de novo genome assembly and annotation with an emphasis on phylogenetic and population genetic studies

Suzy Strickler

Fay-Wei Li

Jacob Landis

Andrew Nelson

Adrian Powell

Jing Zhang

Welcome and introductions

○ Title: *de novo* genome assembly and annotation with an emphasis on phylogenetic and population genetic studies

○ Presenters: Jacob Landis, Fay-Wei Li, Suzy Strickler, Andrew Nelson

○ Length: Full-day workshop, allowing for consulting on individual projects at the end.

○ The costs of generating a draft genome for a study organism of choice are declining each year and more tools are becoming available to help in the endeavor. However, generating the appropriate sequence data is often not a trivial matter and currently, there are many options when it comes to analysis pipelines. The focus of this workshop is to work through the necessary steps for genome assembly and annotation when a closely related reference genome is not available. The workshop will cover assembling genomes using Illumina, Oxford Nanopore, and hybrid assemblers; as well as methods for extracting high-quality DNA suitable for ONT sequencing and library preparation, especially in species that possess secondary compounds which can be especially problematic for nanopore sequencing. Downstream applications/analyses that can be performed after assembly will also be covered while highlighting that a perfect assembly is not necessary to answer many evolutionary questions. Using the CyVerse platform, participants will get hands-on practice with all the necessary steps using either the supplied test data or their own data if previously generated.

Acknowledgements

Topics and Layout

Prerequisites and baseline skills:

- Unix command-line skills (we will review these at the start)
- Familiarly with some basic programming
- CyVerse account
- Atmosphere access
- CoGe account
- Cyberduck

Agenda

- Introductions and introduction to genome sequencing and assembly 9 am - Suzy
- Setting hardware/software baseline skills 9:15 (45 m) - Suzy
- Generate suitable DNA for different platforms: tips, tricks, and experimental design 10 am (30min) - Fay-Wei
- Break 10:30 (15m)
- Sequencing data and genome assembly options 10:45 am (1.75h) - Jacob (Fay-Wei assisting)
- Break -- lunch 12:30 am (1h)
- Genome annotation 1:30 (1.5h) - Suzy
- Break 3:00 pm (15min)
- Downstream evolution analyses: genome management, genome visualization, comparative genomics, reference-based SNP calling, transcriptomics, and phylogenomics 3:15 (2h) - Jacob and Andrew
- Break 5:15 pm (15min)
- Consulting Services 5:30 - 6:30

Where things are

Slides and scripts: https://github.com/bcbc-group/Botany2020NMGWorkshop

Data files: CyVerse DE /iplant/home/shared/Botany2020NMGWorkshop

Terminal: Atmosphere https://atmo.cyverse.org/application/dashboard

https://user.cyverse.org/services/mine

Intro to genome assembly

Why?

How?

Things to consider: genome size, ploidy, heterozygosity, cost, etc