An introduction to Nanopore DNA sequencing and data analysis

Asad Prodhan

Aims

❑ DNA extraction to Data generation

❑ Some key points in each step

❑ Now I have my data, how do I get started?

❑ Some resources

DNA extraction to Nanopore sequencing: an example

MinION
Hight Throughput Sequencing

Bacterial IDs

DNA extraction to Nanopore sequencing: an example

DNA Extraction ⇒ Library Preparation

MinION
Hight Throughput Sequencing

Sequencing Reads

Read 1: ATGCCGGTATTAA …
Read 2: CGGATTAACGAGTAT …
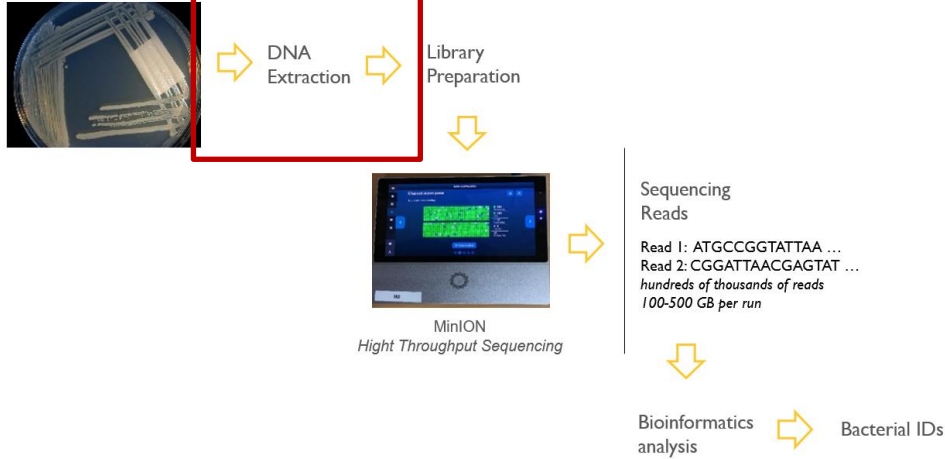*hundreds of thousands of reads*
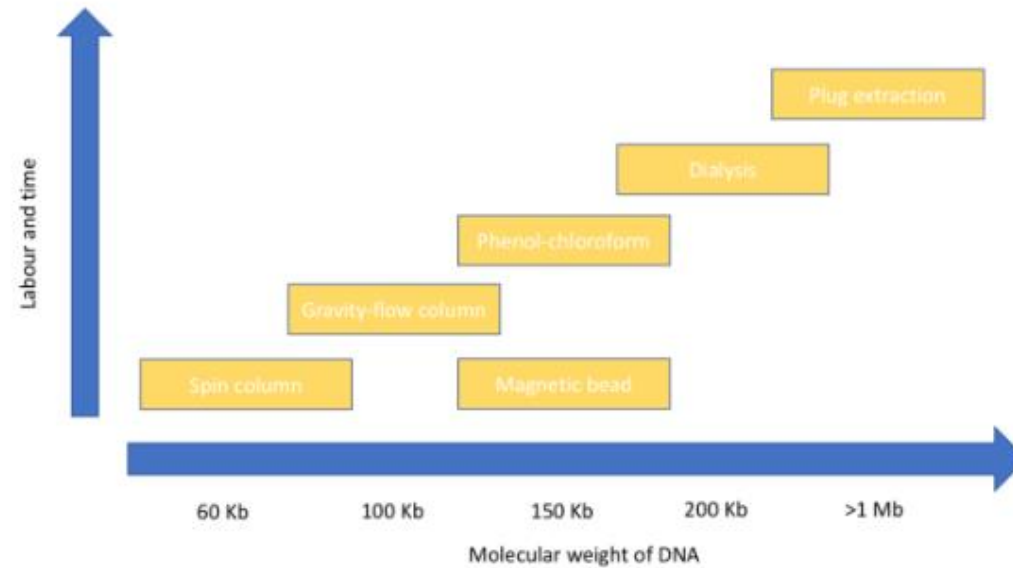*100-500 GB per run*

Bioinformatics analysis ⇒ Bacterial IDs

Dissecting Each Step

DNA Extraction

DNA extraction to Nanopore sequencing: an example

https://lab.loman.net/2018/05/25/dna-extraction-book-chapter/#__RefHeading___Toc505877552

- High Molecular Weight (HMW) DNA = ~ 20-100 Kb

- longer fragments = longer reads

- longer reads are suitable for *De novo* assembly and structural variations (deletion, insertion, duplication, inversion, SNP, translocation)

- Pipet slowly to minimise shearing DNA

- QC is required

- DNA fragment size selection is possible.

