The Journal of Open Source Software

# genomepy: download genomes the easy way

**Simon J. van Heeringen**[1]

**1** Radboud University, Nijmegen, the Netherlands

**DOI:** 10.21105/joss.00320

**Software**

- Review
- Repository
- Archive



## Summary

The technical advances in high-throughput DNA sequencing have resulted in data sets of increasing size and complexity. It is essential that experiments are analyzed in a reproducible manner. Computational workflow systems allow for automation of analysis pipelines, that scale from personal computers to HPC and cloud environments (Koster and Rahmann 2012; Di Tommaso et al. 2017; Vivian et al. 2017). While standardized solutions exist for installation of software and analysis tools ("Pip," n.d.; "Bioconda," n.d.), data download often has to either be performed manually or be scripted on a case-by-case basis.

In many analysis workflows, sequencing reads will need to be mapped to a reference genome. Here we present genomepy, a simple software package to automate the download of genomic sequences. It contains both command-line tools as well as a Python API. Supported providers for genomes include UCSC, NCBI and Ensembl. Downloaded genome sequences can be soft- or hard-masked and specific chromosomes or scaffolds can be either included or excluded based on regular expressions. Genomepy is free and open source software and can be installed through standard package managers ("Pip," n.d.; "Bioconda," n.d.).

In short, genomepy enables simple, straightforward and automatic downloads of genomic sequences.

## References

"Bioconda." n.d. https://bioconda.github.io/.

Di Tommaso, Paolo, Maria Chatzou, Evan W Floden, Pablo Prieto Barja, Emilio Palumbo, and Cedric Notredame. 2017. "Nextflow enables reproducible computational workflows." *Nature Biotechnology* 35 (4): 316–19. doi:10.1038/nbt.3820.

Koster, J., and S. Rahmann. 2012. "Snakemake–a scalable bioinformatics workflow engine." *Bioinformatics* 28 (19): 2520–2. doi:10.1093/bioinformatics/bts480.

"Pip." n.d. https://pip.pypa.io/en/stable/.

Vivian, John, Arjun Arkal Rao, Frank Austin Nothaft, Christopher Ketchum, Joel Armstrong, Adam Novak, Jacob Pfeil, et al. 2017. "Toil enables reproducible, open source, big biomedical data analyses." *Nature Biotechnology* 35 (4): 314–16. doi:10.1038/nbt.3772.