Summary - what we did

Galaxy tutorials

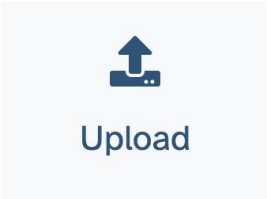
- Learned how to use Galaxy
- Used some common genomics tools for:
  - Data QC
  - Assembly
  - Assembly QC
  - Finding repeats
  - Annotation

Key Galaxy icons in the Activity bar

Activity bar

Getting data into Galaxy

- Upload data
  - e.g. Choose local file (from your computer)
  - e.g. Paste/Fetch data (e.g. a URL)



- For sequencing reads from NCBI Sequence Read Archive:
  - Galaxy Tool: `Faster Download and Extract Reads in FASTQ format from NCBI SRA`
  - You can find sequence reads you want on NCBI, copy the SRR number, and use it as input into the tool

Data QC

Extra links:

- For more about base qualities:
  - https://training.galaxyproject.org/training-material/topics/introduction/tutorials/galaxy-intro-ngs-data-managment/tutorial.html#what-are-base-qualities
- For more about the plots produced by FastQC and what they mean:
  - https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Assembly and annotation

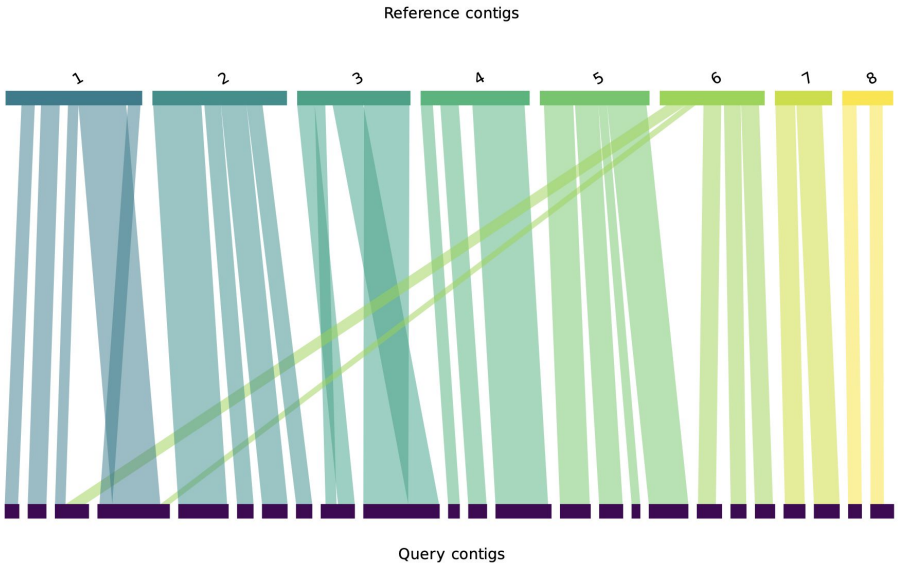
Extra links:

- More about assembling metagenomic data:
  - https://training.galaxyproject.org/training-material/topics/microbiome/tutorials/metagenomics-assembly/tutorial.html
- More about assembly decontamination:
  - https://training.galaxyproject.org/training-material/topics/assembly/tutorials/assembly-decontamination/tutorial.html

Synteny tools

- Align assembled contigs against a reference genome
  - Tool: `RagTag`, then plot with `Plot RagTag output`

Example



Metagenomics, proteomics, phylogenetics

Galaxy curated tutorial sets for:

- Metagenomics data processing and analysis for microbiome: https://training.galaxyproject.org/training-material/learning-pathways/metagenomics.html
- Clinical metaproteomics workflows within Galaxy: https://training.galaxyproject.org/training-material/learning-pathways/clinical-metaproteomics.html
- Proteogenomics: https://training.galaxyproject.org/training-material/learning-pathways/proteogenomics.html

Biosynthetic gene clusters tutorial:

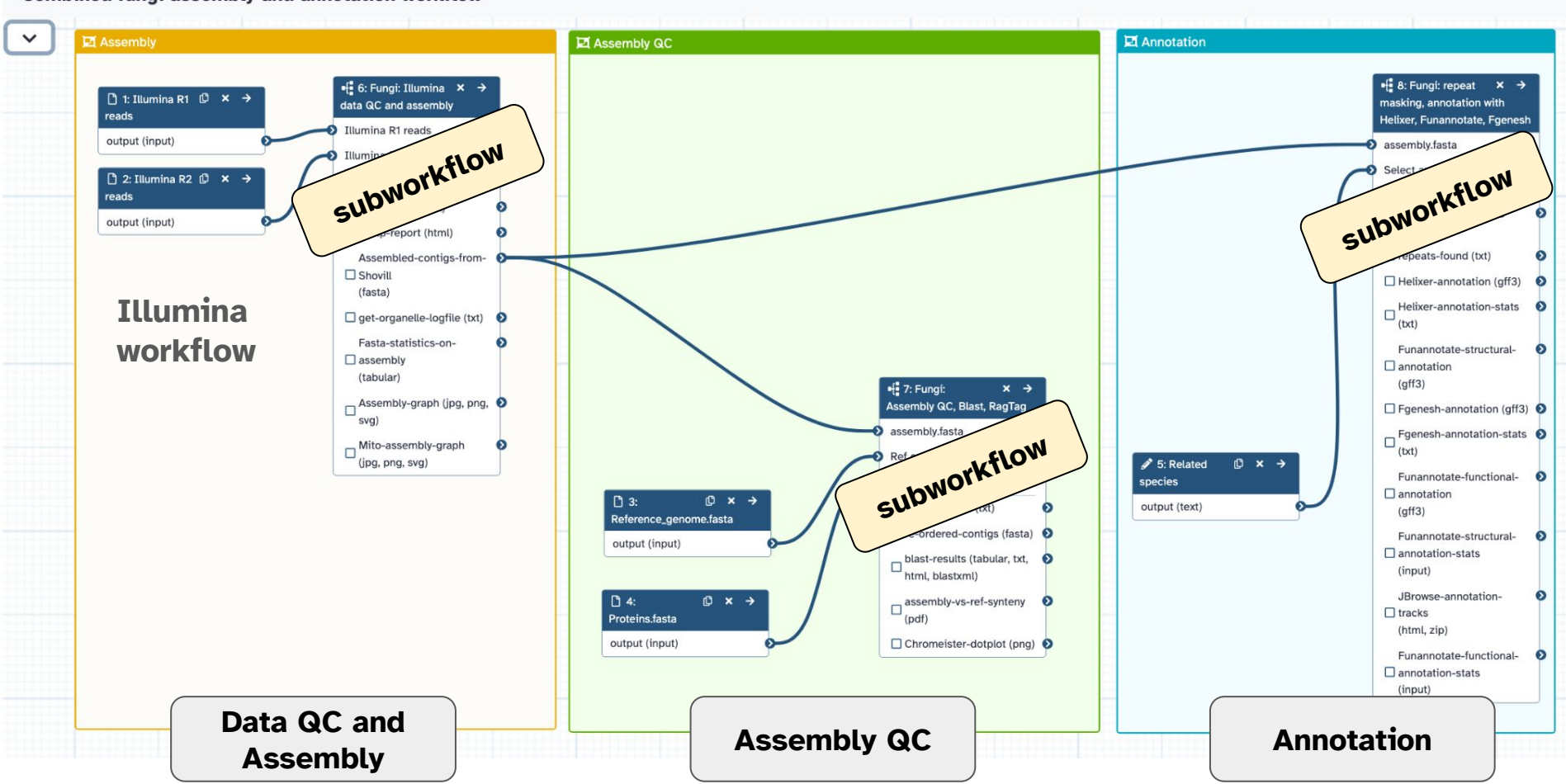
- https://training.galaxyproject.org/training-material/topics/ecology/tutorials/marine_omics_bgc/tutorial.html

