

## The GEN-ERA toolbox: unified and reproducible workflows for research in microbial genomics

Luc Cornet<sup>1</sup>

<sup>1</sup>*University Of Liege*

The GEN-ERA toolbox: unified and reproducible workflows for research in microbial genomics

Luc Cornet, Benoit Durieu, Frederik Baert, Elizabet D'hooge, David Colignon, Loic Meunier, Valérien Lupo, Ilse Cleenwerck, Heide-Marie Daniel, Leen Rigouts, Damien Sirjacobs, Stéphane Declerck, Peter Vandamme, Annick Wilmotte, Denis Baurain, Pierre Becker

Microbial culture collections play a key role in taxonomy by studying the diversity of their strains and providing well-characterized biological material to the scientific community for fundamental and applied research. These microbial resource centers thus need to implement new standards in species delineation, including whole-genome sequencing and phylogenomics. In this context, the genomic needs of the Belgian Coordinated Collections of Microorganisms (BCCM) were studied, resulting in the GEN-ERA toolbox, a unified cluster of bioinformatic workflows dedicated to both bacteria and small eukaryotes (e.g., yeasts). This public toolbox is designed for researchers without a specific training in bioinformatics (launched by a single command line). Hence, it facilitates all steps from genome downloading and quality assessment, including genomic contamination estimation, to tree reconstruction. It also offers workflows for average nucleotide identity comparisons and metabolic modeling. All the workflows are based on Singularity containers and Nextflow to increase reproducibility. The GEN-ERA toolbox can be used to infer completely reproducible comparative genomic and metabolic analyses on prokaryotes and small eukaryotes. Although designed for routine bioinformatics of culture collections, it can also be used by all researchers interested in microbial taxonomy, as exemplified by our case study on Gloeobacterales (Cyanobacteria).

This study is published at <https://doi.org/10.1093/gigascience/giad022>.