



Institut Pasteur

A very brief history of Sequana

Thomas Cokelaer

Institut Pasteur

Jan 24th 2017, Institut Pasteur (HUB talk)

Why Sequana ?

Provide a toolbox to parse and analyse NGS data sets

- Simplified version to external dependencies
- Include pandas for data mining
- Matplotlib for visualisation

Enforce a common framework

- Using Snakemake as a common language to design new pipelines
- Provide reusable snakemake rules and modules

An entry point to a high quality software

- Quality (doc + tests)
- Diffusion
- Reproducibility

Sequana: a Python library

Python behind the scene

Python as a glue language

- Make use of anaconda for installation and various dependencies.
- For low-level computations, use existing libraries (e.g., pysam)

Dev of original tools

e.g. coverage or quick taxonomy (see talk on sequana_coverage)

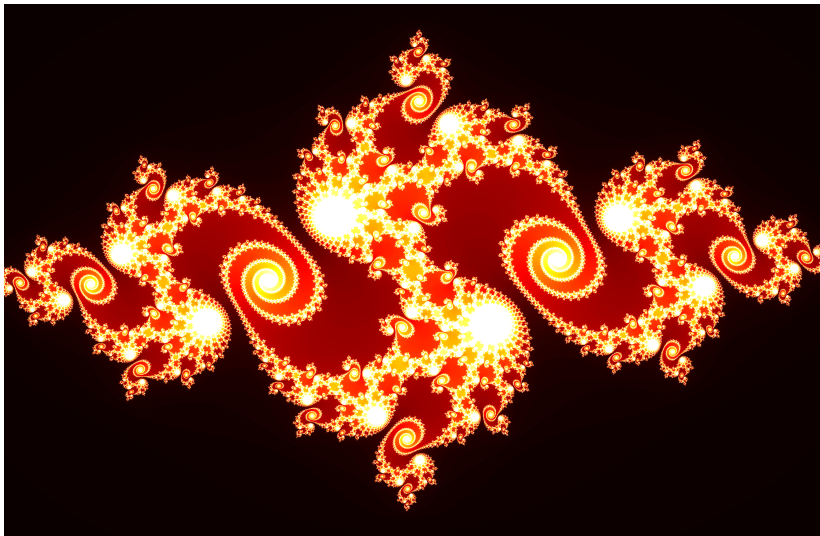
add missing bricks

- kraken2krona - ...

HTML Reporting

Based on Jinja and Sequana reports.

Sequana: snakemake as a workflow manager



Snakemake workflow = Snakefile + config file

pseudo working example of a simple workflow:

```
configfile: "fractal.yaml"
config = config['fractal']
N = config['N']

rule all:
    input: expand("julia_{index}.png", index=range(N))

rule image:
    output: image="julia_{index}.png"
    run:
        xy = "%.3f +%.3f j" % (random(), random())
        cmd = "python -m fractal julia %s" % xy
        cmd += " --size=%(size)s --depth=%(depth)s "
        cmd += " -o julia_%s.png" % wildcards.index
        shell(cmd % config)
```

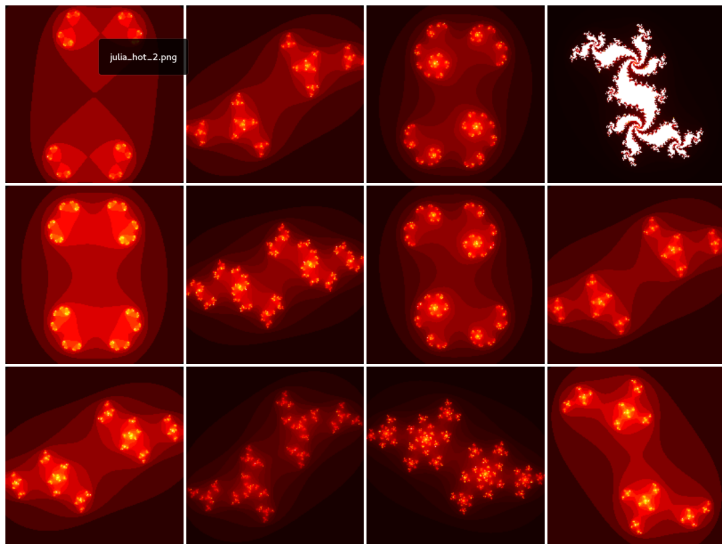
configuration file

```
#####  
# Input parameters for the fractal analysis  
#  
# :Parameters:  
#  
# - size: output image size formatted as NxM where N and M  
#       are integers  
# - depth: a integer (e.g. 200)  
# - zoom: a positive value e.g. 0.5  
# - N: number of random sets  
fractal:  
  - size: 200x200  
  - depth: 200  
  - zoom: 0.7  
  - N: 20
```



```
snakemake -s fractal.rules  
          --configfile fractal.yaml
```

```
snakemake -s fractal.rules  
          --configfile fractal.yaml
```

