



UNIVERSITY of the  
WESTERN CAPE



SANBI  
South African National  
Bioinformatics Institute

# Harnessing Scientific Workflows for AMR Surveillance

Peter van Heusden (@pvanheus)

South African National Bioinformatics Institute  
(SANBI)

# Antibiotics & Resistance

## b Mechanisms of antibiotic resistance

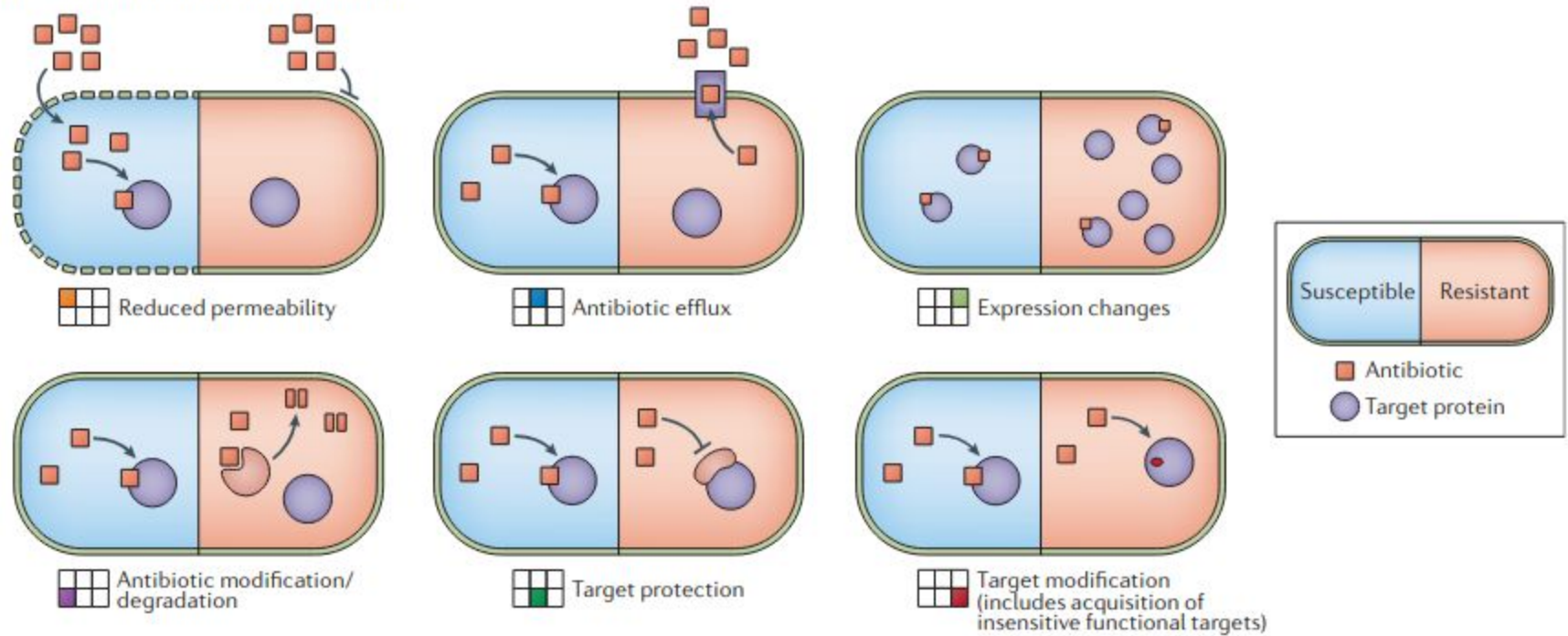
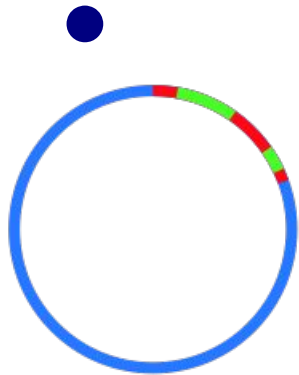


Figure from doi:[10.1038/s41576-019-0108-4](https://doi.org/10.1038/s41576-019-0108-4)

