

# Overview

- Quick recap of the day
- Deeper dives into a few topics
- Bactopia
- Recap

We've thrown a lot at you today!



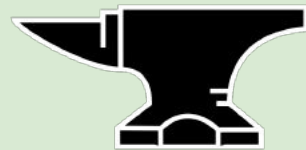
docker



nextflow



CONDA<sup>®</sup>



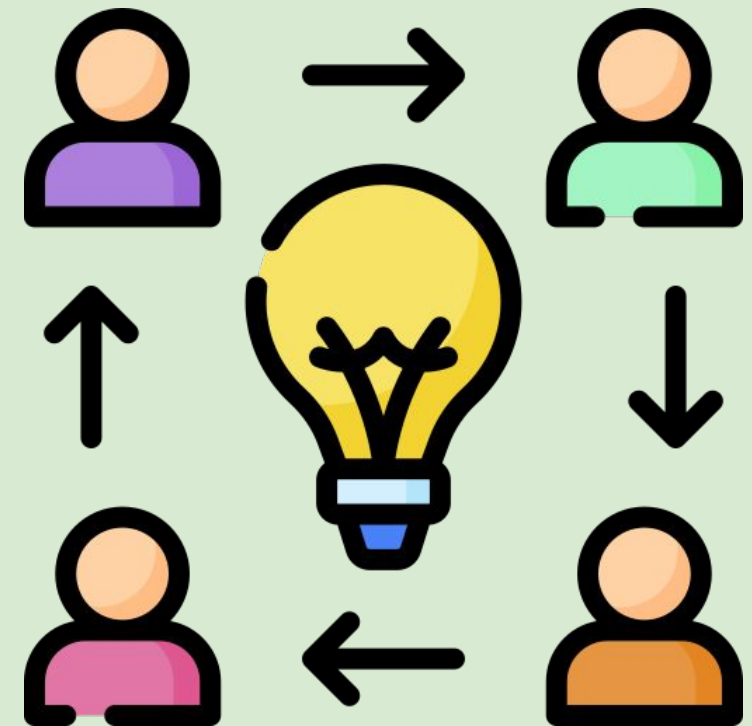
CONDA-FORGE

BIOCONDA<sup>®</sup>

Deeper dives into a few topics

# BIOCONDA<sup>®</sup> for all your bioinformatic tools

- Truly a community driven repository
  - More than 1,300 people have contributed
- Makes bioinformatics accessible
- Downstream containerization
  - Docker - Biocontainers
    - <https://biocontainers.pro/>
  - Singularity Images - Galaxy Project
    - <https://depot.galaxyproject.org/singularity/>
- Currently ~4,000 recipes are available
  - Includes tools and libraries
- Reach out if you need help getting a tool on Bioconda!



# nf-core 🍏 pushing Nextflow to the limits

- Community effort to collect curated Nextflow pipelines
  - 2400 Slack users, 1000+ GitHub contributors
- Includes 58 hi-quality bioinformatic pipelines
  - [rnaseq](#), [mag](#), [bactmap](#), [many more](#)
- Standardized [guidelines](#) for developers
- Thorough review process produces robust pipelines
- **Event notice!** [nf-core Hackathon March 16th-18th](#)



# Repositories to make your workflow-life easier!

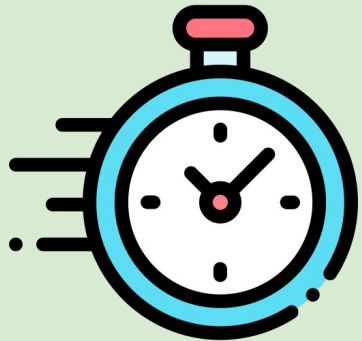
**nf-core/**  
**modules**



- nf-core modules
  - <https://github.com/nf-core/modules>
  - 200+ DSL2 modules available



- Snakemake Wrappers
  - <https://github.com/snakemake/snakemake-wrappers>
  - 100+ wrappers available

