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

Robert A. Petit III[1,2], Davi J. Marcon[3,4], Abhinav Sharma[5], and Timothy D. Read[6]

[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
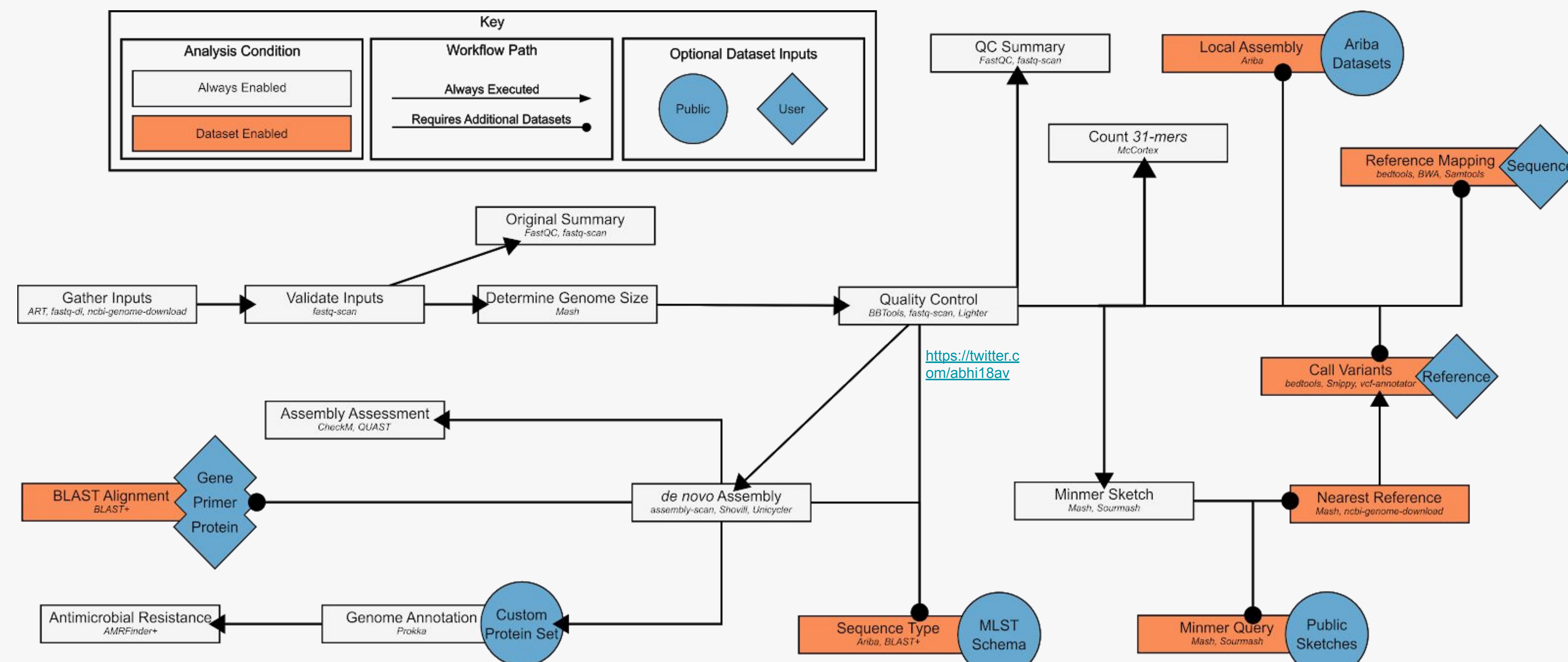
Correspondence: robert.petit@wyo.gov
Bactopia Documentation: bactopia.github.io
Bactopia v2 Overview: https://github.com/bactopia/bactopia/issues/233
Previous Bactopia Posters: https://bit.ly/3gqL1fj (full overview)
https://bit.ly/3sAslOT (processing 67,000 genomes)

Highlights

- Rewritten in Nextflow's DSL2, greatly increasing 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 70 bioinformatics tools into a single pipeline
- Combine Bactopia with Bactopia Tools to create and share reproducible custom workflows
- A framework for community curated species-specific datasets

Summary

Here we introduce version 2 of Bactopia, a significant step forward, where we have rewritten the original pipeline to take advantage of Nextflow DSL2.

You can now easily customize Bactopia to fit your species-specific needs. For example, we implemented Staphopia as a custom Bactopia workflow for analyzing *Staphylococcus aureus* genomes. This workflow supplements the standard Bactopia analysis with additional *S. aureus* specific analyses such as methicillin-resistance (MRSA), *spa* typing, and *agr* typing. Using this as a model, we expect to work alongside you and others to create additional species-specific workflows.

Bactopia (v1 and v2) is open source and has been shown to scale from a single bacterial genome to tens of thousands of genomes. Bactopia v2 maintains the scalability, while also introducing even greater flexibility.

Bactopia is a flexible pipeline for complete analysis of bacterial genomes, that allows you can to get to the fun part of analysis quicker!

🐦 @rpetit3   @tdread_emory   @abhi18av   @rpetit3   @Mxrcon   @abhi18av

Bactopia v2 introduces many changes to improve your experience as well as our ability to continue to maintain it.