# Genomic Mechanisms of AMR



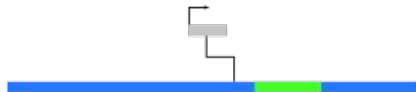
Horizontal  
Gene Transfer



Point Mutation



Insertion or  
Deletion



Change in Promoter



Gene Duplication

# Why use WGS?

---

- One Assay, Many Answers:
  - Species Identification and Classification
    - MLST / cgMLST
    - Contamination ?
    - Lineage
  - Small (single nucleotide) variants
    - Clustering / transmission investigation
  - Sample characteristics
    - AMR gene / mutation

# Sequence Based Methods for AMR Detection

---

- Define the search:
  - Surveillance:
    - How much of (known) X are we seeing?
    - What's the trend? Is this an outbreak?
  - Research:
    - Are we seeing something new?
    - How did this bug survive???

# What Sequence Looks Like

## FASTQ (from the sequencer)

Identifier	●	@SRR566546.970 HWUSI-EAS1673_11067_FC7070M:4:1:2299:1109 length=50
Sequence	●	TTGCCTGCCTATCATTTTAGTGCCTGTGAGGTGGAGATGTGAGGATCAGT
'+' sign	●	+
Quality scores	●	hhhhhhhhhhghhghhhhhfhhhhhfffffe'ee['X]b[d[ed'[Y[~Y
Identifier	●	@SRR566546.971 HWUSI-EAS1673_11067_FC7070M:4:1:2374:1108 length=50
Sequence	●	GATTTGTATGAAAGTATACAACATAAACTGCAGGTGGATCAGAGTAAGTC
'+' sign	●	+
Quality scores	●	hhhhghfhcgghghggfcffdhfehhhhcehdchhdhahehffffde'bVd

## FASTA

Header	●	>VIT_201s0011g03530.1
Sequence	●	AATTAAGCATAAATACTCACTCTTACCCCCTTATTTTCTTATCTCTCATCACTTTTGGTGCGAAG
	●	GACCATGAGAACAAGCTGCAATGGGTGTAGGGTTCTTCGCAAGGCATGCAGCCAAGACTGCATCA
Header	●	>VIT_201s0011g03540.1
Sequence	●	CAGGTAGCGTGAAGTTAAACCCTAGCGCTTTAGACAAACAGCTGTAGTCACCGCCCACAAACACC
	●	AGCCTCTGAGACACCACCTCAAACCTTTCCACTTAAATACACATCCCTCACACCCTTTTCAATTC
Header	●	>VIT_201s0011g03550.1
Sequence	●	CATGCAAAGCTGAACGCGATGCTGTGATTGGTGGTAAGTGGTAGTTGAGTAAATTTGACAGTGAA
	●	GCCGAAATGGTAAAAGACTAAGGCTAGAAGTAGAATACCACTGTTCTTCTCATCACGTGGGCCCA

# What Sequence Looks Like



Short Reads (Illumina, 250bp or less)



Long Reads  
(Oxford Nanopore, 1000s of bp)



Assemblies ("full" genome, multiple contigs)