Dissecting Each Step
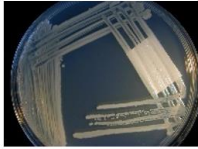
Library Preparation

DNA extraction to Nanopore sequencing: an example

DNA Extraction → Library Preparation → MinION *Hight Throughput Sequencing*

Sequencing Reads
Read 1: ATGCCGGTATTAA …
Read 2: CGGATTAACGAGTAT …
hundreds of thousands of reads
100-500 GB per run

Bioinformatics analysis → Bacterial IDs
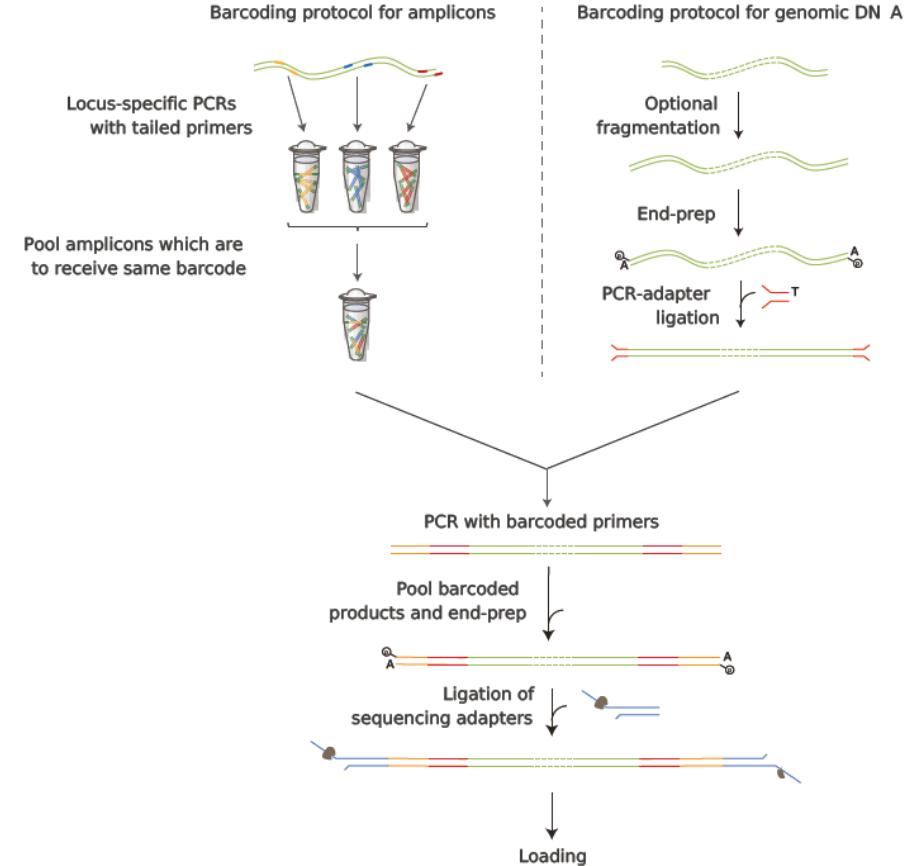
dsDNA: µg to pmol

DNA Length (bp)
1000
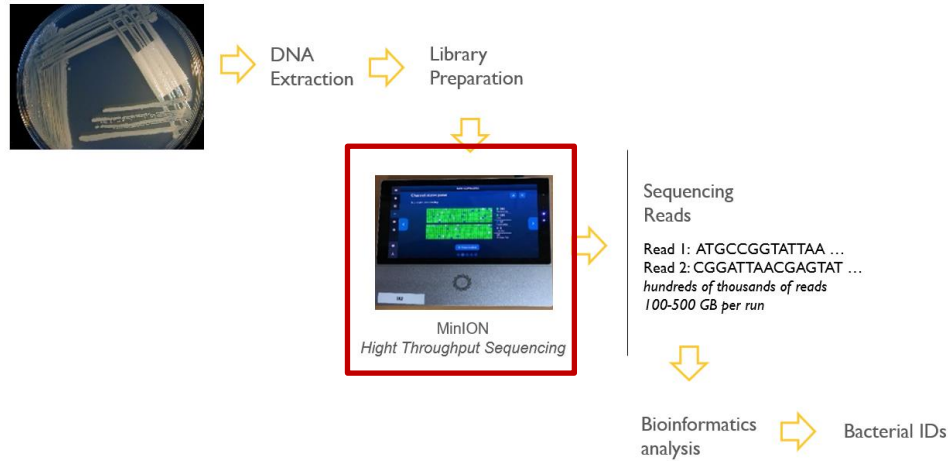
DNA Amount (µg)
0.066

0.1pmols of DNA

Clear   Calculate

- Take the recommended amount of input DNA

- How do we calculate fmol?

