Expanding Bactopia as a framework for bacterial, fungal, and metagenomic analysis

Taylor Fearing, Chayse Rowley, Robert A. Petit III, Jim Mildenberger, Rob Christensen, Joseph M. Reed

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

Wyoming Department of Health

PUBLIC HEALTH DIVISION

BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

Legend

- Process uses FASTQs
- Process uses Contigs
- Process uses Contigs and Proteins
- Minimum QC not met, 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
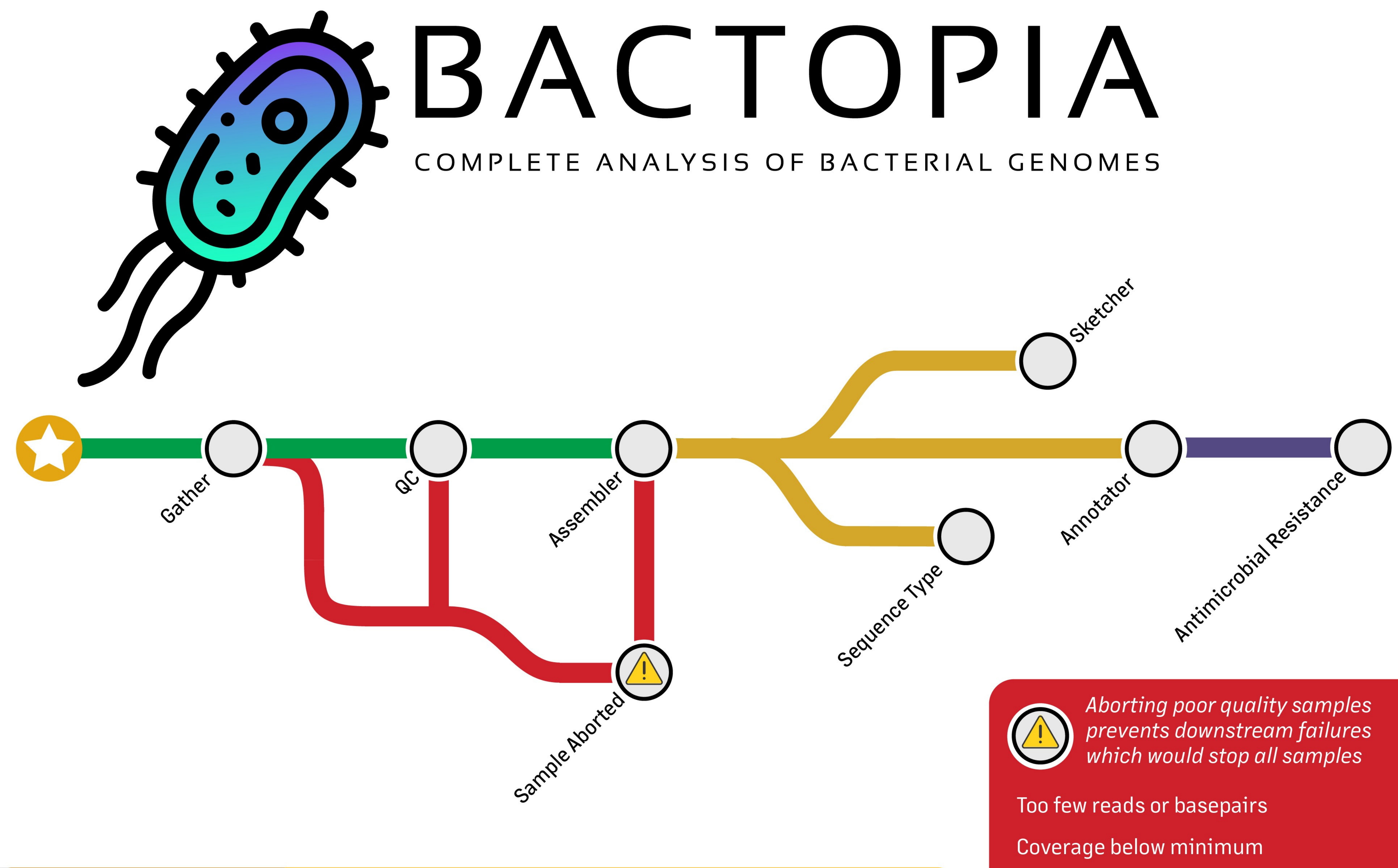Create a de novo assembly (standard, hydbrid, or short read polished) and summary metrics

