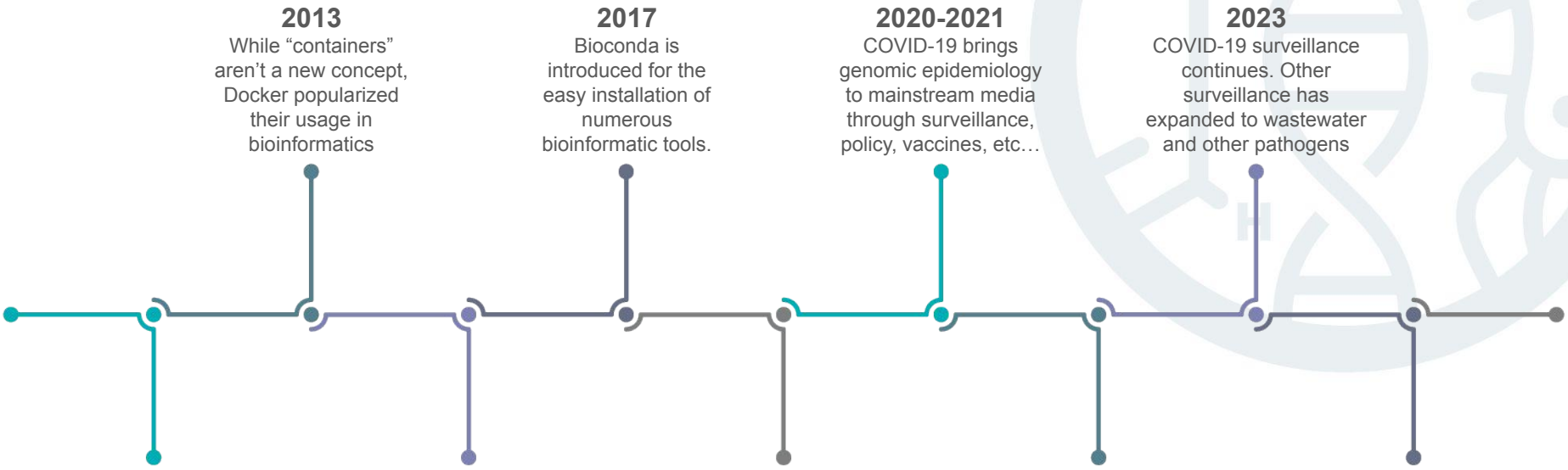
A brief history of workflow managers in bioinformatics

Robert A. Petit III, PhD
AMD Days 2024
September 24th, 2024

Wyoming Department of Health
PUBLIC HEALTH DIVISION
WYOMING PUBLIC HEALTH LABORATORY

Bioinformatics

A public health bioinformatician's view of the last 15 years



Where is bioinformatics accessibility lagging?

- Useability
  - Many tools are not intuitive or lack example use cases
- Documentation
  - For most tools the “--help” text is the documentation
- Sustainability
  - Many bioinformatic tools have a very short lifecycle

Accessibility in the context of US Public Health Labs (PHLs)

Many PHLs are looking for “practical” bioinformatics

Ready to use workflows, that are simple to use, well documented and a track record of long-term maintenance.

PHLs, *especially rural PHLs*, need stability due the time required for training staff and validating tests, as well as length of state processes

What do I mean by "rural PHLs"?

Let's look at Wyoming, home to a rural PHL

Wyoming

Population: 581,381

Area: 97,813 square miles

Largest County: Laramie, 2688 sq mi, 100k

Largest City: Cheyenne, 65k

Metro Atlanta

Population: 6,307,261

Area: 8,376 square miles

This county: Cobb, 345 sq mi, 776k

This city: Smyrna, 56k

Rural PHLs are consistently understaffed with current staff wearing many different hats

To help build our in-house bioinformatics skills we've been working on developing accessible bioinformatic tools

Bactopia for accessible bioinformatics analysis

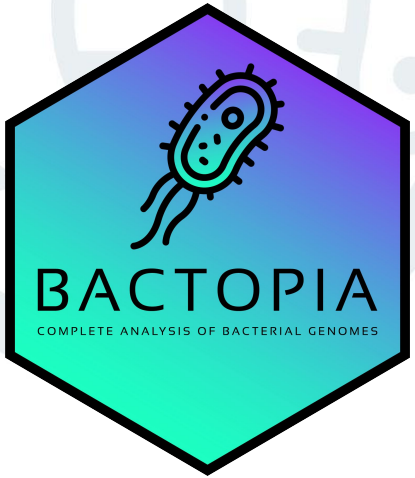
A Nextflow pipeline for end-to-end analysis of bacterial* genomes

