2 = http://huttenhower.sph.harvard.edu/kneadData_databases/Homo_sapiens_hg38_transcriptome_Bowtie2_v0.1.tar.gz
ribosomal_RNA : bowtie2 = http://huttenhower.sph.harvard.edu/kneadData_databases/SILVA_128 LSUParc SSUParc_ribosomal_RNA_v0.2.tar.gz
mouse_C57BL : bowtie2 = http://huttenhower.sph.harvard.edu/kneadData_databases/mouse_C57BL_6NJ_Bowtie2_v0.1.tar.gz

```

metaphlan数据库



http://cmprod1.cibio.unitn.it/biobakery4/metaphlan_databases/

	mpa_vOct22_CHOCOPhAnSGB_202212.md5	2023-02-27 08:49	70
	mpa_vOct22_CHOCOPhAnSGB_202212.tar	2023-02-27 08:49	2.8G
	mpa_vOct22_CHOCOPhAnSGB_202212_marker_info.txt.bz2	2023-02-22 16:29	31M
	mpa_vOct22_CHOCOPhAnSGB_202212_species.txt.bz2	2023-02-22 16:29	502K

这是最新的数据库，以后会更新

接着下载另外两个，网址：

http://cmprod1.cibio.unitn.it/biobakery4/metaphlan_databases/bowtie2_indexes/

	mpa_vOct22_CHOCOPhAnSGB_202212_bt2.md5	2023-02-22 16:27	74
	mpa_vOct22_CHOCOPhAnSGB_202212_bt2.tar	2023-02-22 16:24	20G

文件位置：

http://cmprod1.cibio.unitn.it/biobakery4/metaphlan_databases/

http://cmprod1.cibio.unitn.it/biobakery4/metaphlan_databases/bowtie2_indexes/

将文件用迅雷下载，用filezilla上传到下面的文件夹（按需修改）：

```
1 /home/ubuntu/miniconda3/envs/mgs/lib/python3.7/site-packages/metaphlan/metaphlan_databases/
```

eggNOG功能数据库下载

```
1 db=/home/ubuntu/database/
2 ### eggNOG直接安装
3 # 新建环境并进入
4 conda create -n eggnog -y
5 conda activate eggnog
6 # 安装eggnog比对工具emapper
7 conda install eggnog-mapper -y -c bioconda -c conda-forge
8 ### eggNOG安装测试
9 emapper.py --version # 2.1.12
10 # Expected eggNOG DB version: 5.0.2 / Installed eggNOG DB version: 5.0.2 /
11 # Diamond version found: diamond version 2.0.15 / MMseqs2 version found: 13.45111
12 ### eggNOG数据库安装
13
14 # 下载常用数据库，注意设置下载位置
15 mkdir -p ${db}/eggnog && cd ${db}/eggnog
16 # -y默认同意，-f强制下载，eggnog.db.gz 6.3G+4.9G，解压后48G
17 download_eggnog_data.py -y -f --data_dir ${db}/eggnog
18
```

输入emapper.py --version命令后会出现eggNOG的数据库默认位置：

```
(eggNog) ubuntu@ubuntu-PowerEdge-T440:~$ emapper.py --version
There was an error retrieving eggNog-mapper DB data: not a valid file /home/ubuntu/miniconda3/envs/eggNog/lib/python3.12/site-packages/data/eggNog.db"
Maybe you need to run download_eggNog_data.py
emapper-2.1.12 / Expected eggNOG DB version: 5.0.2 / Installed eggNOG DB version: unknown / Diamond version found: diamond version 2.1.9 / MMseqs2 version found: 15.6f452 / Compatible novel families DB version: 1.0.1
```

将自己下载的数据库位置链接到默认位置（没有data文件夹，需要新建一个）

```
1 ln -sf ${db}/eggNog /home/ubuntu/miniconda3/envs/eggNog/lib/python3.12/site-packages/data/
```

/home/ubuntu/miniconda3/envs/eggNog/lib/python3.12/site-packages/data/					
Name	Size ...	Last modified	Owner	Group	Access
..					
eggNog	1	2024-04-09...	ubuntu	ubuntu	lrwxrwxrwx