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

Bactopia Tools
60+ additional workflows for seamless downstream analyses such as pan-genome contruction, serotyping, AMR, and phylogenies

(Pipeline diagram nodes: Gather → QC → Assembler → Sequence Type / Sketcher → Annotator → Antimicrobial Resistance; Sample Aborted)

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'

Abort Reasons

⚠️ *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

Bactopia Highlights

Broad Sequencing Support
Compatible with both Illumina and Nanopore sequencing technologies

Modular Pipeline
A comprehensive, end-to-end modular pipeline for flexible genomic analysis

Extensive Toolset
Incorporates over 150 bioinformatic tools for diverse research needs

Reliable Updates
Actively supported and consistently maintained for over 5 years

Widespread Adoption
Achieved over 1,000,000 downloads, reflecting global trust and utility

Community-Driven
Contributed over 180 enhancements back to the open-source community

AllTheBacteria Support
Supports further analysis of the 2,000,000 assemblies from the AllTheBacteria project

Shape Your Science

Each process in Bactopia functions as an independent module, enabling straightforward reshaping to develop custom workflows tailored to your specific needs
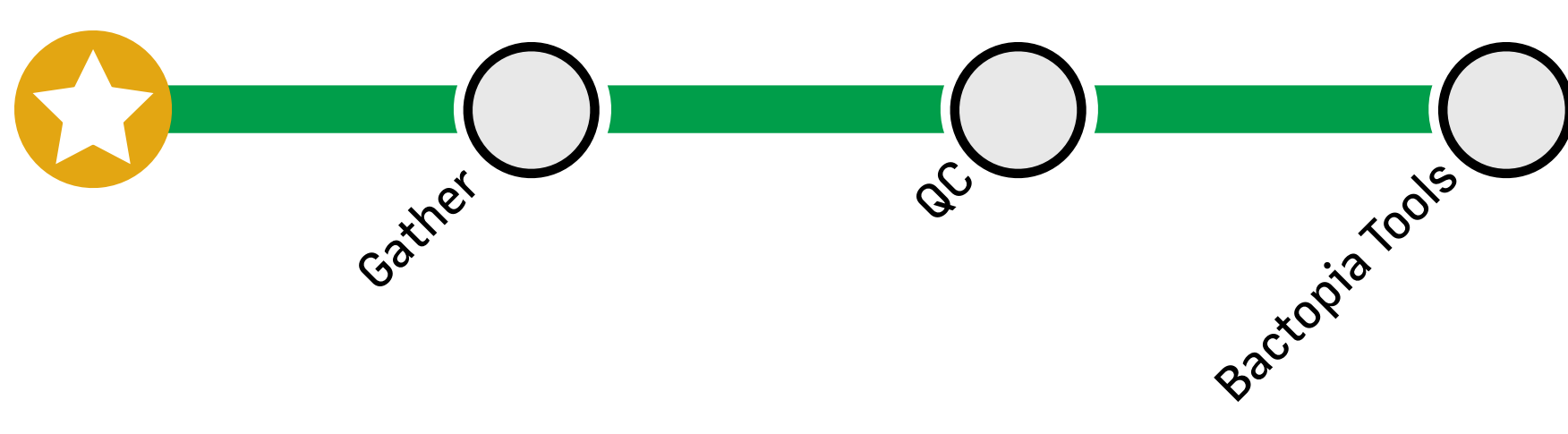
Here are just a few examples
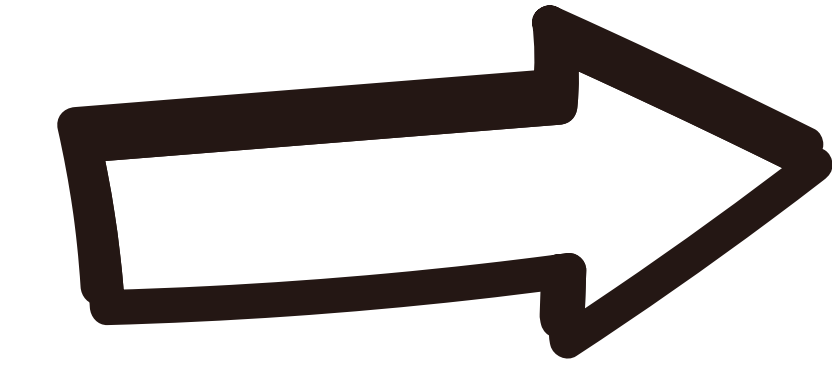
clean-yer-reads

