

# Asking for help! GenErode run using test dataset on rhinos fails in Uppmax

---

I should start by saying that this is the first time I am using snakemake (I am a user of nf-core nextflow pipelines though), so apologies in advance if there is something obvious that I am missing out...

I followed all the instructions provided in the wiki to [install the pipeline](#) in Uppmax (see the dedicated section below for a step-by-step description).

To make sure the pipeline was installed correctly, I first tried to run it on the test dataset on rhinos provided in the [wiki](#).

I uploaded the [rhino](#) files to Uppmax, which were located in the directory [rhino\\_example](#):

```
GenErode$ ll
total 2.1M
drwxrwsr-x 3 angela naiss2023-23-265 4.0K Jul 16 15:15 config/
drwxrwsr-x 5 angela naiss2023-23-265 4.0K Jul 16 15:20 data/
drwxrwsr-x 4 angela naiss2023-23-265 4.0K Jul 15 23:09 docs/
-rw-rw-r-- 1 angela naiss2023-23-265 166 Jul 15 23:09 environment.yml
-rw-rw-r-- 1 angela naiss2023-23-265 35K Jul 15 23:09 LICENSE.md
-rw-rw-r-- 1 angela naiss2023-23-265 2.1K Jul 15 23:09 README.md
drwxrwsr-x 3 angela naiss2023-23-265 4.0K Jul 16 15:20 results/
drwxr-xr-x 6 angela naiss2023-23-265 4.0K Jul 15 21:02 rhino_example/
drwxrwsr-x 2 angela naiss2023-23-265 4.0K Jul 16 14:45 slurm/
-rw-rw-r-- 1 angela naiss2023-23-265 15K Jul 15 23:09 Snakefile
drwxrwsr-x 2 angela naiss2023-23-265 4.0K Jul 15 23:09 utilities/
drwxrwsr-x 7 angela naiss2023-23-265 4.0K Jul 15 23:09 workflow/
```

I placed the config and metadata files of the rhinos in the [config/](#) directory:

```
GenErode/config$ ll
total 128K
-rw-rw-r-- 1 angela naiss2023-23-265 27 Jul 15 23:09
chrX_candidate_scaffolds.txt
-rw-r--r-- 1 angela naiss2023-23-265 22K Jul 15 21:49
config_mitogenomes.yaml
-rw-rw-r-- 1 angela naiss2023-23-265 22K Jul 16 15:15
config_original.yaml
-rw-r--r-- 1 angela naiss2023-23-265 22K Jul 15 21:48
config_sum_rhino.yaml
-rw-rw-r-- 1 angela naiss2023-23-265 22K Jul 16 15:15 config.yaml
-rw-rw-r-- 1 angela naiss2023-23-265 233 Jul 15 23:09
historical_samples_paths.txt
-rw-rw-r-- 1 angela naiss2023-23-265 230 Jul 15 23:09
modern_samples_paths.txt
-rw-r--r-- 1 angela naiss2023-23-265 8.3K Jul 15 21:50
```

```
rhino_3_historical_samples.txt
-rw-r--r-- 1 angela naiss2023-23-265 766 Jul 15 21:47
rhino_3_modern_samples.txt
drwxrwsr-x 3 angela naiss2023-23-265 4.0K Jul 15 23:09 slurm/
```

Note that in the config and metadata files I updated the paths in Uppmax to the fastq files and the reference mitogenome, respectively. I also made sure that the paths were correct (pointed to the right files), using the commands `zcat /path/filename.fastq.gz | head` or `head /path/mitogenome.fasta`.

Then, I followed the instructions on [how to run the pipeline](#) for the dry and actual runs of the pipeline (I first tried the mitogenome example):

```
# Make a copy of the original config.yaml file for the records
cd GenErode/config
cp config.yaml config_original.yaml

# Replace the config.yaml file for the one you want to run first (e.g. on
mitogenome)
cp config_mitogenomes.yaml config.yaml

# Start a detached Terminal session to be able to run GenErode in the
background
cd .. # now at GenErode/ dir
screen -S rhino # rackham1

# Activate the conda environment
conda activate generode

# Start a dry run to check if the pipeline run would work as expected
snakemake --profile slurm -npr &> 230716_dry_run.out
# (replace YYYYMMDD with the current date) and checked the file
YYYYMMDD_dry_run.out for any errors

# Start the main run of the pipeline
snakemake --profile slurm &> 230716_main_run_mt_rhino.out # and checked
the file YYYYMMDD_main_run.out regularly during the run.
```

I did not notice any evident error message in the dry run output file `230716_dry_run.out` (a bunch of missing files but I guess that is expected as no files are generated??). However, in the output file of the actual run `230716_main_run_mt_rhino.out` there were lots of error messages.

