



Prelab: Finding AMRs

Purpose

To use Galaxy tools to perform *de novo* genome assembly of sequencing reads into contigs, visualize the contigs, and find antimicrobial resistance (AMR) genes.

Learning Objectives

Use Galaxy platform to:

1. Perform *de novo* genome assembly of long reads into **contigs**, using the **Flye** tool.
2. Visualize contigs with the **Bandage Image** tool.
3. Identify AMRs in the contig assemblies using the **ABRicate** tool.

Introduction

Genome assembly is the process of reconstructing genomes from DNA sequencing reads. Accurate and continuous genome assembly from sequenced fragments, even very long fragments, is challenging. **Flye**, is a long-read assembly algorithm that aims to produce highly contiguous genome assemblies and overcome some of the assembly challenges, like repetitive DNA sequences. Using Flye we will reconstruct bacterial genomes and plasmids and enable detection of important genes implicated in antibiotic resistance. Ideally, after assembly we want to get back circular contigs as that would typically indicate the identification of entire microbial genome(s) or plasmid(s)!

Antimicrobial resistance is the ability of microbes to evade one or more antibiotics, and can lead to multidrug resistance and ability to survive and even thrive in the presence of antibiotics. Detecting and studying antibiotic-resistant pathogens is therefore extremely important to human health. However, the environmental reservoirs of resistance determinants are poorly understood. Certainly the indiscriminate and sometimes inappropriate use of antibiotics by humans (e.g. in the hospitals, in food production) has contributed to the emergence of resistant bacterial strains, but there are many other ways microbes can acquire AMRs. For example, the environment like soil is emerging as a key reservoir of these antibiotic resistance genes. For more information on AMRs see the following review articles: [10.1038/nrmicro2312](https://doi.org/10.1038/nrmicro2312) and [10.3390/antibiotics13121112](https://doi.org/10.3390/antibiotics13121112).

Overview of the approximate minimum times for a job to be completed on Galaxy using specified tools.

- Note, these times apply only to the specific input file we will be using in this activity, the Zymo_Gut_Standard_D6331_subset that is ~342.5MB, and will take longer (or much longer) for larger input files.

Flye	Bandage Image	ABRicate
5 min	< 5 min	< 5 min

Activity 1 – Flye assembly

Estimated time: 45 min

The sample used in this activity is the [Zymo Gut Microbiome Standard](#), sequenced by Pacific Biosciences using PacBio Sequel II Instrument, and corresponds to sequencing read file SRR13128014. A subset of this data is used in this activity.

Instructions

- Run Flye assembly tool in Galaxy using Zymo Gut Standard D6331 subset2 to assemble gut microbial genomes.
 - Obtain **.fastq** file from Zymo_Gut_Standard_D6331_subset2
 - <https://usegalaxy.org/u/valerie-g/h/zymo-gut-standard-d6331-subset2>
 - Name your new history “**Finding gut AMRs**”.
 - Run **Flye** tool to assess sequence quality using the following Tool Parameters:
 - Under **Input Reads**: select your aymo_gut_standard_D6331_subset2 **.fastq** dataset.
 - Under **Mode**: select PacBio HiFi (--pacbio-hifi) option, since the sequences were obtained using PacBio HiFi sequencing technology.
 - Under **Perform metagenomic assembly**: select Yes.
 - Under **Generate a log file**: select Yes.
- View Flye results - Explore Flye output files and answer questions below.

Questions

- Explore Flye tool purpose and output.

1A. In your own words describe the purpose of Flye based on the Introduction section, and from the ‘Purpose’ section of Flye tool description.

1B. How many Flye output files did you get back, and what are they?

1C. What are the file extensions (formats) for the following Flye output files?

consensus:

graphical fragment assembly:

2. Explore Flye log report file.

2A. Based on your Flye output `log_report_file`: At the very bottom of the very long file find how many bases were assembled. What is the `Total_length`?