clean-yer-reads is a simple reshaping of Bactopia that enables you to QC sequences across any organism, including metagenomic samples.

With clean-yer-reads you still have full access to all available Bactopia Tools.

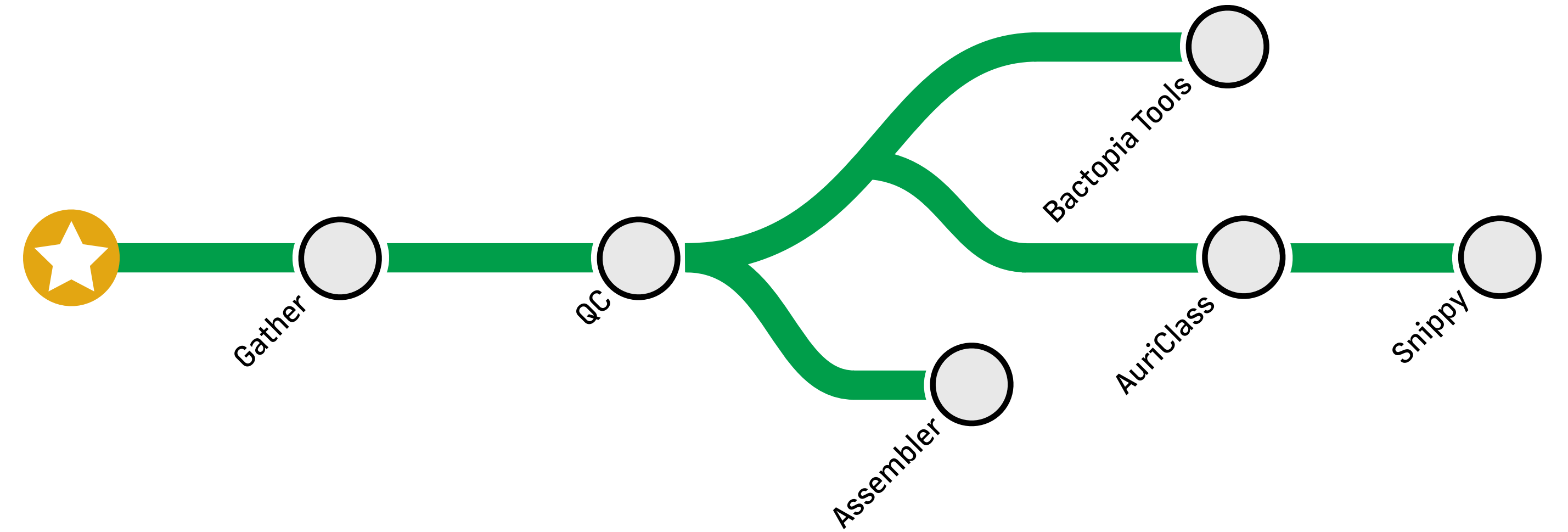*(Pipeline: Gather → QC → Bactopia Tools)*

teton

teton reshapes Bactopia with a focus on human read scrubbing and taxonomic classification, by utilizing the existing Scrubber and the Bracken modules.

teton makes for a useful first step for all sequence analysis.