In the very first pipeline installation, the common error was:

```
Traceback (most recent call last):
  File "slurm/slurm-submit.py", line 9, in <module>
    import requests
ModuleNotFoundError: No module named 'requests'
Error submitting jobscript (exit code 1):
```

So I fix it by installing the missing package called `requests` following the steps described [here](#).

Then I tried a new run but the pipeline kept failing, so I decided to reinstall everything from scratch, after deleting the pipeline `rm -R GenErode/` and the conda environment with:

```
conda deactivate
conda remove -n generode --all
```

Now the common error message is:

```
Traceback (most recent call last):
  File "slurm/slurm-submit.py", line 10, in <module>
    from snakemake.utils import read_job_properties
ModuleNotFoundError: No module named 'snakemake'
Error submitting jobscript (exit code 1):
```

Not sure why the package `snakemake` is missing if the conda environment of `generode` is enabled?

As I run out of ideas, I examined the files generated in the `data/` and `results/` directories, trying to figure out when the pipeline failed (see the "Exploring outputs" section below for detailed descriptions).

It seems that no jobs were submitted to slurm and the pipeline stopped after the symbolic links to the fastq files were created...but I cannot pinpoint what may be causing the problem. I attach the output.log files, so perhaps you could help me figure out what may be causing this problem? Thanks in advance for any help, I really need to run this pipeline ASAP.

## Exploring outputs

- No results were generated:

```
GenErode/results$ ll
total 4.0K
drwxrwsr-x 4 angela naiss2023-23-265 4.0K Jul 16 15:21 logs/
```

- The symbolic link to the mitogenome was created, but not the indexes and dict files:

```
GenErode/mitogenomes/:
total 104K
-rw-rw-r-- 1 angela naiss2023-23-265 17K Jul 15 23:09
chicken_NC_001323.fasta
-rw-rw-r-- 1 angela naiss2023-23-265 17K Jul 15 23:09 cow_NC_006853.fasta
-rw-rw-r-- 1 angela naiss2023-23-265 17K Jul 15 23:09
human_NC_012920.fasta
-rw-rw-r-- 1 angela naiss2023-23-265 17K Jul 15 23:09
mouse_NC_005089.fasta
```

```
-rw-rw-r-- 1 angela naiss2023-23-265 17K Jul 15 23:09 pig_NC_000845.fasta
lrwxrwxrwx 1 angela naiss2023-23-265 86 Jul 16 15:21
rhino_NC_012684.fasta -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/reference/rhino_NC_012684.fasta
```