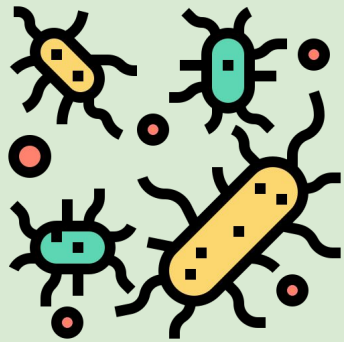


- Allows rapid prototyping, leading to lots of time saved

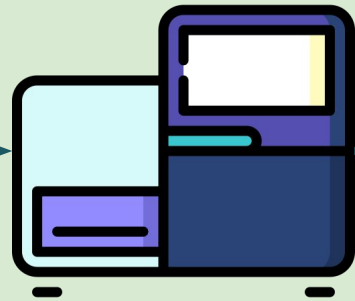
Workflows and analyzing bacterial genomes



# Bacterial genomics in a nutshell



Extract DNA



Sequence DNA

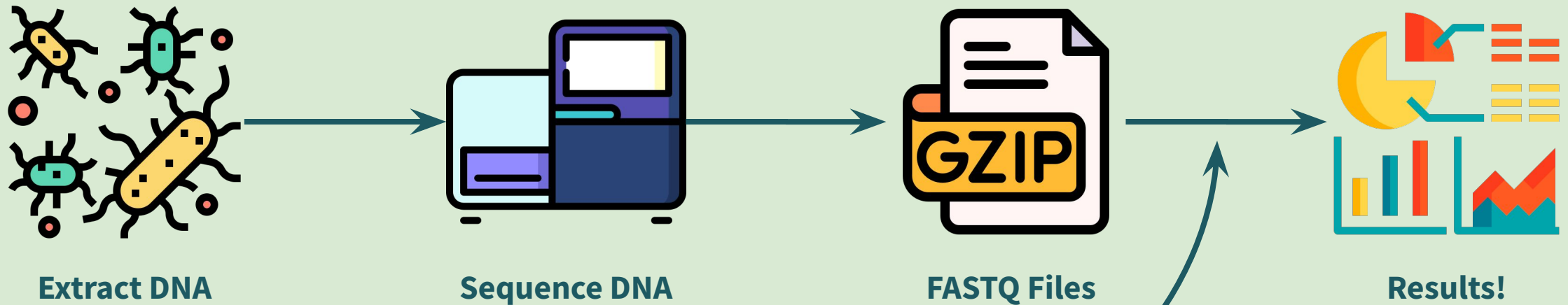


FASTQ Files



Results!

# Bacterial genomics in a nutshell



What options are there for analysis?