  - https://www.promega.com.au/resources/tools/biomath/

  - 1 pmol = 1000 fmol

  - 100 fmol = 1 Kb X 66 ng

- Prepare the ends of the DNA fragments

- Attach the barcodes and adapter sequences (need to be trimmed)

Barcoding protocol for amplicons

Locus-specific PCRs with tailed primers

Pool amplicons which are to receive same barcode

Barcoding protocol for genomic DN A

Optional fragmentation

End-prep

PCR-adapter ligation

PCR with barcoded primers

Pool barcoded products and end-prep

Ligation of sequencing adapters

Loading

Oxford
NANOPORE
Technologies

Dissecting Each Step

MinION

DNA extraction to Nanopore sequencing: an example

DNA Extraction → Library Preparation

MinION
Hight Throughput Sequencing

Sequencing Reads

Read 1: ATGCCGGTATTAA …
Read 2: CGGATTAACGAGTAT …
hundreds of thousands of reads
100-500 GB per run

Bioinformatics analysis → Bacterial IDs

Channel states panel

Run state: sequencing

● 346 Sequencing
● 131 Pore
● 16 Recovering
● 3 Inactive
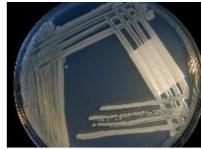16 Unclassified

+ Show detailed

Fig. MinION Mk1C

- Run time parameters

- Basecalling mode: fast, hac (high accuracy) or sup (supper accuracy)

- Min qscore = quality threshold for the reads to pass

- Minimum length filter = length threshold for the reads to pass

- Data output format: fast5 or pod5 (before basecalling), fastq (after basecalling)

Dissecting Each Step

DNA extraction to Nanopore sequencing: an example

Sequencing Reads

DNA Extraction → Library Preparation

MinION
Hight Throughput Sequencing

Sequencing Reads

Read 1: ATGCCGGTATTAA ...
Read 2: CGGATTAACGAGTAT ...
hundreds of thousands of reads
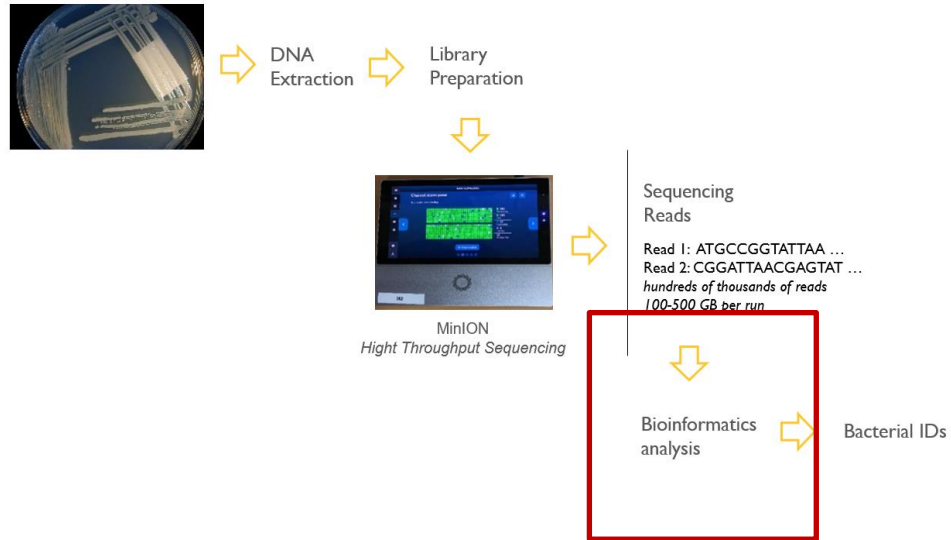*100-500 GB per run*