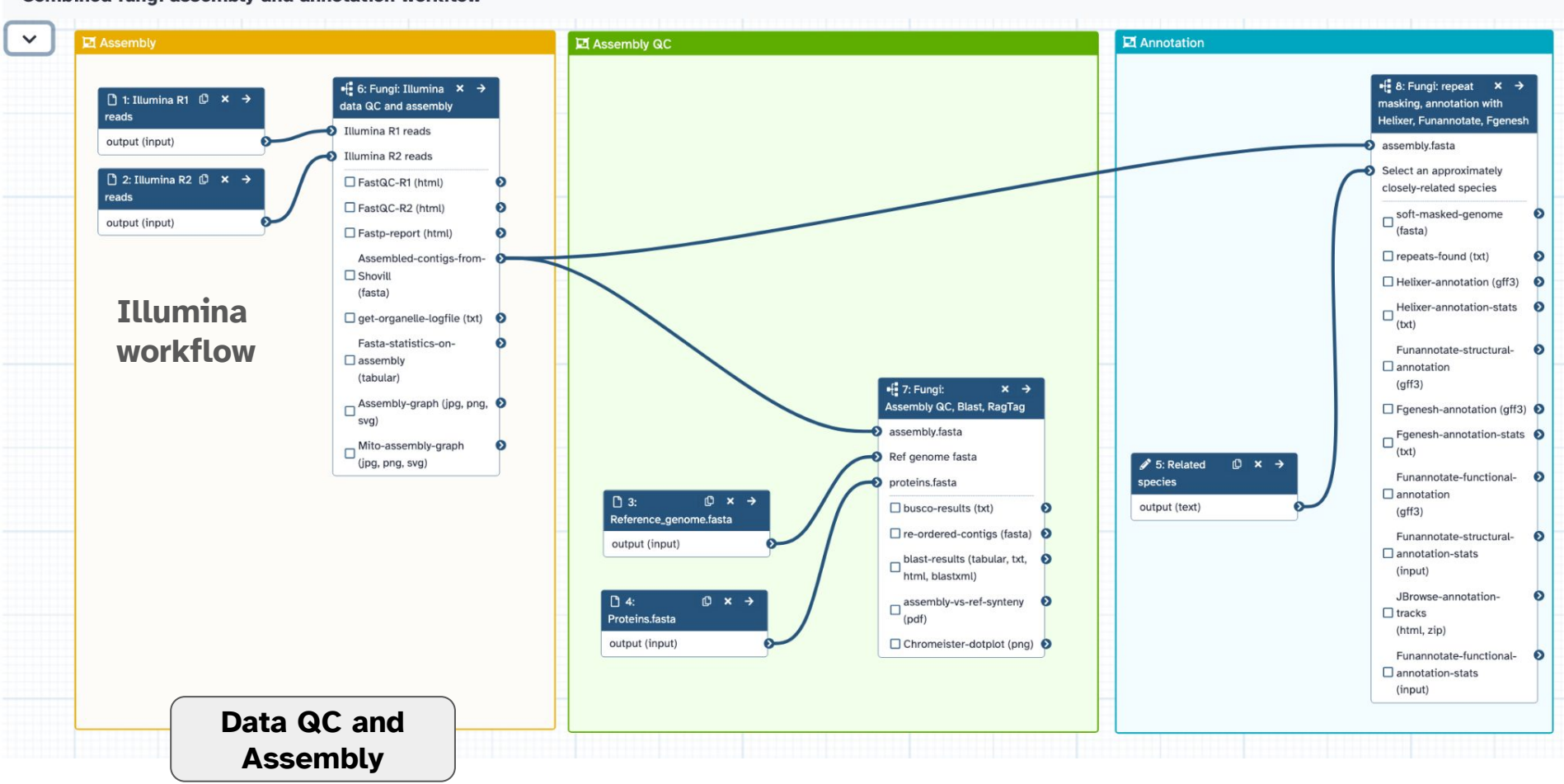
Phylogenetics tutorial:

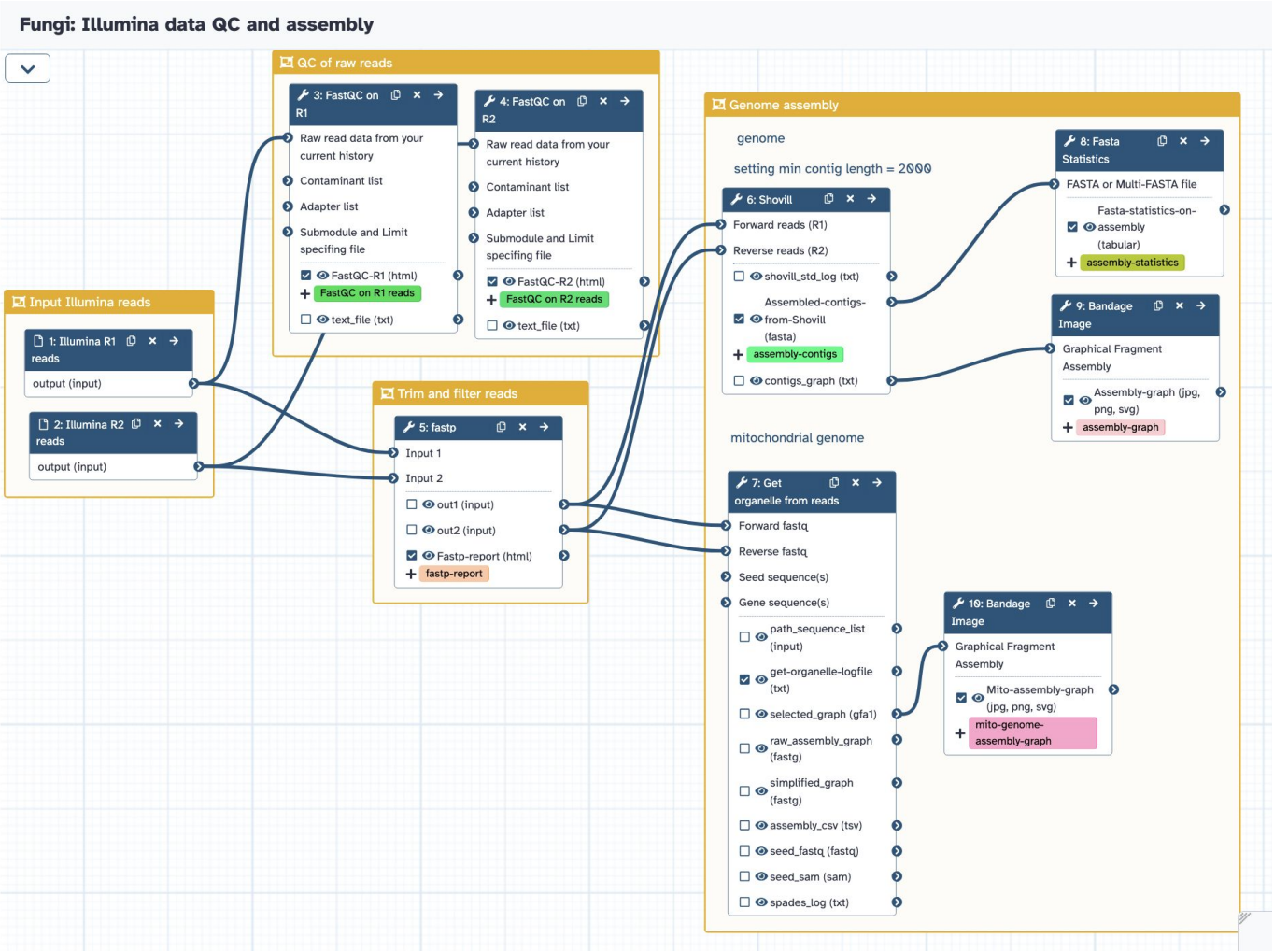
- https://training.galaxyproject.org/training-material/topics/evolution/tutorials/abc_intro_phylo/tutorial.html

What tools and workflows to use?