- The symbolic link to the fastq files were created:

```
GenErode/data$ ll */*/
logs/1.1_fastq_processing/:
total 24K
drwxrwsr-x 2 angela naiss2023-23-265 20K Jul 16 16:16 historical/
drwxrwsr-x 2 angela naiss2023-23-265 4.0K Jul 16 16:16 modern/

raw_reads_symlinks/historical/:
total 288K
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS008_08_L2_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_08_L2.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS008_08_L2_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_08_L2.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS008_08_L6_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_08_L6.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS008_08_L6_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_08_L6.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS008_10_L2_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_10_L2.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS008_10_L2_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_10_L2.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS008_10_L6_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_10_L6.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS008_10_L6_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_10_L6.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS008_11_L2_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS008_11_L2.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
```

```
JvS008_11_L2_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS008_11_L2.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS008_11_L6_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS008_11_L6.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS008_11_L6_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS008_11_L6.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS009_09_L3_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_09_L3.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS009_09_L3_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_09_L3.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS009_09_L7_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_09_L7.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS009_09_L7_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_09_L7.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS009_15_L3_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_15_L3.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS009_15_L3_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_15_L3.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS009_15_L7_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_15_L7.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS009_15_L7_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_15_L7.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS009_19_L3_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_19_L3.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS009_19_L3_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_19_L3.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS009_19_L7_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS009_19_L7.extr.R1.fastq.
```

```
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS009_19_L7_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS009_19_L7.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_01_L1_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_01_L1.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_01_L1_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_01_L1.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_02_L1_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_02_L1.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_02_L1_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_02_L1.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_03_L1_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_03_L1.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_03_L1_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_03_L1.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_04_L1_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_04_L1.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_04_L1_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_04_L1.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_05_L1_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_05_L1.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_05_L1_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_05_L1.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_06_L1_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_06_L1.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_06_L1_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_06_L1.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
```

```
JvS022_07_L1_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_07_L1.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS022_07_L1_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_07_L1.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_08_L1_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_08_L1.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_08_L1_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_08_L1.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS022_09_L1_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_09_L1.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS022_09_L1_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_09_L1.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_10_L1_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_10_L1.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_10_L1_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_10_L1.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_11_L1_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_11_L1.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_11_L1_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_11_L1.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_12_L1_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_12_L1.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_12_L1_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_12_L1.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_74_L8_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_74_L8.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_74_L8_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_74_L8.extr.R2.fastq.gz
```

```
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_75_L8_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_75_L8.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_75_L8_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_75_L8.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_76_L8_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_76_L8.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_76_L8_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_76_L8.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_77_L8_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_77_L8.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_77_L8_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_77_L8.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_78_L8_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_78_L8.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_78_L8_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_78_L8.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_79_L8_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_79_L8.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_79_L8_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_79_L8.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_80_L8_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_80_L8.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_80_L8_R2.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_80_L8.extr.R2.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
JvS022_81_L8_R1.fastq.gz -> /proj/naiss2023-23-
265/nobackup/GenErode/rhino_example/historical/JvS022_81_L8.extr.R1.fastq.
gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:22
```



```
JvS022_81_L8_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_81_L8.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS022_82_L8_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_82_L8.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS022_82_L8_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_82_L8.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_83_L8_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_83_L8.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_83_L8_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_83_L8.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS022_84_L8_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_84_L8.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:20
JvS022_84_L8_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_84_L8.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_85_L8_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_85_L8.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 95 Jul 16 15:21
JvS022_85_L8_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/historical/JvS022_85_L8.extr.R2.fastq.gz

raw_reads_symlinks/modern/:
total 24K
lrwxrwxrwx 1 angela naiss2023-23-265 91 Jul 16 15:22
JvS033_01_L7_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/modern/JvS033_01_L7.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 91 Jul 16 15:22
JvS033_01_L7_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/modern/JvS033_01_L7.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 91 Jul 16 15:21
JvS034_02_L8_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/modern/JvS034_02_L8.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 91 Jul 16 15:21
JvS034_02_L8_R2.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/modern/JvS034_02_L8.extr.R2.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 91 Jul 16 15:22
JvS035_14_L7_R1.fastq.gz -> /proj/naiss2023-23-265/nobackup/GenErode/rhino_example/modern/JvS035_14_L7.extr.R1.fastq.gz
lrwxrwxrwx 1 angela naiss2023-23-265 91 Jul 16 15:22
```



- The directory `results/logs/slurm/` does not exist, does it mean that no jobs were actually submitted to uppmx?

```
GenErode/results/logs$ ll
total 8.0K
drwxrwsr-x 4 angela naiss2023-23-265 4.0K Jul 16 15:22
1.1_fastq_processing/
drwxrwsr-x 2 angela naiss2023-23-265 4.0K Jul 16 15:22
1.2_map_to_mitogenomes/
```

- all the logs files were empty:

```
GenErode/results/logs/1.1_fastq_processing$ ll */
historical/:
total 0
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS008_08_L2_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS008_08_L6_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS008_10_L2_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS008_10_L6_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS008_11_L2_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS008_11_L6_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS009_09_L3_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS009_09_L7_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS009_15_L3_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS009_15_L7_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS009_19_L3_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS009_19_L7_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_01_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_02_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_03_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_04_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
```

```
JvS022_05_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_06_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_07_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_08_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_09_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_10_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_11_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_12_L1_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_74_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_75_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_76_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_77_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_78_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_79_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_80_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_81_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_82_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_83_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_84_L8_fastp_historical.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS022_85_L8_fastp_historical.log

modern/
total 0
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS033_01_L7_fastp_modern.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS034_02_L8_fastp_modern.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 16:16
JvS035_14_L7_fastp_modern.log

GenErode/results/logs/1.2_map_to_mitogenomes$ ll
total 0
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 15:22
chicken_NC_001323_bwa_index_mito_ref.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 15:22
```

```
cow_NC_006853_bwa_index_mito_ref.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 15:22
human_NC_012920_bwa_index_mito_ref.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 15:22
mouse_NC_005089_bwa_index_mito_ref.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 15:22
pig_NC_000845_bwa_index_mito_ref.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 15:22
rhino_NC_012684_bwa_index_mito_ref.log
-rw-rw-r-- 1 angela naiss2023-23-265 0 Jul 16 15:21
symlink_mito_genome.log
```

## Current software versions

```
/GenErode$ cat environment.yml
cat environment.yml
channels:
  - bioconda
  - conda-forge
dependencies:
  - python=3.8.15
  - snakemake=7.20.0
  - biopython=1.76
  - matplotlib=3.4.2
  - pandas=1.0.3
  - numpy=1.18.4
```

