

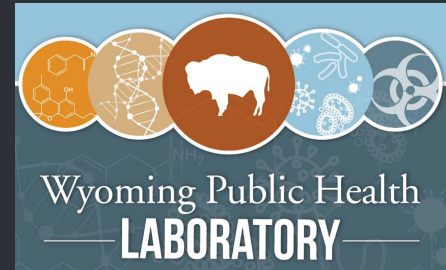
BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

v3 enhancements for genomic surveillance

Nextflow Summit – Boston

Robert A. Petit III, PhD
Wyoming Public Health Laboratory
November 29th, 2023



What is Bactopia?

Bactopia is an end-to-end Nextflow pipeline for the complete analysis of bacterial genomes. In a few simple steps, go from raw sequences to an extensive set of results to investigate.



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

STEP 02

Install Bactopia

Bactopia is easily installed using Conda and analyses can be run 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 05

Choose Compute Environment

Built using Nextflow, Bactopia supports many compute and execution environments. Simple parameter changes, allow efficient switching between local, HPC, and major cloud environments.



BACTOPIA STEP 06

Process Genomes

Bactopia includes more than 150 bioinformatic tools in the pipeline, with extensive version control and data provenance. Upon completion outputs are available in a standardized structure for easy access.



STEP 04

Include Public Samples

`bactopia search` can query ENA/SRA to identify accessions to further supplement analyses.



STEP 01

Sequence Bacterial Genomes

Bactopia supports sequences from Illumina and Nanopore technologies, or a combination of the two.



STEP 08

Investigate Results

With analysis completed, numerous output files are available to assist in your genomic investigations



BACTOPIA TOOLS STEP 07

Additional Workflows

More than 60 Bactopia Tools allow seamless post-analysis of Bactopia outputs, including additional workflows for pangenome, taxonomic classification, phylogenies, AMR, among many more.





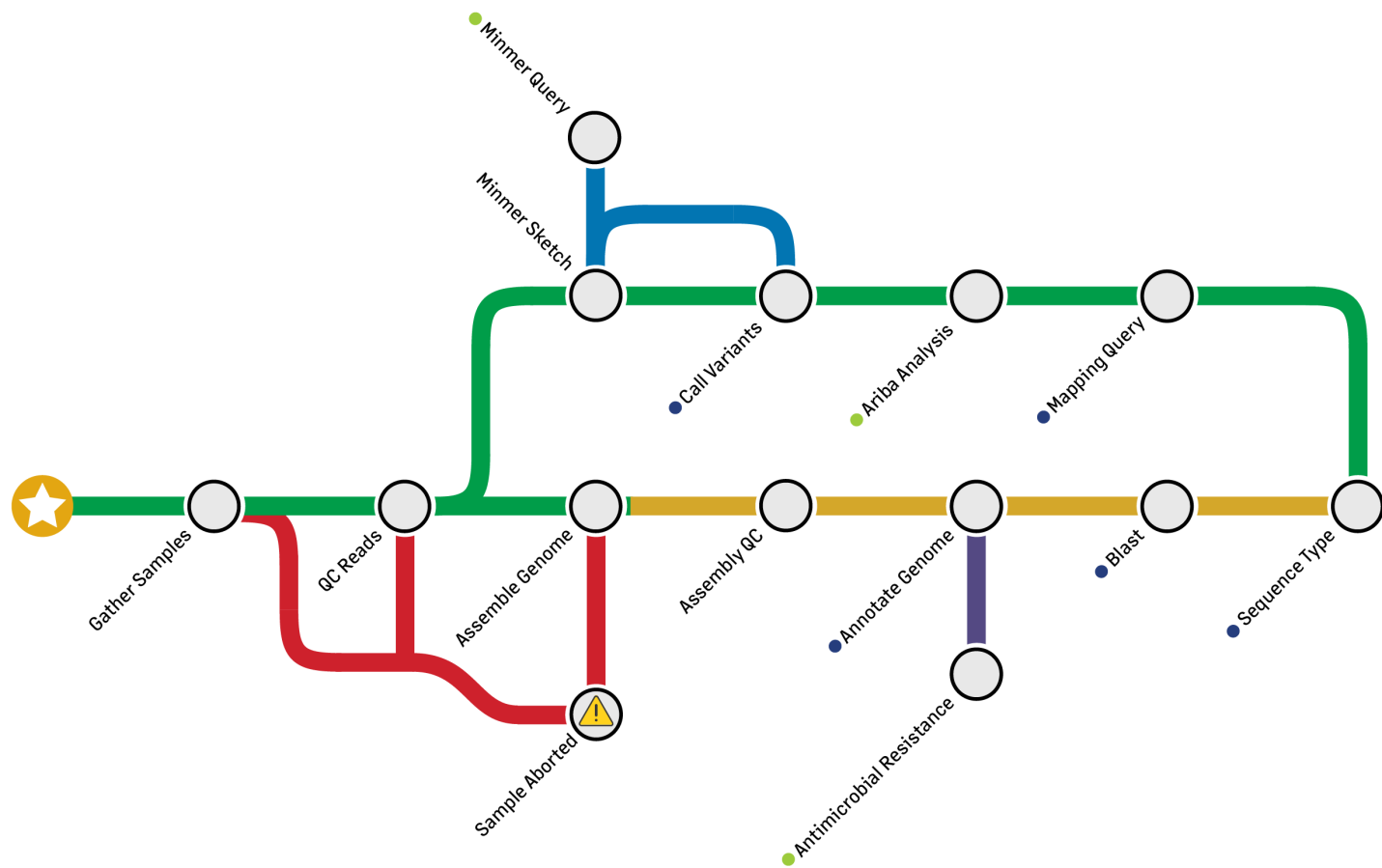
Bactopia streamlines analysis, enabling users to focus on their science and outcomes

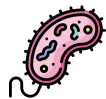


“

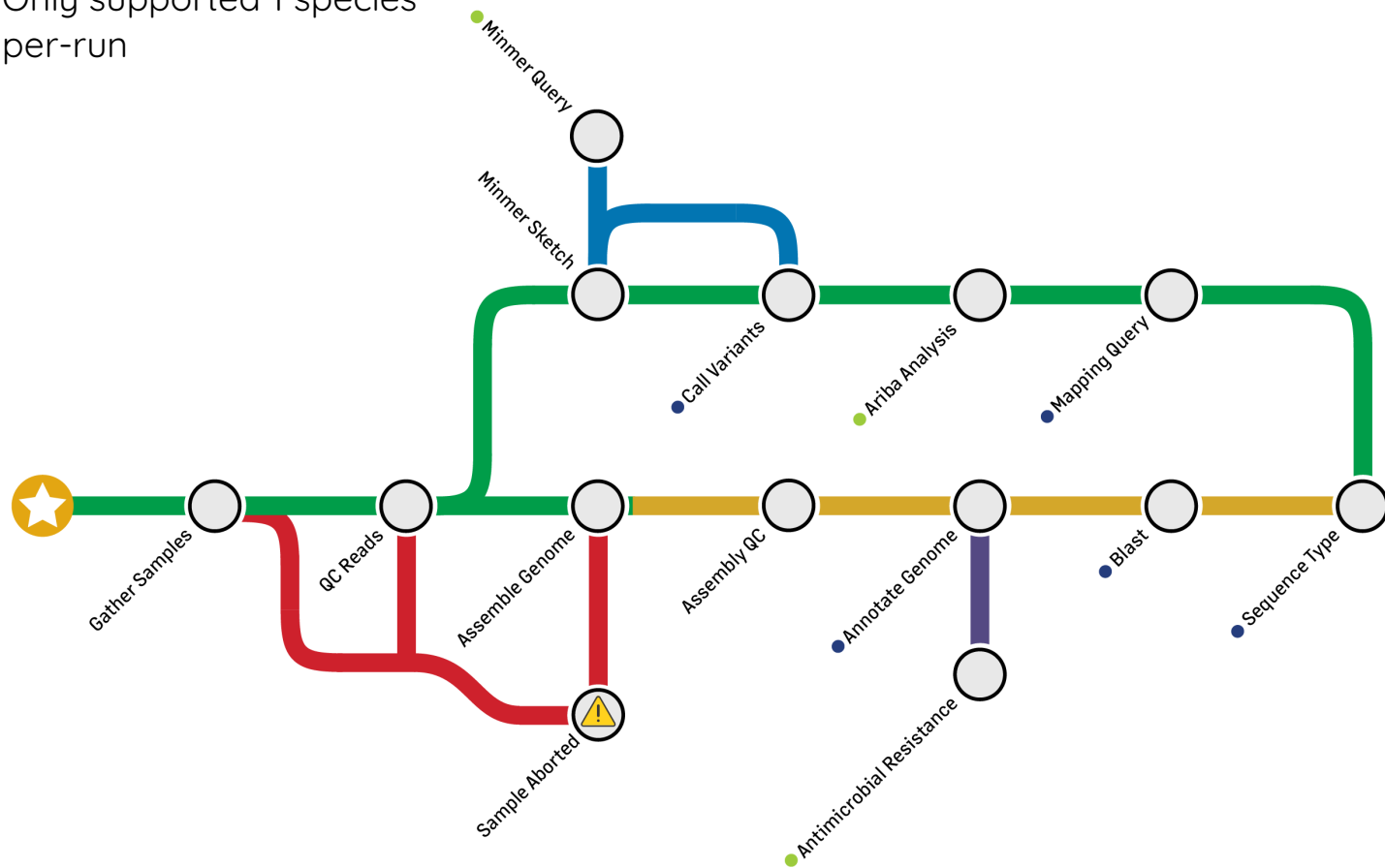
Bactopia streamlines analysis, enabling users to focus on their science and outcomes

