

Module Planning template

NGS Bioinformatics - Remote classrooms

Module Title: Genome Assembly/Annotation

→ **Module Lead/Co-leads/Assistants** - Eugene, Narendar

→ **Summary or objectives** – This module will demonstrate to course participants how to utilize various sequencing technologies discussed in previous modules of this course to generate a complete, annotated genome. Learning will be divided into two parts over two days: **Assembly** and **Annotation**. Skills learned will include the fundamental theory behind genome assembly, short and long read assembly and quality control, and the annotation of newly assembled genomes with information such as repeats and genes.

→ **List of learning outcomes specific for this module**

- ◆ Understanding the fundamentals of how sequencing reads can be used to assemble genomes
- ◆ Utilising Illumina short read technology to generate genome assemblies
- ◆ Utilising Pacific Biosciences long read technology to generate genome assemblies
- ◆ Utilising a combined (i.e. short and long read) approach to further improve genome assembly
- ◆ Comparing versions of different genome assemblies for the same organism
- ◆ Identifying repeat elements in assembled genomes
- ◆ Discovering protein coding genes
- ◆ Using comparative genomics for gene functional annotation

→ **Total number of hours/days to be spent on this content:** 8 hours over 2 days

→ **Tools/software and resources to be used**

- ◆ Software:
 - Assembly:
 - canu, jellyfish, velvet, assembly-stats, wtdbg2
 - Annotation:
 - mummer, repeatmasker, hisat2, samtools, augustus, genomethreader
- ◆ Datasets to be used:
 - Short and long read DNA sequencing of *Plasmodium falciparum* made available by the Wellcome Sanger Institute.
 - RNA-sequencing from Chappell et al. ***Refining the transcriptome of the human malaria parasite Plasmodium falciparum using***

amplification-free RNA-seq (2020). (DOI:
<https://doi.org/10.1186/s12864-020-06787-5>)

- ◆ Other resources, or readings: Please see the “practical exercise” PDFs for additional reading.

→ **Overview of activities and exercises** (list specific practical activities which participants will do)

Assembly Lecture: Explanation of the methods used for genome assembly and how they are being employed as part of the “Tree of Life” project at the Wellcome Sanger Institute

Assembly Practical: Assembly of a *Plasmodium falciparum* chromosome

Activity 1: Assembly with Pacific Biosciences long-read sequencing data

Activity 2: Understanding assembly algorithms

Activity 3: Assembly with Illumina short-read sequencing data

Activity 4: Estimating assembly quality

Annotation Lecture: Explanation of the methods used for genome annotation.

Annotation Practical: Annotation of a *Plasmodium falciparum* chromosome

Activity 1: Reference genome comparison

Activity 2: Identifying repetitive DNA in assembled genomes

Activity 3: Gene discovery in Eukaryotic genomes

Activity 4: Comparative genomics to determine gene function

→ **Assessment**

MCQ will assess lecture understanding

Each practical activity has a write-up which the students can submit on completion.

→ **Competencies**

If you are comfortable with competencies, please list the competency/ies this module will address:

-

Appendix (include any guideline documents related to content development such as breaking down content for remote classroom format, how to write LO's, designing assignments etc...)

See links to other guidelines

-Overall trainer guideline and template documents will include information on lecture formats and recordings, contact sessions schedules, checklists, assignments and assessments, quizzes, feedback, audio-video conferencing, learning management platform, teaching assistants