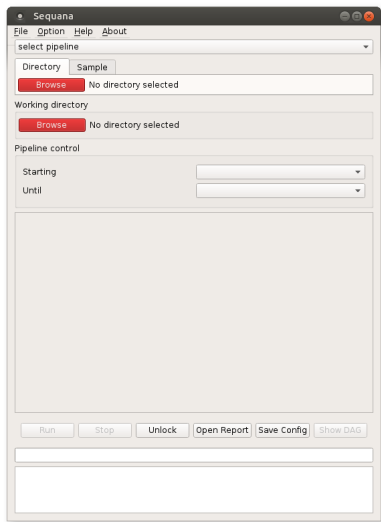


Pipelines available in Sequana



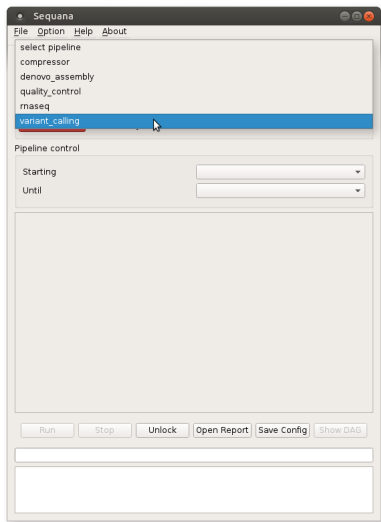
Sequana: for end users with Sequanix

GUI to simplify the usage of snakemake



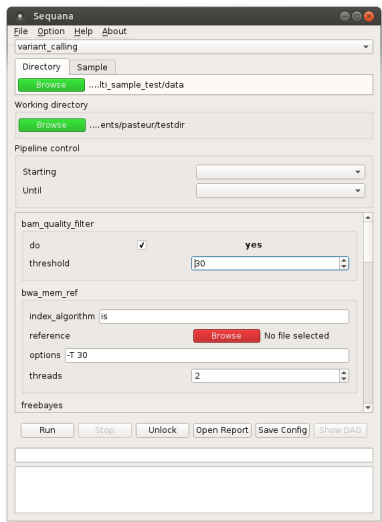
- Interface developed with PyQT5 and python
- Wrap our snakemake pipelines to ease the usage
- Usable on our cluster, which allows X11

GUI to simplify the usage of snakemake



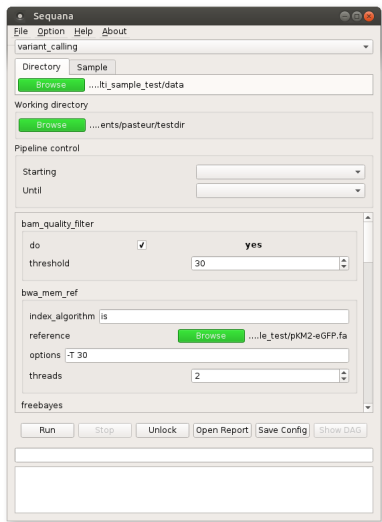
① Choose a pipeline

GUI to simplify the usage of snakemake



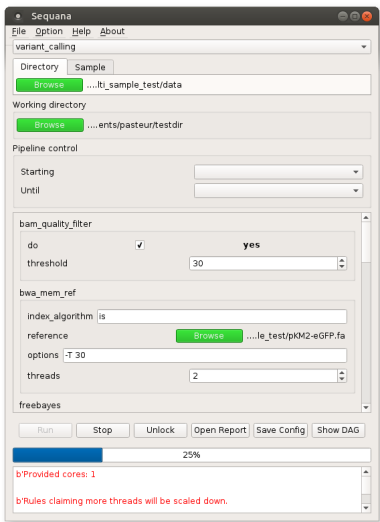
- 1 Choose a pipeline
- 2 Set input and output