But there was a need for a revamp





Only supported 1 species
per-run

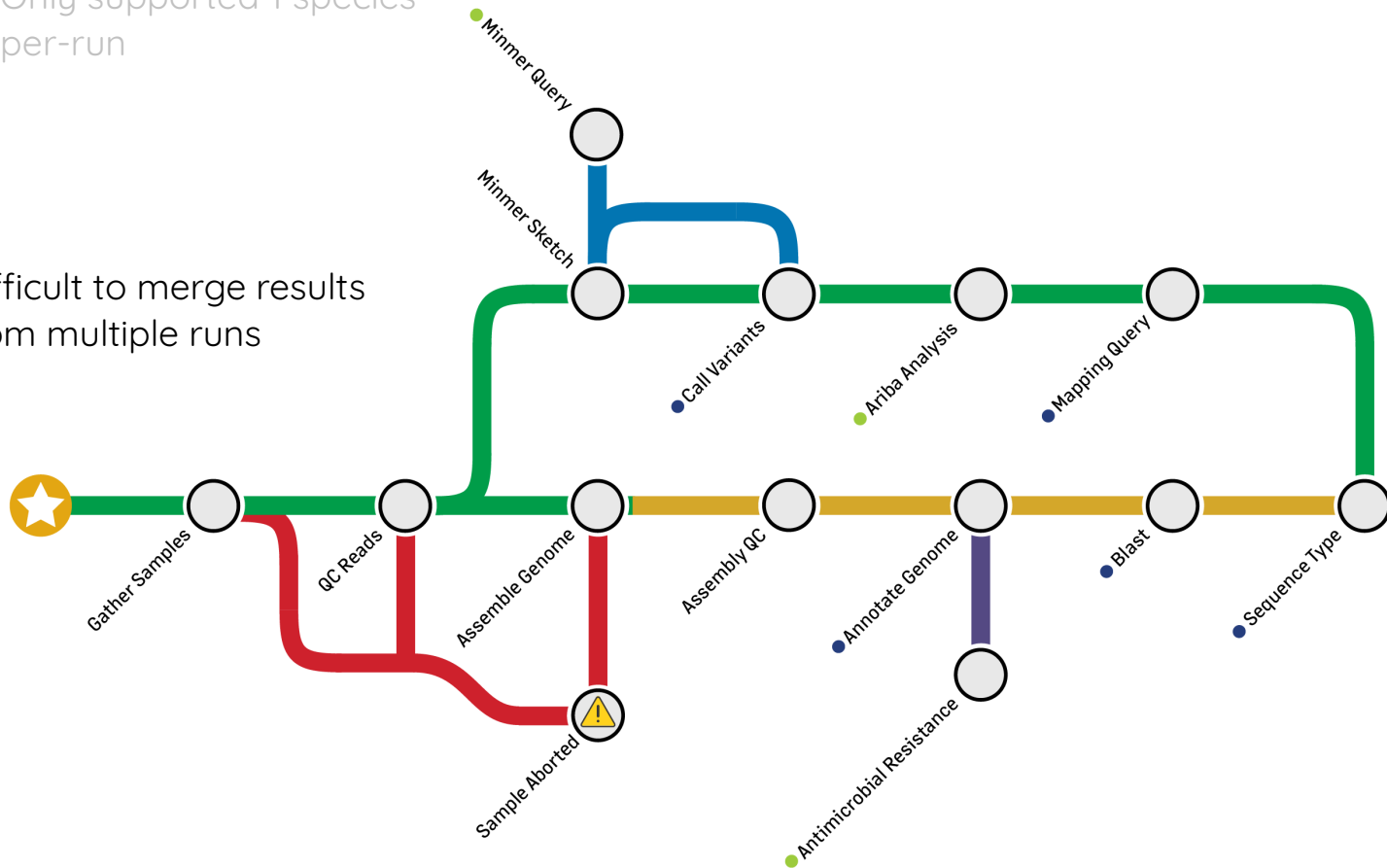




Only supported 1 species
per-run



Difficult to merge results
from multiple runs





Only supported 1 species
per-run



8 steps relied on input datasets



Difficult to merge results
from multiple runs





Only supported 1 species
per-run



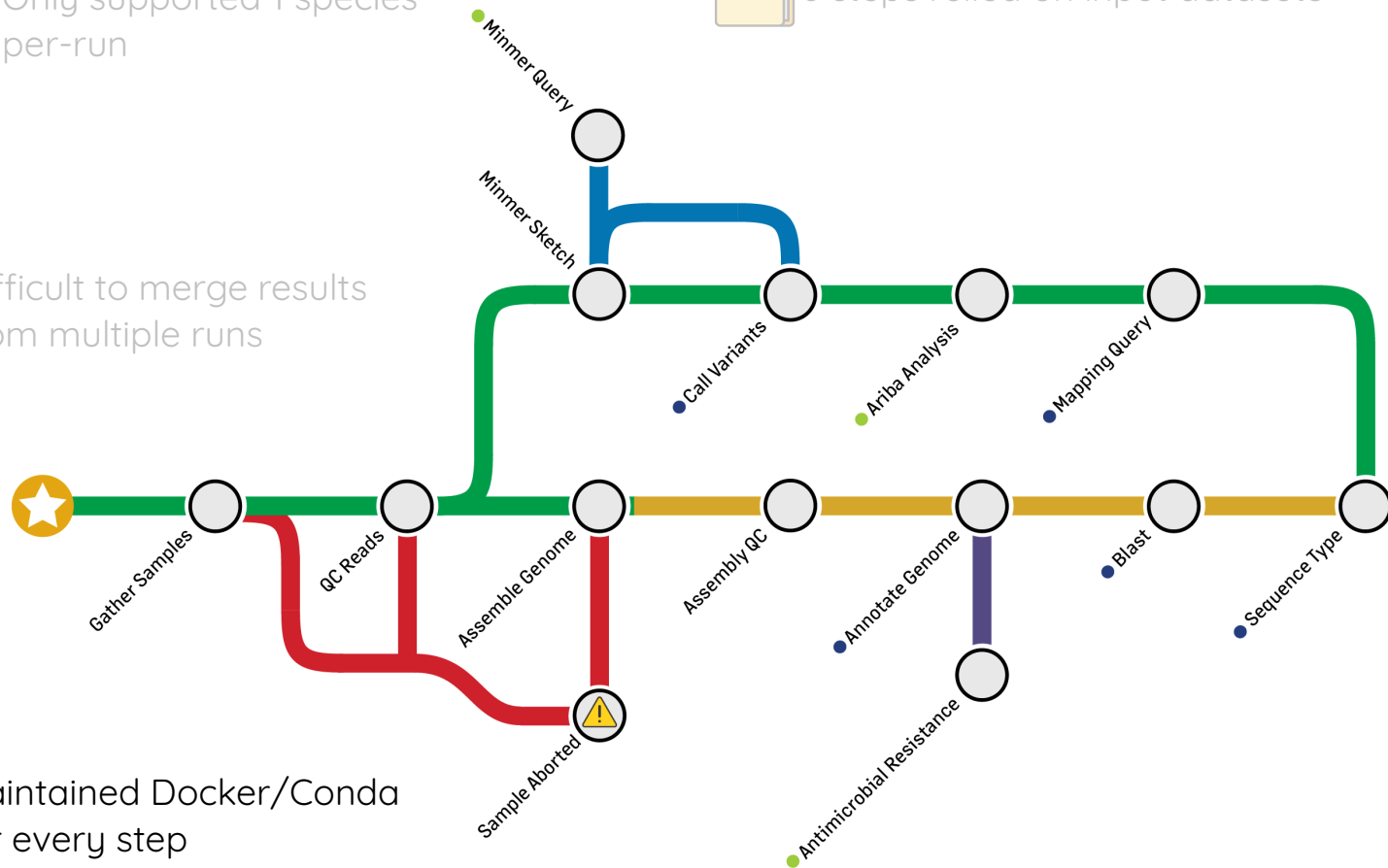
8 steps relied on input datasets



Difficult to merge results
from multiple runs



Maintained Docker/Conda
for every step





Only supported 1 species per-run



8 steps relied on input datasets



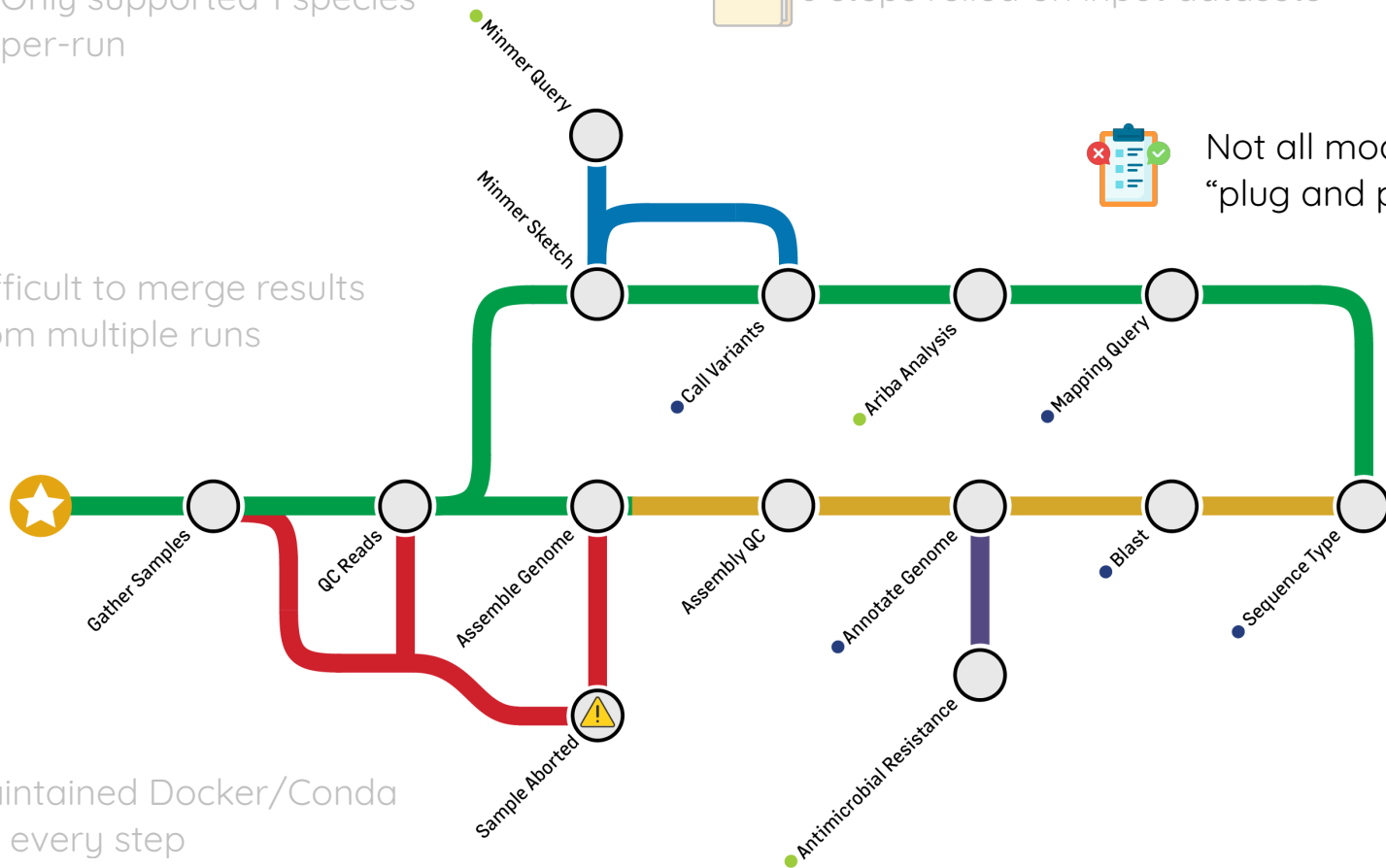
Difficult to merge results from multiple runs



Not all modules were “plug and play”



