

BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

Workshop for using Bactopia

CDC Enteric Diseases Laboratory Branch

Robert A. Petit III, PhD
Wyoming Public Health Laboratory
March 16th, 2023



Wyoming Public Health
LABORATORY



Disclaimer

The views and opinions expressed in today's workshop are mine and do not necessarily reflect the views or positions of the Wyoming Public Health Laboratory

- Today's Workshop Outline

- **Background**

First, we'll learn about Bactopia as a pipeline. We'll investigate the many moving parts of Bactopia.

- **Bactopia**

We'll install Bactopia and process a few genomes through the main pipeline. During this time we will take a deeper look into a few steps.

- **Bactopia Tools**

We'll run Bactopia Tools to learn how these independent workflows can boost your analyses.

Outline for Bactopia Introductions



People

Meet the people behind Bactopia and how they are helping to improve it.



Bactopia

An introduction into Bactopia and how Bactopia Tools help streamline complex analyses.



Design Decisions

A quick glimpse into some decisions that were made ease on-going development.



Enhancements to OSS

Learn how Bactopia is helping to further enhance open-source science.



Future Directions

A look into what is on the horizon for Bactopia. Many new changes coming soon.



Wrap Up

Not much to say here, we'll close the first part of this session.

1

People behind Bactopia

Let's put some faces to Bactopia



Yo! 🖐️

I AM Robert

The developer and maintainer of Bactopia

Supporting Roles



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 lab at WPHL. She has played a tremendous and critical role in helping to expand Bactopia into public health.

Many scientists around the world that provide feedback

- Wyoming Public Health Lab

- 15-20 people at WPHL

Led nation in % SC2+
cases sequenced

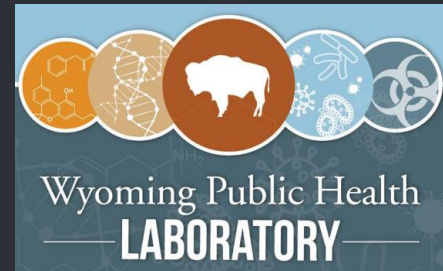
Domestic and
International trainings

Strong relationship with
Vet Lab at Ag Lab

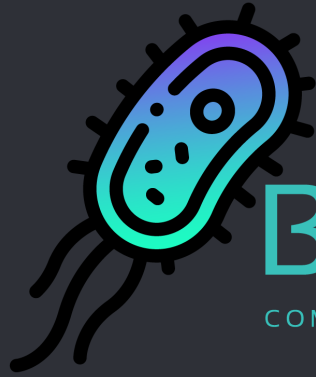


“

My colleagues at the Wyoming Public Health Laboratory and Emory University have played an incredibly supportive role in the advancement of Bactopia.



