

nf-core/



bytesize

Bactopia & it's usage of nf-core components

Robert A. Petit III, PhD

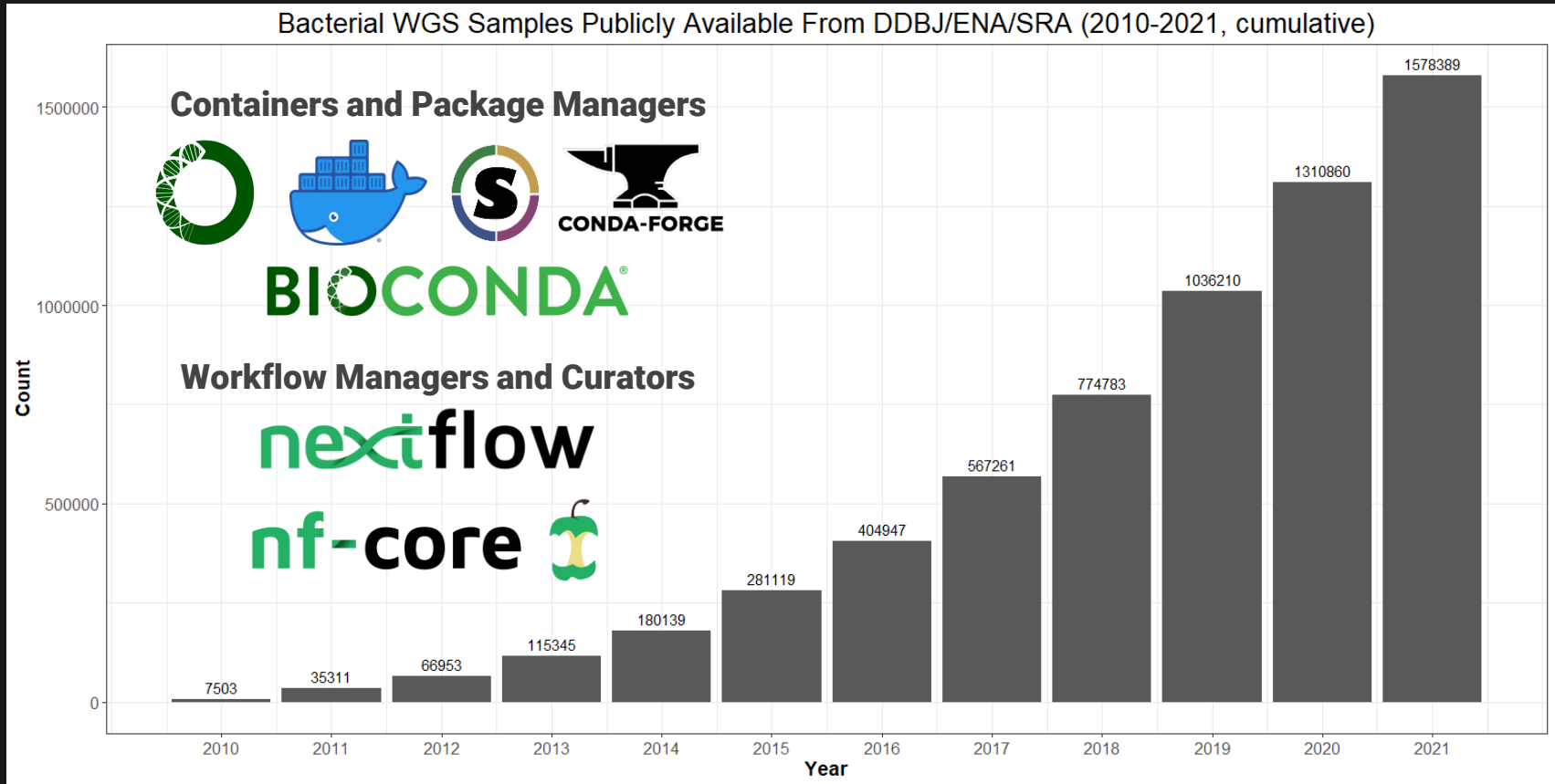


@rpetit3



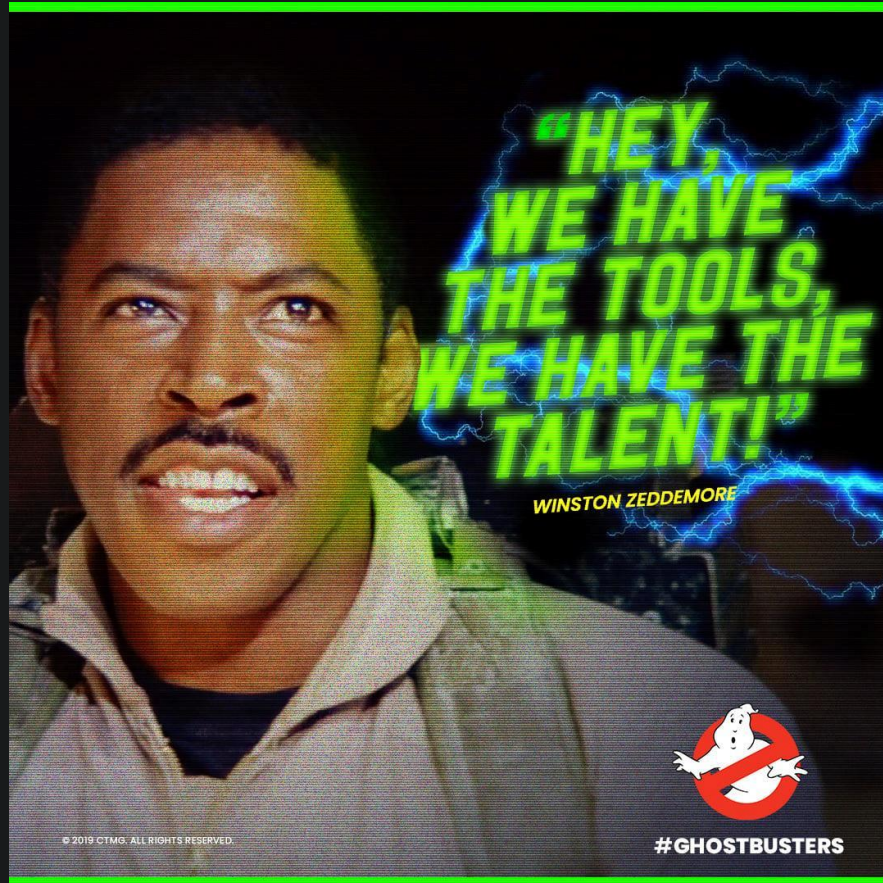
@rpetit3

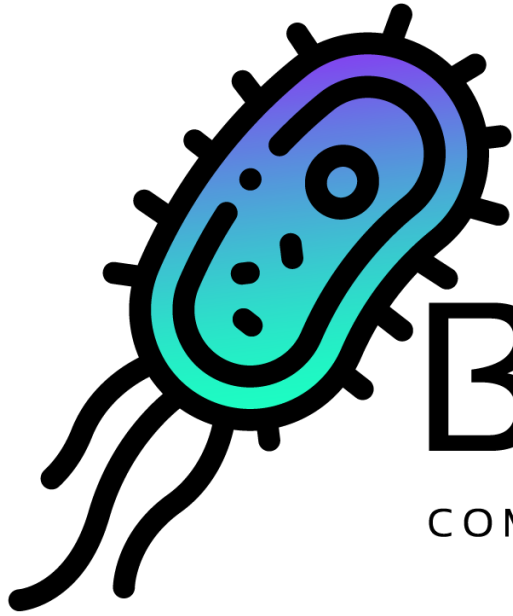
Lots of bacterial genomes and bioinformatic growth in the last decade



Can we make use of all these publicly available bacterial genomes?

Yes, we can!





BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

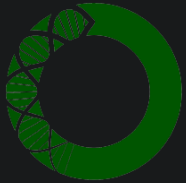
Bactopia is a scalable, reproducible, and portable all-in-one pipeline.

A few Bactopia highlights

- Supports Illumina and Nanopore reads
 - Local or from public databases
- Includes more than [140 bioinformatic tools](#)
- [45+ Bactopia Tools](#) provide more workflows for more science
- Extensively tested with 100+ tests for 10,000+ output files
- Available on Bioconda, Docker, and Singularity
- Well documented at bactopia.github.io



Bactopia design principles



Bactopia requires tools be available from Bioconda



All modules in Bactopia Tools must be on nf-core/modules



Bactopia should be easy to install and adaptable to user needs

Bactopia Helpers, help get started using Bactopia



- `bactopia citations`
 - Print citations for all datasets, tools and Bactopia
- `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

Accepted Inputs



Illumina and/or Nanopore Reads
--R1/--R2, --SE, --ont, --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