Maintained Docker/Conda for every step





Only supported 1 species per-run



8 steps relied on input datasets



Difficult to merge results from multiple runs



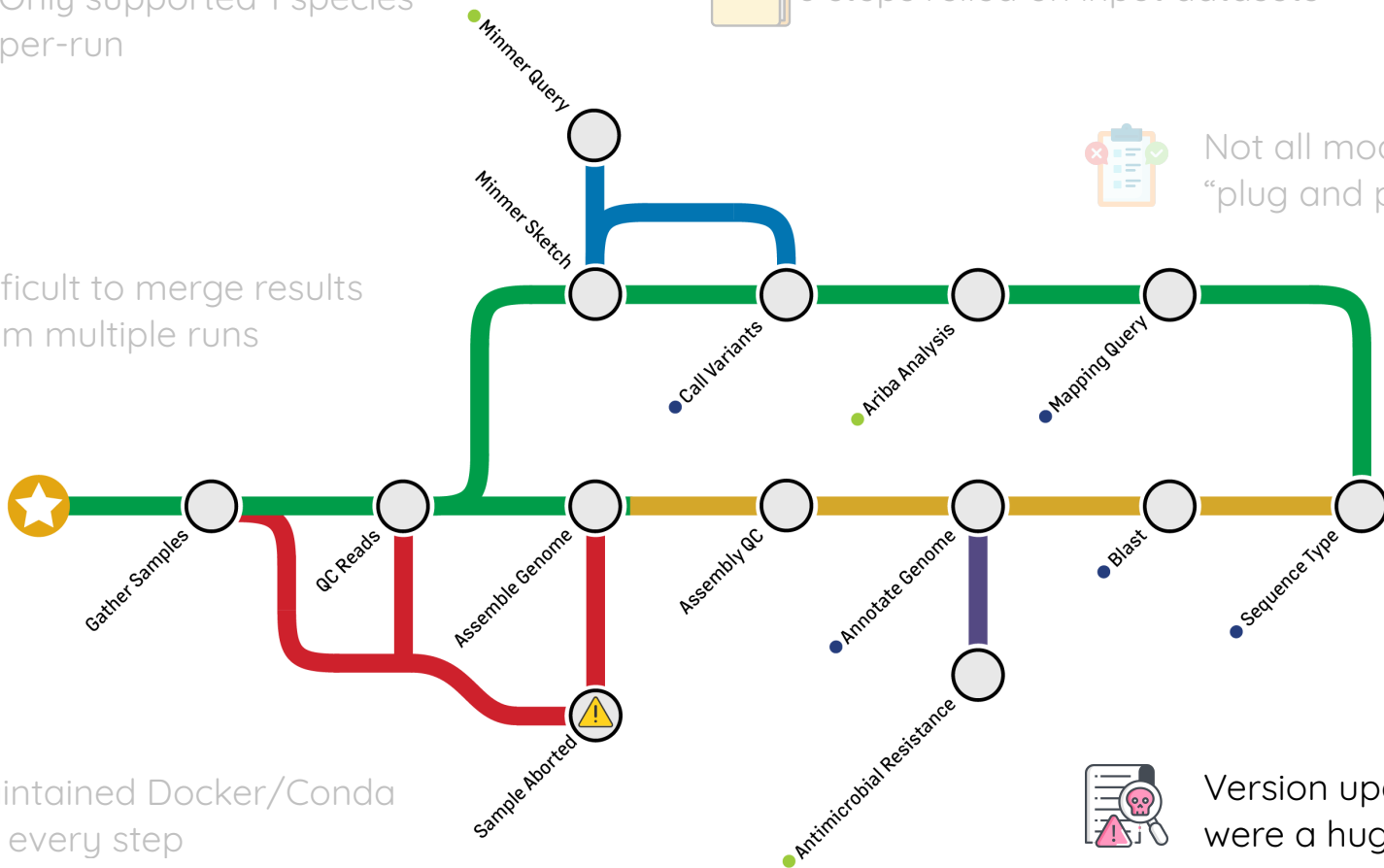
Not all modules were "plug and play"



Maintained Docker/Conda for every step



Version updates were a huge hassle





Only supported 1 species
per-run

Minmer



8 steps relied on input datasets



Diff
fror



dules were
play”



Maintained Docker/Conda
for every step

Sample Abn.

Antimicrobial



Version updates
were a huge hassle



v3 enters the chat...

A vertical white line on the left side of the slide, with three small white circles positioned at regular intervals along it.

