

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

Gisel De La Cerda

Jacob Landis

Fay-Wei Li

Suzy Strickler

Welcome and introductions

- Title: *de novo* genome assembly and annotation with an emphasis on phylogenetic and population genetic studies
- Length: Full-day workshop (8:00 am - 5:00 pm)
- 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

- CyVerse especially Amanda Cooksey
- Boyce Thompson Institute
- Cornell University
- NSF

Topics and Layout

Prerequisites and baseline skills:

- Unix command-line skills
- Familiarly with some basic programming
- CyVerse account
- Cyberduck, Atom, GitHub Desktop (optional)

Syllabus

- Introductions and overview genome sequencing and assembly 8 am (30 min) - Suzy & Jacob
- Setting hardware/software baseline skills 8:30 (45 min) - Gisel
- Generate suitable DNA for different platforms: tips, tricks, and experimental design 9:15 am (30 min) - Fay-Wei
- Break 9:45 am (45 min)
- Sequencing data and genome assembly options 10:30 am (1.5 h) - Jacob
- Break for lunch 12:00 pm (1 h)
- Genome assembly wrap up 1:00 (30 min) - Jacob
- Genome annotation 1:30 (1.5 h) - Suzy
- Break 3:00 pm (30 min)
- Downstream evolution analyses: genome management, genome visualization, comparative genomics, reference-based SNP calling, transcriptomics, and phylogenomics 3:30 (1 h) - Jacob
- Wrap-up and discussion 4:30 - All

Where things are

Slides and scripts: <https://github.com/bcbc-group/Botany2022NMGWorkshop>

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

Terminal: CyVerse Vice

Let's get things set up!



The Open Science Workspace for Collaborative Data-driven Discovery

[Create Account](#)

As CyVerse evolves toward a diversified funding model, please help us understand how this change may impact you.

[Give Feedback](#)





Welcome to the CyVerse User Portal

The User Portal is the place to customize your CyVerse experience. Request access to services and explore learning materials and workshops.

LEARNING CENTER

GETTING STARTED WEBINAR

CLOSE AND DON'T SHOW AGAIN

My Services



CoGe

Online system for quick and easy retrieval and comparison of genomic information and sequences.

LAUNCH [↗](#)



Data Commons

A unified system for managing and sharing your data across CyVerse's tools and services

LAUNCH [↗](#)



Discovery Environment

Use hundreds of bioinformatics apps and manage data in the CyVerse Data Store from a simple web interface

LAUNCH [↗](#)



Atmosphere

Cloud computing platform for CyVerse

LAUNCH [↗](#)



DE - VICE

Visual Interactive Computing Environment in the Discovery Environment



Featured Apps

Filter

Manage Tools

Instant Launches

Create

<input type="checkbox"/>	Name ↑	Integrated B	Rating	System ID	
<input type="checkbox"/>	<input checked="" type="checkbox"/> Cloud Shell	Tyson Swetr	★★★★★(3)	de	⋮
<input type="checkbox"/>	<input checked="" type="checkbox"/> DataHog	Upendra K	★★★★★(2)	de	⋮
<input type="checkbox"/>	<input checked="" type="checkbox"/> JupyterLab Datascience Latest	Ty	★★★★★(1)	de	⋮
<input type="checkbox"/>	<input checked="" type="checkbox"/> Jupyter Lab Geospatial Latest	tyson Swetnam	★★★★★(1)	de	⋮
<input type="checkbox"/>	<input checked="" type="checkbox"/> Rocker RStudio Geospatial Latest	Tyson Swetnam	★★★★★(1)	de	⋮
<input type="checkbox"/>	<input checked="" type="checkbox"/> Rocker RStudio Latest	Tyson Swetnam	★★★★★(1)	de	⋮
<input type="checkbox"/>	<input checked="" type="checkbox"/> Rocker RStudio Verse Latest	Tyson Swetnam	★★★★★(1)	de	⋮
<input type="checkbox"/>	<input checked="" type="checkbox"/> VS Code Latest	Tyson Swetnam	★★★★★(1)	de	⋮
<input type="checkbox"/>	<input checked="" type="checkbox"/> Xpra Desktop Ubuntu Geospatial 20.04	Tyson Swetnam	★★★★★(1)	de	⋮