2



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

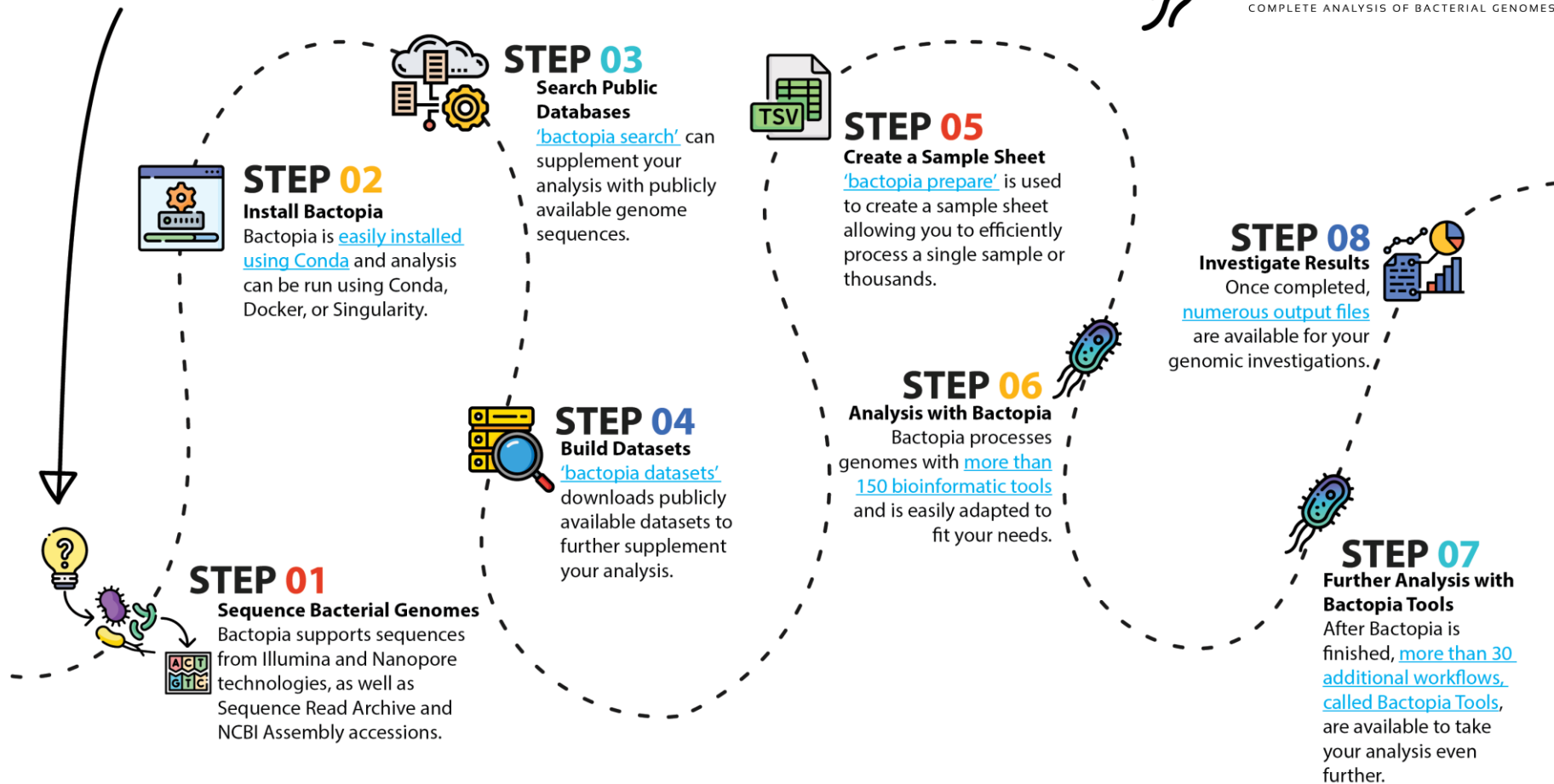
Let's learn about the Bactopia and Bactopia Tools

“

In a few steps, Bactopia allows you to go from raw data to investigating your results

What is Bactopia?

Bactopia is an extensive Nextflow pipeline for the complete analysis of bacterial genomes. To learn more, follow the step-by-step guide below.



Accepted Inputs

- Illumina and/or Nanopore Reads**
 --RL/--R2/--SE/--ant/--hybrid/--sample
 --fastqs with 'bactopia prepare' file-of-filenames
- DDBJ/ENA/SRA Accessions**
 --accession 'Experiment Accession'
 --accessions with 'bactopia search' results
- NCBI Assembly Accessions**
 --accession 'Assembly Accession'
 --accessions 'file with accessions'
- Assemblies**
 --assembly, --sample
 --fastqs with 'bactopia prepare' file-of-filenames



Legend

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

Supplemented By Bactopia Datasets

- Generic datasets (--datasets)
- Species-specific datasets (--species)

Bactopia Processes

Gather Samples
 Collect local files and/or download SRA or NCBI Assembly

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

Minmer Sketch and Minmer Query
 Create minmer sketches and query them against RefSeq and GenBank

Call Variants
 Determine SNPs and InDels against a reference genome

Mapping Query
 Align to a reference and determine per-base coverage

Assemble Genome & Assembly QC
 Create a de novo assembly and summary metrics, then assess the quality of the assembly

Annotate Genome
 Predict genes and proteins from the assembled contigs

Antimicrobial Resistance
 Identify presence of AMR and/or virulence genes

Blast
 Align genes, proteins, or primers to assembled contigs

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 mismatching IDs
- Paired-end with different read counts
- Paired-end with skewed proportions
- Genome size below minimum
- Genome size exceeds maximum
- 0 assembled contigs
- Assembled size below minimum

Abort Reasons

Bactopia Tools

More workflows for more science

Easy comparative analysis of Bactopia outputs

Two Types:

