Bactopia v3 - Enhancements for on-going genomic surveillance of bacterial pathogens

Robert A. Petit III[1], Taylor Fearing[1], Chayse Rowley[1], Jim Mildenberger[1], and Timothy D. Read[2]

[1]Wyoming Public Health Laboratory, Wyoming Department of Health, Cheyenne, Wyoming, USA
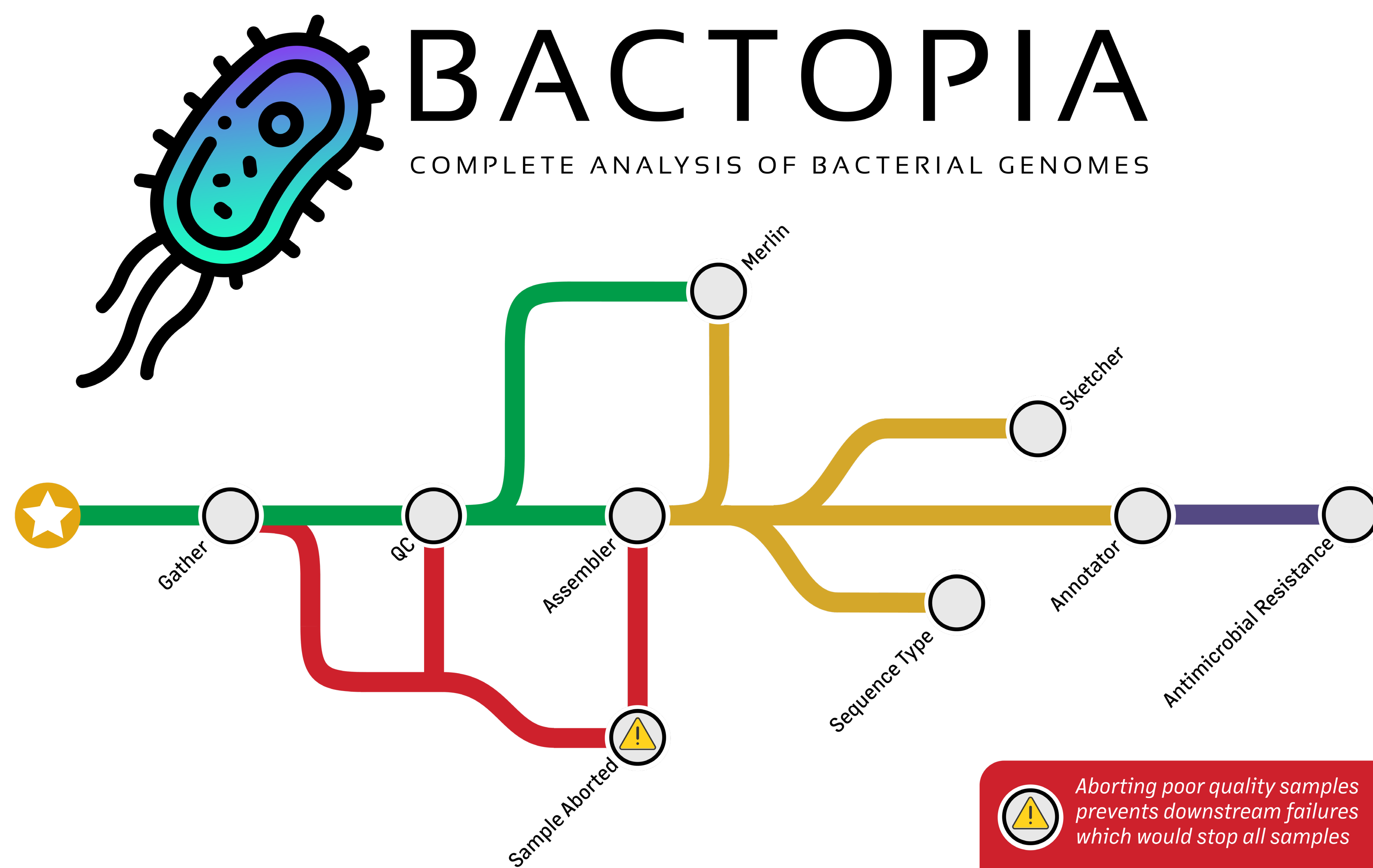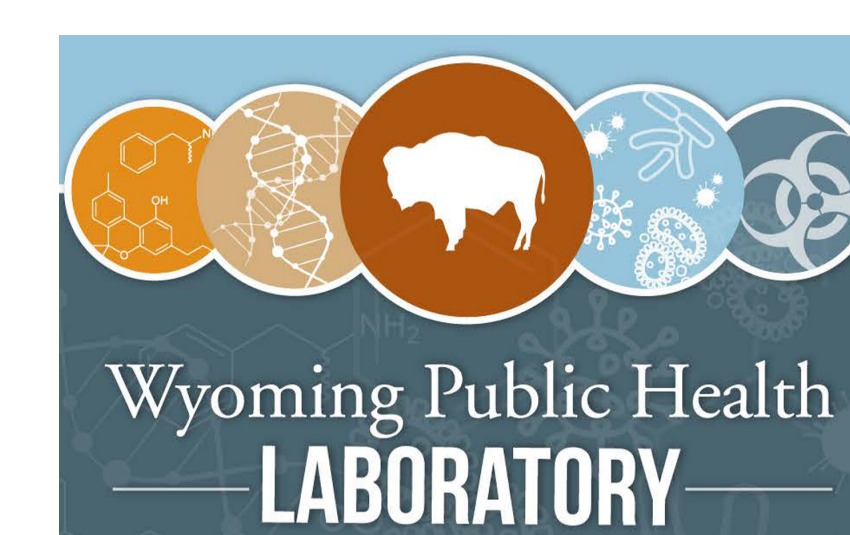[2]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia, USA

