Introduction to Bactopia for Bacterial Genome Analysis

Robert A. Petit III, PhD
2025 AMD Molecular Epidemiology Training Series
August 15th, 2025

Wyoming Department of Health
PUBLIC HEALTH DIVISION
WYOMING PUBLIC HEALTH LABORATORY

What is "Bactopia"?

Let's introduce Bactopia and its impact

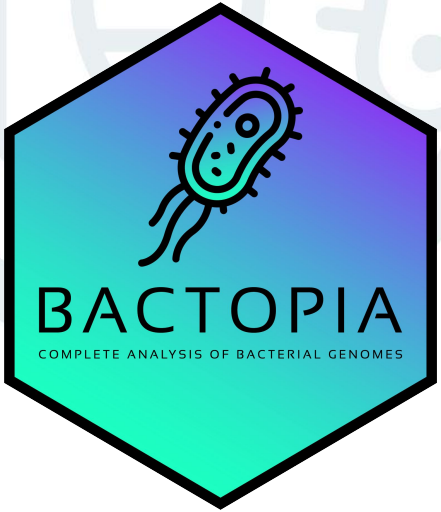
Bactopia streamlines bioinformatics analysis

- ***End-to-end pipeline for bacterial genome analysis***
  - Written in Nextflow, following nf-core standards
  - Available from Conda, Docker, or Singularity
  - Supports numerous compute infrastructures

- ***Wraps 150+ bioinformatic tools into stand-alone modules***

- ***Active user-base providing regular feedback***
  - Includes folks at WPHL, as well as global users from academia, government, and commercial

- ***Consistently maintained for 5+ years***



bactopia.github.io

150+ citations, 467 GitHub Stars

Used in all sorts of bacterial genomic studies

50,000+ unique visitors

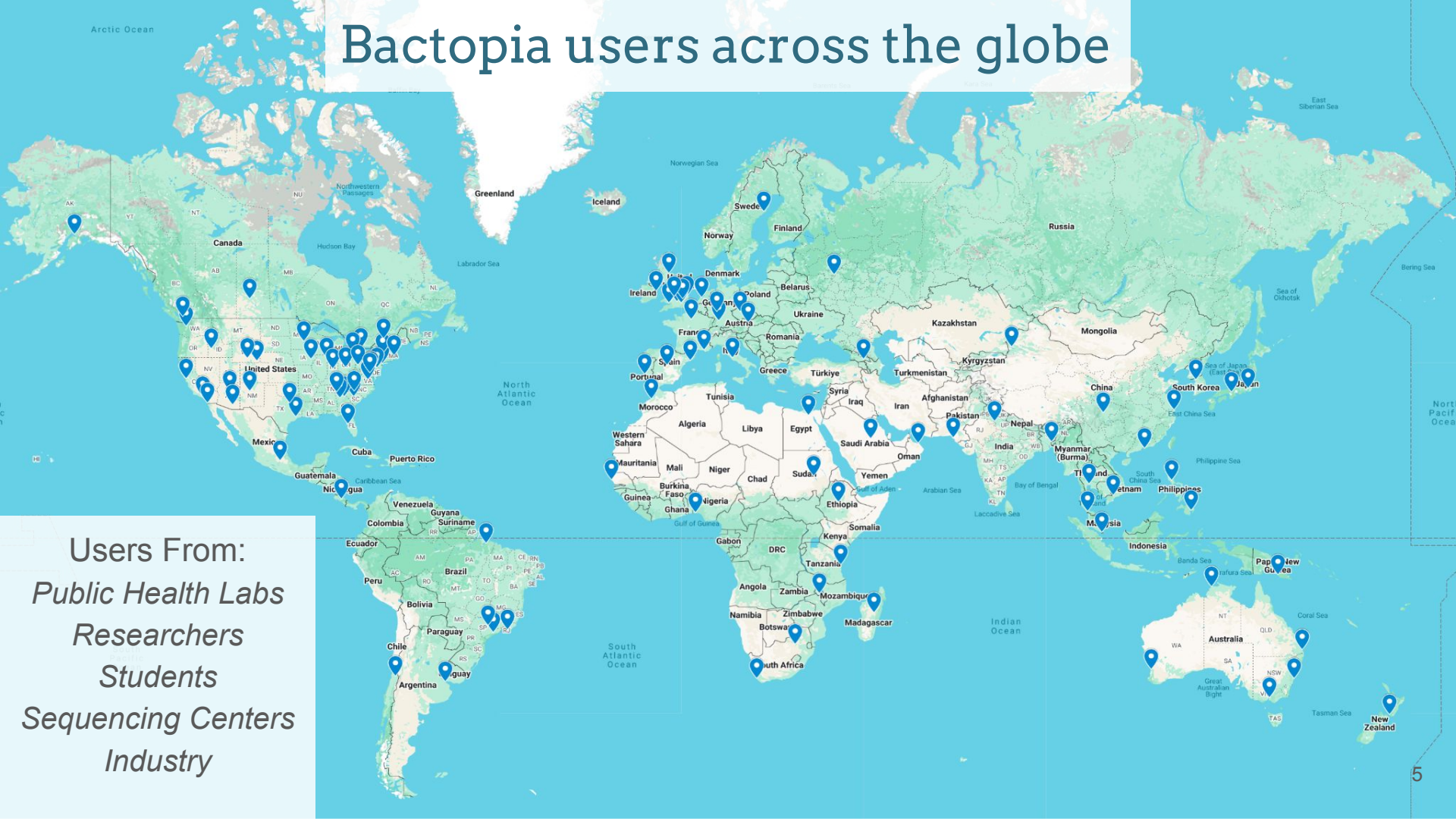
Many users from around the world are visiting the docs

1,000,000+ downloads

That's nearly twice as many people here in Wyoming!

