First, install miniconda:

<https://conda.io/miniconda.html>

Then, we create two different environments, one for python 2 and one for python 3:

conda create -n snakemake\_python2\_env python=2.7

conda create -n snakemake\_python3\_env python=3

Add channel bioconda:

<https://bioconda.github.io/>

conda config --add channels bioconda

source activate snakemake\_python3\_env

conda install snakemake

conda install cookiecutter

#can’t connect to conda homepage (due to expired certificate?) → instead using:

pip install snakemake

pip install cookiecutter

cookiecutter --checkout only\_snakemake\_workflow <https://github.com/EnvGen/cookiecutter-binning-project>

<https://github.com/EnvGen/cookiecutter-binning-project/tree/only_snakemake_workflow>

cd snakemake-workflow

git pull origin master

snakemake --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile --dryrun megahit\_coassemble\_all

mkdir finished\_reads

# Project specific bin

export PATH=/proj/b2010008/bin:$PATH

~/.bash\_profile

which Snakefile\_sbatch.py

tmux

snakemake --cluster "Snakefile\_sbatch.py {dependencies}" --immediate-submit -j 99 --latency-wait 120 --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile assembly/megahit\_coassembly/meta-sensitive/final.contigs.fa

<http://www.cazy.org/>

snakemake --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile --dryrun assembly/megahit\_coassembly/meta-sensitive/parts/contigs.0.fasta

snakemake --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile --cluster "Snakefile\_sbatch.py {dependencies}" --immediate-submit -j 99 --latency-wait 120 assembly/megahit\_coassembly/meta-sensitive/parts/contigs.0.fasta

snakemake --cluster "Snakefile\_sbatch.py {dependencies}" --immediate-submit -j 99 --latency-wait 120 --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile annotation/prokka\_extended/prokka\_output/megahit\_coassembly.{0..6}/PROKKA.gff

replace after successful --dryrun

--cluster "Snakefile\_sbatch.py {dependencies}" --immediate-submit -j 99 --latency-wait 120

interactive -A b2010008 -n 1 -t 2:00:00

module load bioinfo-tools Metaxa2

metaxa2 --plus T -i /proj/b2010008/nobackup/projects/rene\_roundup/metag\_glyph\_workflow/assembly/megahit\_coassembly/meta-sensitive/parts/contigs.0.fasta

-o /proj/b2010008/nobackup/projects/rene\_roundup/metag\_glyph\_workflow/taxonomy/metaxa/

git clone <https://github.com/BinPro/CONCOCT.git>

snakemake --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile --cluster "Snakefile\_sbatch.py {dependencies}" --immediate-submit -j 99 --latency-wait 120 annotation/prokka\_extended/prokka\_output/megahit\_coassembly.{0..6}/PROKKA.gff concoct/meta-sensitive\_coassembly/input\_kallisto/concoct\_inputtableR.tsv

less annotation/prokka\_extended/prokka\_output/megahit\_coassembly.4/PROKKA.gff-slurm.out

less concoct/meta-sensitive\_coassembly/cutup/contigs\_10K.fasta-slurm.out

<https://github.com/johnne/parse-db/tree/master/metacyc>

Bio/SearchIO/hmmer3.pm

source activate snakemake\_python2\_env && module load bioinfo-tools BioPerl blast/2.2.28+ hmmer prodigal/2.6.3 gnuparallel

conda install pandas

k141\_1288.0 10000 9786.7 145 2.14778

k141\_1288.1 10000 9786.7 24 0.355495

k141\_1288.2 11798 11584.7 2 0.0250267

zegrep "k141\_1288\\.[0-2]" < ./A1/abundance.tsv.gz >> compare.tsv

/pica/v14/b2010008\_nobackup/database/metacyc

tar -xvzf {file}

for i in {0..6}; do

echo $i;

grep "eC\_number=" ./megahit\_coassembly.${i}/PROKKA.gff | cut -f9 | cut -f1,2 -d ';'| sed 's/ID=//g'| sed 's/;eC\_number=/\t/g' >> PROKKA.ec;

done

cd CONCOCT

source activate snakemake\_python2\_env

conda install matplotlib scipy numpy

python setup.py install

concoct --help

conda install checkm-genome

snakemake --dryrun --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile concoct/meta-sensitive\_coassembly/output\_kallisto/cutoff-1000-maxcluster-1000/clustering.csv

concoct/meta-sensitive\_coassembly/cutup/contigs\_10K.fasta

snakemake --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile --cluster "Snakefile\_sbatch.py {dependencies}" --immediate-submit -j 99 --latency-wait 120 concoct/meta-sensitive\_coassembly/output\_kallisto/cutoff-1000-maxcluster-1000/checkm\_output/stats.tsv

export MinPath=/proj/b2010008/src/MinPath\_common

python /proj/b2010008/src/MinPath\_common/MinPath1.2.py -h

python /proj/b2010008/src/MinPath\_common/MinPath1.2.py -any PROKKA.ec -map /pica/v14/b2010008\_nobackup/database/metacyc/metacyc\_20.5/20.5/metacyc.ec2pathcats.tab -report PROKKA.metacyc.minpath > MinPath.Metacyc.PROKKA.log

snakemake --snakefile snakemake-workflows/workflows/large\_coassembly/Snakefile --cluster "Snakefile\_sbatch.py {dependencies}" --immediate-submit -j 99 --latency-wait 120 mapping/bowtie2/default/meta-sensitive\_megahit\_coassembly\_contigs/samples/A{1..7}.sorted.bam mapping/bowtie2/default/meta-sensitive\_megahit\_coassembly\_contigs/samples/B{8..10}.sorted.bam

cd /proj/b2010008/nobackup/database

chgrp -R b2010008 metacyc/

/pica/v14/b2010008\_nobackup/database/metacyc/metacyc\_20.5/20.5/ec.to.pwy

### CheckM

source activate snakemake\_python2\_env && module load bioinfo-tools prodigal/2.6.3 hmmer/3.1b2 MUMmer/3.23

checkm lineage\_wf -t 8 -x fa -f concoct/meta-sensitive\_coassembly/output\_kallisto/cutoff-1000-maxcluster-1000/checkm\_output/stats.tsv concoct/meta-sensitive\_coassembly/output\_kallisto/cutoff-1000-maxcluster-1000/all\_bins\_nocutup concoct/meta-sensitive\_coassembly/output\_kallisto/cutoff-1000-maxcluster-1000/checkm\_output/

/pica/v14/b2010008\_nobackup/database/checkm\_v1.05

checkm data update