

Kraken2

Kraken

Kraken is a metagenomic mapping tool that calculates abundances of taxonomic labels in a mixed sample. To use it you can load the existing module **kraken/2.1.1**. For a pre-built database refer to the section below.

Demo call:

```
kraken2 --db /gpfs/project/databases/Kraken2_refseq+hfmv --threads 8 --report kraken.report myReads.fastq
```

Bracken

Bracken is a complementary tool for species level re-estimation and a compatible module is available: **bracken/2.6.0**.

Demo call:

```
est_abundance.py -i kraken.report -k /gpfs/project/databases/Kraken2_refseq+hfmv/kmerDistribution.txt -o bracken.report --out-report bracken.krakenstyle.report
```

Database

We have a shared database for Kraken2 available at

/gpfs/project/databases/Kraken2_refseq+hfmv

that contains the following organisms:

- Archaea: 264 genomes (403 contigs), 640.15mb.
- Bacteria <prokaryotes>: 10753 genomes (20883 contigs), 40490.27mb.
- Eukaryota: 2192 genomes (7279859 contigs), 78719.21mb.
 - davon:
 - Apicomplexa: 2 genomes (21 contigs), 30.84mb.
 - Ascomycota: 1536 genomes (2311006 contigs), 45008.14mb.
 - Basidiomycota: 502 genomes (4496104 contigs), 20111.28mb.
 - Blastocladiomycota: 2 genomes (610 contigs), 93.84mb.
 - Chordata: 2 genomes (6556 contigs), 8498.23mb.
 - > Mensch und Maus
 - Chytridiomycota: 14 genomes (58021 contigs), 742.89mb.
 - Cryptomycota: 2 genomes (1281 contigs), 18.17mb.
 - Microsporidia: 36 genomes (78447 contigs), 239.41mb.
 - Mucoromycota: 74 genomes (193583 contigs), 2985.22mb.
 - Undefined: 6 genomes (2654 contigs), 209.03mb.
 - Zoopagomycota: 16 genomes (131576 contigs), 782.16mb.
 - Viroids: 95 genomes (95 contigs), 0.03mb.
 - Viruses: 27827 genomes (67295 contigs), 939.23mb.

Update: Two newer databases have been uploaded to:

One is taken from Ben Langmeads K2 Index page, the other (plus) is custom built containing all NCBI clades and in addition a selection of interesting gut microbiome species, a mouse genome and representative genomes from the fungi and protozoa clade even if they are not marked as "full genome" or "chromosome". The folders contain .txt files giving a detailed breakdown of the contents.

/gpfs/project/databases/Kraken2-2022-09-28