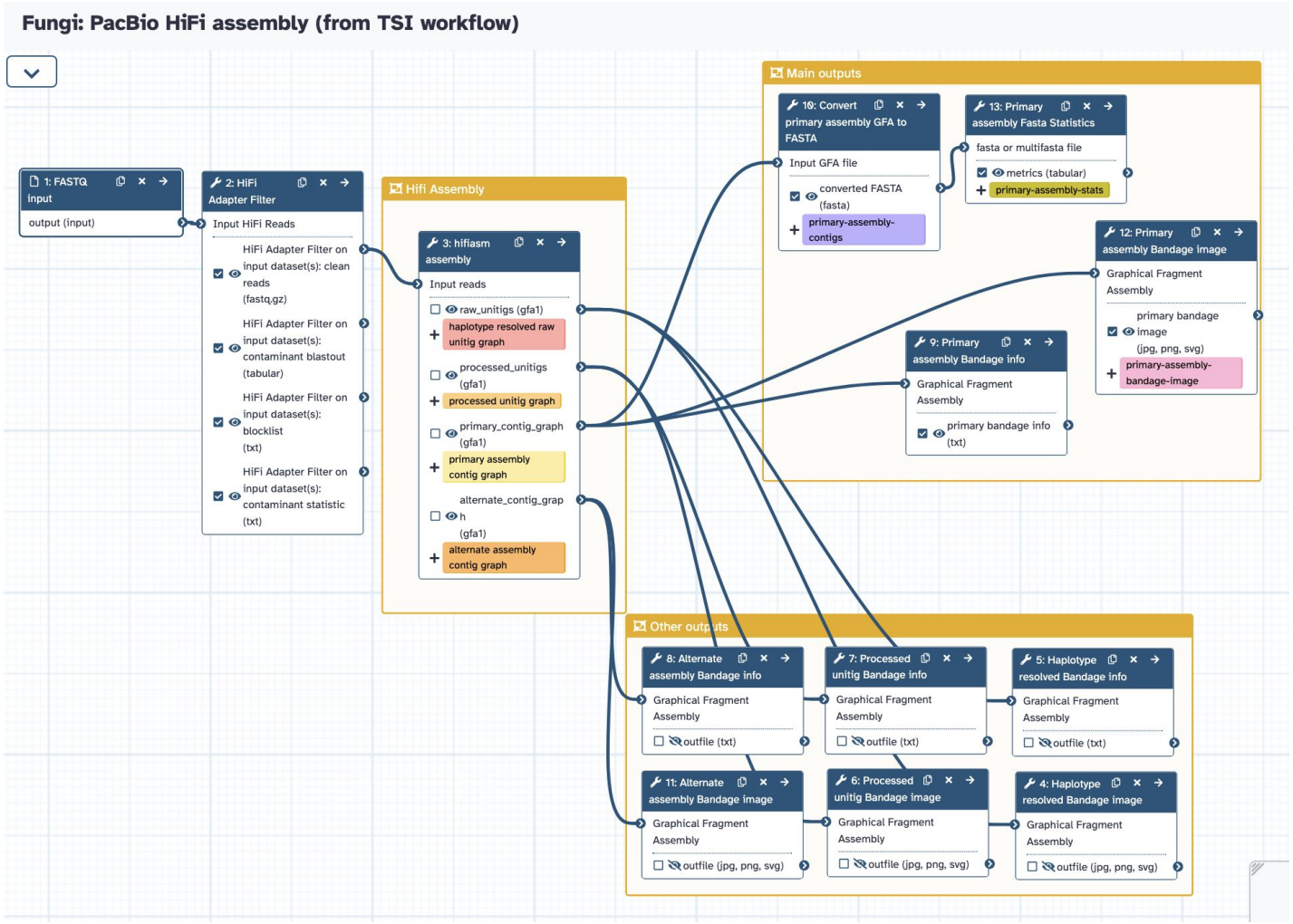
- Will depend on your data and research questions
- Taking some of the main tools we used or talked about this week:
  - We made some workflows to demonstrate how these tools could be used
  - Don't worry about the detail in these slides following, the idea is to show the range of tools available
  - You can import a workflow to your account, then zoom in, and change as needed (links provided)


Combined fungi assembly and annotation workflow
Assembly
Assembly QC
Annotation
1: Illumina R1 reads
output (input)
2: Illumina R2 reads
output (input)
6: Fungi: Illumina data QC and assembly
Illumina R1 reads
Illumina R2 reads
subworkflow
Assembled-contigs-from-Shovill (fasta)
get-organelle-logfile (txt)
Fasta-statistics-on-assembly (tabular)
Assembly-graph (jpg, png, svg)
Mito-assembly-graph (jpg, png, svg)
Illumina workflow
3: Reference_genome.fasta
output (input)
4: Proteins.fasta
output (input)
7: Fungi: Assembly QC, Blast, RagTag
assembly.fasta
subworkflow
blast-results (tabular, txt, html, blastxml)
assembly-vs-ref-synteny (pdf)
Chromeister-dotplot (png)
5: Related species
output (text)
8: Fungi: repeat masking, annotation with Helixer, Funannotate, Fgenesh
assembly.fasta
subworkflow
Helixer-annotation (gff3)
Helixer-annotation-stats (txt)
Funannotate-structural-annotation (gff3)
Fgenesh-annotation (gff3)
Fgenesh-annotation-stats (txt)
Funannotate-functional-annotation (gff3)
Funannotate-structural-annotation-stats (input)
JBrowse-annotation-tracks (html, zip)
Funannotate-functional-annotation-stats (input)
Data QC and Assembly
Assembly QC
Annotation


Combined fungi assembly and annotation workflow
Assembly
Assembly QC
Annotation
1: Illumina R1 reads
output (input)
2: Illumina R2 reads
output (input)
6: Fungi: Illumina data QC and assembly
Illumina R1 reads
Illumina R2 reads
FastQC-R1 (html)
FastQC-R2 (html)
Fastp-report (html)
Assembled-contigs-from-Shovill (fasta)
get-organelle-logfile (txt)
Fasta-statistics-on-assembly (tabular)
Assembly-graph (jpg, png, svg)
Mito-assembly-graph (jpg, png, svg)
Illumina workflow
3: Reference_genome.fasta
output (input)
4: Proteins.fasta
output (input)
7: Fungi: Assembly QC, Blast, RagTag
assembly.fasta
Ref genome fasta
proteins.fasta
busco-results (txt)
re-ordered-contigs (fasta)
blast-results (tabular, txt, html, blastxml)
assembly-vs-ref-synteny (pdf)
Chromeister-dotplot (png)
5: Related species
output (text)
8: Fungi: repeat masking, annotation with Helixer, Funannotate, Fgenesh
assembly.fasta
Select an approximately closely-related species
soft-masked-genome (fasta)
repeats-found (txt)
Helixer-annotation (gff3)
Helixer-annotation-stats (txt)
Funannotate-structural-annotation (gff3)
Fgenesh-annotation (gff3)
Fgenesh-annotation-stats (txt)
Funannotate-functional-annotation (gff3)
Funannotate-structural-annotation-stats (input)
JBrowse-annotation-tracks (html, zip)
Funannotate-functional-annotation-stats (input)
Data QC and Assembly