robert.petit@wyo.gov   @rpetit3   @rpetit3   Bactopia Documentation: bactopia.github.io

BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

Scan to get started with Bactopia

Legend
- Process uses FASTQs
- Process uses Contigs
- Process uses Contigs and Proteins
- Minimum QC not met, sample aborted

Pipeline nodes: Gather, QC, Assembler, Merlin, Sketcher, Sequence Type, Annotator, Antimicrobial Resistance, Sample Aborted

Bactopia Processes

Gather
Collect local files and/or download sequences from SRA/ENA/DDBJ or NCBI Assembly accessions

QC
Trim and filter low quality reads, subsample to specified coverage, and generate summary metrics

Assembler
Create a de novo assembly (standard, hydrbid, or short read polished) and summary metrics

Merlin
Use Mash distances to automatically execute species specific tools, requires --ask_merlin

Sketcher
Create minmer sketches and query them against GTDB and RefSeq

Annotator
Predict genes and proteins from the assembled contigs

Antimicrobial Resistance
Identify presence of AMR and/or virulence genes

Sequence Type
Determine sequence type base on PubMLST profiles

⚠ Aborting poor quality samples prevents downstream failures which would stop all samples

- Too few reads or basepairs
- Coverage below minimum
- Paired-end with different read counts, mismatched IDs, or skewed proportions
- Genome size outside expectation
- 0 assembled contigs
- Assembled size below minimum

Abort Reasons

Accepted Inputs

Illumina and/or Nanopore Reads (fastq)
--R1/--R2, --SE, --ont. --hybrid, --sample
--samples with 'bactopia prepare' file-of-filenames

Assemblies (fasta)
--assembly, --sample
--samples with 'bactopia prepare' file-of-filenames

DDBJ/ENA/SRA Accessions (SRX)
--accession 'Experiment Accession'
--accessions with 'bactopia search' results

NCBI Assembly Accessions (GCF)
--accession 'Assembly Accession'
--accessions 'file with accessions'

Bactopia v3 Highlights

Version 3 of Bactopia brings many new changes, including:

- Simpler installation and a streamlined workflow to quickly get started

- Sample centric runs for on-going surveillance

- Many new Bactopia Tools

- Additional workflows for non-bacterial samples

- Greatly reduced maintenance burden

ACCESSIBLE
Bactopia is open source and available from Conda, Docker and Singularity. Bactopia has been downloaded over 1,000,000 times from users around the world.

COMPREHENSIVE
Bactopia is a start-to-finish pipeline which includes over 150 numerous tools and 60 workflows commonly used for bacterial genome analysis.

PORTABLE
With a simple profile change you can go from processing genomes on your laptop to an HPC system or any of the major cloud providers (AWS, GCP, Azure).

REPRODUCIBLE
Bactopia was developed following nf-core best practices which ensures a robust pipeline with strict version control and an extensive audit trail.

RESILIENT
More than 100 tests, testing 10,000+ variables, assist in identifying potential bugs and downstream changes, before users are affected.

SCALABLE
Bactopia allows you to easily scale from a few genomes to thousands of genomes. For example processing 67,000 genomes in 5 days on AWS.

STEP 01
Sequence Bacterial Genomes
Bactopia supports sequences from Illumina and Nanopore technologies, as well as Sequence Read Archive and NCBI Assembly accessions

STEP 02
Install Bactopia
Bactopia is easily installed using Conda, and anlyses can be executed using Conda, Docker or Singularity

STEP 03
Create a Sample Sheet
bactopia prepare is used to create a sample sheet allowing a single sample or thousands of samples to be processed

STEP 04
Include Publc Samples
bactopia search can query ENA/SRA to identify accessions to further supplement analyses

STEP 05
Choose Compute Environment
Built with Nextflow, Bactopia supports many compute environments such as local, HPC, and major cloud providers

BACTOPIA STEP 06
Process Genomes
Bactopia includes 150+ bioinformatic tools, with robust version control, data provenance, and readily accessible outputs in a standardized format

BACTOPIA TOOLS STEP 07
Additional Workflows
More than 60 Bactopia Tools enable effortless post-analysis of outputs, covering pangenome, phylogenies, AMR, and more

STEP 08
Investigate Results
Once finished with Bactopia, a wide range of output files are accessible to support your genomic investigations

Wyoming Public Health LABORATORY

EMORY UNIVERSITY SCHOOL OF MEDICINE