# Some options for analyzing your samples

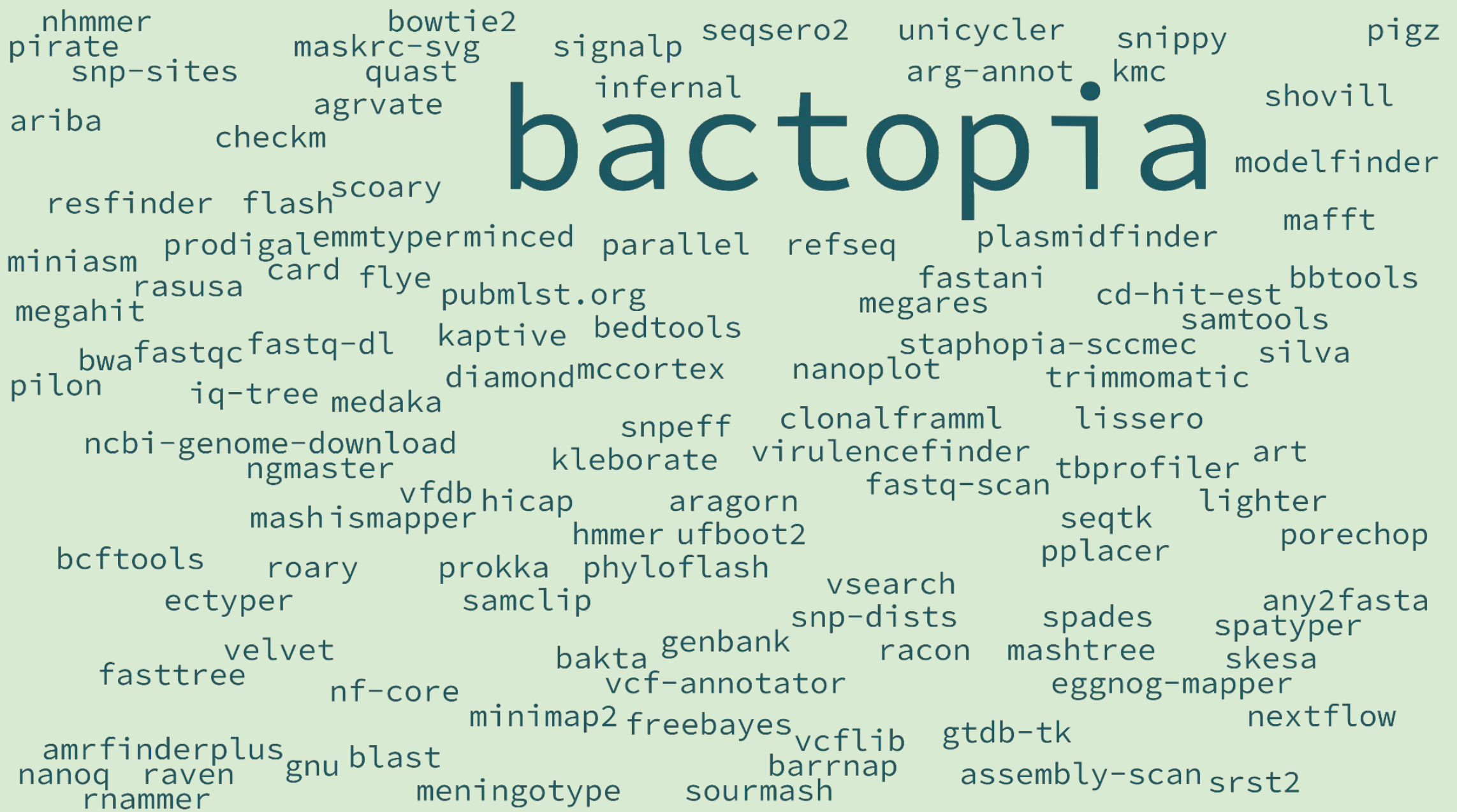
- Ad-hoc analysis on the command-line
  - Useful for one-off, quick answers
- Toss the commands into a bash script
  - Gets the job done, but we can do better
- Create workflow with a workflow manager
  - Can be time consuming, you're in complete control
- Use an existing workflow
  - Saves time, standardized, but might be obstacles



I'm biased, but consider using an existing workflow!

But which one should you consider? There are many

Again I'm biased, but consider using Bactopia!



**Bactopia is a pipeline for the complete analysis of bacterial genomes!**

**Includes 100+ bioinformatic tools!**

# Bactopia is an all-in-one workflow



- Built using Nextflow DSL2 and available from Bioconda
- Allows inputs from:
  - Cloud storage buckets or local file systems
  - Can download from SRA or Assembly (FASTA) database
  - Illumina or Nanopore FASTQs
- Includes steps for:
  - QC, assembly, annotation, MLST, Antimicrobial Resistance, SNP and InDel calling, Minmer sketching and querying, BLAST, mapping, etc...