BACTOPIA

COMPLETE ANALYSIS OF BACTERIAL GENOMES

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

- Allow easy comparative analysis of Bactopia outputs
- Two Types:
 - Single tool (Kleborate, TB Profiler)
 - Multiple tools connected together
 - pangenome: Prokka -> PIRATE -> IQ-Tree
- 45+ Bactopia Tools are available
 - Frame-worked for easy addition
 - Example: [pangenome Bactopia Tool](#)



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

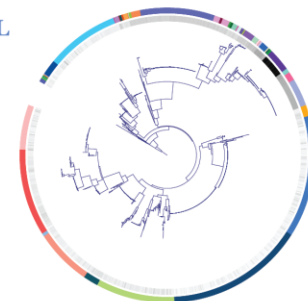
Assess the assembly quality of your samples

mlst

Automatic MLST calling from assembled contigs

BACTOPIA TOOLS

More workflows for more science



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

emntyper

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

staphlopia-scemec

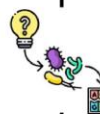
Primer based SCCmec typing of Staphylococcus aureus genomes

TBProfiler

Detect resistance and lineages of Mycobacterium tuberculosis

In a few steps, go from raw data to investigating results

Bactopia Step-By-Step



STEP 01

Sequence Bacterial Genomes

Bactopia supports sequences from Illumina and Nanopore technologies, as well as Sequence Read Archive and NCBI Assembly accessions.



STEP 02

Install Bactopia

Bactopia is [easily installed using Conda](#) and analyses can be run using Conda, Docker, or Singularity.

STEP 03

Search Public Databases

'[bactopia_search](#)' can supplement your analysis with publicly available genome sequences



STEP 05

Create a Sample Sheet

'[bactopia_prepare](#)' is used to create a sample sheet allowing you to efficiently process thousands of genomes

STEP 04

Build Datasets

'[bactopia_datasets](#)' downloads publicly available datasets to further supplement your analyses



STEP 06

Process Genomes with Bactopia

Bactopia processes genomes with more than 130 bioinformatic tools and is easily adapted to fit your needs.



Note: Bactopia is described above in the metro-style map



STEP 07

BACTOPIA TOOLS

Note: [Bactopia Tools](#) are described to the left

Further Analyze Genomes With Bactopia Tools
After Bactopia is finished, more than 30 additional workflows, called [Bactopia Tools](#), are available to take your analyses even further.

STEP 08

Investigate Results

Once completed, [numerous output files](#) describing each of your samples are available. You can use these outputs in your genomic investigations.



Adopting nf-core components has made much of this more achievable in Bactopia, and you should consider the same for your non-nf-core pipeline

A group of volunteers always ready to share their vast amount of experience



A few ways Bactopia is making use of
nf-core components

nf-core library in Bactopia



- Bactopia has 45+ workflows in from a single-entry point ([main.nf](#))
- [Dynamic imports](#) of JSON schemes based on [workflow config](#)
 - Each module gets its own [params.config](#) and [params.json](#)
- Required script to aggregate into a “[nextflow_schema.json](#)” (Nextflow Tower)

nf-core/modules in Bactopia

- Every Bactopia Tool is available from nf-core modules
 - Slight modifications on the Bactopia-side
 - Ex. AgrVATE: [nf-core/modules](#) vs [Bactopia](#)
- Similar pytest framework as nf-core/modules
 - Tests every step in [Bactopia](#) and [Bactopia Tools](#)
 - Uses self-hosted GitHub Action runner



nf-core meta.yaml template for documentation



- Bactopia adds [additional metadata](#)
 - [Citations](#), [markdown tables](#), [output tree](#)
- Yaml is used in [Bactopia docs](#)
 - Built with Jinja2, Mkdocs and GitHub Actions
- *Writing docs while building*

What's next for Bactopia?

MultiQC Modules

More nf-core/modules

Eyes on future of configs

rich-click everything



Report generation

More Bactopia Tools

Custom workflows for surveillance

Enhanced CLI

Don't hesitate to reach out if you think I can help!

Need help?

Repository: [bactopia/bactopia](https://github.com/bactopia/bactopia)

Documentation: bactopia.github.io

Always looking for feedback!



Taylor Fearing
Jim Mildenerger
Chayse Rowley

Follow nf-core on



**Chan
Zuckerberg
Initiative** 

<https://nf-co.re>

Icons:

openmoji.org



EMORY
UNIVERSITY
SCHOOL OF
MEDICINE

Department of Medicine

Timothy D. Read, PhD