Single tool

Kleborate, SeqSero2, TB Profiler

Multiple tools connected together

pangenome: Prokka -> PIRATE -> IQ-Tree

50+ Bactopia Tools are available

Frame-worked for easy addition

Ex. [pangenome Bactopia Tool](#)



BACTOPIA TOOLS

More workflows for more science



ANTIMICROBIAL RESISTANCE

Abricate

Mass screening of contigs for antimicrobial and virulence genes

AMRFinder+

Identify antimicrobial resistance in genes or proteins

Resistance Gene Identifier

Predict antibiotic resistance from assemblies

ANNOTATION

Bakta

Rapid annotation of bacterial genomes and plasmids

eggNOG-Mapper

Functional annotation of proteins using orthologous groups and phylogenies

DISTANCE

FastANI

Fast alignment-free computation of Average Nucleotide Identity (ANI)

mash dist

Calculate Mash distances between sequences

mashtree

Quickly create a tree using Mash distances

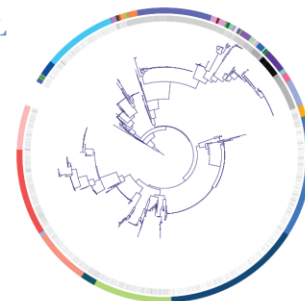
SEQUENCE SURVEY

CheckM

Assess the assembly quality of your samples

mlst

Automatic MLST calling from assembled contigs



PANGENOME

Create a pan-genome and core-genome phylogeny of your samples. Additionally, supplement your samples by including publicly available assemblies.

TAXONOMIC CLASSIFICATION

GTDB

Identify marker genes and assign taxonomic classifications

Kraken2

Taxonomic classifications of sequence reads

MOBILE GENETIC ELEMENTS

ISMMapper

Identify insertion site positions in bacterial genomes

MOB-suite

Reconstruct and annotate plasmids in bacterial assemblies

MERLIN

Use Merlin to automatically run species-specific tools for the following organisms.

Escherichia	Mycobacterium
Haemophilus	Neisseria
Klebsiella	Salmonella
Legionella	Staphylococcus
Listeria	Streptococcus