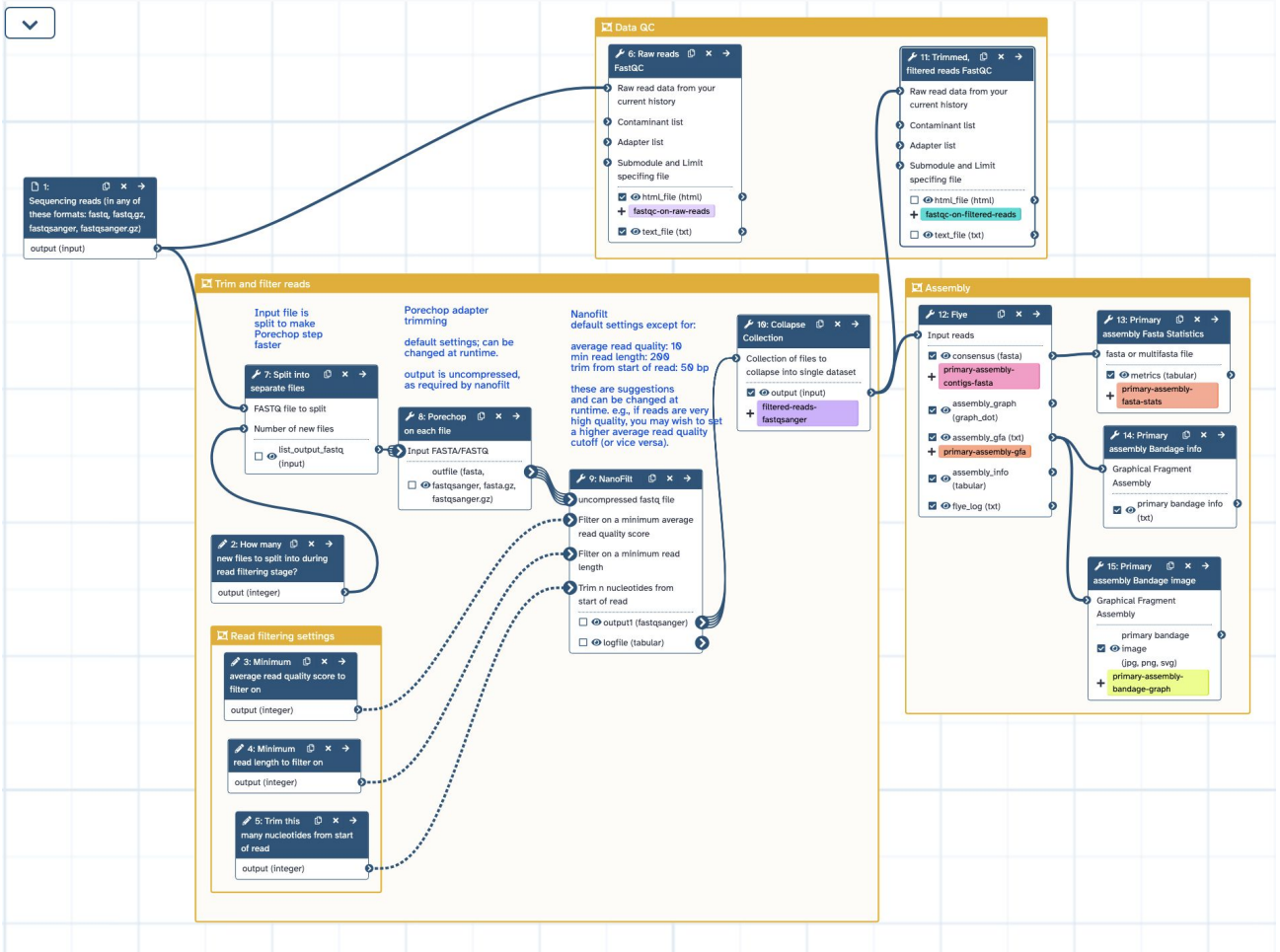


Tools:
FastQC
Fastp
Shovill
get-organelle
Fasta Statistics
Bandage

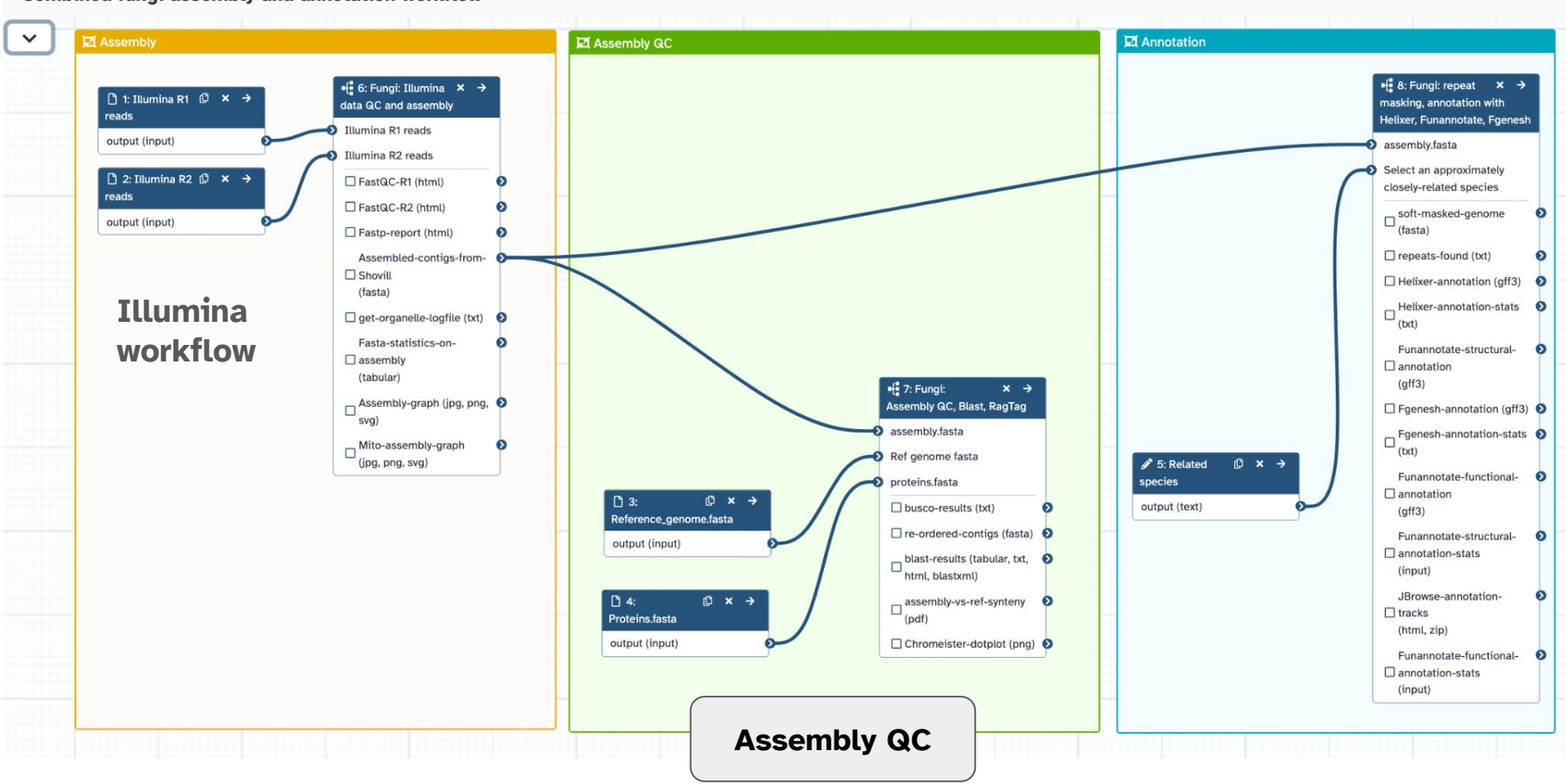


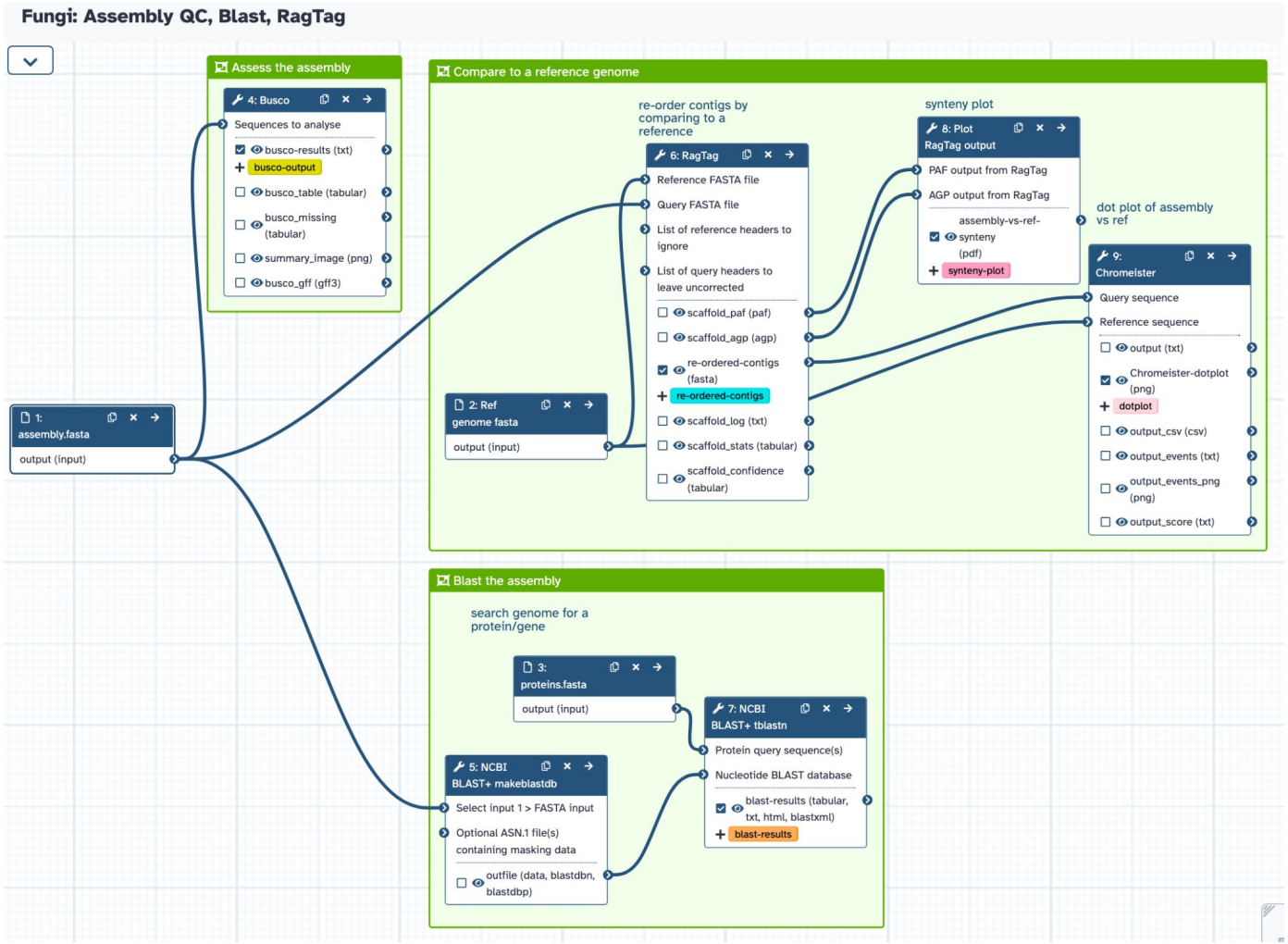
Tools:
Hifi Adapter Filter
Hifiasm
Bandage

Fungi: Nanopore assembly (from TSI workflow)



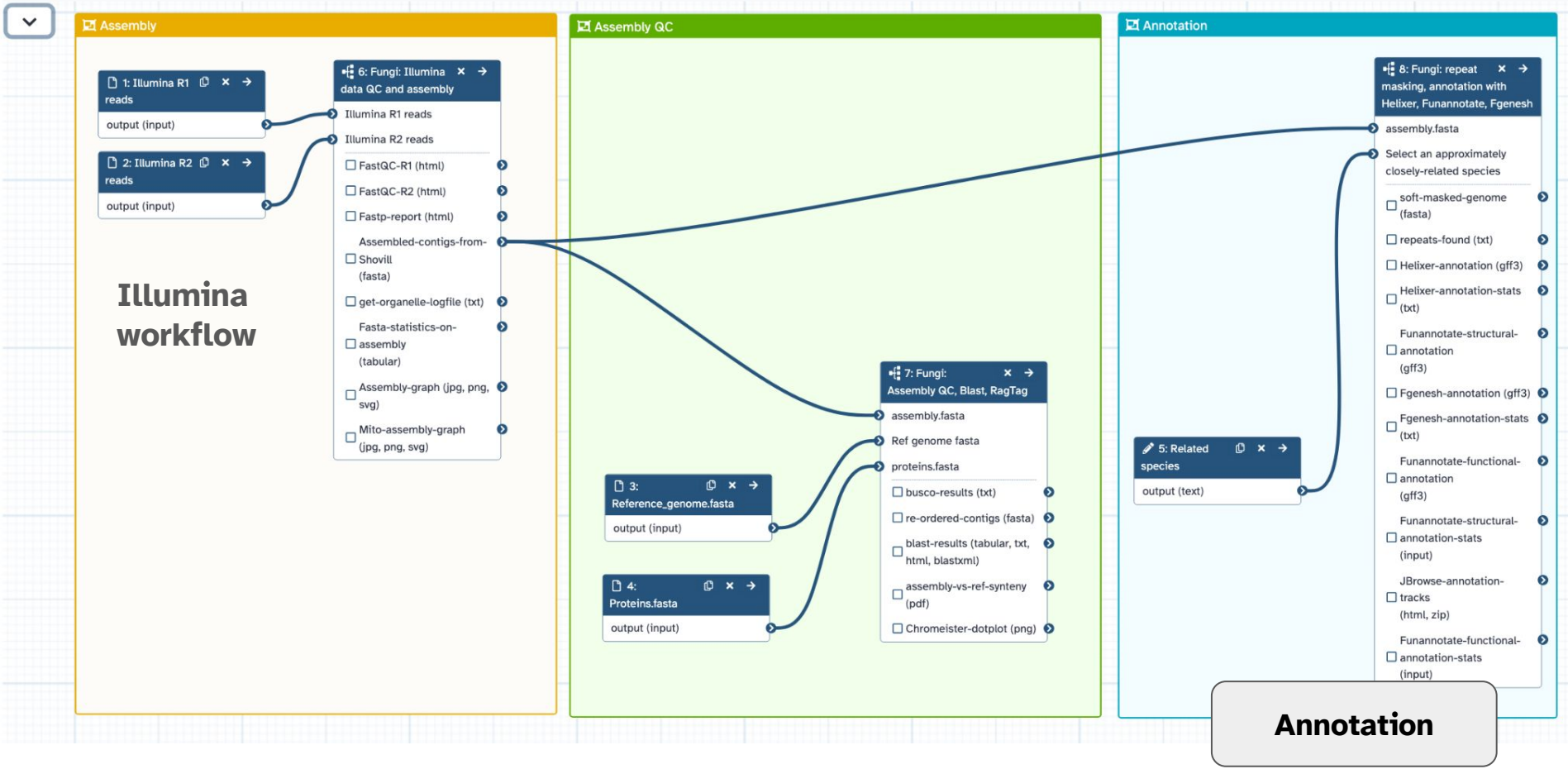
Tools:
Porechop
Nanofilt
FastQC
Flye
Fasta Statistics
Bandage


Combined fungi assembly and annotation workflow
Assembly
Assembly QC
Annotation
1: Illumina R1 reads
output (input)
2: Illumina R2 reads
output (input)
6: Fungi: Illumina data QC and assembly
Illumina R1 reads
Illumina R2 reads
FastQC-R1 (html)
FastQC-R2 (html)
Fastp-report (html)
Assembled-contigs-from-Shovill (fasta)
get-organelle-logfile (txt)
Fasta-statistics-on-assembly (tabular)
Assembly-graph (jpg, png, svg)
Mito-assembly-graph (jpg, png, svg)
Illumina workflow
3: Reference_genome.fasta
output (input)
4: Proteins.fasta
output (input)
7: Fungi: Assembly QC, Blast, RagTag
assembly.fasta
Ref genome fasta
proteins.fasta
busco-results (txt)
re-ordered-contigs (fasta)
blast-results (tabular, txt, html, blastxml)
assembly-vs-ref-synteny (pdf)
Chromeister-dotplot (png)
5: Related species
output (text)
8: Fungi: repeat masking, annotation with Helixer, Funannotate, Fgenesh
assembly.fasta
Select an approximately closely-related species
soft-masked-genome (fasta)
repeats-found (txt)
Helixer-annotation (gff3)
Helixer-annotation-stats (txt)
Funannotate-structural-annotation (gff3)
Fgenesh-annotation (gff3)
Fgenesh-annotation-stats (txt)
Funannotate-functional-annotation (gff3)
Funannotate-structural-annotation-stats (input)
JBrowse-annotation-tracks (html, zip)
Funannotate-functional-annotation-stats (input)
Assembly QC