# Multiple “helper” commands are included



- bactopia citations
  - Print citations for all datasets, tools and Bactopia
  - <https://bactopia.github.io/acknowledgements/>
- bactopia datasets
  - Download and setup useful datasets for Bactopia
- bactopia download
  - Build Conda, Docker or Singularity environments for all steps in bactopia
- bactopia prepare
  - Create a file of filenames for analysis
- bactopia search
  - Generate a list of SRA accessions for analysis

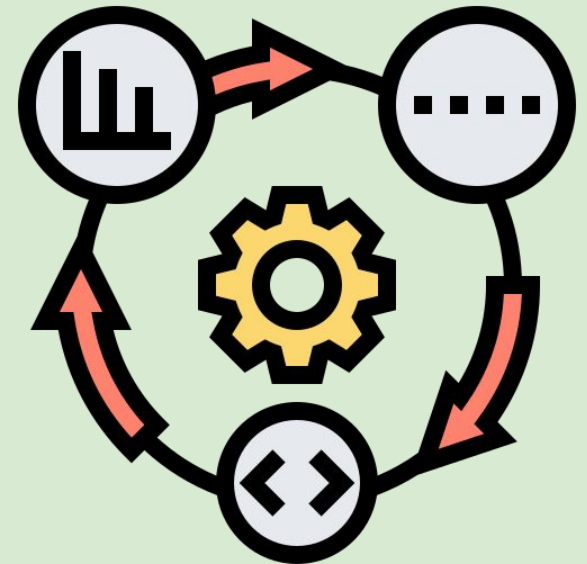
# Bactopia Datasets supplement analyses

- Including datasets is easy and really brings a lot of value to your analysis
- Includes:
  - Sketches for all RefSeq and GenBank
  - Reference genomes
  - Custom proteins for annotation
  - MLST schemas
  - Ariba datasets
  - Genes, proteins and primers



# Bactopia Tools provide independent workflows

- Allow easy comparative analysis of Bactopia outputs
  - e.g. pan-genome analysis with core-genome phylogeny
- Two Types:
  - Modules: single tool workflows (ECTyper, Kleborate, TB Profiler)
  - Subworkflows: Connect multiple modules together
    - Prokka -> PIRATE -> IQ-Tree -> SNP-Dists
- Frame-worked for easy addition
  - Simple command to create new Bactopia Tool
- **Because of DSL2, every Bactopia Tool can be reused**



## Available Bactopia Tools

Subworkflows (5)	Description
<a href="#">eggnog</a>	Functional annotation of proteins using orthologous groups and phylogenies
<a href="#">gtdb</a>	Identify marker genes and assign taxonomic classifications
<a href="#">merlin</a>	MinmER assisted species-specific bactopia tool seLectIoN
<a href="#">pangenome</a>	Pangenome analysis with optional core-genome phylogeny
<a href="#">staphtyper</a>	Determine the agr, spa and SCCmec types for <i>Staphylococcus aureus</i> genomes

## Use existing Bactopia outputs

## All built with DSL2 Modules

Modules (17)	Description
<a href="#">agrivate</a>	Rapid identification of <i>Staphylococcus aureus</i> agr locus type and agr operon variants.
<a href="#">bakta</a>	Rapid annotation of bacterial genomes and plasmids
<a href="#">ectyper</a>	In-silico prediction of <i>Escherichia coli</i> serotype
<a href="#">emmtyper</a>	emm-typing of <i>Streptococcus pyogenes</i> assemblies
<a href="#">fastani</a>	fast alignment-free computation of whole-genome Average Nucleotide Identity (ANI)
<a href="#">hicap</a>	Identify cap locus serotype and structure in your <i>Haemophilus influenzae</i> assemblies
<a href="#">ismapper</a>	Identify insertion sites positions in bacterial genomes
<a href="#">kleborate</a>	Screening Klebsiella genome assemblies for MLST, sub-species, and other related genes of interest
<a href="#">lissero</a>	Serogroup typing prediction for <i>Listeria monocytogenes</i>
<a href="#">mashdist</a>	Calculate Mash distances between sequences
<a href="#">mashtree</a>	Quickly create a tree using Mash distances
<a href="#">meningotype</a>	Serotyping of <i>Neisseria meningitidis</i>
<a href="#">ngmaster</a>	Multi-antigen sequence typing for <i>Neisseria gonorrhoeae</i>
<a href="#">seqsero2</a>	Salmonella serotype prediction from reads or assemblies
<a href="#">spatyper</a>	Computational method for finding spa types in <i>Staphylococcus aureus</i>
<a href="#">staphopiasccmec</a>	Primer based SCCmec typing of <i>Staphylococcus aureus</i> genomes
<a href="#">tbprofiler</a>	Detect resistance and lineages of <i>Mycobacterium tuberculosis</i> genomes

