Genome Assembly using NanoDJ Notebooks in Cyverse

Presented by: Tamara Hernández-Beeftink
Hospital Nuestra Señora de Candelaria, Spain

September 13th, 2019
Outline of Today’s Focus Forum

• Next generation sequencing steps

• Oxford nanopore technology sequencing

• Assembly process

• NanoDJ bacteria genome assembly demo

*Due to time constraint and the complexity of the assembly process, we will not perform the assembly of a complete genome in real time.
Main steps of Next Generation Sequencing

1. Fragmentation
2. Sequencing
3. Assembly
4. De Novo
5. Alignment
Nanopore sequencing
Nanopore sequencing

- Portable
- Long reads
- Sequencing in real time
- Assemble complex genome
File formats (FAST5 → FASTQ)
Some assemblers

- Shorts reads: Velvet, SPAdes
- Long reads: Canu, Flye, Miniasm
- Short and long reads: Unicycler, MASurCa
Overlap Layout Consensus assembly (OLC)

AGCGATGAGCGTGACGAA
AGCGATGANCGTGACGAA
ATGAGCGTGACGAAGTT
GTGACGAAGTTGACCAG
AAGTTGACCNNGGCA

CCAGGCAGTGAGC
CAGGCAGTGAAGCT
CAGTGAGCTAGAC
AGCTAGACTCGAC

AGCGATGAGCGTGACGAAAGTTGACCAGGCAGTGAGCTAGACTCGAC

De Bruijn graph (DBG)

TCGG → CGGC → GGCA → GCAT
9X

CATT → ATTA → TTAC

AGTC → GTCG

TCGTC → CGTC → GTCA → TCAT
1X

AGTCG → GTCG

AGTTCG → GTCTC → GTCTAGC → TCTAGC
Input, basecalling, and simulations

Fast5 (signal)

Fast5 (basecalls)

Reference

Albacore

NanoSim

Fasta/q

https://github.com/genomicsITER/NanoDJ
### Input, basecalling, and simulations
- **Fast5 (signal)**
- **Fast5 (basecalls)**
- **Reference**

#### Albacore
- **NanoSim**
- **Fasta/q**

#### NanoSim
- **Fasta/q**

### Summary, QC and filtering
- **PoreChop**
  - **Demultiplex**
  - **Fasta/q**
  - **Biopython**

### Genome assembly and comparisons
- **Reference**
  - **BWA Rebaler**
  - **Fasta/BAM**

- **Unicycler MasuRCA**
  - **BLAST**
  - **Summary**
  - **Bandage**

- **Miniasm Canu Flye**
  - **QUAST**
  - **Summary**

- **Short-reads**

---

https://github.com/genomicsITER/NanoDJ
Genome assembly demo
Things to keep in mind

• The assembly depends on different factors such as the nature of the sequenced organism and the run performance.

• The type of assembler depends on the objective of your study and the genome you want to assemble.

• Assembly of larger genomes requires greater computational performance and longer execution time.

• We are testing new tools to incorporate in NanoDJ.
Thank you for attending

CyVerse is supported by the National Science Foundation under Grants No. DBI-0735191, DBI-1265383 and DBI-1743442.