Wyoming Department of Health | PUBLIC HEALTH DIVISION | WYOMING PUBLIC HEALTH LABORATORY

Bactopia users across the globe



Users From:

Public Health Labs

Researchers

Students

Sequencing Centers

Industry

People behind Bactopia

Let’s put some faces to Bactopia



Yo! 👋

I AM ROBERT

WPHL's bioinformatician and developer of Bactopia

Wyoming Department of Health | PUBLIC HEALTH DIVISION | WYOMING PUBLIC HEALTH LABORATORY

People behind the scenes of Bactopia



Tim Read, PhD

Emory University, Professor

Tim has played a role in Bactopia since its inception. Through the years Tim has provided feedback and ideas to help shape Bactopia.



Joseph Reed, PhD

WPHL, Laboratory Administrator

As the WPHL lab administrator, Joe encourages the lab to pursue the development of skills and tools like Bactopia to strengthen WPHL.



Jim Mildenberger

WPHL, Molecular Lab Supervisor

Jim keeps the molecular lab running. Like Joe, Jim's support has helped introduce many new features (e.g., ONT support) into Bactopia.



Taylor Fearing

WPHL, EID & NGS Supervisor

Taylor oversees the sequencing at WPHL. She has played a tremendous and critical role in helping to expand Bactopia into public health.

BACTOPIA

Bactopia Users

All Sorts of Fields across the World

Bactopia user's are regularly providing feedback such as new features, improvements and bugs identified.

What's the Bactopia pipeline?

Let's take a look at the Bactopia pipeline

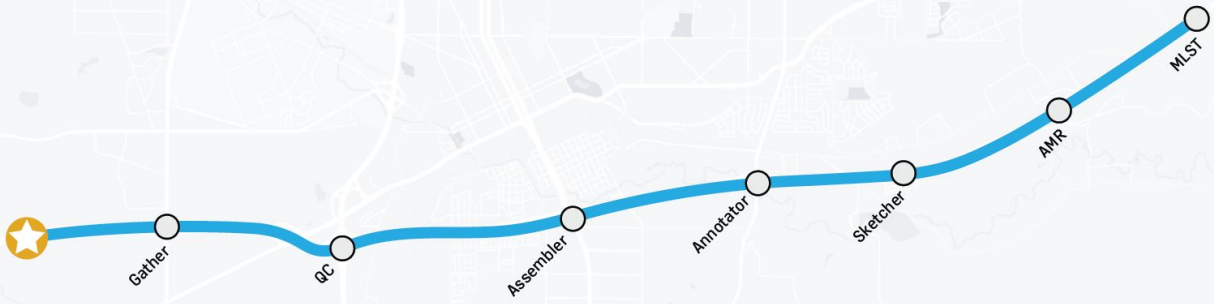


Every pipeline needs a metro map

But here in the U.S., metros are like unicorns...


So... Bactopia gives you a road map instead!

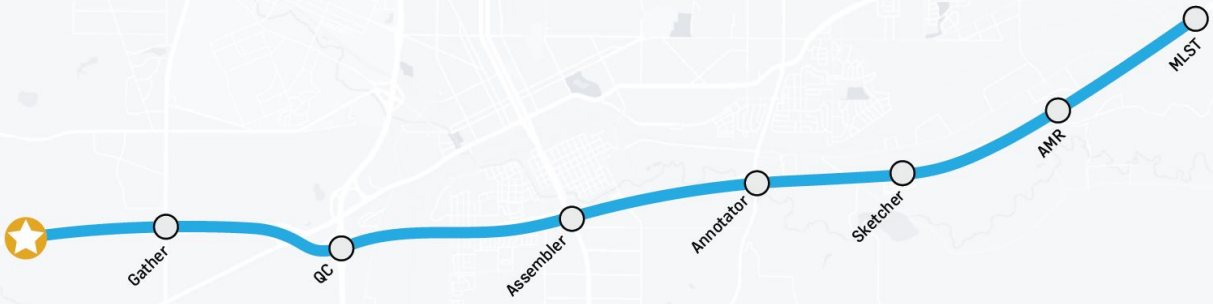
Bactopia includes routine yet essential steps in bacterial genome analysis.



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

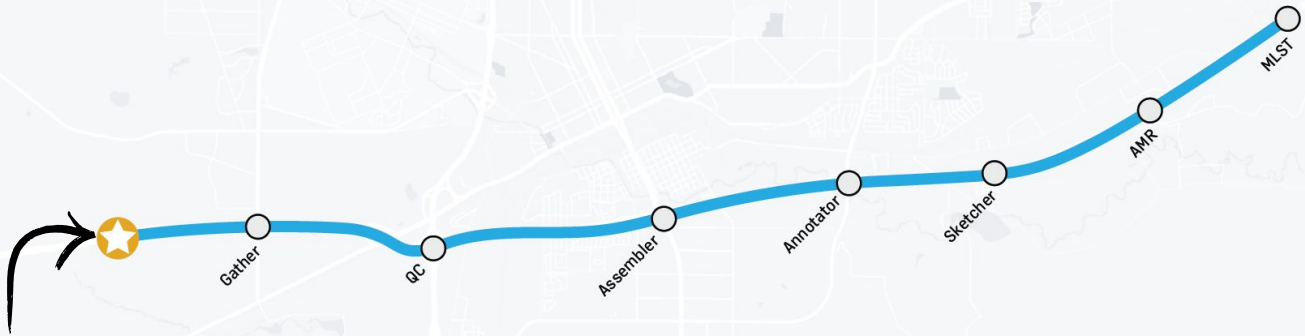
From the start, Bactopia is flexible, supporting short reads, long reads, and assembled genomes.