Meet Wyoming's pathogen surveillance needs

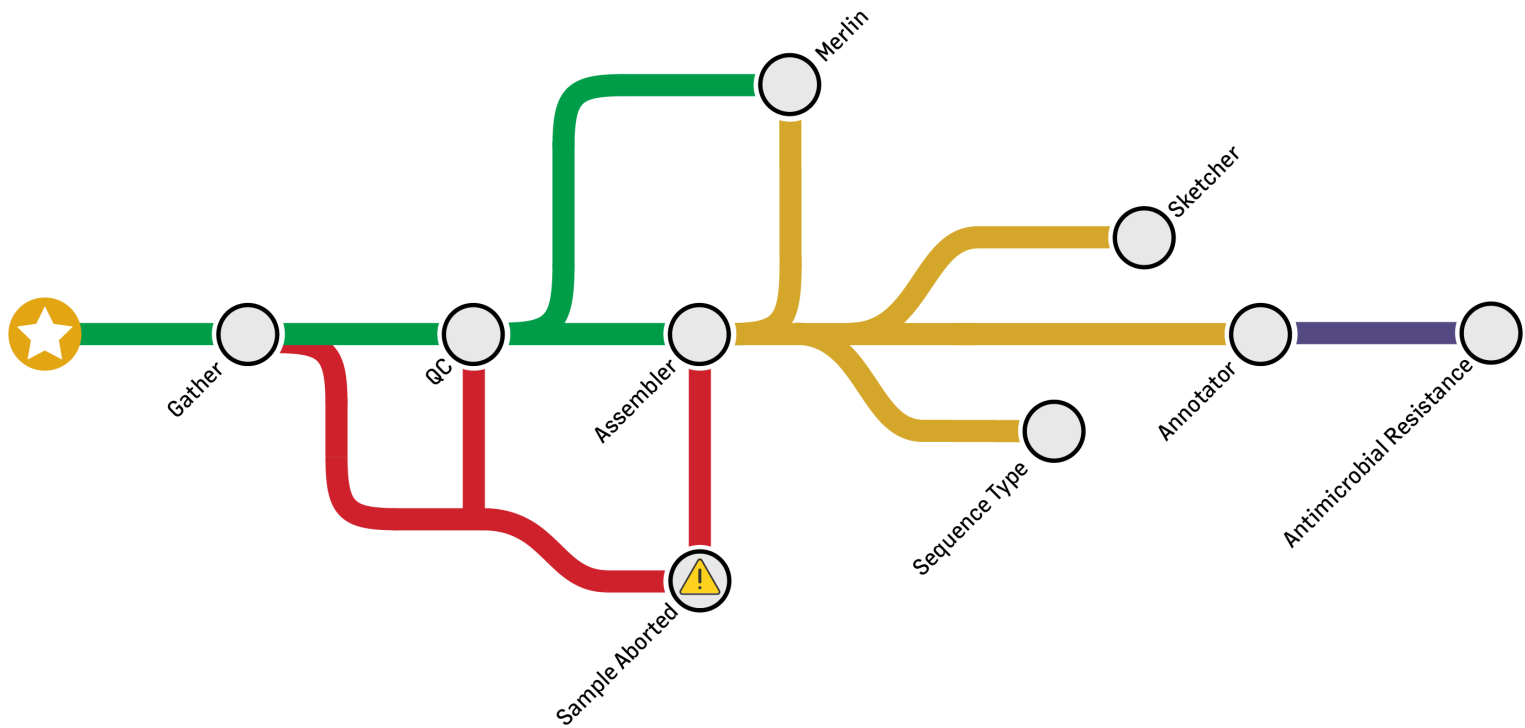
Lower the barrier of entry for users

A vertical white line on the left side of the slide, with three white circles positioned at regular intervals along it.