SPECIES SPECIFIC

AgrVATE

Rapid identification of *Staphylococcus aureus* agr locus type

ECTyper

In-silico prediction of *Escherichia coli* serotype

emuntyper

emm-typing of *Streptococcus pyogenes* assemblies

hicap

cap locus serotype and structure in *Haemophilus influenzae* assemblies

HpsuisSero

Serotype prediction of *Haemophilus parasuis* assemblies

Kleborate

Screen *Klebsiella* assemblies for MLST, sub-species, and genes of interest

legsta

Typing of *Legionella pneumophila* assemblies

LisSero

Serogroup typing prediction for *Listeria monocytogenes*

meningotype

Serotyping of *Neisseria meningitidis* assemblies

ngmaster

Multi-antigen sequence typing for *Neisseria gonorrhoeae*

SeqSero2

Salmonella serotype prediction from reads or assemblies

SISTR

Serovar prediction of *Salmonella* assemblies

spaTyper

Computational method for finding spa types in *Staphylococcus aureus*

SsuisSero

Serotype prediction of *Streptococcus suis* assemblies

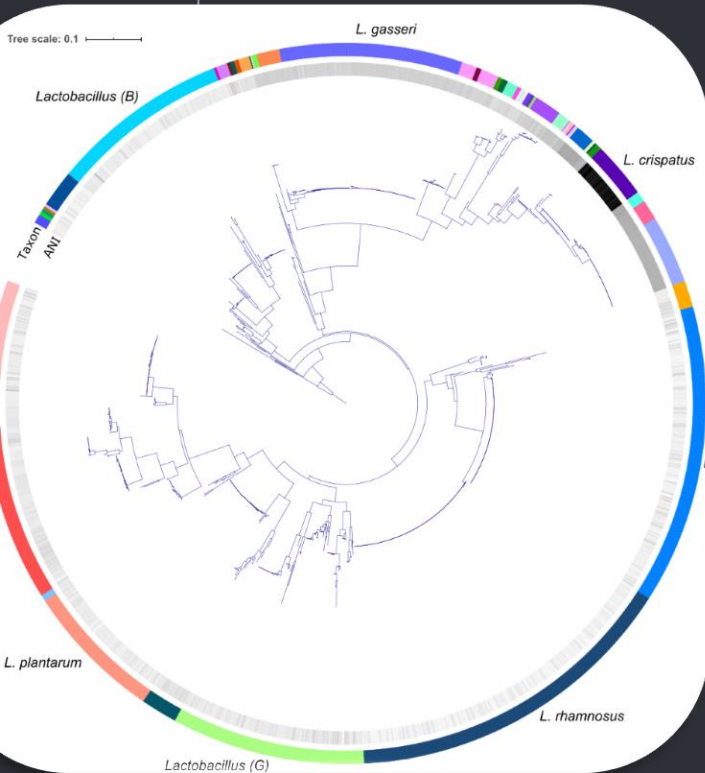
staphopia-secmec

Primer based SCCmec typing of *Staphylococcus aureus* genomes

TBProfiler

Detect resistance and lineages of *Mycobacterium tuberculosis*

- Bactopia Tools can simplify complex tasks



For example, you can quickly generate a phylogeny based on a core-genome, core-snps, 16S rRNA, or sketches.

3

Design descisions

Let's take a look at a few fundamental values behind Bactopia

“

Needed to set some ground rules to reduce the maintenance burden of Bactopia



Tools must be free and open-source



Tools must be available from Bioconda or Conda-Forge



Bactopia Tools must be available from nf-core/modules

“

These help ease the maintenance burden of Bactopia while unexpectedly facilitating contributions back to the community.

4

Enhancements to OSS

Let's learn how Bactopia contributes back to the community

“

As a developer of a pipeline making use of hundreds of open-source tools, it is very import to me that I find ways to contribute back to the community

● Bactopia Enhancements to OSS



Many tools originally developed for Bactopia have been [made available as stand-alone tools](#), including [dragonflye](#), [pbptyper](#), [fastq-df](#), and others. These tools have been downloaded more than 100,000 times from Conda.



By using tools from Conda, it has facilitated [contributions to Bioconda and Conda-Forge](#). To date [29 new recipes](#) have been added, [35 recipes updated](#), and more than [2,000 Bioconda pull requests reviewed](#).



Requiring Bactopia Tools be available from [nf-core/modules](#), has also facilitated [contributions to nf-core/modules](#). To date, [62 contributions](#) have been made including [46 new modules](#) and [16 modules updated](#).



Occasionally users or CI testing may identify bugs in tools used Bactopia. If a fix is identified, it is submitted upstream to the tool. This has led to [18 contributions](#) to tools including: [Ariba](#), [Bowtie2](#), [Kleborate](#), [Seroba](#), [Shovill](#), [ShigaTyper](#), and others.



BACTOPIA Enhancements to Open Source Science

“

This separation simplifies the on-going maintenance of Bactopia and these tools.



Through Bactopia contributions, I was invited to join the Bioconda Core Team, and the nf-core Modules Team

5

Future Directions

Let's see what's on the radar for Bactopia



Bactopia v3!

Major changes are inbound, and you'll get a glimpse today!

● Evolution of Bactopia

Circa 2011 –
Development of
Staphopia begins

1

Circa 2017 – Staphopia
adopts Nextflow, Conda,
and Containerization

3

Circa 2021 – Bactopia v2 is
released being rewritten in
Nextflow DSL2

5

Circa 2014 – Staphopia
adopts a Python
workflow manager
called Ruffus

2

Circa 2019 – Staphopia
is generalized to all
bacteria, and thus
Bactopia v1 is born

4

Today – Bactopia is nearing
version 3. With many
improvements geared towards
on-going surveillance

6

- Major changes inbound

- Bactopia Datasets no longer needed
- Multiple species per-run
- Improved directory structure
- A new Python package
 - <https://github.com/bactopia/bactopia-py>
- Additional workflows and Bactopia Tools
- Everything is a Bactopia Tool
- Many fixes and improvements
 - <https://github.com/bactopia/bactopia/blob/dev/CHANGELOG.md>

New Named Workflows



Art by Joan Stark

tetodon v3.0.0

Host removal and taxon classification with estimated abundances

Typical pipeline command:

```
tetodon --fastqs samples.txt -profile singularity
```

Clean Yr Reads

clean-yr-reads v3.0.0

Use Bactopia's read QC steps to Clean-Yr-Reads

Typical pipeline command:

```
clean-yr-reads --fastqs samples.txt -profile singularity
```

No significant changes required, just reshuffling Bactopia Tools

- On the radar

- - Customizable reports, starting with MultiQC
 - Opened the door to metagenomics
 - Full Nextflow Tower support
 - Terra.bio support depending on demand
 - R Shiny app to view results interactively
 - Additional features to the Python package
 - Documentation updates
 - Always more Bactopia Tools

6

Wrap Up

Let's see close this out and get the workshop started!