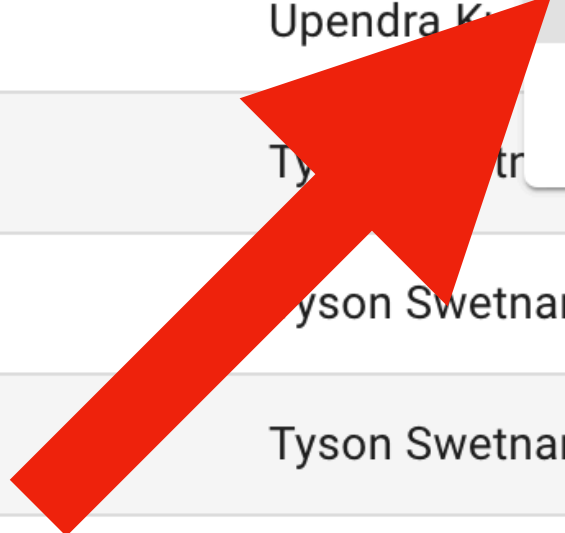
<

1

>

100 ▼

- HPC
- DE
- VICE
- OSG



2022

Apps

?

46

S

Filter

VICE

Manage Tools

Instant Launches

+ Create

	Integrated By	Rating	System ID	
	Tyson Swetnam	★★★★★ (3)	de	⋮
	Upendra Kumar Devisetty	★★★★★ (2)	de	⋮
	Tyson Swetnam	★★★★★ (1)	de	⋮
	Tyson Swetnam	★★★★★ (1)	de	⋮
	Tyson Swetnam	★★★★★ (1)	de	⋮
	Tyson Swetnam	★★★★★ (1)	de	⋮
<input type="checkbox"/>	Rocker RStudio Verse Latest	★★★★★ (1)	de	⋮
<input type="checkbox"/>	VS Code Latest	★★★★★ (1)	de	⋮
<input type="checkbox"/>	Xpra Desktop Ubuntu Geospatial 20.04	★★★★★ (1)	de	⋮

<

1

>

100

Apps Under Development

Favorite Apps

My Public Apps

Shared With Me

Featured Apps

High-Performance Computing

My Collections

Browse All Apps



Browse All Apps

Filter

VICE

Manage Tools

Instant Launches

Create

<div><input type="checkbox"/></div>	<div>Name ↑</div>	<div>Integrated By</div>	<div>Rating</div>	<div>System ID</div>	
<input type="checkbox"/>	<div>3.1 dev GEA Leptin RNA Seq Lab</div>	williams	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>🔒</div><div>Apollo</div></div>	REETU TUTEJA	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>Asc-Seurat</div></div>	REETU TUTEJA	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div>bash</div>	Tyson Swetnam	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>Botany workshop 2022</div></div>	Amanda Cooksey	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>✓</div><div>Cloud Shell</div></div>	Tyson Swetnam	<div>★ ★ ★ ★ ★ (3)</div>	de	<div></div>
<input type="checkbox"/>	<div>Cloud Shell</div>	Tyson Swetnam	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>DataCarpentry_Bieser</div></div>	Kayla Bieser	<div>★ ★ ★ ★ ★ (6)</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>FIELDimageR</div></div>	Chris Schnaufer	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>Firefly 1.0</div></div>	Tom Hicks	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>GEA Leptin RNA Seq Lab dev 3_0</div></div>	williams	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>GEA RNA-Seq QC Demo 0.2</div></div>	williams	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>genepattern-notebook</div></div>	Sarah Roberts	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>Genomics Education Alliance RNA-Seq Lesson - Leptin</div></div>	williams	<div>★ ★ ★ ★ ★</div>	de	<div></div>
<input type="checkbox"/>	<div><div>β</div><div>IDEAS</div></div>	Janet Yuan	<div>★ ★ ★ ★ ★ (1)</div>	de	<div></div>





← Back

Documentation

Details

Botany workshop 2022

App for the Botany workshop 2022

1

Analysis Info

2

Parameters

3

Advanced Settings (optional)

4

Review and Launch

Step 1: Analysis Info

← Back

Next →

Analysis Name *

Botany_workshop_2022_analysis1

Enter analysis name

Comments

Output Folder *

/iplant/home/srs57/analyses

Browse

← Back

Next →





← Back

Documentation

Details

Botany workshop 2022

App for the Botany workshop 2022



Analysis Info



Parameters



Advanced Settings (optional)