Bioinformatics analysis → Bacterial IDs

- fast5 & fastq format

- SFTP

- Transfer data from MinION to your Linux computer

- sftp minit@MinION_Number.local:/data; then get -r *

- fast5 will allow you to basecall again later if needed

  (https://github.com/asadprodhan/GPU-accelerated-guppy-basecalling)

- fast5 takes a lot of storage space

- storage space can be reduced by converting fast5 to slow5

  (https://github.com/asadprodhan/Reducing-storage-space-by-converting-Nanopore-fast5-to-slow5-using-slow5tools)

Dissecting Each Step

Bioinformatics Analysis

DNA extraction to Nanopore sequencing: an example

DNA Extraction → Library Preparation

MinION
Hight Throughput Sequencing

Sequencing Reads

Read 1: ATGCCGGTATTAA ...
Read 2: CGGATTAACGAGTAT ...
hundreds of thousands of reads
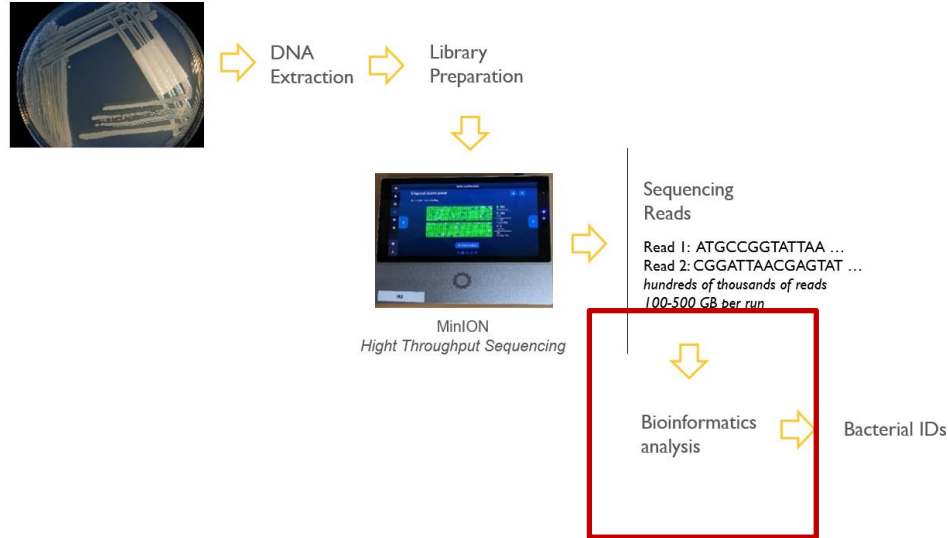*100-500 GB per run*

Bioinformatics analysis → Bacterial IDs

- Read relevant papers

- Identify bioinformatics tools

- How to install the tools?

you get an answer

Dissecting Each Step

DNA extraction to Nanopore sequencing: an example

DNA
Extraction

Library
Preparation

MinION
Hight Throughput Sequencing

Sequencing
Reads

Read 1: ATGCCGGTATTAA …
Read 2: CGGATTAACGAGTAT …
hundreds of thousands of reads
100-500 GB per run

Bioinformatics
analysis

Bacterial IDs

Bioinformatics Analysis

Step 1: Install conda
- ❑ How to check if conda already installed
- ❑ If not, how to install conda in your computer

Step 2: Create conda environment
- ❑ Why conda environment
- ❑ How to create, use, and share conda environment

Step 3: Install softwares within your conda environment
- ❑ How to install softwares using conda

Step 4: Perform your bioinformatics analysis
- ❑ How to download sequences from NCBI
- ❑ How to perform BLAST search
- ❑ How to sort out the BLAST hits
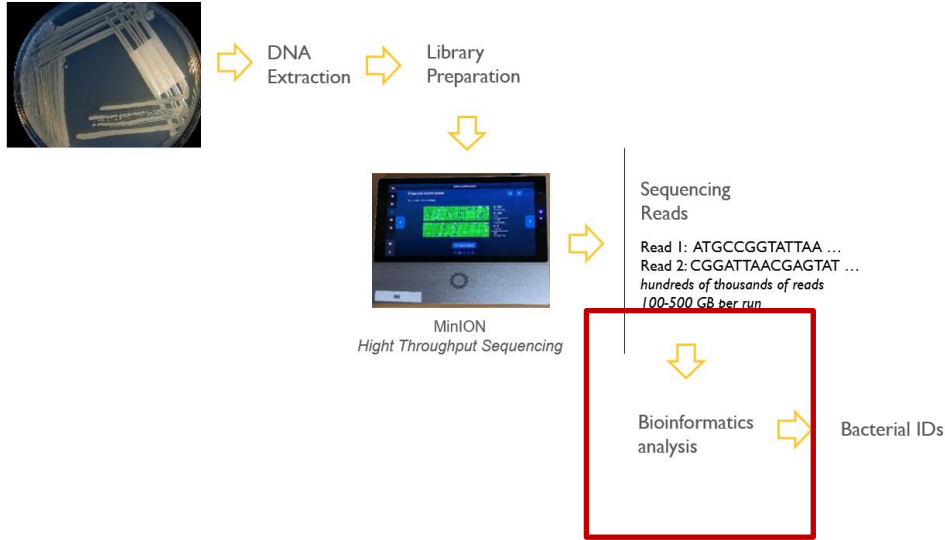- ❑ How to extract sequences of the BLAST hits

