

Bactopia v2: Highly scalable, portable and customizable bacterial genome analyses

Robert A. Petit III^{1,2}, Davi J. Marcon^{3,4}, Abhinav Sharma⁵, and Timothy D. Read⁶

¹ Public Health Laboratory, Wyoming Department of Health, Cheyenne, Wyoming, USA

² Theiagen Genomics, Highlands Ranch, CO, USA

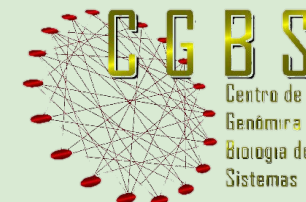
³ Center of Genomics and Systems Biology, Institute of Biological Sciences, Universidade Federal do Pará, 66075-110 Belém, PA, Brazil

⁴ Laboratory of Genetic Engineering, Guamá Science and Technology Park, 66075-750 Belém, PA, Brazil

⁵ Faculty of Engineering and Technology, Liverpool John Moores University, Liverpool, United Kingdom.

⁶ Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia, USA

Correspondence: robert.petit@theiagen.com Bactopia Documentation: bactopia.github.io

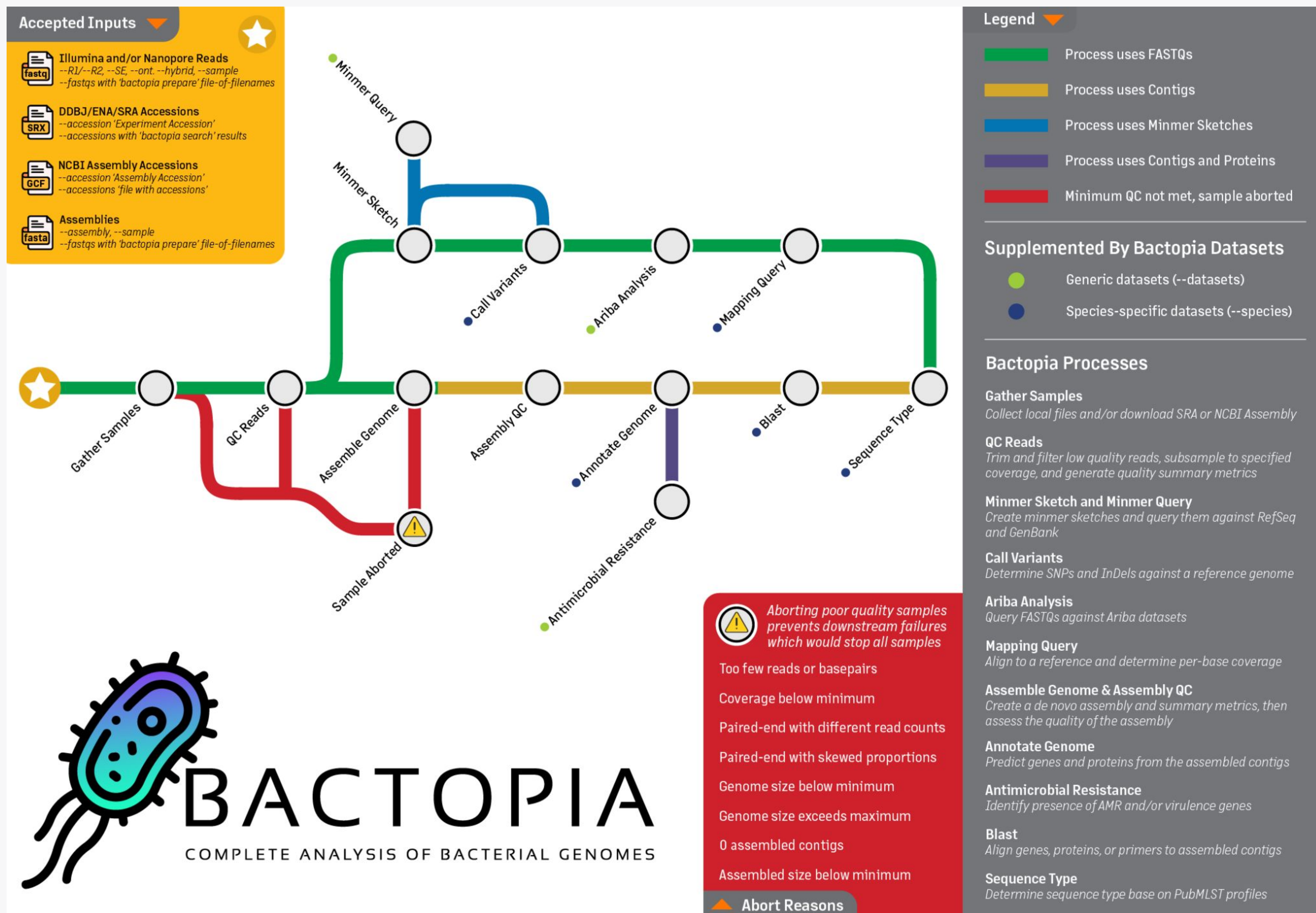


Highlights

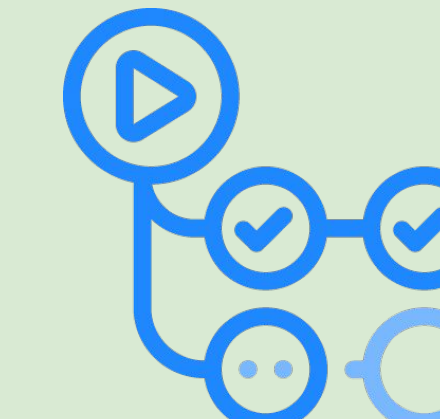
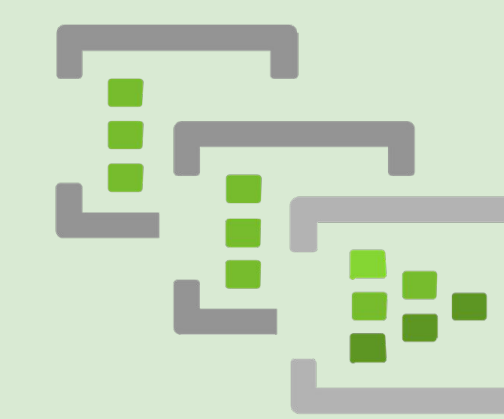
- Written in Nextflow DSL2, greatly increases Bactopia's ability to fit your needs
- Supports Illumina, Oxford Nanopore technologies as well as NCBI's Sequence Read Archive and Assembly databases
- Includes more than [130 bioinformatics tools](#) into a single pipeline
- [33 Bactopia Tools](#) allow easy comparative analyses of Bactopia outputs
- Available on Bioconda, Docker, and Singularity

Bactopia is platform Independent

By using Nextflow, Bactopia can run on multiple platforms, including:

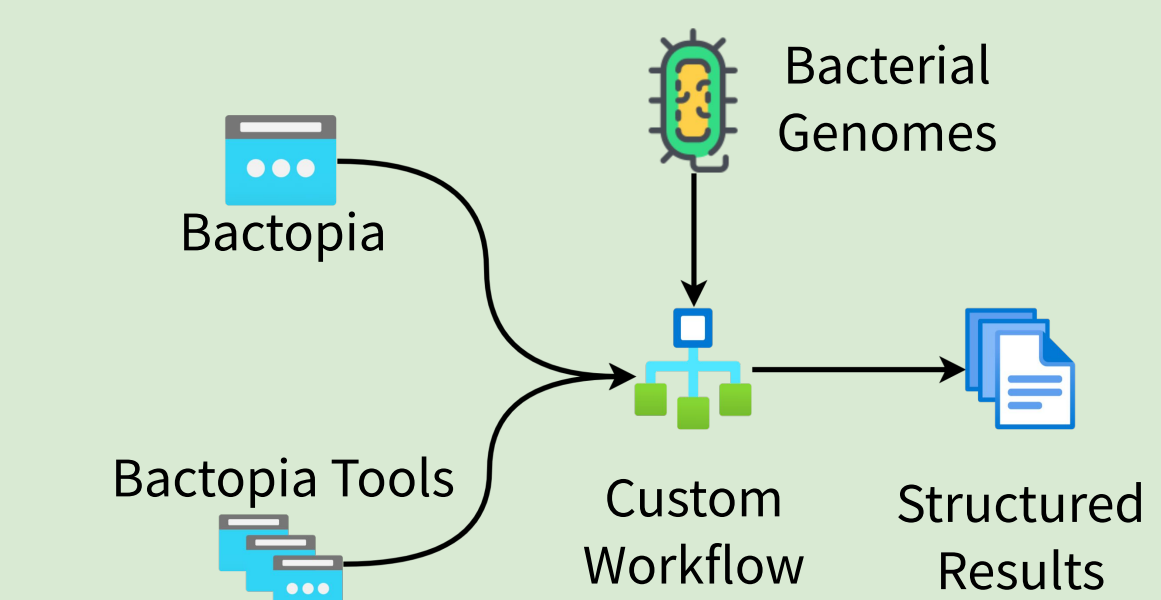


Bactopia v2, written in Nextflow DSL2, supports Nanopore reads and includes more than 130 Bioinformatic tools!



More Bactopia Tools!

Bactopia Tools are a set of additional Nextflow workflows you can conveniently run on your Bactopia outputs. Currently there are [33 Bactopia Tools](#) for analyses like pan-genome construction and serotyping. A framework was created to make the addition of new Bactopia Tools easier than ever.