Review and Launch

Step 2: Analysis Parameters

← Back

Next →

Input

Section 1 of 1

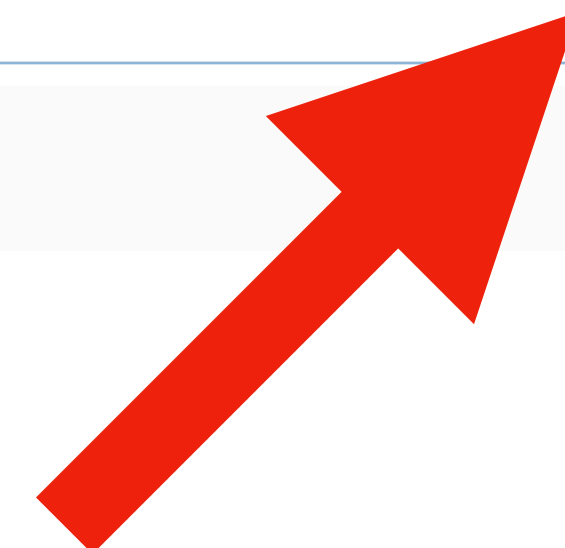


Input Folder

Browse

← Back

Next →



Resource Requirements



Resource Requirements may be adjusted so that the analysis can be requested to run with resources that match your need. Note that this may cause the analysis to wait longer in the submission queue until a node that matches those resource requirements becomes available. So generally it is advisable to leave the resource requests set to the defaults that the tool or app integrator has set.

Select Minimums

Minimum CPU Cores

4

Minimum Memory

16.0 GiB

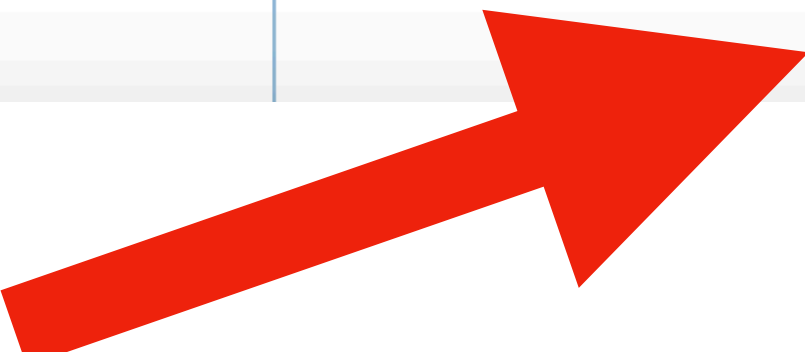
Minimum Disk Space

512.0 GiB

Select Maximums

Maximum CPU Cores

8



← Back

Documentation

Details

Botany workshop 2022

App for the Botany workshop 2022



Analysis Info



Parameters



Advanced Settings
(optional)



Review and Launch

Step 4: Launch or Save

← Back

Next →

☒ Show All Parameters

Turn off to hide parameters without values entered or selected.

Input Folder

Resource Requirements




← Back

Create Saved Launch

▶ Launch Analysis





Botany_workshop_2022_analysis1


Running for 1 minute (2022-07-24 02:31:47.8)

Go to Analysis

Refresh

Request Help

More Actions

App	Botany workshop 2022
Output folder	/iplant/home/srs57/analyses/Botany_workshop_2022_analysis1-2022-07-24-02-31-47.8 
Start date	2022-07-23 18:31:47
End date	-
User	srs57

Status History

c5734cf6-0af8-11ed-bef8-008cfa5ae621

Copy Analysis ID

1: Interactive - Running

Parameters

Name ↓	Type	Value
Input Folder	FolderInput	

* This page will refresh automatically. * You can bookmark this page to come back to it later.

 [Contact Support](#)



Launching VICE app: cloud-shell

 100%

Last step: waiting for cloud-shell to respond

```
user@a2442f596:~/work$ ls
data  home
user@a2442f596:~/work$ cd ..
user@a2442f596:~$ ls
Augustus  cli          fastp          gmes_linux_64_4.tar.gz  perl5          v2.1.6.tar.gz          WORK
bamtools  diamond      fmlrc2         MaSuRCA-4.0.9           stacks-2.62    vcftools-vcftools-581c231  wtdbg2
bwa       diamond-linux64.tar.gz  gmes_linux_64_4  MaSuRCA-4.0.9.tar.gz    stacks-2.62.tar.gz  vcftools-vcftools-v0.1.16-18-g581c231.tar.gz
user@a2442f596:~$ █
```