2B. Based on your Flye output `log_report_file`: At the very bottom of the file find the length of the longest assembled fragment? Look for `Largest_frg`.

2C. What proportion of input was assembled into contigs?

*At the very top of the log file you will find that the input number of bases was 177,760,975 (Look for `Total read length`). Compare to the total length after assembly from your answer to question 2A above.

3. Explore Flye assembly info file.

3A. Based on your Flye output `assembly_info_file` sorted by contig length (high to low, in base pairs, bp) - What is the longest contig size?

3B. Based on your Flye output `assembly_info_file` sorted by contig length (high to low, in base pairs, bp) - What is the shortest contig size?

4. Explore Flye output consensus file (in FASTA format).

4A. What is the beginning line of the FASTA format?

4B. How does FASTA sequence format differ from FASTQ sequence format?

See this link for a quick summary comparison of the 2 formats

<https://compgenomr.github.io/book/fasta-and-fastq-formats.html>

5. Test your general understanding of genome assembly.

5A. Summarize your experience with assembling a genome with Flye.

E.g., were you surprised at the percentage of assembled input, at the length of the largest contig, at the abundance of linear contigs, or anything else?

5B. What would you want the ideal genome assembly tool to do?

Activity 2 – Visualizing contigs

Estimated time: 15 min

Instructions

Run **Bandage Image** tool in Galaxy on your Flye: **graphical fragment assembly** output (in gfa1 format) using default parameters.

Questions

1. Paste the resulting image below.

2. Describe contig profile based on the Bandage Image.

3. Do you expect to obtain more contigs, including larger and circular contigs, with more sequencing reads? Why?

Activity 3 – Finding AMRs

Estimated time: 30 min

Instructions

1. Run **ABRicate** tool in Galaxy using Flye **consensus** as input using the following Tool Parameters:
 - Under Input Reads: select your Flye: consensus output in **FASTA** format
 - IMPORTANT: Under Advanced Options: select **NCBI Bacterial Antimicrobial Resistance Reference Gene Database** as your database option; the default 'resfinder' may not work well.
2. Explore **ABRicate report file**.

ABRicate output report has the following information:

Column	Description
FILE	The filename this hit came from
SEQUENCE	The sequence in the filename
START	Start coordinate in the sequence
END	End coordinate in the sequence
GENE	ABR gene name
COVERAGE	What proportion of the gene is in our sequence
COVERAGE_MAP	A visual of coverage map (gaps or no gaps)
GAPS	Was there any gaps in the alignment - possible pseudogene?
%COVERAGE	Proportion of gene covered
%IDENTITY	Proportion of exact nucleotide matches
DATABASE	The database this sequence comes from
ACCESSION	The genomic source of the sequence

3. Answer questions below.

Questions

1. Explore ABRicate output report.

1A. How many AMR genes were detected? This is the number of rows in your file.

1B. How many DIFFERENT AMR genes were detected and what are their GENE names?

1C. What are the different AMR genes resistant to? What is their RESISTANCE?

1D. How many DIFFERENT contigs had AMRs?

2. Research an AMR gene.

- Use any search tools for your research, but we encourage you to use PubMed <https://pubmed.ncbi.nlm.nih.gov> where you can find many scientific articles on the topic if you search for e.g. your AMR gene name, or resistance name or using a sentence as input.
- Talk about anything of interest, e.g., which microbes have the AMR of interest, what is the substance to which the gene shows resistance to, where could the resistance to this substance come from, what are possible health implications, etc.

2A. Research and write a small paragraph report on one of the AMR genes.

2B. Ask one question you want to know about AMRs?

Grading Criteria

- Download as Microsoft Word (.docx) and upload on Canvas

Footnotes

Resources

- [Google Doc](#)

Contributions and Affiliations

- Valeriya Gaysinskaya, Johns Hopkins University
- Frederick Tan, Johns Hopkins University

Last Revised: May 2025