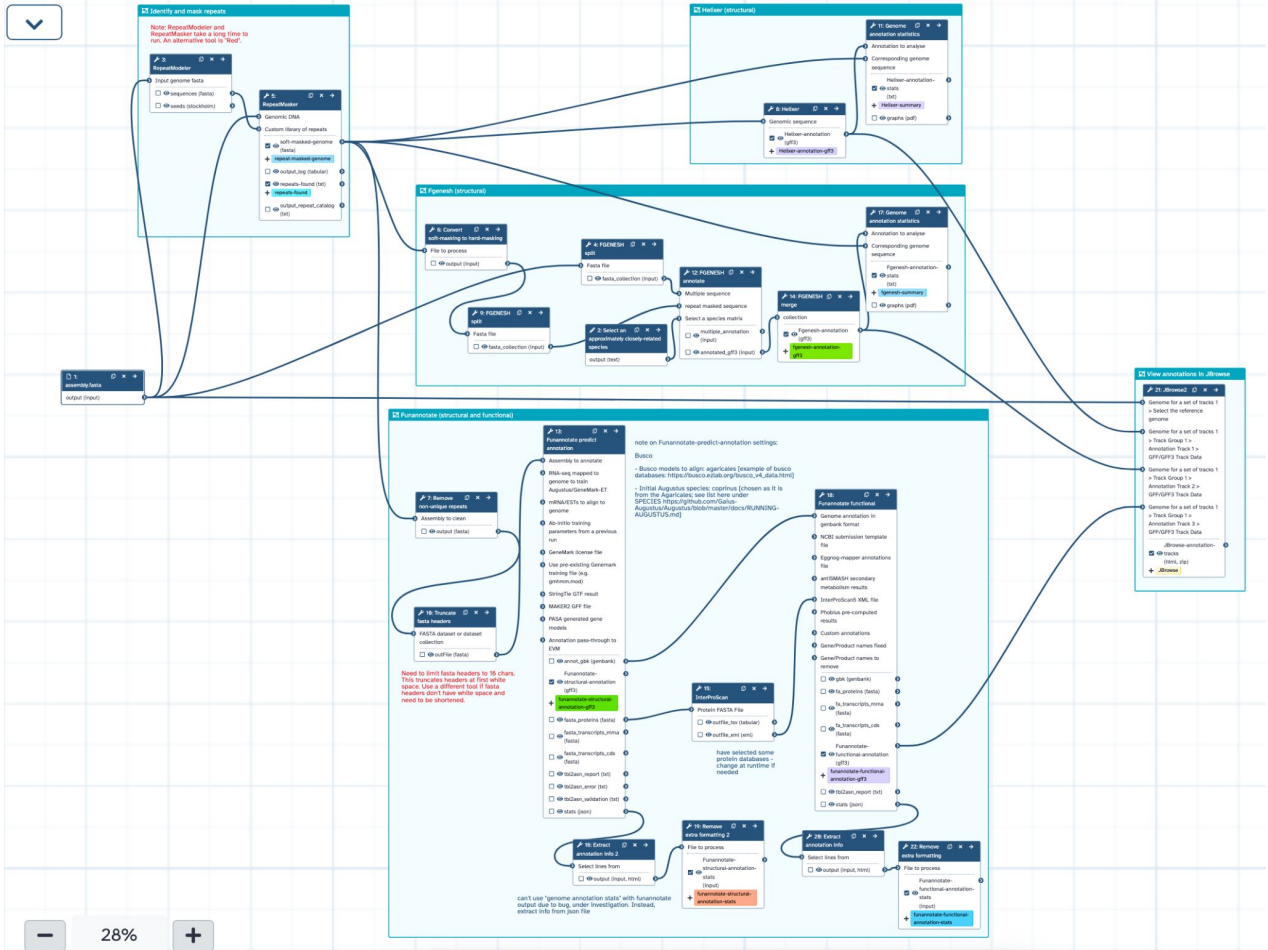


Tools:
Busco
RagTag
Chromeister
Blast

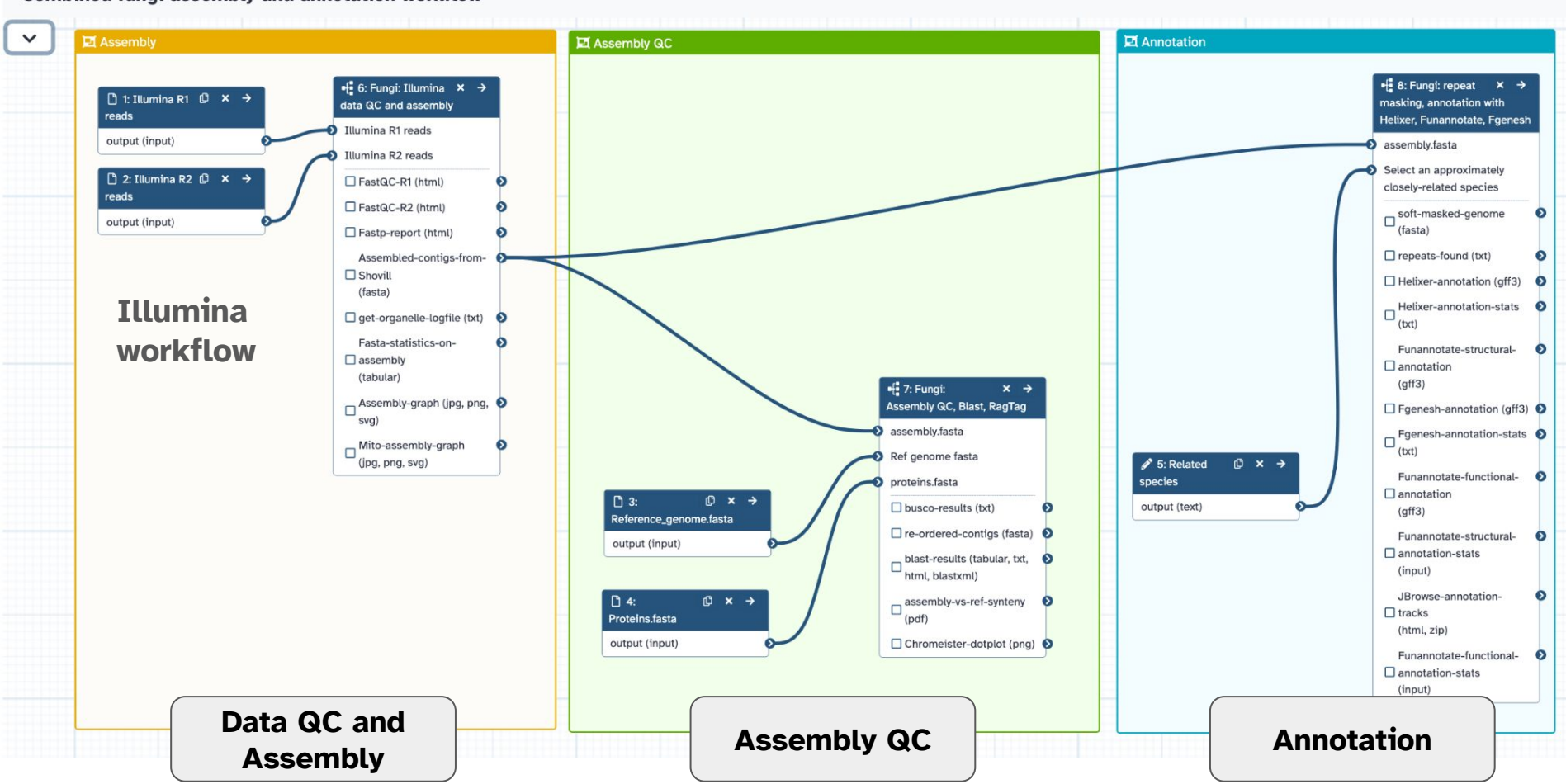
Combined fungi assembly and annotation workflow



Fungi: repeat masking, annotation with Helixer, Funannotate, Fgenesh



Tools:
RepeatModeler
RepeatMasker
Helixer
Fgenesh
Funannotate
InterProScan
JBrowse