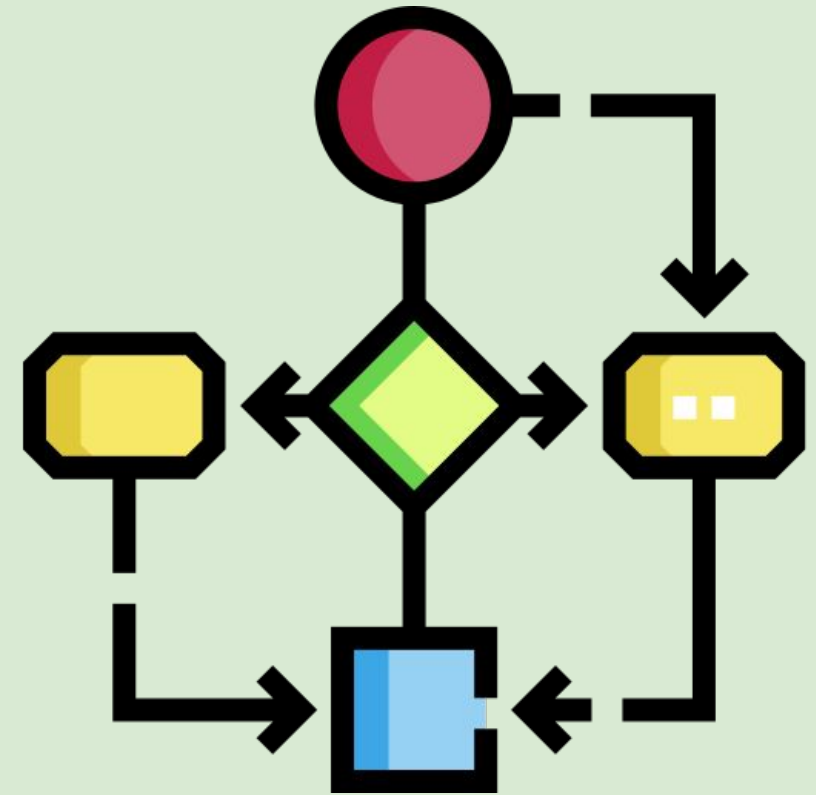
## Types of analyses

- Species-specific tools
- Pan-genome
- Phylogeny
- Annotation

**Because of DSL2, every Bactopia Tool can be reused!**

# Demonstration of DSL2 reusability

- Bactopia includes 10+ species-specific subworkflows
- Sometimes the species is unknown
- Can a new subworkflow be used to automatically execute these subworkflows based on some distance?