Meet Wyoming's pathogen surveillance needs

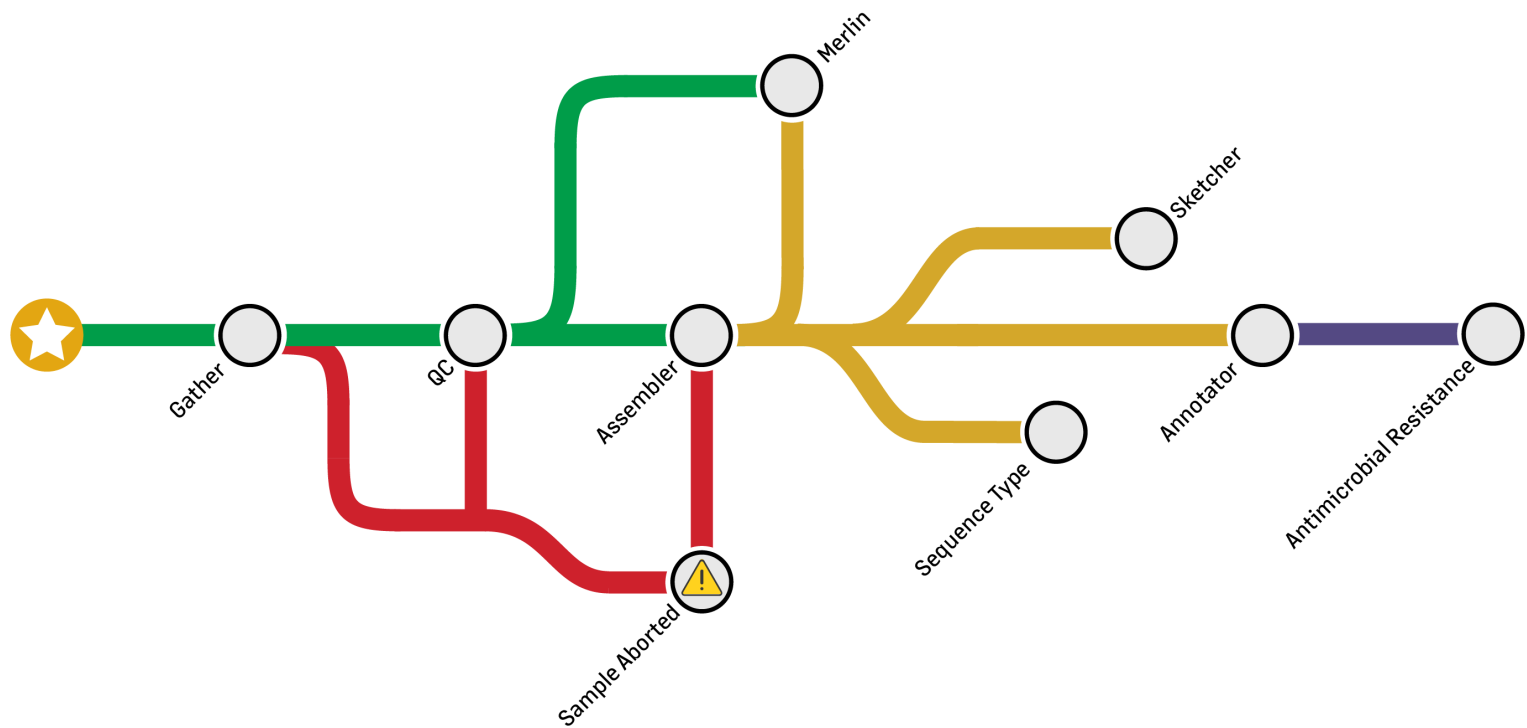
Lower the barrier of entry for users

Reduce the long-term maintenance burden





Mix and match any number
of species per-run

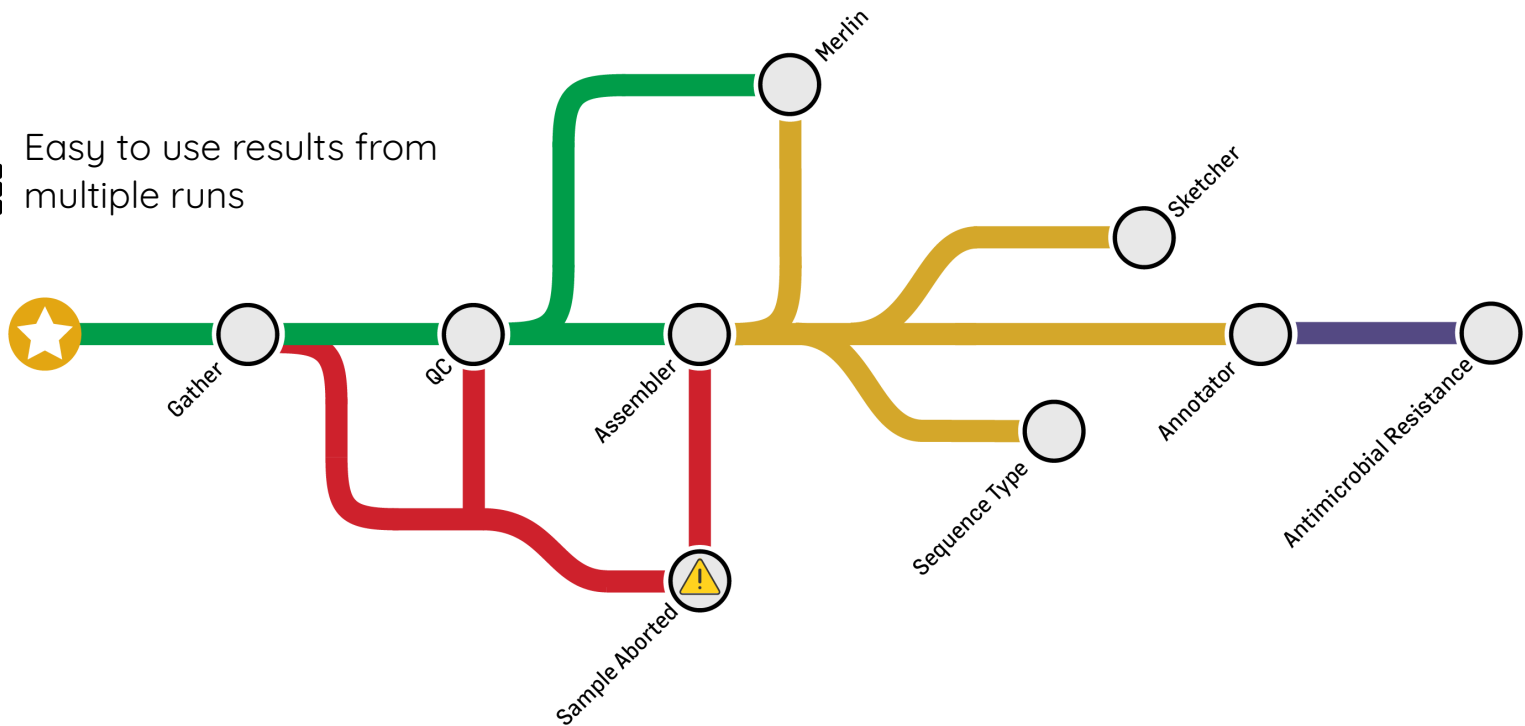




Mix and match any number
of species per-run



Easy to use results from
multiple runs





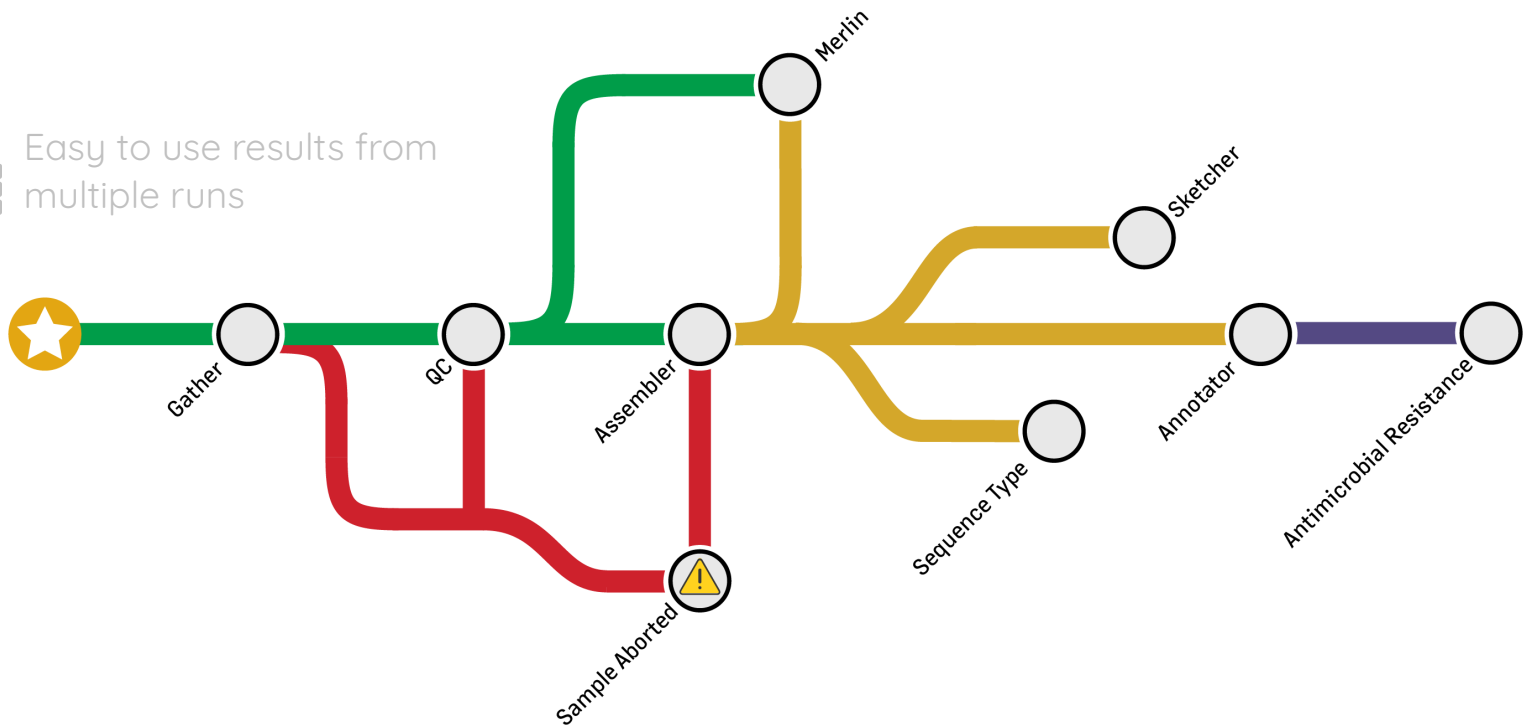
Mix and match any number
of species per-run