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

From the start, Bactopia is flexible, supporting short reads, long reads, and assembled genomes.



fastq **Short and Long Reads**

SRX **Reads from SRA/ENA**

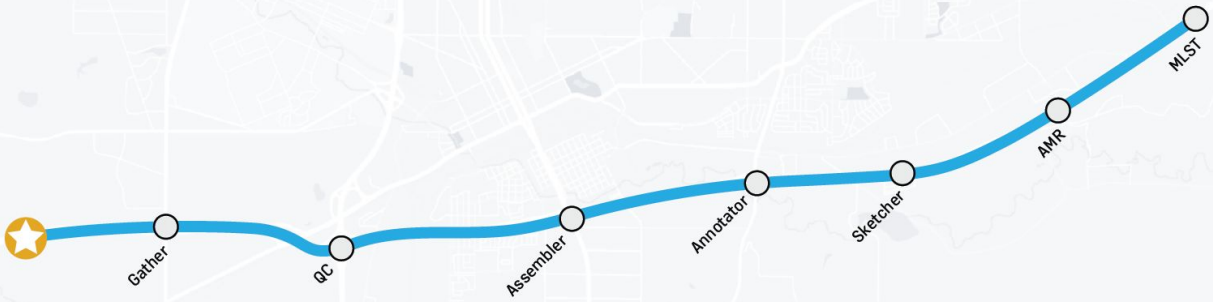
fasta **Assembled genomes**

GCF **Assemblies from NCBI**

BACTOPIA

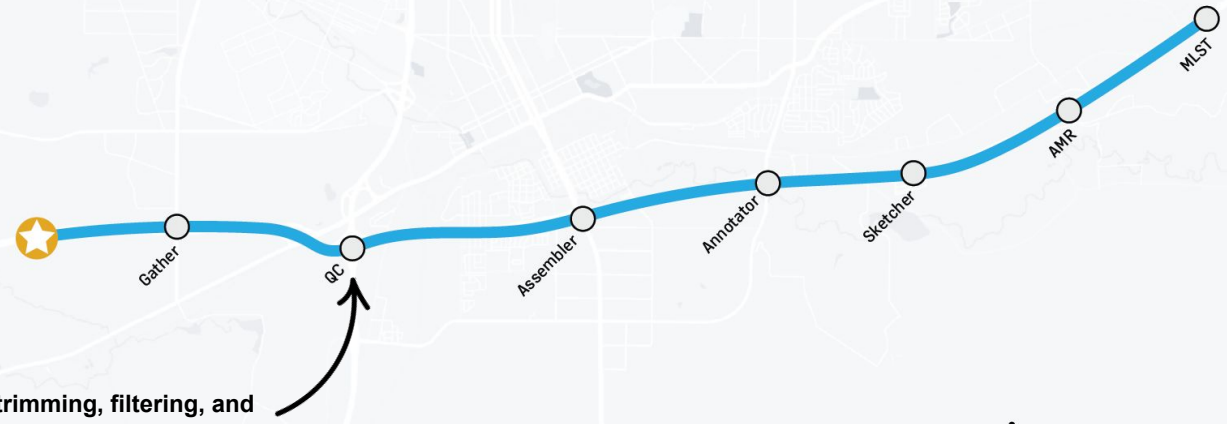
COMPLETE ANALYSIS OF BACTERIAL GENOMES

Sequences are automatically assessed and cleaned to ensure reliable downstream results.



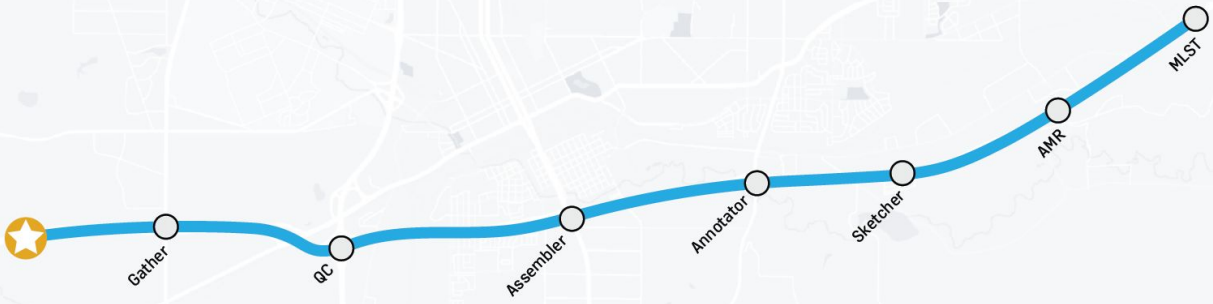
BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

Sequences are automatically assessed and cleaned to ensure reliable downstream results.



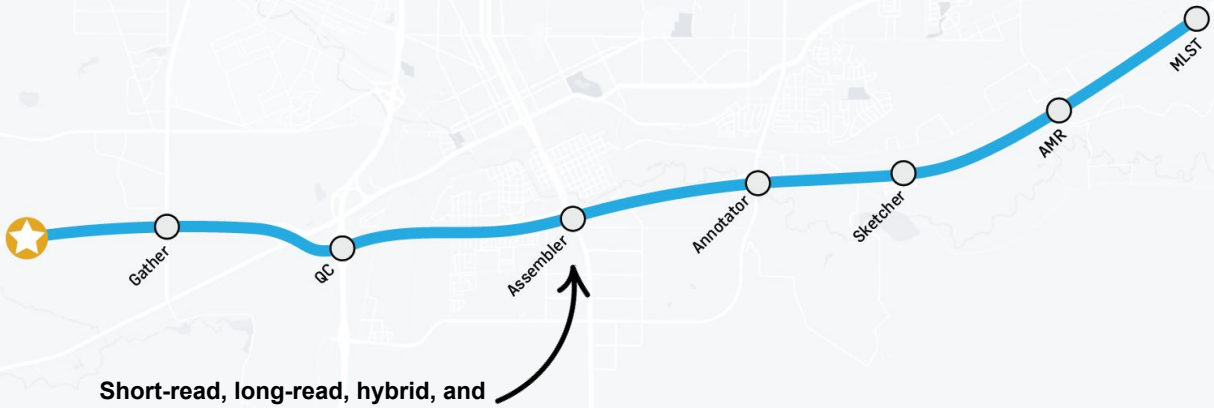
BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

Bactopia adapts to your data, by utilizing multiple assembly strategies out of the box.



BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

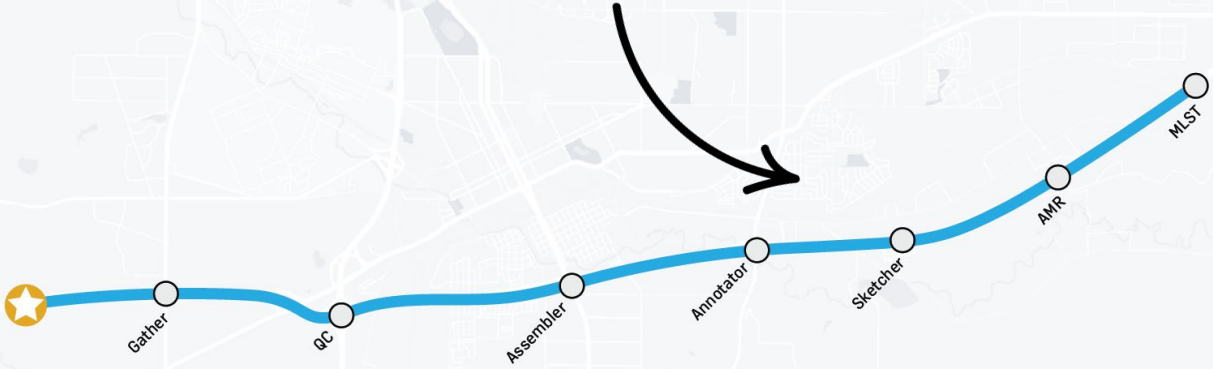
Bactopia adapts to your data, by utilizing multiple assembly strategies out of the box.



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

Bactopia's final steps deliver annotations, AMR predictions, MLST typing, and genome sketches



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

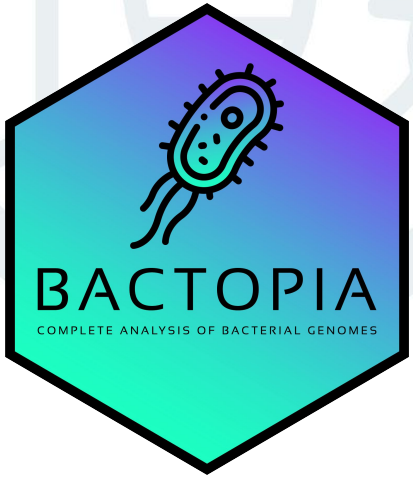
Bactopia – Streamlines bacterial genome analysis

- These steps will take 5-20 minutes
  - *Depending on input sizes and available resources*

- Output files are placed in a standardized directory structure

- Extensive audit trails are available
  - Logs, staging, versions, etc…

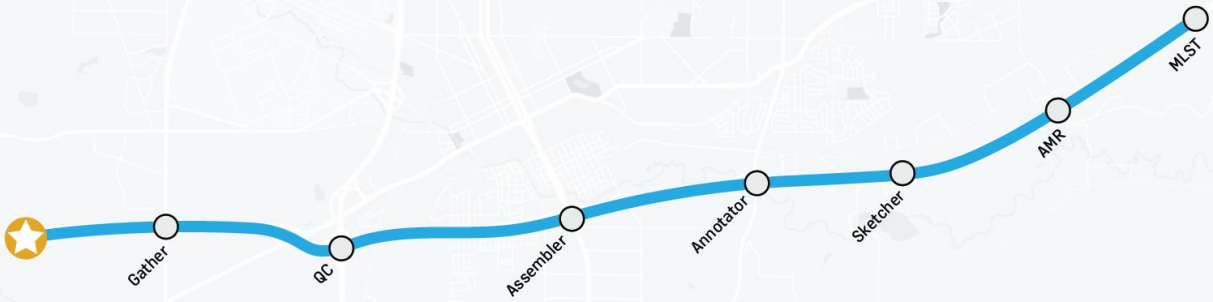
- When possible, outputs are automatically merged



What's else does Bactopia do?

