'Sequana': a Set of Snakemake NGS pipelines

Thomas Cokelaer[1,2], **Dimitri Desvillechabrol**[2], **Rachel Legendre**[1,2], **and Mélissa Cardon**[2]

1 Institut Pasteur - Bioinformatics and Biostatistics Hub - C3BI, USR 3756 IP CNRS - Paris, France
2 Institut Pasteur - Biomics Pole - Paris, France

Summary

Sequana is a Python-based software dedicated to the development of Next Generation Sequencing (NGS) pipelines. We use the Snakemake (Köster and Rahmann 2012) framework to design our pipelines, which eases the decomposition of pipelines into modular sub-units. We currently have 7 pipelines covering quality control, variant calling, long-reads quality, de-novo and RNA-seq analysis (see https://sequana.readthedocs.io for details). Our pipelines are associated with HTML reports based on JINJA templating and Javascript. The reports are used to store the results of a pipelines but also materials required to reproduce the results. **Sequana** is also a Python library that provides tools to perform various analysis tasks (e.g., variant call filtering). Some of the library components provide original tools that are also available as standalone applications. For instance a fast taxonomic analysis based on Kraken (Wood and Salzberg 2014) as well as a tool to perform exhaustive coverage analysis (Desvillechabrol et al. 2016) (bottom right panel in the image here below).

Sequana is an open source project (https://github.com/sequana/sequana). It is developed with the aim of simplifying the development of new tools (for developers) and the deployment of the pipelines (for users). The extended documentation (http://sequana.readthedocs.org) and test suite (on Travis.org) provide a high-quality software that is routinely tested. **Sequana** is now available on bioconda making the installation easier and faster by taking care of the dependencies (e.g., samtools, bwa, or Python libraries).

Finally, for end-users, we also developed a Graphical interface called **Sequanix** (Desvillechabrol et al. 2017) developed with the PyQt framework (see left panel of the image here below). **Sequanix** standalone exposes all **Sequana** pipelines (Snakemake pipelines) within an easy-to-use interface. Within the graphical interface, the configuration file used by Snakemake are automatically loaded and can be edited by end-users with dedicated widgets. We made the interface generic enough that not only Sequana pipelines can be run interactively but also any Snakemake pipelines.

Future works

Sequana is an on-going project. Although the project has reached a mature stage with stable pipelines, new pipelines will be including on demand or based on new technologies.
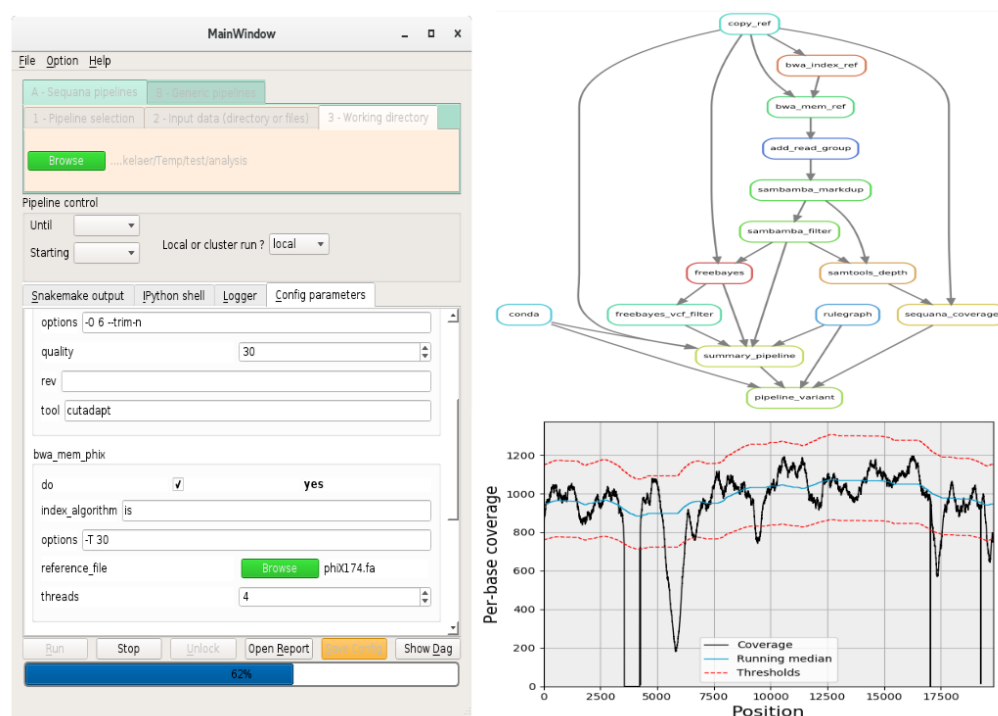
Figure 1

References

Desvillechabrol, Dimitri, Christiane Bouchier, Sean Kennedy, and Thomas Cokelaer. 2016. "Detection and Characterization of Low and High Genome Coverage Regions Using an Efficient Running Median and a Double Threshold Approach." *bioRxiv*. doi:https://doi.org/10.1101/092478.

Desvillechabrol, Dimitri, Rachel Legendre, Claire Rioualen, Christiane Bouchier, Jacques van Helden, Sean Kennedy, and Thomas Cokelaer. 2017. "Sequanix: A Dynamic Graphical Interface for Snakemake Workflows." *bioRxiv*. doi:https://doi.org/10.1101/162701.

Köster, Johannes, and Sven Rahmann. 2012. "Snakemake - a Scalable Bioinformatics Workflow Engine." *Bioinformatics* 28 (19). Oxford University Press: 2520–2.

Wood, Derrick E., and Steven L. Salzberg. 2014. "Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments." *Genome Biology* 15:R46.