Datasets are baked into
the process



Easy to use results from
multiple runs





Mix and match any number of species per-run



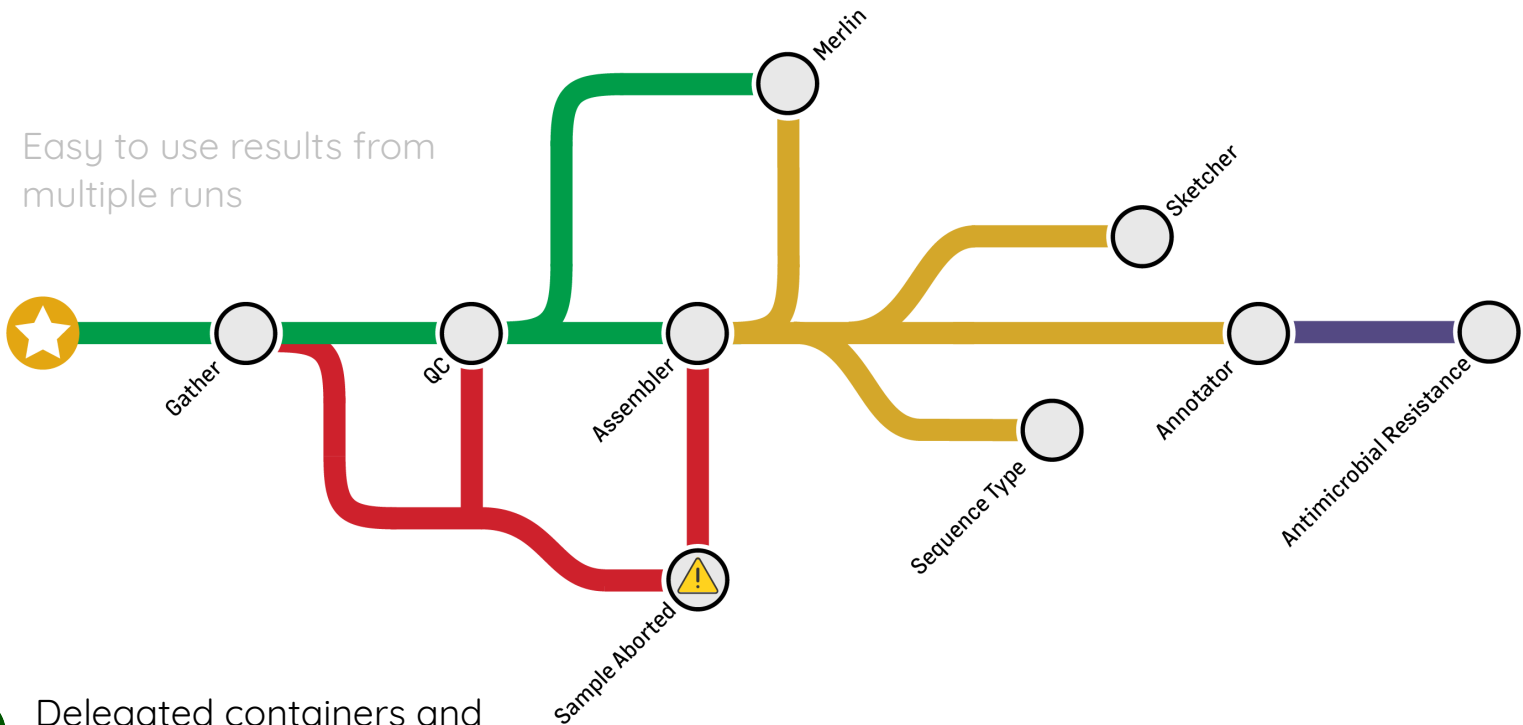
Datasets are baked into the process



Easy to use results from multiple runs



Delegated containers and environments to Bioconda





Mix and match any number of species per-run



Datasets are baked into the process



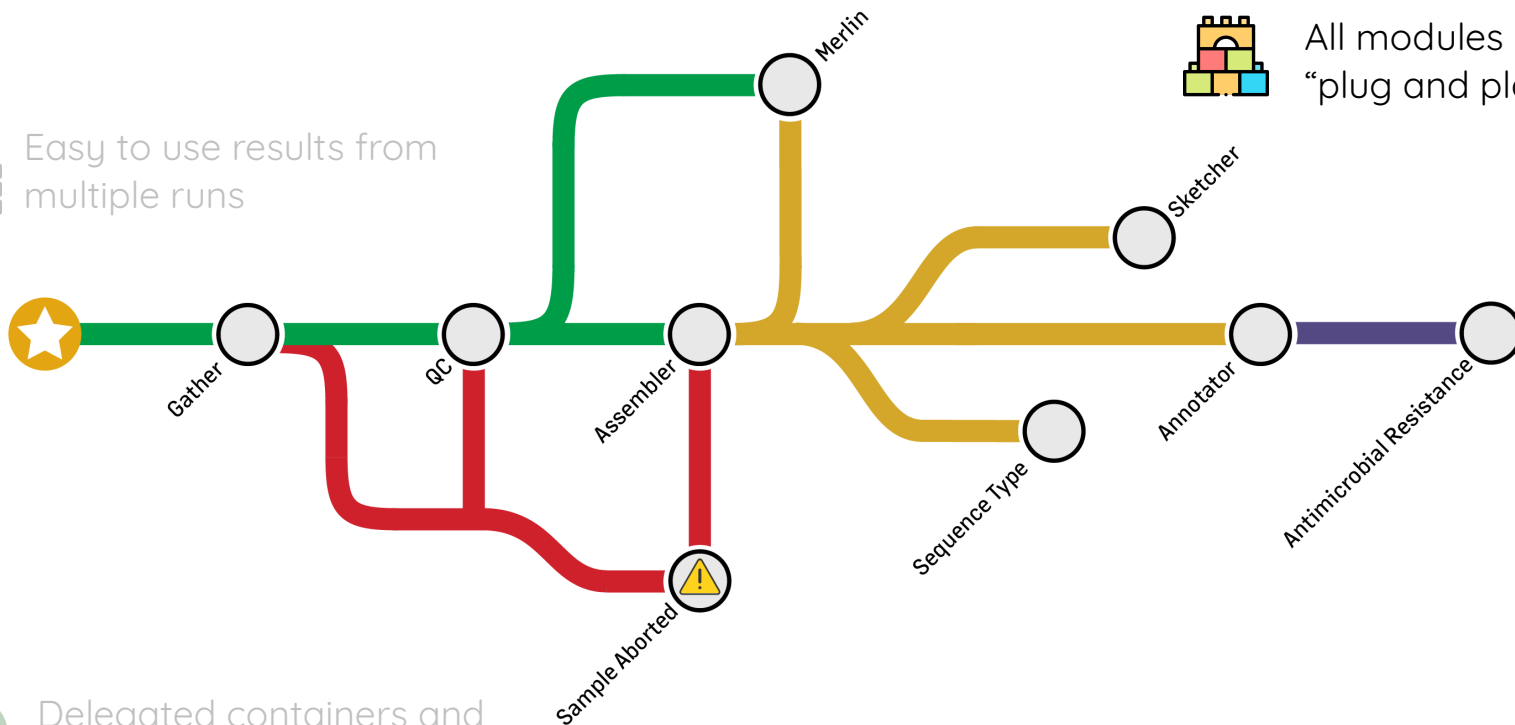
Easy to use results from multiple runs



All modules are now "plug and play"



Delegated containers and environments to Bioconda





Mix and match any number of species per-run



Datasets are baked into the process



Easy to use results from multiple runs



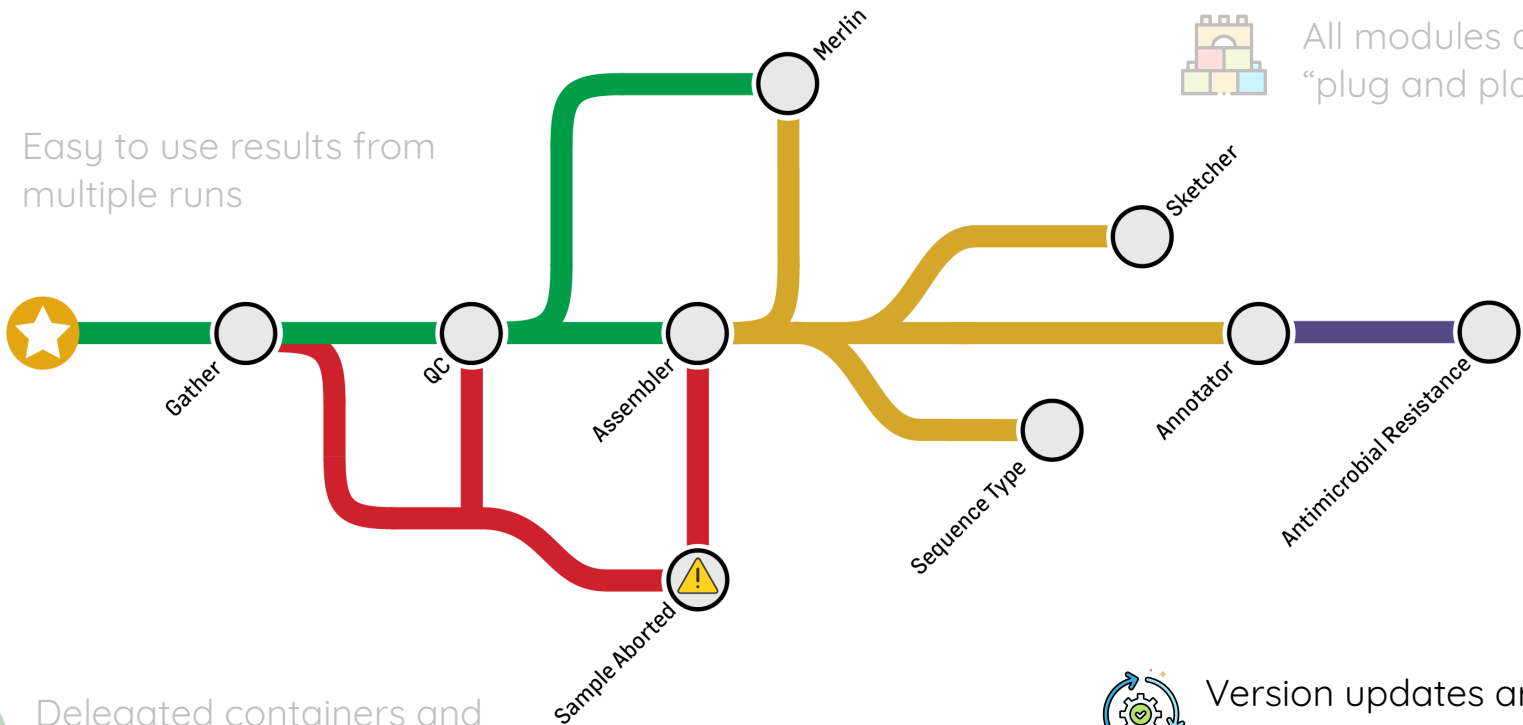
All modules are now "plug and play"



Version updates are easy and seamless



Delegated containers and environments to Bioconda





Mix and match any number of species per-run



Datasets are baked into the process



Easy to use multiple run



Gather



Modules are now and play”