Wraps 150+ bioinformatic tools into stand-alone modules

Bactopia Tools provide 60 additional workflows

Available from Docker, Singularity, and Conda

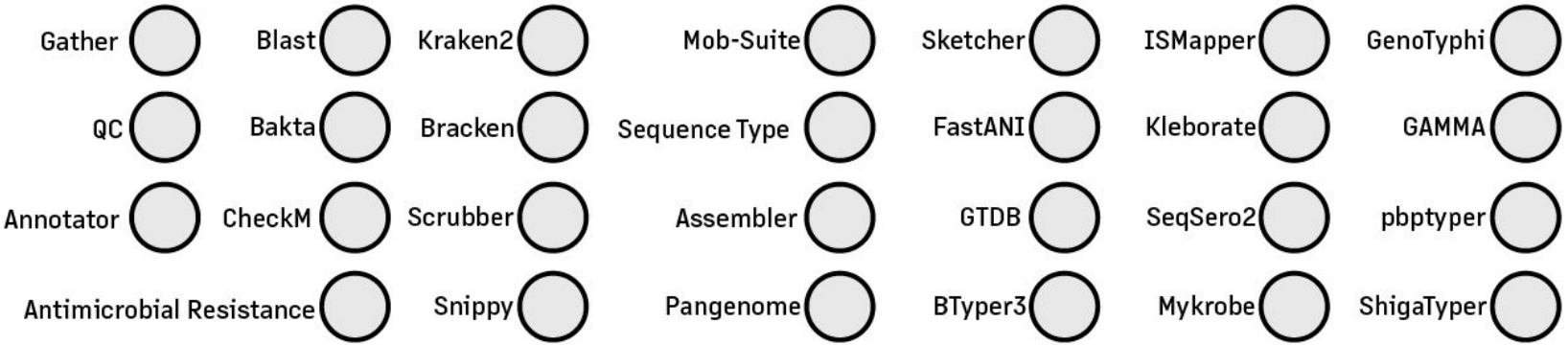
Sustained for 5+ years with support from WPHL, Emory, CAPE



Ask me or Tim for a sticker!

**Hopefully I'll convince you it can be used for more than bacterial genomes*

All modules are “plug and play”

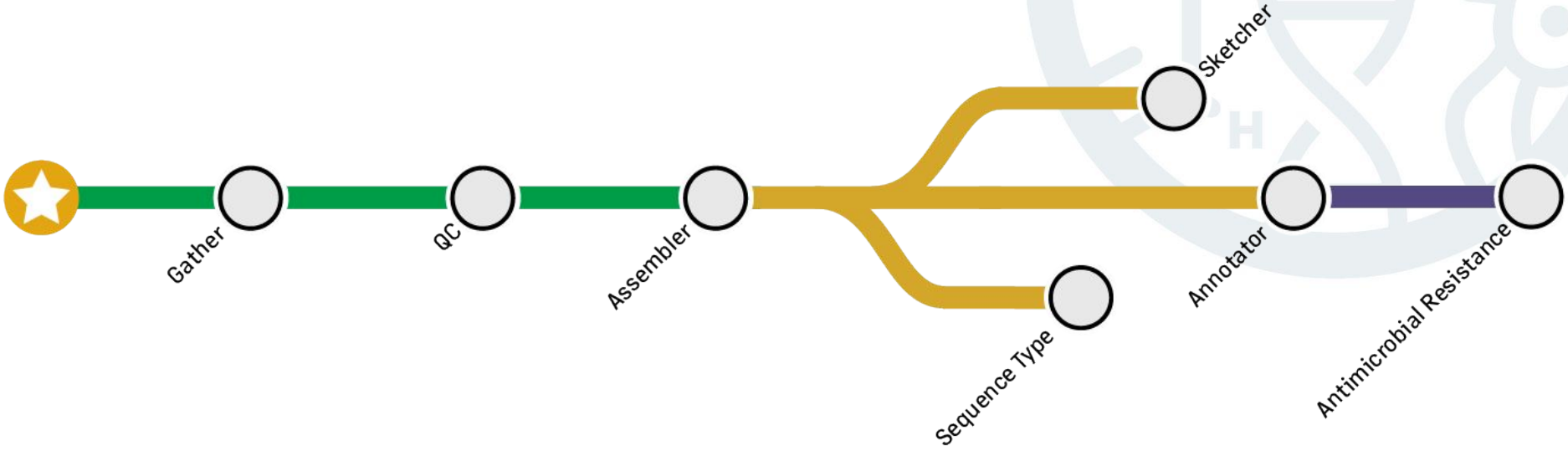


… another 30+ modules not listed

These can all be treated as building blocks

This allows Bactopia to be used as a framework, to rapidly adapt it to user needs, ***without having to learn a new workflow.***