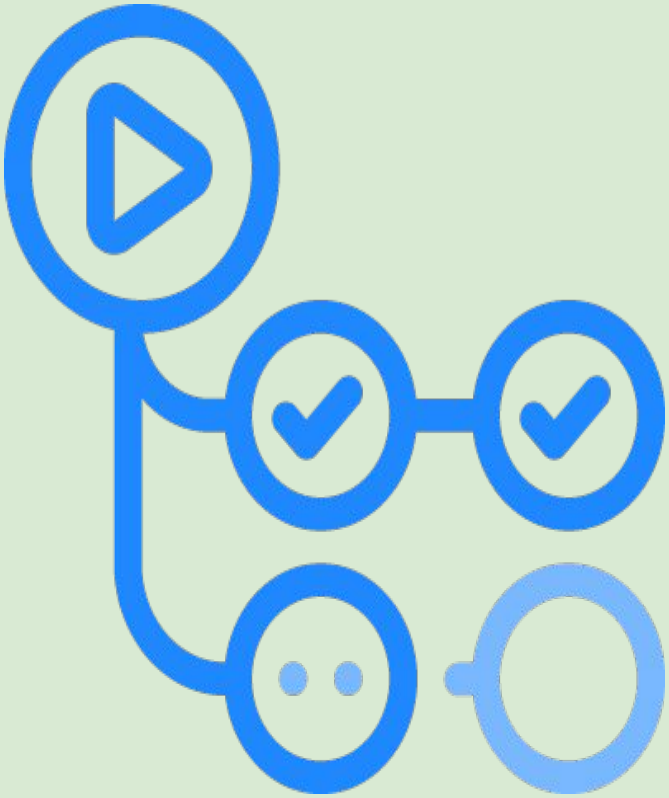


# Merlin - **MinmER** assisted species-specific bactopia tool se**LectioN**

- Use mash distances to select species-specific tools
  - Uses RefSeq Mash Sketch downloaded by Bactopia Datasets
- Currently includes members of:
  - Escherichia, Haemophilus, Klebsiella, Listeria, Mycobacterium, Neisseria, Salmonella, Shigella, Staphylococcus, Streptococcus
- Demonstrates module reuse and condition-based execution
- Name created with ACRONYM
  - Acronym CReatiON for You and Me
  - <https://github.com/bacook17/acronym>



# Tests keep the wheels turning



- Every step is tested in Bactopia using real data
  - Usages, workflows, Subworkflows, modules
- More than 10,000 output files are tested
  - MD5sums match, contains strings, file exists
- Testing conducted by pytest
  - Adapted from nf-core/modules
- All automated through GitHub Actions



# By using Nextflow, Bactopia is platform independent

- Available on:
  - Bioconda, Docker, Singularity
  - Google Cloud Platform
    - Google Life Sciences
  - Amazon Web Services
    - Batch
  - Microsoft Azure
    - Batch
  - Nextflow Tower
  - Terra.bio
  - HPC
- All done with '-profile <PLATFORM>' or your custom config
- Snakemake and Cromwell offer similar portability



# Opportunities to contribute back to the community



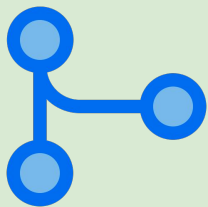
Bactopia requires tools be available from Bioconda

- Easy to install, downstream Docker and Singularity containers
- Submitted 17 recipes, reviewed 1000+ Bioconda pull requests



All Bactopia Tools must be available from nf-core/modules

- Developers can use these modules in their own workflows
- Submitted 25 modules (~6%) to nf-core/modules