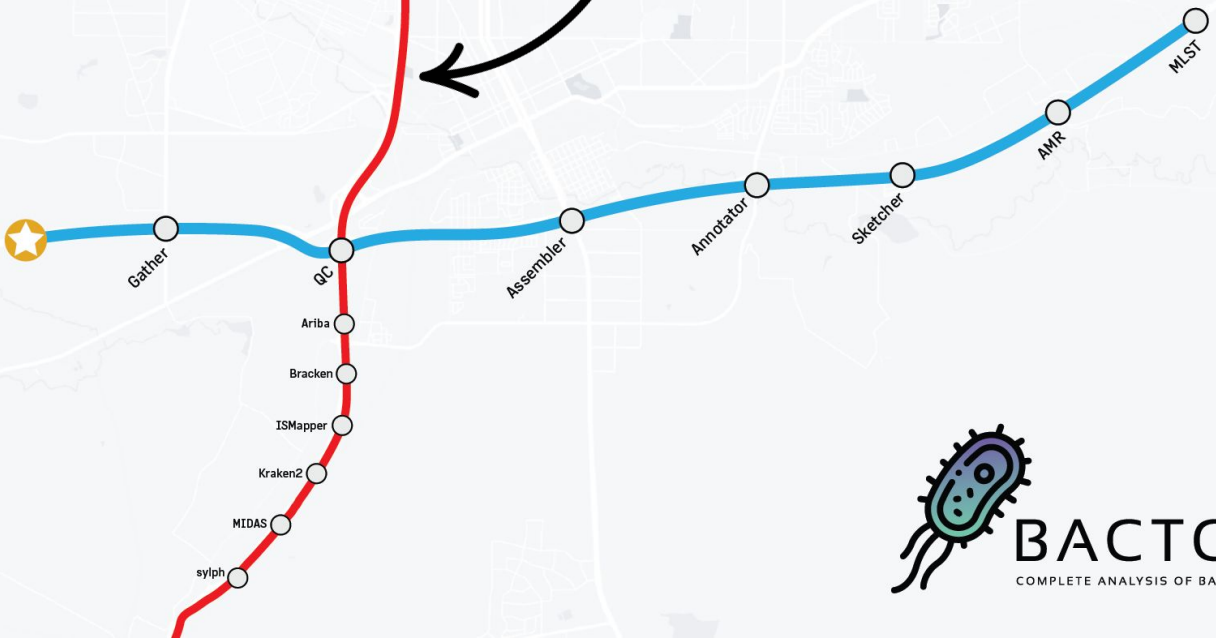
Let's take a look at the Bactopia Tools

Bactopia's core steps structure your data for smooth integration into downstream workflows.



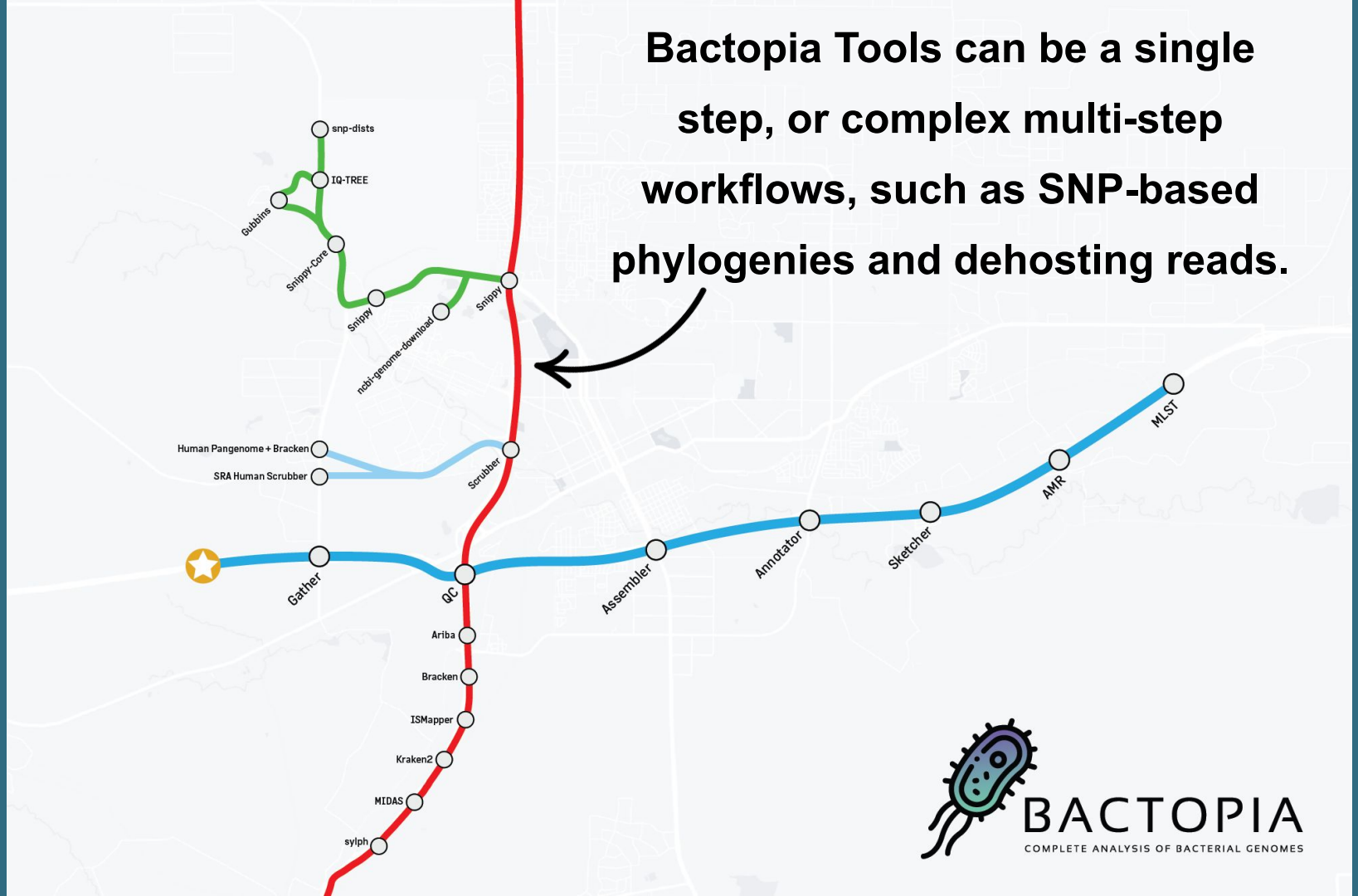
BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

With a simple parameter change, QC'd FASTQs can be reused across multiple downstream workflows called "Bactopia Tools."



BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

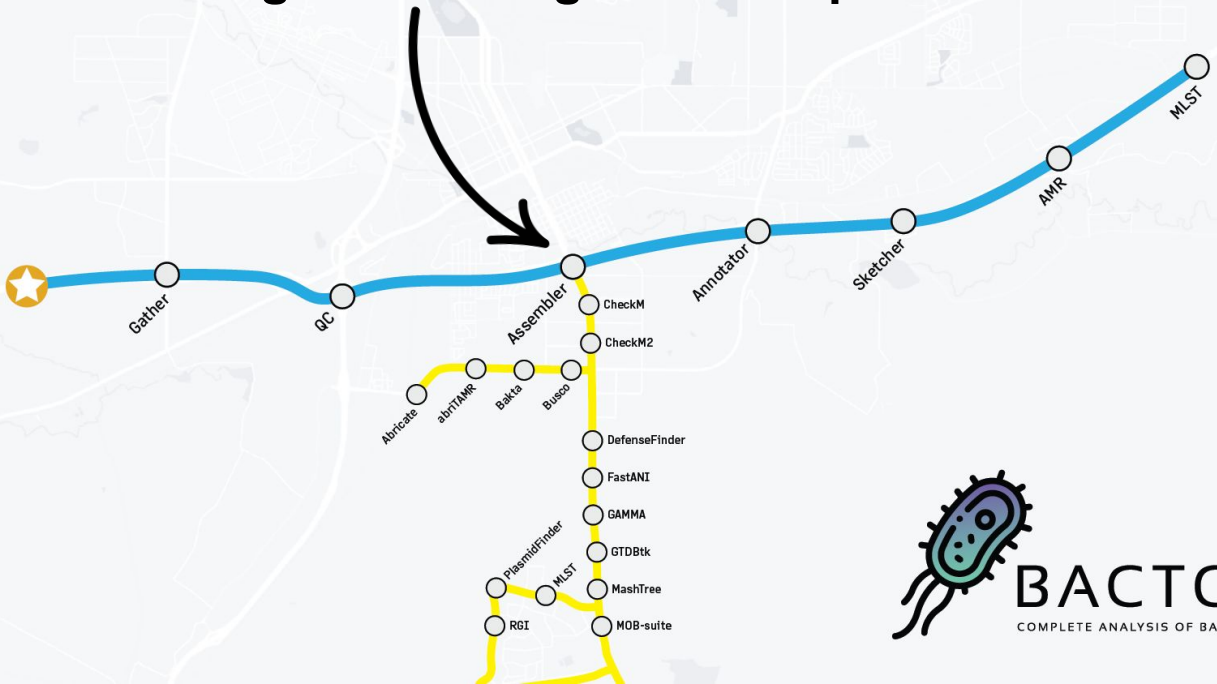
Bactopia Tools can be a single step, or complex multi-step workflows, such as SNP-based phylogenies and dehosting reads.



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

As with QC'd FASTQs, assembled genomes can be processed using a wide range of Bactopia Tools.



BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES

Species-specific tools can be run directly on FASTQs and assemblies, or auto-selected based on genomic relatedness using Merlin.



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

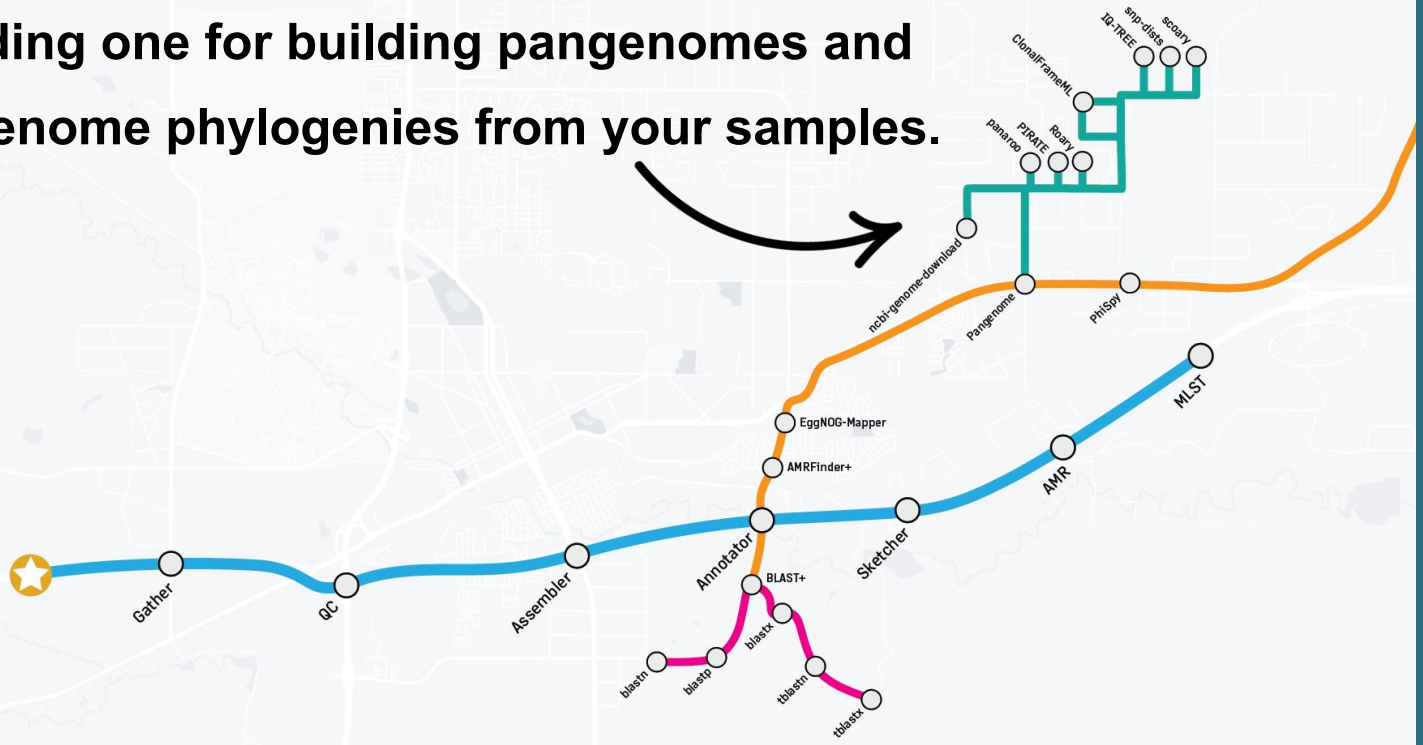
After annotation, there are more Bactopia Tools available for you to make use of.



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

Including one for building pangenomes and core-genome phylogenies from your samples.