## Slurm settings

Confirmed that the naiss account has ben corretly added to the slurm profile:

```
GenErode/slurm$ cat settings.json
{
  "SBATCH_DEFAULTS": "account=naiss2023-22-508",
  "CLUSTER_NAME": "",
  "CLUSTER_CONFIG": ""
}
```

## Pipeline installation in Uppmax

Shared by: John Sundh

Clone the repository in some location on Uppmax *where you have sufficient storage space* (ideally several TB), preferably the **nobackup** part of your project directory:

```
# Directory where the pipeline will be stored in Uppmax
cd /proj/naiss2023-23-265/nobackup

git clone https://github.com/NBISweden/GenErode.git
#git clone git@github.com:NBISweden/GenErode.git
```

then change directory into the newly created GenErode dir:

```
cd GenErode/
```

then create the conda environment with the software required to run the pipeline (it takes >60 minutes...):

```
screen -S test # rackham1
conda env create -n generode -f environment.yml
```

Once the environment is created activate it with:

```
conda activate generode
```

Now to configure the pipeline to run on the SLURM workload manager you can use the `cookiecutter` program. First install cookiecutter by running:

```
conda install -c conda-forge cookiecutter
```

(**note** that this will install cookiecutter into the generode conda environment if you have it activated)

Now you can install the official SLURM profile which includes a set of configuration presets and utility scripts that allows Snakemake to automatically submit jobs to the cluster queue:

```
cookiecutter --output-dir . https://github.com/Snakemake-
Profiles/slurm.git
```

You will now be prompted for a number of setup questions. For each question the default answer is shown in square brackets (e.g. profile name [slurm]:) and you can *accept the default* by just pressing the **Enter** key.

Make sure you answer the questions as suggested under point 5 in the Setup instructions on the Wiki (<https://github.com/NBISweden/GenErode/wiki/2.-Requirements-&-pipeline-configuration>). **IMPORTANT:** For the prompt about "SBATCH\_defaults" make sure you enter your compute account id, for example

`account=SNIC2023-1-001`. Replace `SNIC2023-1-001` with *your actual compute account*, then press **ENTER** and answer the remaining questions. :

```
profile_name [slurm]: slurm
Select use_singularity:
1 - False
2 - True
Choose from 1, 2 [1]: 2
Select use_conda:
1 - False
2 - True
Choose from 1, 2 [1]: 1
jobs [500]: 100
restart_times [0]:
max_status_checks_per_second [10]:
max_jobs_per_second [10]:
latency_wait [5]:
Select print_shell_commands:
1 - False
2 - True
Choose from 1, 2 [1]: 2
sbatch_defaults []: account=naiss2023-22-508 # fill in your compute
account
cluster_sidecar_help [Use cluster sidecar. NB! Requires snakemake >= 7.0!
Enter to continue...]: (press Enter)
Select cluster_sidecar:
1 - yes
2 - no
Choose from 1, 2 [1]: 2
cluster_name []:
cluster_jobname [%r_%w]:
cluster_logpath [logs/slurm/%r/%j]: results/logs/slurm/%r/%j
cluster_config_help [The use of cluster-config is discouraged. Rather, set
snakemake CLI options in the profile configuration file (see snakemake
documentation on best practices). Enter to continue...]: (press Enter)
cluster_config []:
```

This is what I typed (for the records):

```
profile_name [slurm]: slurm
Select use_singularity:
1 - False
2 - True
Choose from 1, 2 [1]: 2
Select use_conda:
1 - False
2 - True
Choose from 1, 2 [1]: 1
jobs [500]: 100
restart_times [0]:
```

```
max_status_checks_per_second [10]:
max_jobs_per_second [10]:
latency_wait [5]:
Select print_shell_commands:
1 - False
2 - True
Choose from 1, 2 [1]: 2
sbatch_defaults []: account=naiss2023-22-508
cluster_sidecar_help [Use cluster sidecar. NB! Requires snakemake >= 7.0!
Enter to continue...]:
Select cluster_sidecar:
1 - yes
2 - no
Choose from 1, 2 [1]: 2
cluster_name []:
cluster_jobname [%r_%w]:
cluster_logpath [logs/slurm/%r/%j]: results/logs/slurm/%r/%j
cluster_config_help [The use of cluster-config is discouraged. Rather, set
snakemake CLI options in the profile configuration file (see snakemake
documentation on best practices). Enter to continue...]:
cluster_config []:
```

Once you've answered all the questions there will be a new `slurm/` folder in your current directory. As a final step copy over the file `config/slurm/profile/config.yaml` to the new `slurm/` directory:

```
cp config/slurm/profile/config.yaml slurm/
```

Now you're ready to use the pipeline with SLURM workload manager support by adding `--profile slurm` to the snakemake command when running the pipeline.