Resistance



Delegated containers and environments to Bioconda

Sample



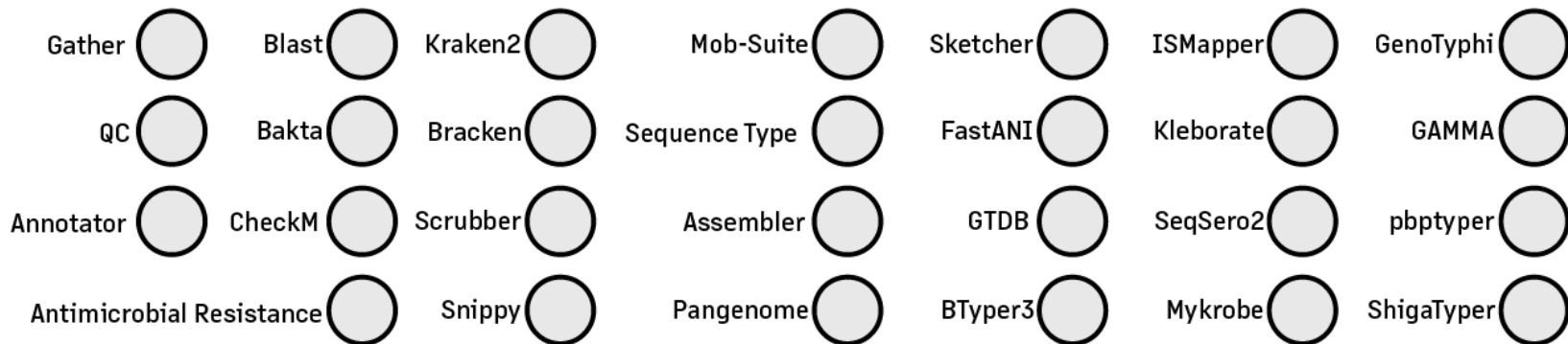
version updates are easy and seamless

A vertical line on the left side of the slide, with a small circle at the level of the main text.

Let's explore “plug & play”



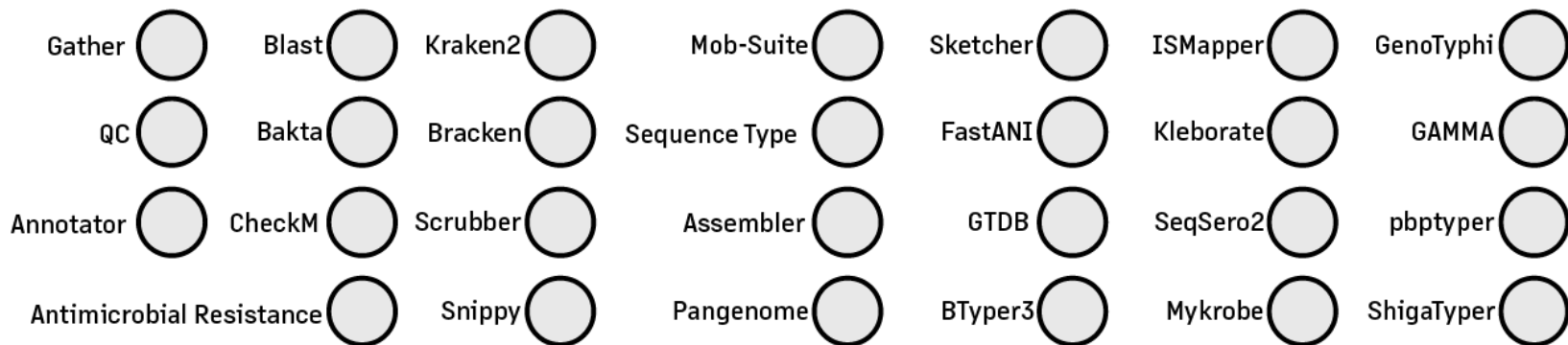
All modules are now “plug and play”



... another 30+ modules not listed



All modules are now “plug and play”

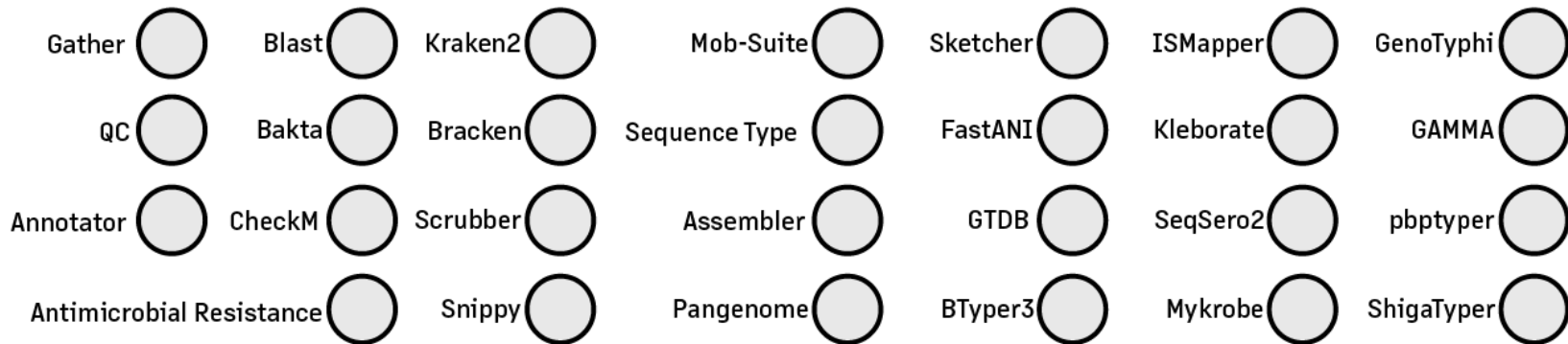


... another 30+ modules not listed

Each module has a Nextflow config file and a params JSON file



All modules are now “plug and play”



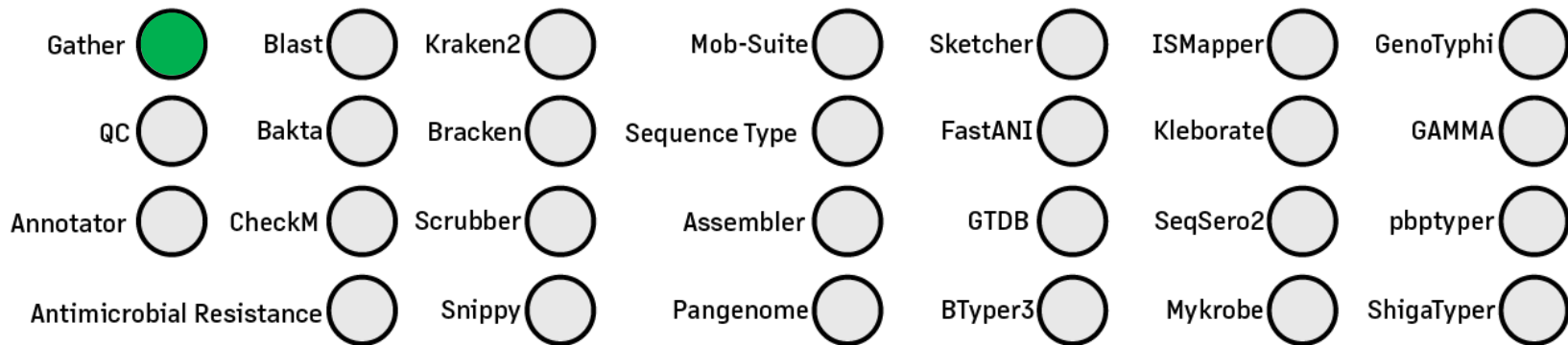
... another 30+ modules not listed

Each module has a Nextflow config file and a params JSON file

These can all be treated as building blocks



All modules are now “plug and play”



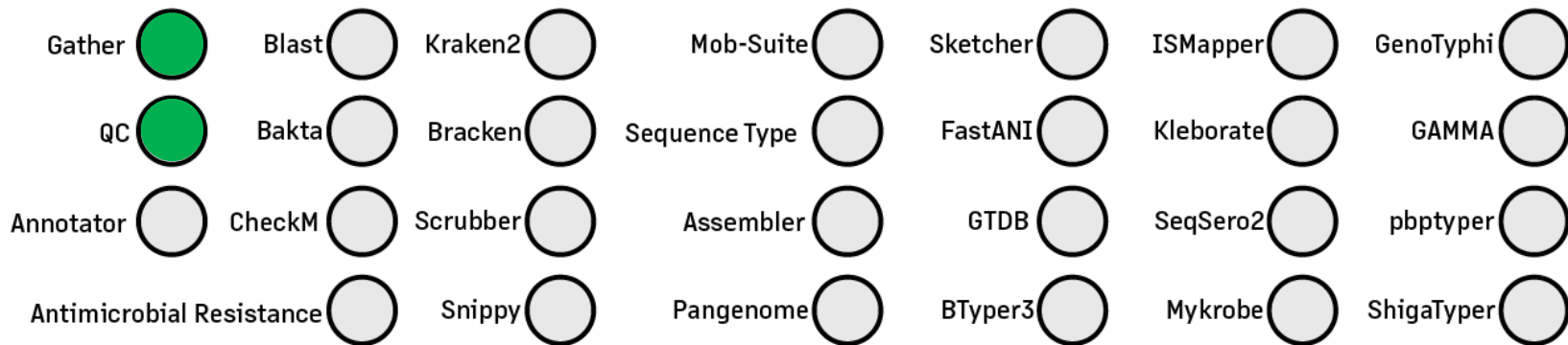
... another 30+ modules not listed

Each module has a Nextflow config file and a params JSON file

These can all be treated as building blocks



All modules are now “plug and play”