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

Bactopia Tools – Simplifying comparative genomics

- 65+ workflows for more science
  - Uses standardized output structure
  - Single parameter change *(--wf)*

- Categories
  - Organism-specific
  - Mobile elements
  - Antimicrobial resistance & virulence
  - Pan-genome
  - SNP/InDel
  - Taxon Classification
  - Phylogeny

Example Bactopia Tool Usage

```
# Process Staphylococcus aureus samples
bactopia --samples saureus.txt

# Run Bactopia Tools

# Staph-specific tools (agr, sccmec, spa)
bactopia --wf staphtyper

# Call SNPs, build core-snp tree
bactopia --wf snippy --accession GCF_000009645
```

Bactopia supports AllTheBacteria

2,500,000 bacterial assemblies at your fingertips!

Bactopia supports AllTheBacteria (ATB)

- 2,500,000 bacterial assemblies
  - allthebacteria.readthedocs.io

- **`bactopia atb-downloader`**
  - *Download ATB assemblies by BioSample, TaxID, or Species*

- **`bactopia atb-downloader`**
  - *Arrange ATB assemblies for Bactopia Tool usage*

Example ATB Usage

```
# Download Staphylococcus aureus samples
bactopia atb-downloader --query 1280

# Format for Bactopia Tool usage
bactopia atb-formatter --path atb-assemblies

# Staph tools, plasmid, and core-genome tree
bactopia --wf sccmec
bactopia --wf mobsuite
bactopia --wf pangenome
```

Bactopia supports AllTheBacteria (ATB)

```
[skipped  ] process > BACTOPIATOOLS:DEFENSEFINDER:DEFENSEFINDER_UPDATE (update)          [100%] 1 of 1, stored: 1 ✓
[6d/12d9ac] process > BACTOPIATOOLS:DEFENSEFINDER:DEFENSEFINDER_RUN (SAMN31104671)       [100%] 644788 of 644788, failed: 89, retries: 89 ✓
[6d/9dbd16] process > BACTOPIATOOLS:DEFENSEFINDER:GENES_CONCAT (defensefinder-genes)     [100%] 1 of 1 ✓
[27/beb48b] process > BACTOPIATOOLS:DEFENSEFINDER:HMMER_CONCAT (defensefinder-hmmer)     [100%] 2 of 2, failed: 1, retries: 1 ✓
[9a/3dd68c] process > BACTOPIATOOLS:DEFENSEFINDER:SYSTEMS_CONCAT (defensefinder-systems) [100%] 3 of 3, failed: 2, retries: 2 ✓
[01/410378] process > BACTOPIATOOLS:CUSTOM_DUMPSOFTWAREVERSIONS (1)                    [100%] 1 of 1 ✓

    Bactopia Tools Execution Summary
    ---------------------------------
    Workflow         : defensefinder
    Bactopia Version : 3.1.1
    Nextflow Version : 24.04.4
    Command Line     : nextflow run /home/rpetit/.conda/envs/bactopia-dev//share/bactopia-3.1.1/main.nf -w /gscratch/rpetit/atb/atb-assemblie
    Profile          : arcc
    Resumed          : false
    Completed At     : 2024-10-31T19:22:06.055328084-06:00
    Duration         : 6d 3h 59m 33s
    Success          : true
    Exit Code        : 0
    Error Report     : -
    Launch Dir       : /pfs/tc1/gscratch/rpetit/atb/atb-assemblies/salmonella_enterica
    Merged Results   : bactopia//bactopia-runs/defensefinder-20241025-152333
```

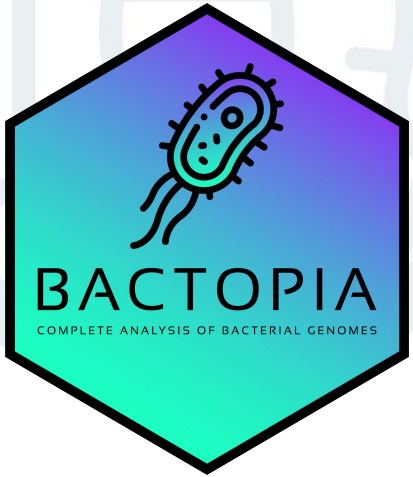
Let's Wrap This Up

What's next and all that fun stuff

Many things are brewing

- Bactopia version 4 is in the works
  - Compatibility with future Nextflow (>25) versions
  - Long overdue visual reports
  - More features and QOL improvements
  - See GitHub at: bactopia/bactopia/issues/628

- Revamp of the docs
  - Getting started on your system
  - Expanded tutorials
  - Workshop materials

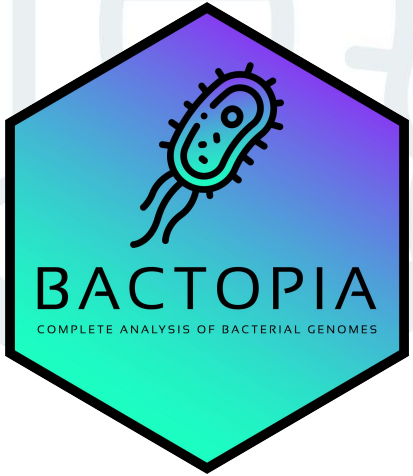


Bactopia streamlines bacterial genome analysis

Bactopia is an extensive pipeline that allows its users to focus more on the science behind their samples.

It will continue to be sustained for 5+ years through support from WPHL, Emory University, CAPE, and CZI.

There are many around the world utilizing Bactopia, and we are still only scratching the surface of where we plan to take it.



WYOMING PUBLIC HEALTH LABORATORY

EMORY UNIVERSITY SCHOOL OF MEDICINE

Center for Applied Pathogen Epidemiology and Outbreak Control

OPEN SCIENCE PROGRAM

Acknowledgements

The many developers of open source software and the users of Bactopia that are regularly providing feedback.