https://github.com/asadprodhan/A-beginner-s-guide-to-Bioinformatics

you get an answer

Dissecting Each Step

Bioinformatics Analysis

DNA extraction to Nanopore sequencing: an example

DNA Extraction → Library Preparation

Sequencing Reads

Read 1: ATGCCGGTATTAA …
Read 2: CGGATTAACGAGTAT …
hundreds of thousands of reads
100-500 GB per run

MinION
Hight Throughput Sequencing

Bioinformatics analysis → Bacterial IDs

Fig. Bioinformatics analysis workflow

you get an answer

Dissecting Each Step

DNA extraction to Nanopore sequencing: an example
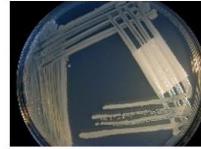
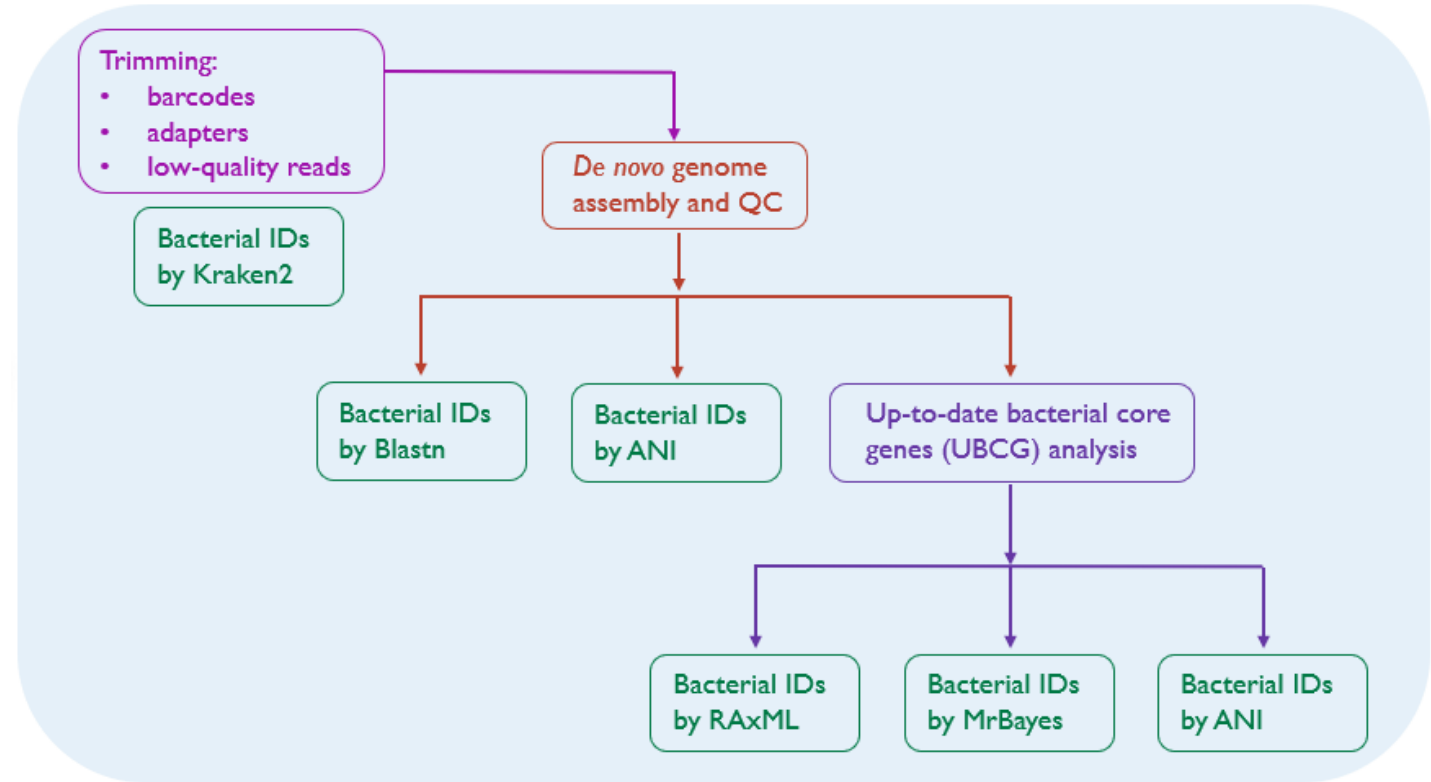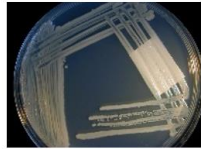DNA Extraction → Library Preparation → MinION *Hight Throughput Sequencing*