# How We Look for Resistance

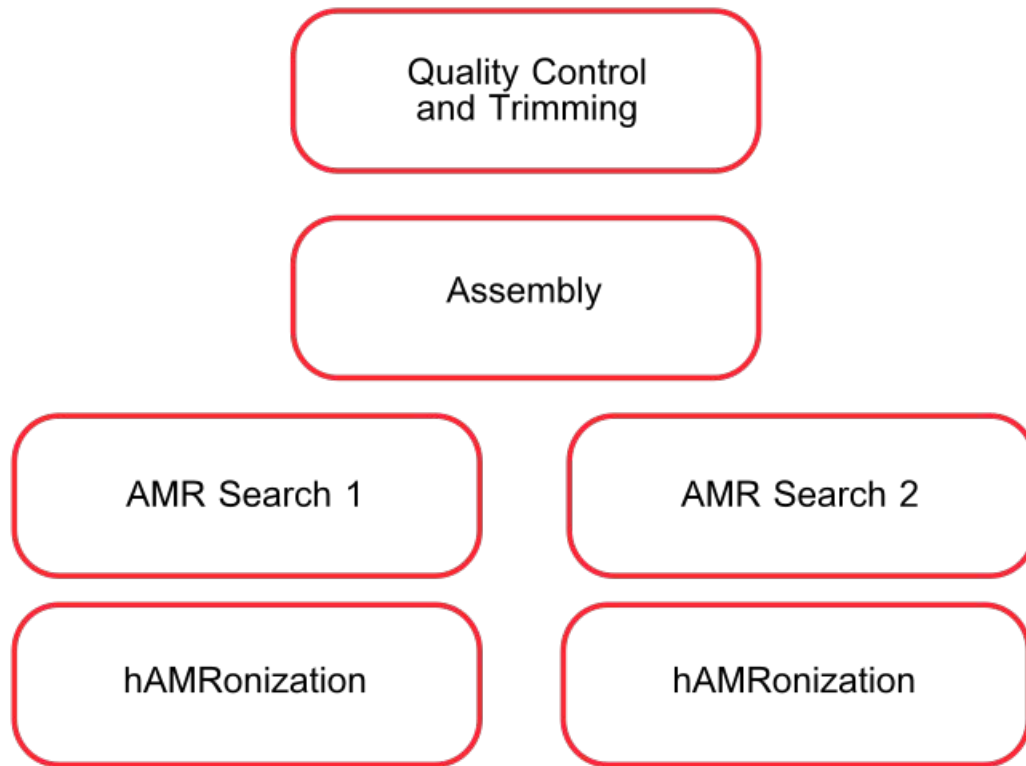
---

- Reads:
  - Quality Control essential
  - Mapped reads: point mutations
  - Per-read error can be high
- From Assemblies:
  - Assemblies need QC too
  - Repetitive sequence challenges assembly
- Learn lots more from the PHA4GE / JPIAMR / CLIMB-BD [workshop](#)



# AMR Workflow

---



# Lots of Tools, Lots of Steps

---

Modern Workflow solutions:


1. Handle software packaging
2. Make user interface more consistent
3. Allow tasks to be chained in workflows









# Galaxy Project

---


- Web based environment:
  - <https://usegalaxy.eu> etc
  - Specialised servers:
    - ARIES Galaxy:
      - <https://galaxyproject.org/use/aries/>
    - GalaxyTrakr:
      - <https://galaxyproject.org/use/galaxytrakr/>
- Project with development community, training resources, etc:
  - <https://training.galaxyproject.org/>
  - Gitter chat (PH, Training, microGalaxy)

# Galaxy UI

 Galaxy Europe

 Workflow  Visualize  Shared Data  Help  User   

Using 28%

 [TIP] Are you uploading large files? Give a try to the FTP service! Easier and faster. Instruction on <https://galaxyproject.eu/ftp>

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Fetch Sequences / Alignments

GENOMICS ANALYSIS

Annotation

COVID-19 Research!


Want to learn the best practices for the analysis of SARS-CoV-2 data using Galaxy? Visit the [Galaxy SARS-CoV-2 portal](#). We mirror **all public SARS-CoV-2 data** from [ENA](#) in a [Galaxy data library](#) for your convenience. The Galaxy community has created [COVID-19 dedicated training materials](#). Please check our [recent activities](#) for more details.

If you need help submitting your data to public archives, like ENA, please [get in touch](#). We will support you in sharing your data.


"Anyone, anywhere in the world should have free, unhindered access to not just my research, but to the research of every great and enquiring mind across the spectrum of human understanding." – Prof. Stephen Hawking