WPHL
Taylor Fearing
Elena Groves
Maren Anderson
Chayse Rowley
Jim Mildenberger
Rob Christensen
Joseph Reed

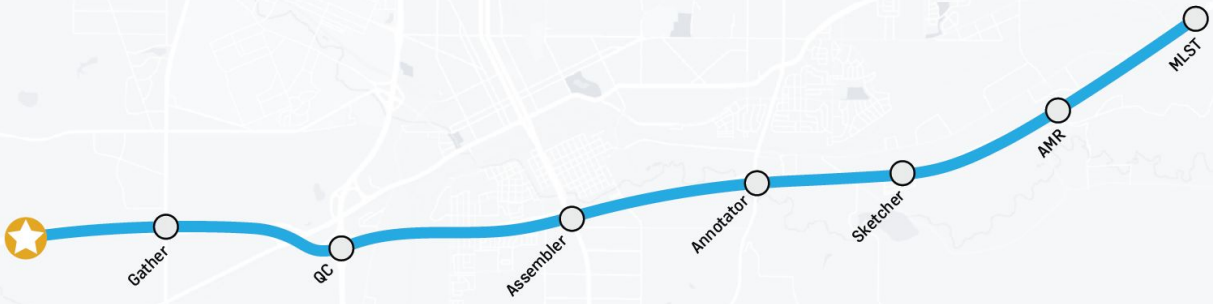
Emory
Tim Read
EMERGENT Group

Rush
Ahmed Babiker

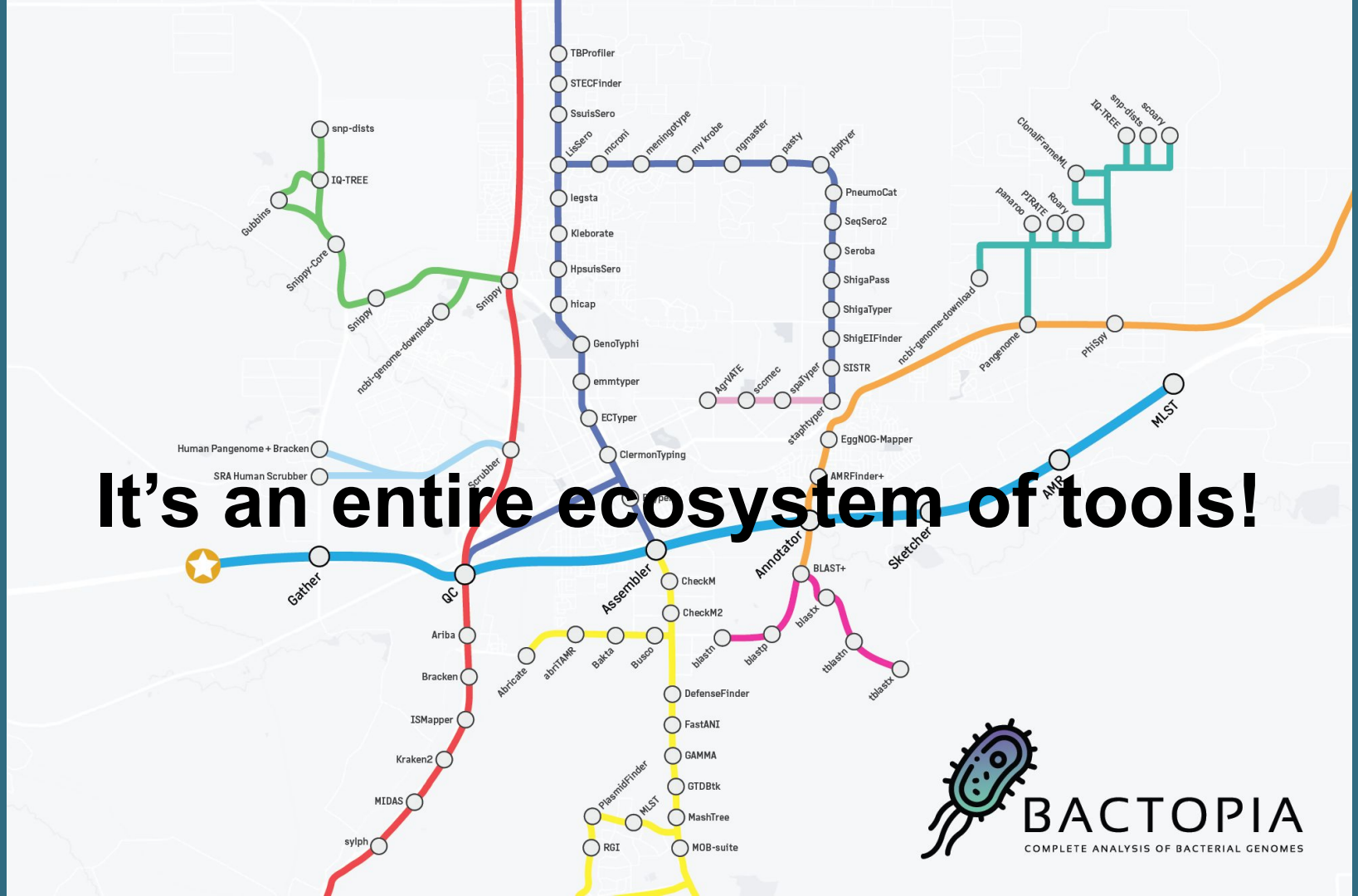


One final message…

Bactopia is more than “just a pipeline”



BACTOPIA
COMPLETE ANALYSIS OF BACTERIAL GENOMES



It’s an entire ecosystem of tools!

BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

Thank you! Happy to take any questions

Learn more about Bactopia at
bactopia.github.io

Contact Info
robert.petit@wyo.gov

Wyoming Department of Health | PUBLIC HEALTH DIVISION | WYOMING PUBLIC HEALTH LABORATORY