Sequencing Reads

Read 1: ATGCCGGTATTAA …
Read 2: CGGATTAACGAGTAT …
hundreds of thousands of reads
100-500 GB per run
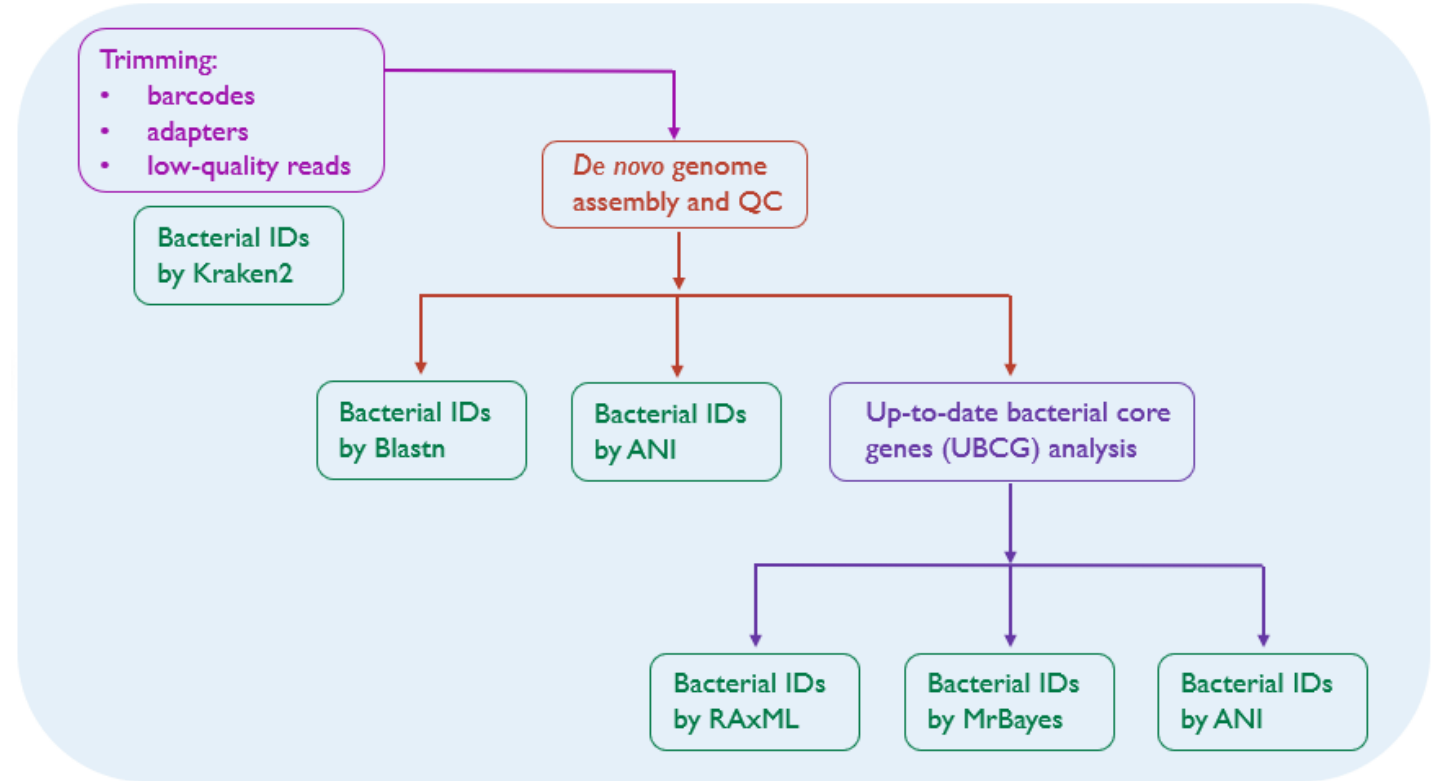
Bioinformatics analysis → Bacterial IDs