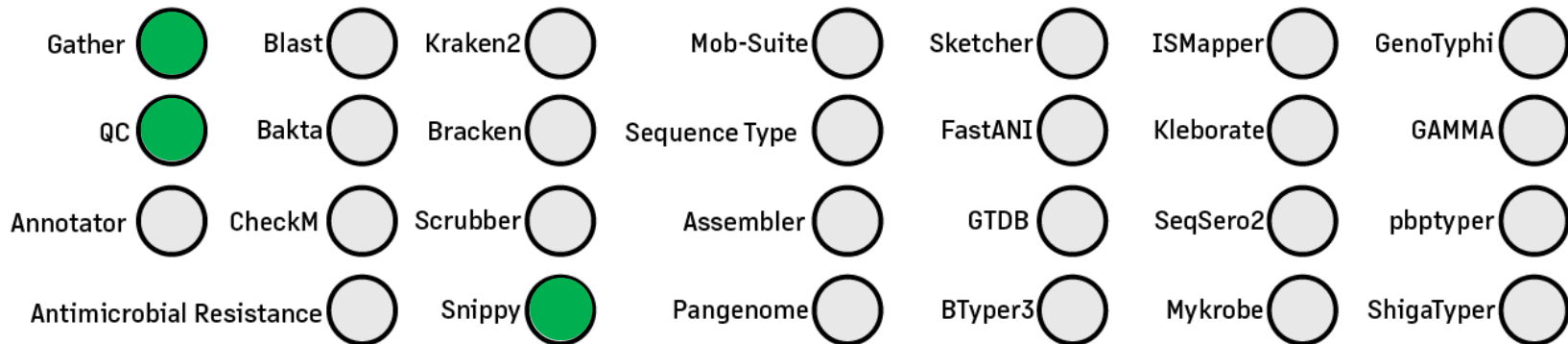
... another 30+ modules not listed

Each module has a Nextflow config file and a params JSON file

These can all be treated as building blocks



All modules are now “plug and play”



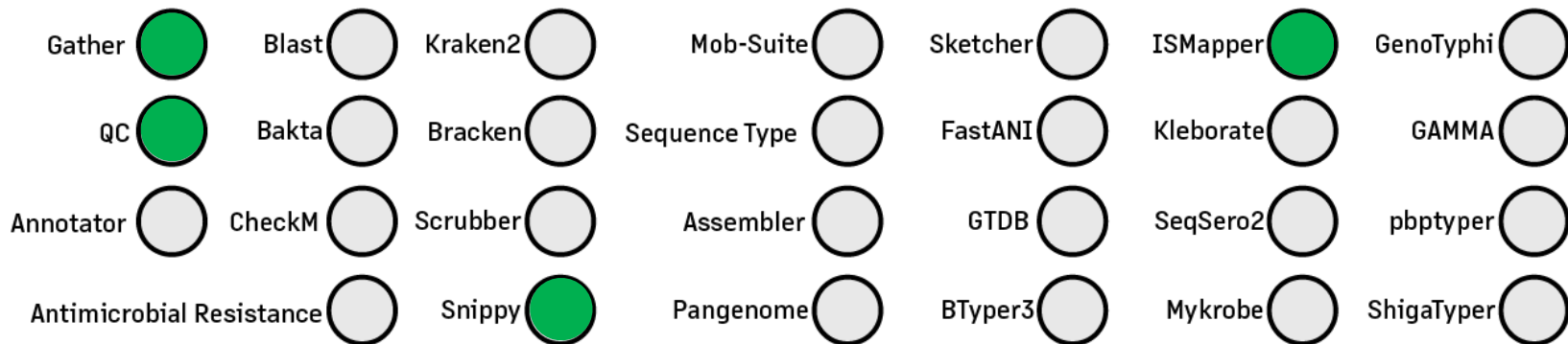
... another 30+ modules not listed

Each module has a Nextflow config file and a params JSON file

These can all be treated as building blocks



All modules are now “plug and play”



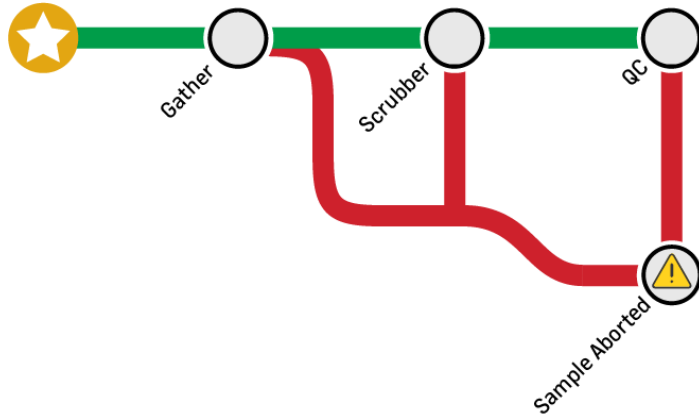
... another 30+ modules not listed

Each module has a Nextflow config file and a params JSON file

These can all be treated as building blocks



Reshuffle modules for new named workflows



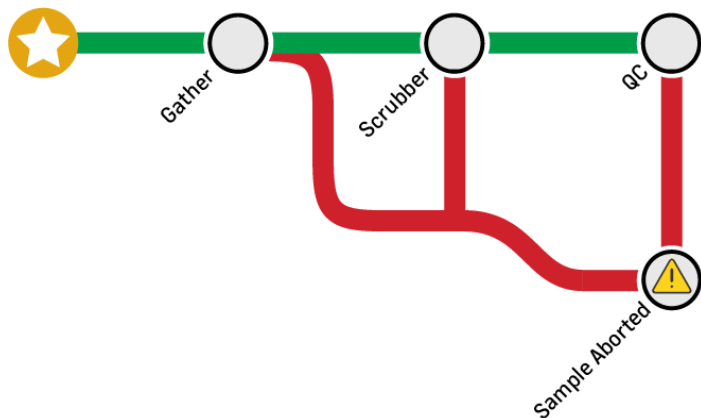
Clean-Yer-Reads

A workflow to gather samples, remove host reads, and finally QC the reads.

Can use applicable Bactopia Tools afterwards



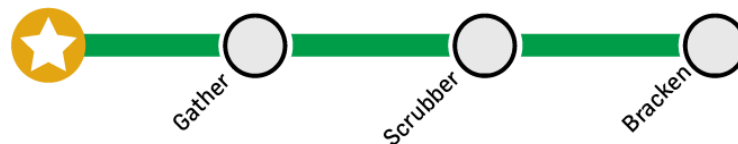
Reshuffle modules for new named workflows



Clean-Yer-Reads

A workflow to gather samples, remove host reads, and finally QC the reads.

Can use applicable Bactopia Tools afterwards



Teton

A workflow to gather samples, remove host reads, and do basic taxonomic classification.

Workflow for newly staged sequences on our cloud infrastructure

“

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



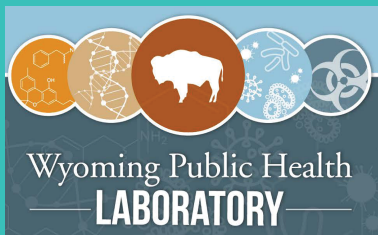
Let's wrap this up



Bactopia v3 has been rebuilt for on-going genomic surveillance of public health pathogens, while improving the user experience and significantly reducing the long-term maintenance burden.

Acknowledgements

All the developers of open-source software used by Bactopia, and the many users regularly providing feedback and suggestions



Taylor Fearing

Chayse Rowley

Jim Mildenberger

Rob Christensen

Joseph Read

Many more at WPHL!



Tim Read and EMERGENT

@ddomman, @egenomics, @harousuz,
@davidcarlson, @Phemarajata,
@dysci2022, @lfenske-93,
@plasminogene, @kruem-ar,
@elbriandominguez, @Happymands,
@incoherentian

BIOCONDA[®]

nextflow

nf-core

seqera



Presentation template by

A vertical line on the left side of the slide, with a small circle at the bottom.

Questions?

Learn more at