Custom Workflows

With DSL2, custom workflows can now be easily created and shared with others. To demonstrate this, we have implemented [Staphopia](#), as a named Bactopia workflow. Staphopia is the combination of the Bactopia and [staph typer](#), a Bactopia Tool for analysis of *Staphylococcus aureus* genomes.

Test Everything

Bactopia v2 features per-module tests using real bacterial sequences as inputs available from [bactopia-tests](#). Currently 100+ tests have been created to verify that more than 10,000 output files. We have integrated these tests into [GitHub Actions](#) to ensure the stability of Bactopia over time.



Curated Datasets

Bactopia includes automated methods to create species-specific datasets, but these methods remain fairly broad. In order to address this we have created the [bactopia datasets](#) organization, for community members to curate species-specific datasets, such as [Staphylococcus aureus](#). Users can then integrate these curated datasets in their own analyses.

Merlin has arrived!

"MinMER assisted species-specific bactopia tool [seLectioN](#)", or *Merlin*, uses distances based on the RefSeq sketch downloaded by "*bactopia datasets*" to automatically run tools for members of *Escherichia*, *Haemophilus*, *Klebsiella*, *Legionella*, *Listeria*, *Mycobacterium*, *Neisseria*, *Salmonella*, *Staphylococcus*, and *Streptococcus*.

You can use Merlin in Bactopia with the "--ask_merlin" parameter, or alternatively you can run Merlin as a Bactopia Tool!



[nf-core](#) is a community effort to curate analysis pipelines built using Nextflow. This effort has created a powerful design framework for developing Nextflow pipelines. Bactopia v2 has followed their lead by implementing modules from [nf-core/modules](#) for Bactopia Tools, per-module tests ([nf-core/modules](#)), and an argument parser ([nf-core/tools](#)).