News

Jan 29, 2022

 UseGalaxy.eu Tool Updates for 2022-01-29


Jan 23, 2022

 UseGalaxy.eu Tool Updates for 2022-01-23


Jan 20, 2022

Usegalaxy.eu surpassed 40.000 registered users

Jan 14, 2022

 UseGalaxy.eu Tool Updates for 2022-01-14

Dec 17, 2021



 UseGalaxy.eu Tool Updates for 2021-12-17

Dec 16, 2021




Usegalaxy.eu project infrastructure and CVE-2021-44228 vulnerability

Events



Feb 21, 2022 - Feb 25, 2022

  Galaxy HTS data analysis workshop for CRC 1425, CRC 992 and MeInBio members



Mar 7, 2022

   Single-cell RNA-seq analysis using Galaxy




Mar 10, 2022

  microGalaxy Community Call




Mar 14, 2022 - Mar 18, 2022

  GTN Smörgåsbord 2

Mar 29, 2022

   Microbial Community Analysis Workshop 2022

Apr 27, 2022

   Galaxy introduction for life scientists

History

search datasets

fastp test

27 shown, 15 deleted, 41 hidden

17.45 MB

76: ABRicate Summary on data 67 and data 66

73: hamronize on collection 65: Output

72: hamronize on data 61: Output

65: ABRicate on collection 32 report file

64: staramr on data 38 and data 35: results.xlsx

63: staramr on data 38 and data 35: settings.txt

62: staramr on data 38 and data 35: plasmidfinder.tsv

61: staramr on data 38 and data 35: resfinder.tsv

60: staramr on data 38 and data 35: detailed\_summ

OPEN CHAT

Galaxy Europe

Workflow
 Visualize
 Shared Data
 Help
 User

Using 28%

[TIP] Are you uploading large files? Give a try to the FTP service! Easier and faster. Instruction on <https://galaxyproject.eu/ftp>

Tools

Inputs

Get Data

Send Data

Collection Operations

Expression Tools

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

Convert Formats

FASTA/FASTQ

Quality Control

SAM/BAM

BED

VCF/BCF

Nanopore

COMMON GENOMICS TOOLS

M. tuberculosis Variant Analysis tutorial

Read 1

output (input)

FastQC

Short read data from your current history

Contaminant list

Adapter list

Submodule and Limit specifying file

FastQC on input dataset(s): Webpage (html)

FastQC on input dataset(s): RawData (txt)

Read 2

output (input)

FastQC

Short read data from your current history

Contaminant list

Adapter list

Submodule and Limit specifying file

FastQC on input dataset(s): Webpage (html)

FastQC on input dataset(s): RawData (txt)

Trimmomatic

Input FASTQ file (R1/first of pair)

Input FASTQ file (R2/second of pair)

fastq\_out\_r1\_paired (input)

fastq\_out\_r2\_paired (input)

fastq\_out\_r1\_unpaired (input)

fastq\_out\_r2\_unpaired (input)

MultiQC

Results 1 > FastQC output 1 > FastQC output

MultiQC on input dataset(s): Stats (input)

MultiQC on input dataset(s): Plots (input)

mtbva\_multiqc\_webpage (html)

Kraken2

Forward strand

Reverse strand

mtbva\_kraken\_report (tabular)

Kraken2 on input dataset(s): Classification (tabular)

snippy

Use the following dataset as the reference sequence

Select first set of reads

Select second set of reads

mtbva\_snippy\_vcf (vcf)

Genbank

reference genome

TB Variant Filter

VCF file to be filter

output1 (vcf)

Genome annotation (GFF3)

output (input)

FASTA reference genome

output (input)

Text transformation

File to process

output (input)

Name

M. tuberculosis Variant Analysis tutorial

Version

3: Feb 19th 2021, 16 steps

Annotation

M. tuberculosis Variant Analysis

These notes will be visible when this workflow is viewed.

License

Specify a license for this workflow.

Creator

Add a new creator - either a person or an organization.

Tags

variant-analysis

prokaryote

Apply tags to make it easy to search for and find items with the same tag.

Apply tags to make it easy to search for and find items with the same tag.

# Getting The Most From Galaxy

---

- Multiple Samples grouped into Collections
- Tools: many (not all) tools available
  - contributions welcome (to tools-iuc, etc)
- Training:  
<https://training.galaxyproject.org>
  - [Smörgåsbord 2: 14-18 March](#)

# SAGESA & Galaxy

---

- We have a Slack, use it!
- Link with other communities:
  - microGalaxy
  - Galaxy Public Health
  - Galaxy Training Network
- Many roles in the Galaxy Community:
  - See [video](#)