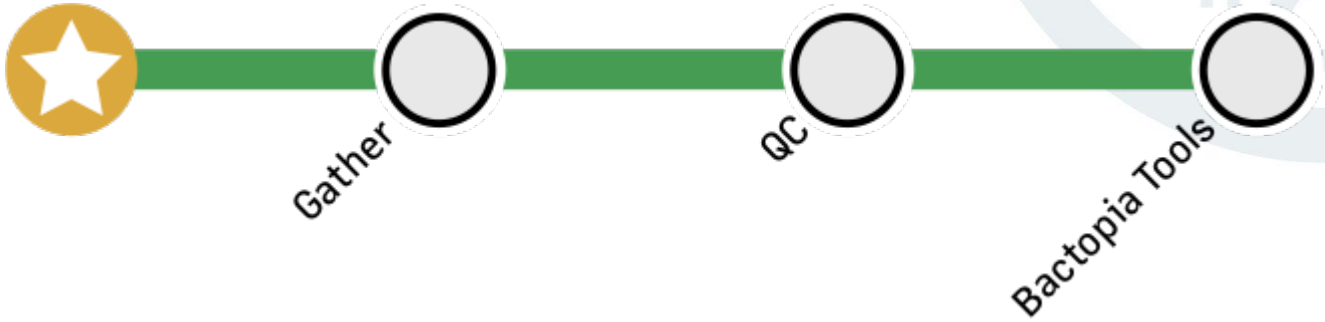
For example, the main Bactopia workflow



What if we have metagenomic sequencing?

clean-yer-reads *for access to Bactopia Tools*

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

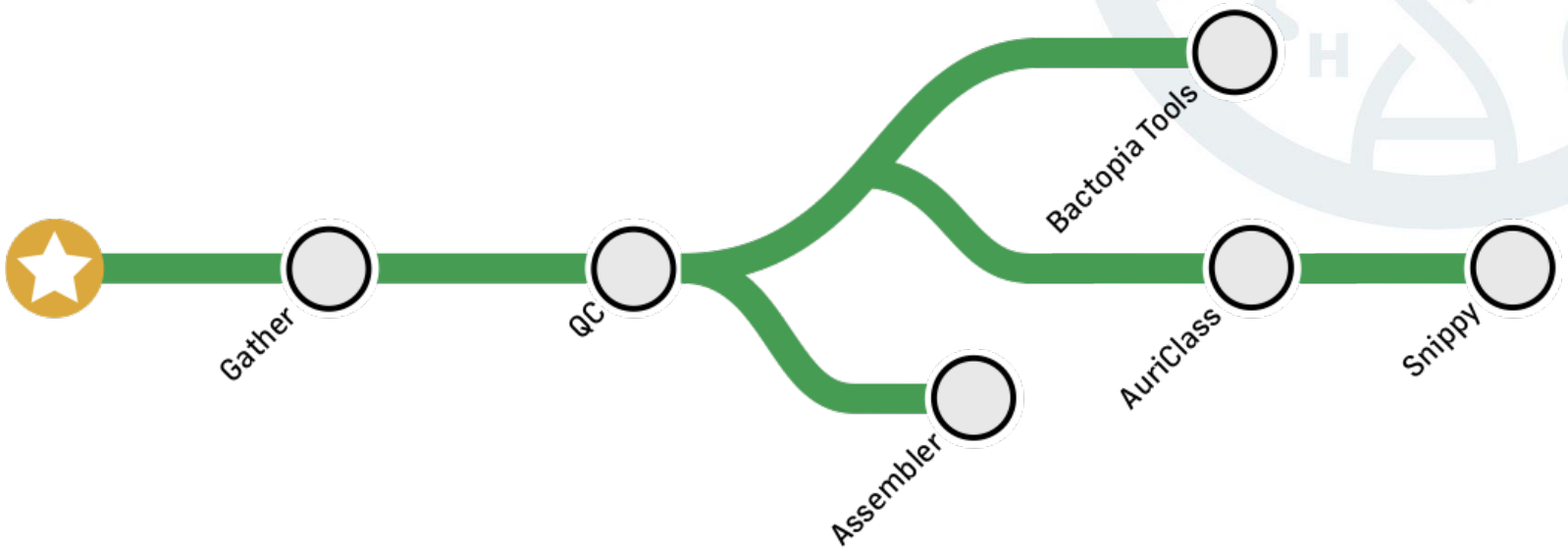


With clean-yer-reads you still have full access to all available Bactopia Tools.

What if we have fungal sequencing?

mycotopia *for haploid fungal genomes*

mycotopia represents a specialized adaptation of Bactopia, specifically for haploid fungal organisms, including the notable fungal pathogen *Candida auris*.



In each of these, staff that were already accustomed to Bactopia required no additional training

Acknowledgements

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

WPHL
Taylor Fearing
Chayse Rowley
Jim Mildenberger
Rob Christensen
Joseph Reed

Emory
Tim Read
EMERGENT Group
Ahmed Babiker

Questions?

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