Bioinformatics Analysis

Fig. Bioinformatics analysis workflow

you get an answer

- Genome assembly
  (https://github.com/asadprodhan/Setting-up-and-running-Miniasm-assembler)
- Blastn
  (https://github.com/asadprodhan/A-beginner-s-guide-to-Bioinformatics)
- Kraken2
  (https://github.com/asadprodhan/How-to-run-Kraken2-on-HPC-using-Singularity-container-and-Nextflow)
- ANI
  (https://github.com/asadprodhan/Average-Nucleotide-Identity-ANI-analysis)
- UBCG & Phylogeny
  (https://github.com/asadprodhan/Bacterial-phylogenetic-tree-reconstruction-using-UBCG-pipeline)

Dissecting Each Step

Bioinformatics Analysis

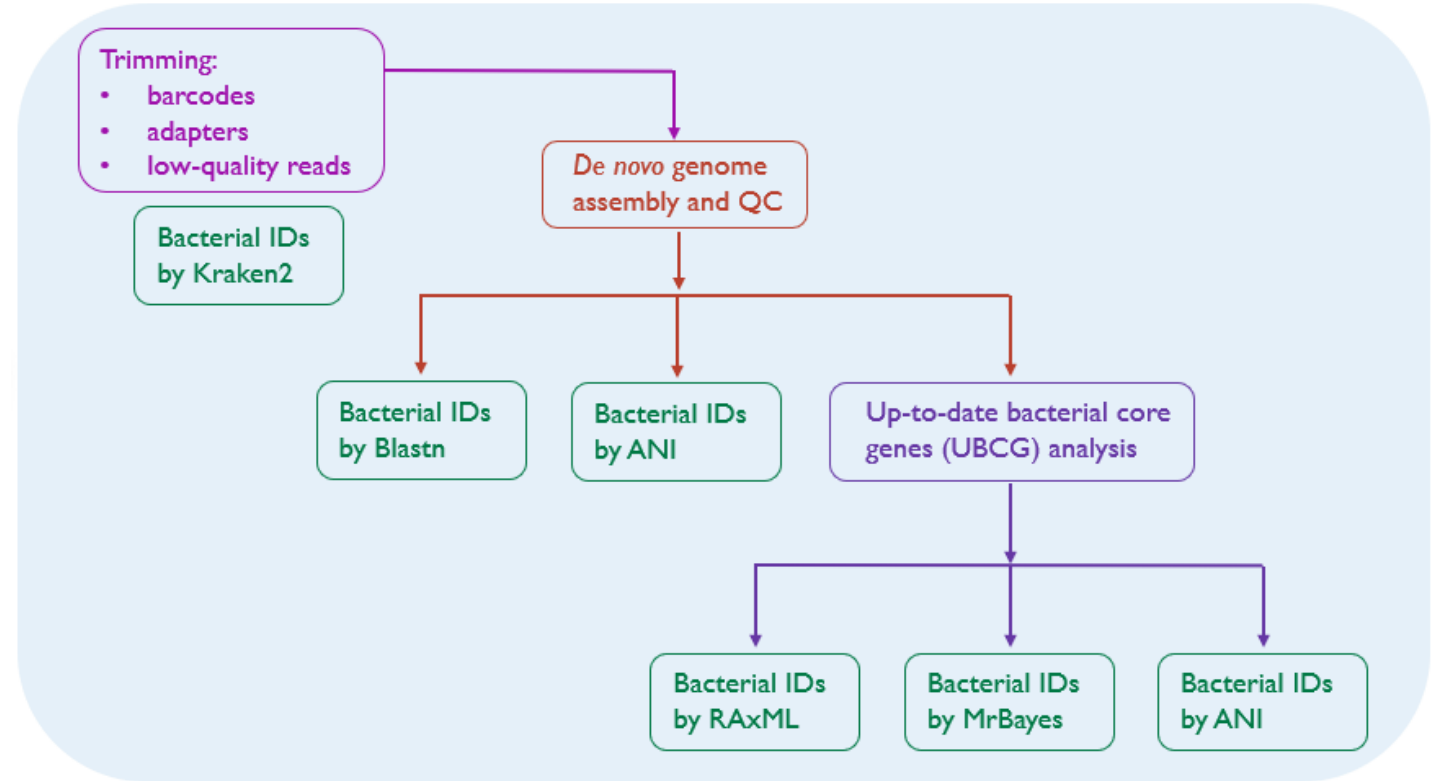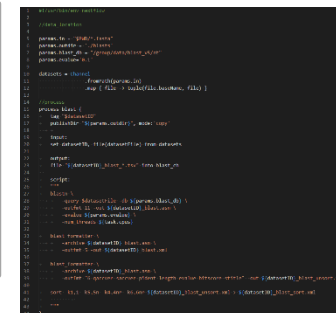**DNA extraction to Nanopore sequencing: an example**