Bactopia is a robust pipeline for bacterial genome analysis



Accessible

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



Comprehensive

Bactopia is a start-to-finish pipeline which includes numerous tools and 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.



Who and where is Bactopia being used?

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

849,200

Docker container pulls – that's a lot!



47,500

Conda environments built

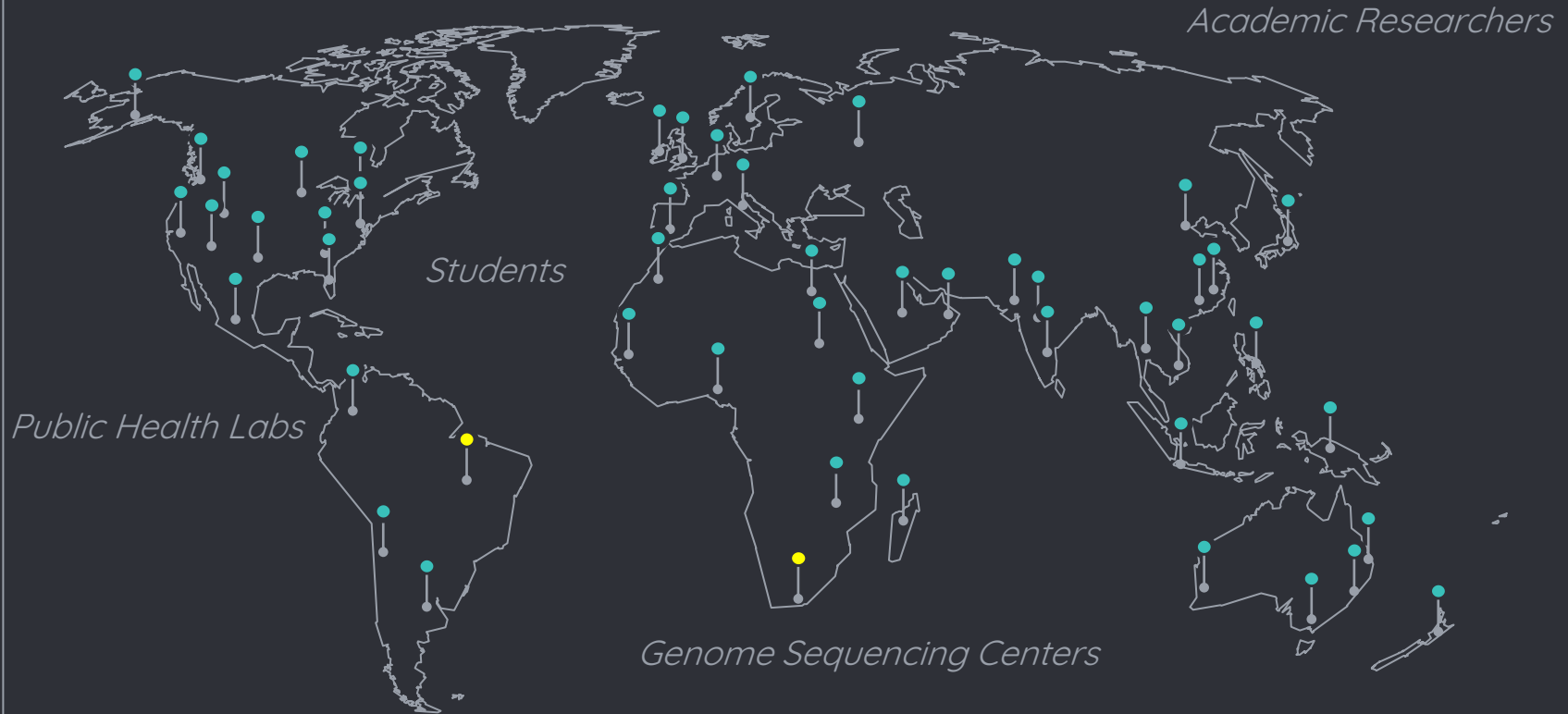
34,500 visitors

Many users around the world visiting the docs

229

GitHub stars!