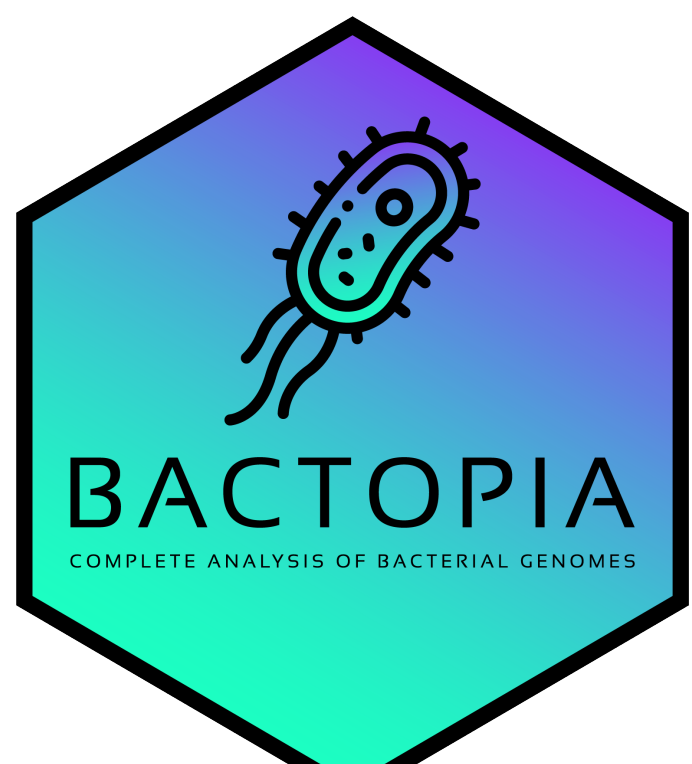*(Pipeline: Gather → Scrubber → Bracken)*
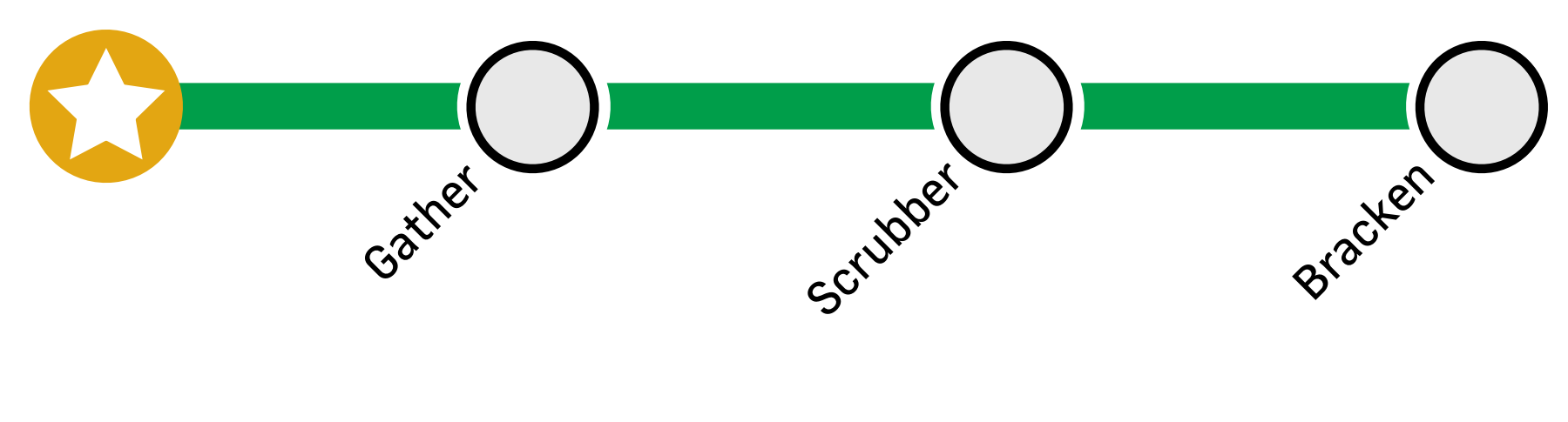
mycotopia

mycotopia represents a specialized adaptation of Bactopia, specifically for the analysis of non-bacterial samples. It has been reshaped for haploid fungal organisms, including the notable fungal pathogen *Candida auris*. It serves as a prime example of Bactopia's flexibility in meetings its users' needs

(Pipeline: Gather → QC → Assembler / Bactopia Tools → AuriClass → Snippy)

Scan to learn more!

BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

Ask for a sticker!