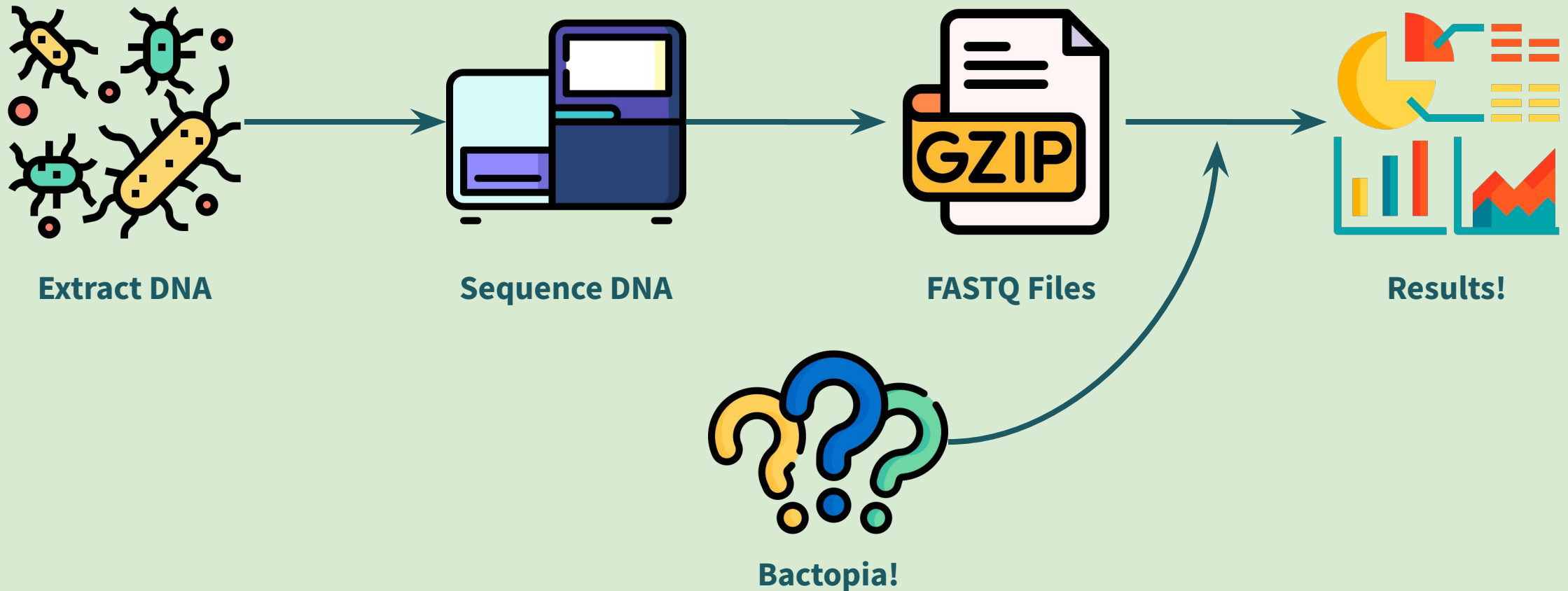
Multiple PRs submitted, and issues resolved

- Find bug in “X” tool, submit PR with fix
- Bowtie2, Prokka, Shovill, PIRATE, Ariba, ISMapper, PhyloFlash, etc....

# Bactopia Wrap Up

- Nextflow DSL2, following nf-core standards and practices
- Complete analysis of bacterial genomes
  - Includes 100+ (and growing) bioinformatic tools
- Still have a lot in the works!
  - Standard reports, many more Bactopia Tools, example outputs
- You can use Bactopia modules in your own workflows!
- A tutorial is available at <https://bactopia.github.io/tutorial/>

# Try Bactopia for your bacterial genomics needs!



We threw even more at you!

**Bactopia**



**docker**



**nextflow**



**{wd1}**

**CONDA®**



**CONDA-FORGE**

**BIOCONDA®**



**nf-core**



# Useful Links

- Bactopia
  - <https://bactopia.github.io/>
- Nextflow:
  - <https://www.nextflow.io/>
- nf-core:
  - <https://nf-co.re/>
- nf-core/modules
  - <https://github.com/nf-core/modules>
- Snakemake
  - <https://snakemake.github.io/>
- Snakemake Wrappers
  - <https://github.com/snakemake/snakemake-wrappers>
- Workflow Development Language (WDL)
  - <https://openwdl.org/>
- Docker
  - <https://www.docker.com/>
- Singularity
  - <https://sylabs.io/singularity>
- Bioconda
  - <https://bioconda.github.io/>
- Biocontainers
  - <https://biocontainers.pro/>
- Galaxy Project
  - <https://galaxyproject.org/>