● Bactopia Users Across the Globe



Many of these dots are in response to direct communications

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



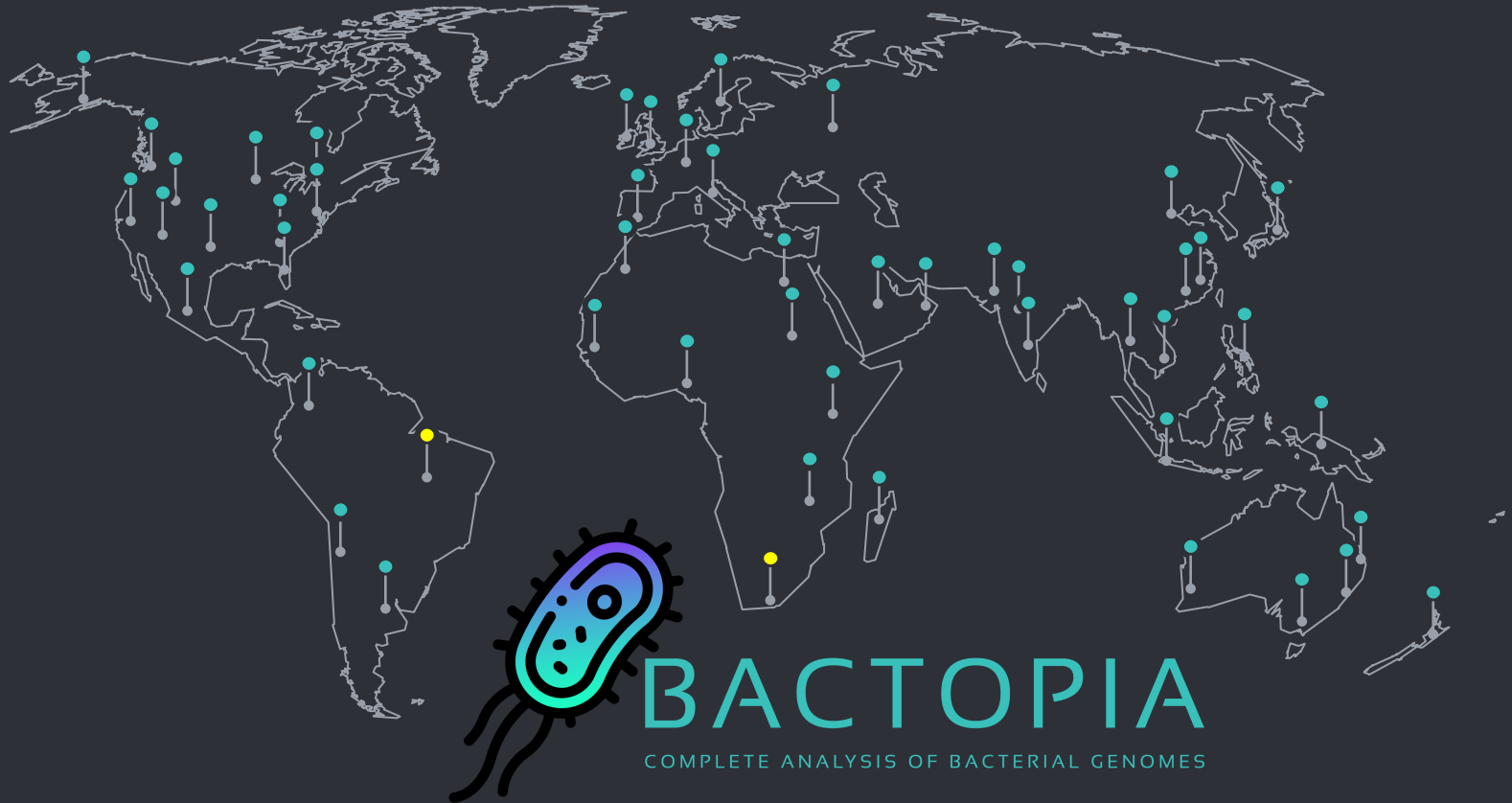
BIOCONDA®

nextflow

nf-core 

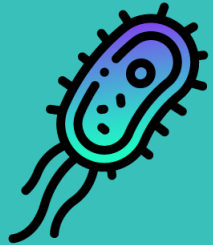
Presentation template by [SlidesCarnival](#)

Any Questions?





Bactopia Workshop



Bactopia Tools Workshop