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.
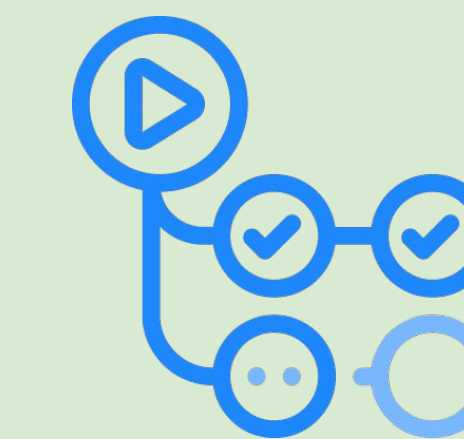
Test Everything

Bactopia v2 features per-module tests using real bacterial sequences as inputs available from bactopia-tests. Currently 65+ tests have been created to verify that more than 2,000 output files. We have integrated these tests into GitHub Actions to ensure the stability of Bactopia over time.

Custom Workflows

With Bactopia's conversion to DSL2, custom workflows can now be easily created and shared with others. These custom workflows can execute the main Bactopia pipeline, along with any additional Bactopia Tools you might want to include. For example, you might include the PIRATE Bactopia Tool to create a core-genome alignment and phylogeny of all your samples.

### nf-core/ influences

nf-core is a community effort to curate analysis pipelines built using Nextflow. This effort has created a powerful design framework when developing Nextflow pipelines, and Bactopia v2 has followed their lead by implementing:

- Modules from nf-core/modules for Bactopia Tools
- Per-module tests modeled after nf-core/modules pytest implementation
- Argument parser based off nf-core/tools
- Single meta variable for general value storage

Ultimately adoption of these practices has made Bactopia a much better pipeline to use and to maintain overtime.

(Thank you nf-core team!)
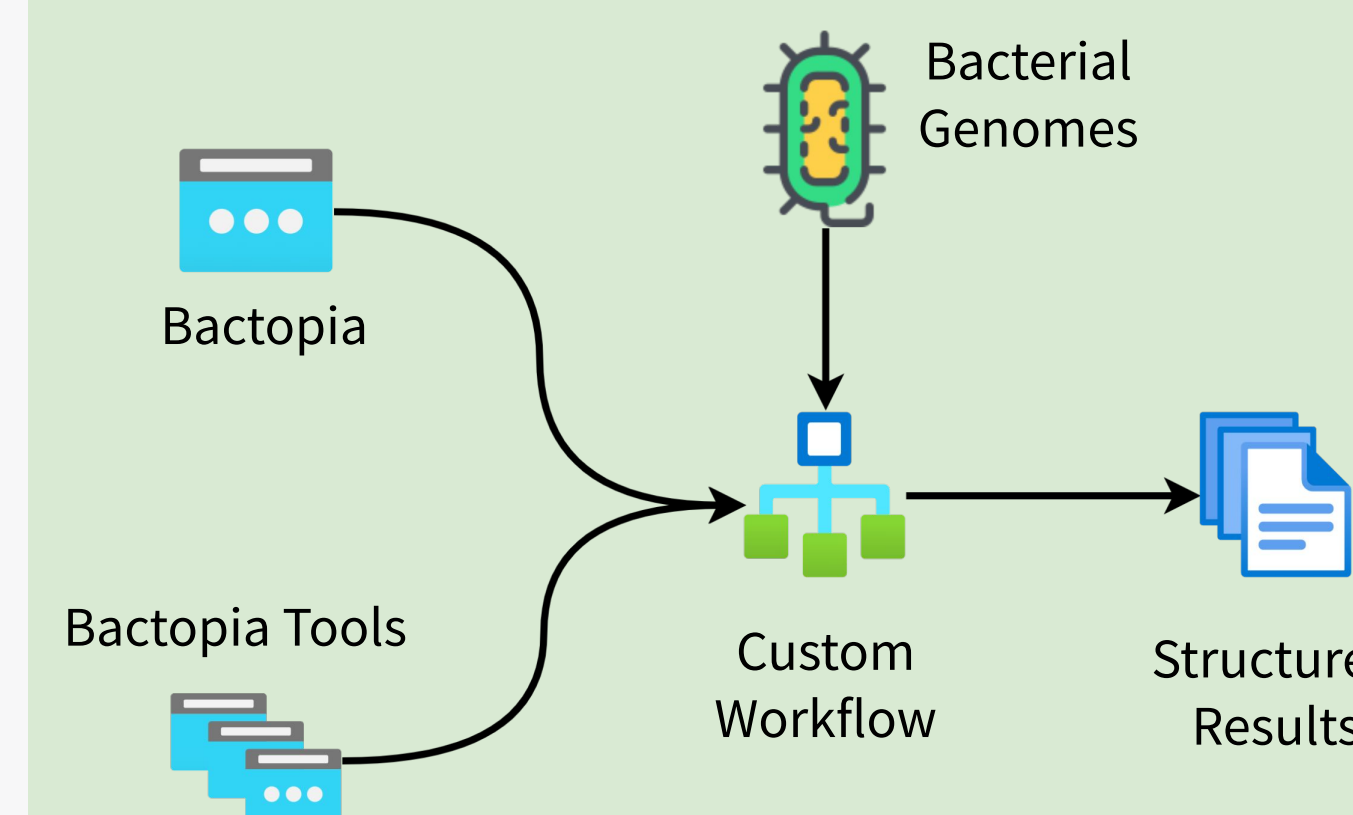
Bactopia v2 to be released soon!
Learn full details and follow progress at Bactopia v2 Overview