GUI to simplify the usage of snakemake



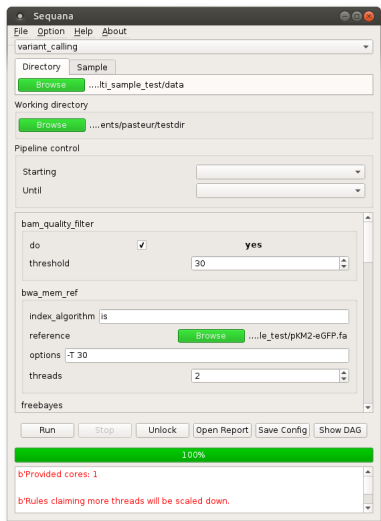
- 1 Choose a pipeline
- 2 Set input and output
- 3 Fill the config formular

GUI to simplify the usage of snakemake



- 1 Choose a pipeline
- 2 Set input and output
- 3 Fill the config formular
- 4 Run the pipeline

GUI to simplify the usage of snakemake



- 1 Choose a pipeline
- 2 Set input and output
- 3 Fill the config formular
- 4 Run the pipeline
- 5 Finished !

Sequana: Continuous integration

Versioning, Test and Documentation



<https://github.com/sequana/sequana>



Travis CI

Continuous Integration on Travis with 120 tests with 75% coverage



Uses Sphinx (RST syntax) to document the source code and provides user guide.



Read *the* Docs

Updated after each commits on sequana.readthedocs.io

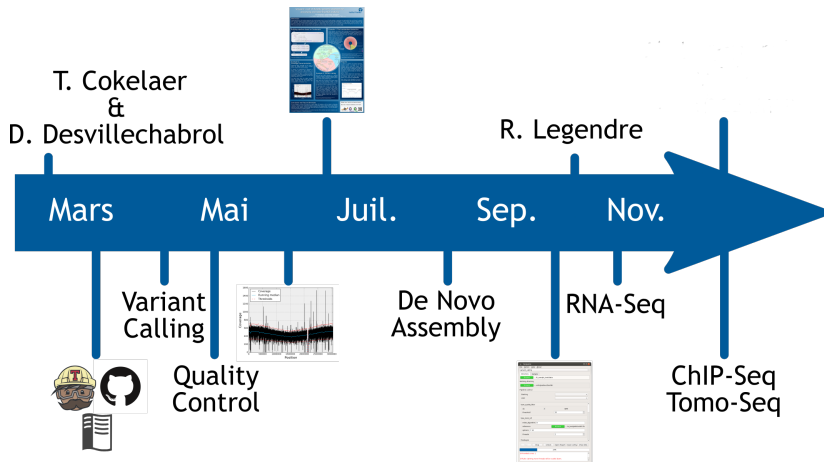
Summary

Summary

Sequana is a versatile tool that provides

- ① A set of snakemake workflows dedicated to NGS
- ② A GUI to execute them easily with Sequanix
- ③ A Python library dedicated to NGS analysis (e.g., tools to visualise standard NGS formats).
- ④ HTML reports
- ⑤ Standalone applications:
 - **sequana_coverage** ease the extraction of genomic regions of interest and genome coverage information
 - **sequana_taxonomy** get a quick overview of read contents
 - ...

Sequana history



¹Detection and characterization of low and high genome coverage regions using an efficient running median and a double threshold approach. Dimitri Desvillechabrol, Christiane Bouchier, Sean Kennedy, Thomas Cokelaer bioRxiv 092478; doi: <http://dx.doi.org/10.1101/092478>