Fig. Bioinformatics analysis workflow
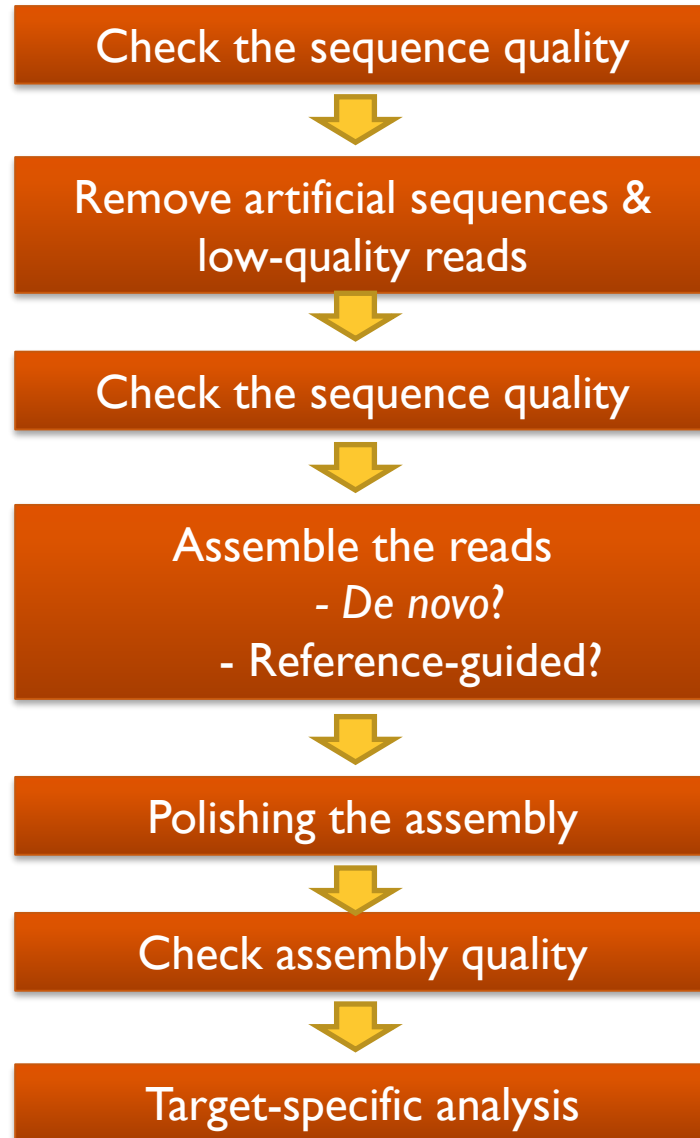
HPC Cluster
Nextflow & Singularity container
A collection of tutorials to learn coding from basics to bioinformatics.

(https://www.linkedin.com/in/asadprodhan/)

Summary

Analysis Steps	Example tools
Check the sequence quality	NanoPlot
Remove artificial sequences & low-quality reads	NanoFilt
Check the sequence quality	NanoPlot
Assemble the reads - *De novo*? - Reference-guided?	Miniasm, Flye, Canu, Spades etc
Polishing the assembly	Medaka, NanoPolish
Check assembly quality	Quast
Target-specific analysis	Blastn, Kraken2, ANI, Phylogeny etc

Acknowledgements

- Monica Kehoe

- Tonny Kinene

- Dominie Wright

- Cuiping Wang

Thank you

dpird.wa.gov.au