Combined fungi assembly and annotation workflow
Assembly
Assembly QC
Annotation
1: Illumina R1 reads
output (input)
2: Illumina R2 reads
output (input)
6: Fungi: Illumina data QC and assembly
Illumina R1 reads
Illumina R2 reads
FastQC-R1 (html)
FastQC-R2 (html)
Fastp-report (html)
Assembled-contigs-from-Shovill (fasta)
get-organelle-logfile (txt)
Fasta-statistics-on-assembly (tabular)
Assembly-graph (jpg, png, svg)
Mito-assembly-graph (jpg, png, svg)
Illumina workflow
3: Reference_genome.fasta
output (input)
4: Proteins.fasta
output (input)
7: Fungi: Assembly QC, Blast, RagTag
assembly.fasta
Ref genome fasta
proteins.fasta
busco-results (txt)
re-ordered-contigs (fasta)
blast-results (tabular, txt, html, blastxml)
assembly-vs-ref-synteny (pdf)
Chromeister-dotplot (png)
5: Related species
output (text)
8: Fungi: repeat masking, annotation with Helixer, Funannotate, Fgenesh
assembly.fasta
Select an approximately closely-related species
soft-masked-genome (fasta)
repeats-found (txt)
Helixer-annotation (gff3)
Helixer-annotation-stats (txt)
Funannotate-structural-annotation (gff3)
Fgenesh-annotation (gff3)
Fgenesh-annotation-stats (txt)
Funannotate-functional-annotation (gff3)
Funannotate-structural-annotation-stats (input)
JBrowse-annotation-tracks (html, zip)
Funannotate-functional-annotation-stats (input)
Data QC and Assembly
Assembly QC
Annotation

Workflow tests

- We ran these workflows on some fungi data from BPA and NCBI (details next slide)
  - You can import/view the Galaxy histories to see how the tools worked and what the results look like

- (Note: these workflows and their results are just examples, so don't use the results as input data for real analyses)

Links to these workflows and histories

Workflow name	Workflow link	Data	Galaxy history
Combined workflow: Illumina assembly + QC + annotation	https://usegalaxy.org.au/u/anna/w/combined-fungi-assembly-and-annotation-workflow	*Psilocybe subaeruginosa* BPA sample 465877 Ref genome *P. cubensis* Protein: psilocybin synthase	https://usegalaxy.org.au/u/anna/h/combined-fungi-workflows-illumina-psilocybe
Fungi: Illumina data QC and assembly	https://usegalaxy.org.au/u/anna/w/fungi-illumina-assembly	*Psilocybe subaeruginosa* BPA sample 465877	https://usegalaxy.org.au/u/anna/h/psilocybe-illumina-assembly
Fungi: PacBio assembly	https://usegalaxy.org.au/u/anna/w/fungi-hifi-assembly	*Rhynchosporium commune* BPA sample 395386	https://usegalaxy.org.au/u/anna/h/rhynchosporium-pacbio-assembly
Fungi: Nanopore assembly	https://usegalaxy.org.au/u/anna/w/fungi-nanopore-assembly	*Aspergillus fumigatus* NCBI SRR23337894 SRR23337893 SRR23337895	https://usegalaxy.org.au/u/anna/h/aspergillus-nanopore-assembly
Fungi: Assembly QC, Blast, RagTag	https://usegalaxy.org.au/u/anna/w/fungi-assembly-qc	*Psilocybe subaeruginosa* assembly from Illumina workflow Ref genome *P. cubensis* (as in fewer contigs than ref genome for *P. subaeruginosa*). Protein: psilocybin synthase	https://usegalaxy.org.au/u/anna/h/psilocybe-assembly-qc
Fungi: repeat masking, annotation with Helixer, Funannotate, Fgenesh	https://usegalaxy.org.au/u/anna/w/fungi-annotation	*Psilocybe subaeruginosa* assembly from Illumina workflow	https://usegalaxy.org.au/u/anna/h/repeat-masking-and-annotation---psilocybe

Galaxy Training! Contributors Learning Pathways Help Settings Search Tutorials

Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

Search for more tutuorials in the Galaxy Training Network

Galaxy for Scientists

We have separated the tutorials into several categories based on field and technology. We are exploring other ways to organise the tutorials going forward!

Start Here

Topic	Tutorials
Introduction to Galaxy Analyses	13
Using Galaxy and Managing your Data	23

Not sure where to start?

Quickstart

Learning Pathways

Galaxy for SysAdmins

Galaxy for Developers

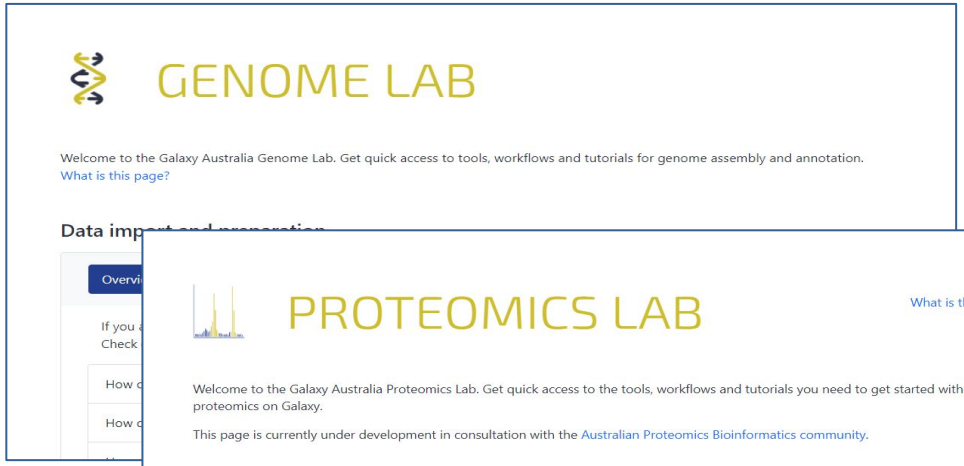
Galaxy for Teachers

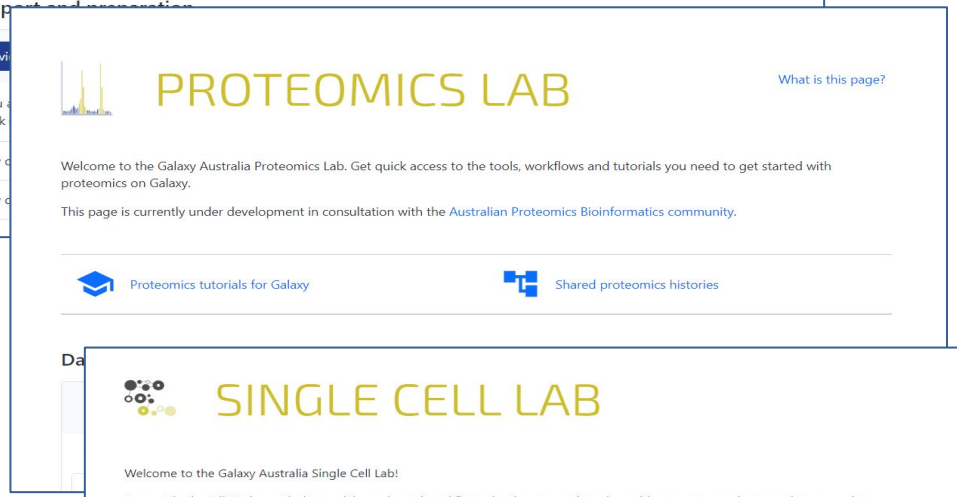
Reminder on help:

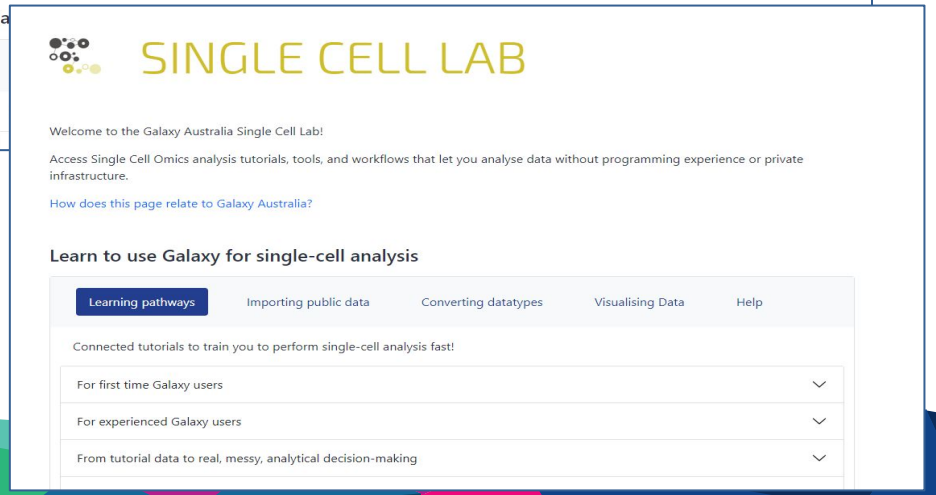
- In Galaxy: top menu bar: Help
- Report bugs (bug icon under file) 
- Email: help@genome.edu.au
- Matrix: https://gitter.im/galaxyproject/Lobby
- Global help: https://help.galaxyproject.org/
- Galaxy Training Network: https://training.galaxyproject.org/
- Toolshed: https://toolshed.g2.bx.psu.edu/
- How-To-Guides: https://australianbiocommons.github.io/how-to-hub/index
- Australian BioCommons youtube channel: https://www.youtube.com/channel/UC5WIFNBSfmt3e8Js8o2fFqQ

Galaxy or: How I Learned to Stop Worrying and Love the GUI

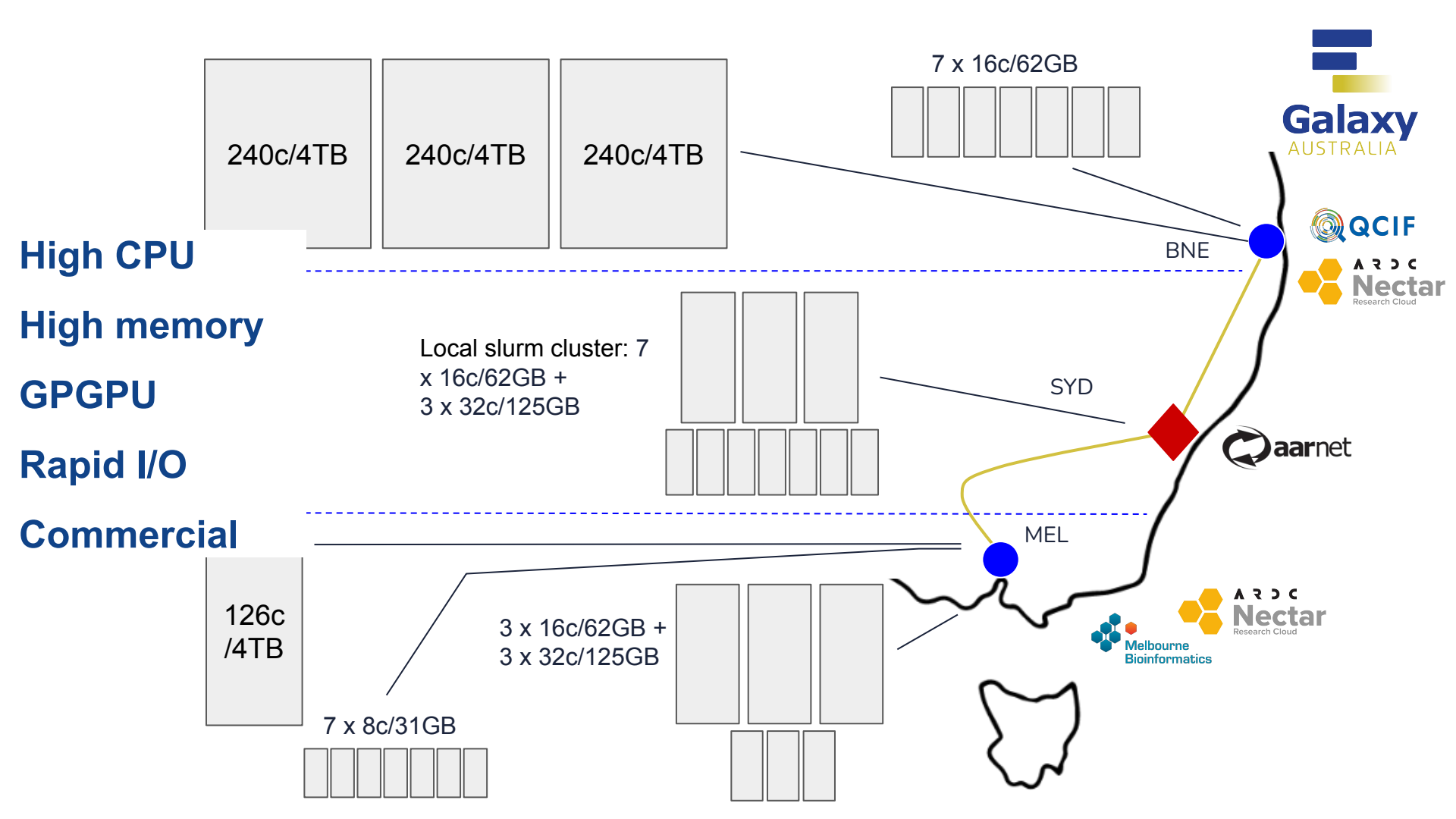
Galaxy Labs







Australian BioCommons


High CPU
High memory
GPGPU
Rapid I/O
Commercial
240c/4TB
240c/4TB
240c/4TB
7 x 16c/62GB
Galaxy
AUSTRALIA
QCIF
ARDC
Nectar
Research Cloud
BNE
Local slurm cluster: 7
x 16c/62GB +
3 x 32c/125GB
SYD
aarnet
MEL
126c
/4TB
7 x 8c/31GB
3 x 16c/62GB +
3 x 32c/125GB
Melbourne
Bioinformatics
ARDC
Nectar
Research Cloud

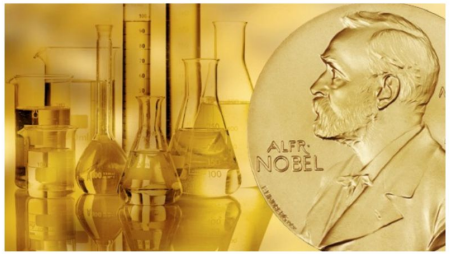
3D protein prediction in Galaxy Australia - Alphafold2

nature > collection

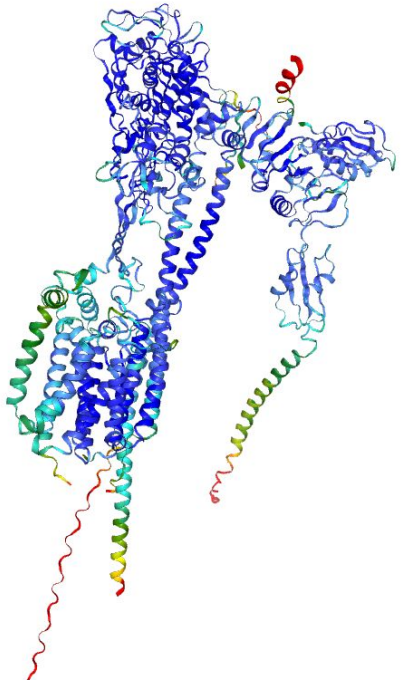
Collection | **09 October 2024**

Nobel Prize in Chemistry 2024

The 2024 Nobel Prize in Chemistry has been awarded to David Baker “for computational protein design” and to Demis Hassabis and John M. Jumper “for protein structure prediction”. Proteins are life’s essential building blocks, nature’s most ingenious molecular machines and the basis of all living organisms. Hassabis and Jumper have developed a series of artificial intelligence models to address the decades-long structural biology problem of how to predict the complex 3D structures of proteins solely from their linear amino acid sequences, while Baker has dedicated his scientific career to designing and constructing proteins that are not, and even can not, be found in nature. In recognition of this award, Nature Portfolio presents a collection of research, review and opinion articles that celebrates both contributions by the awardees and the advances they have inspired.



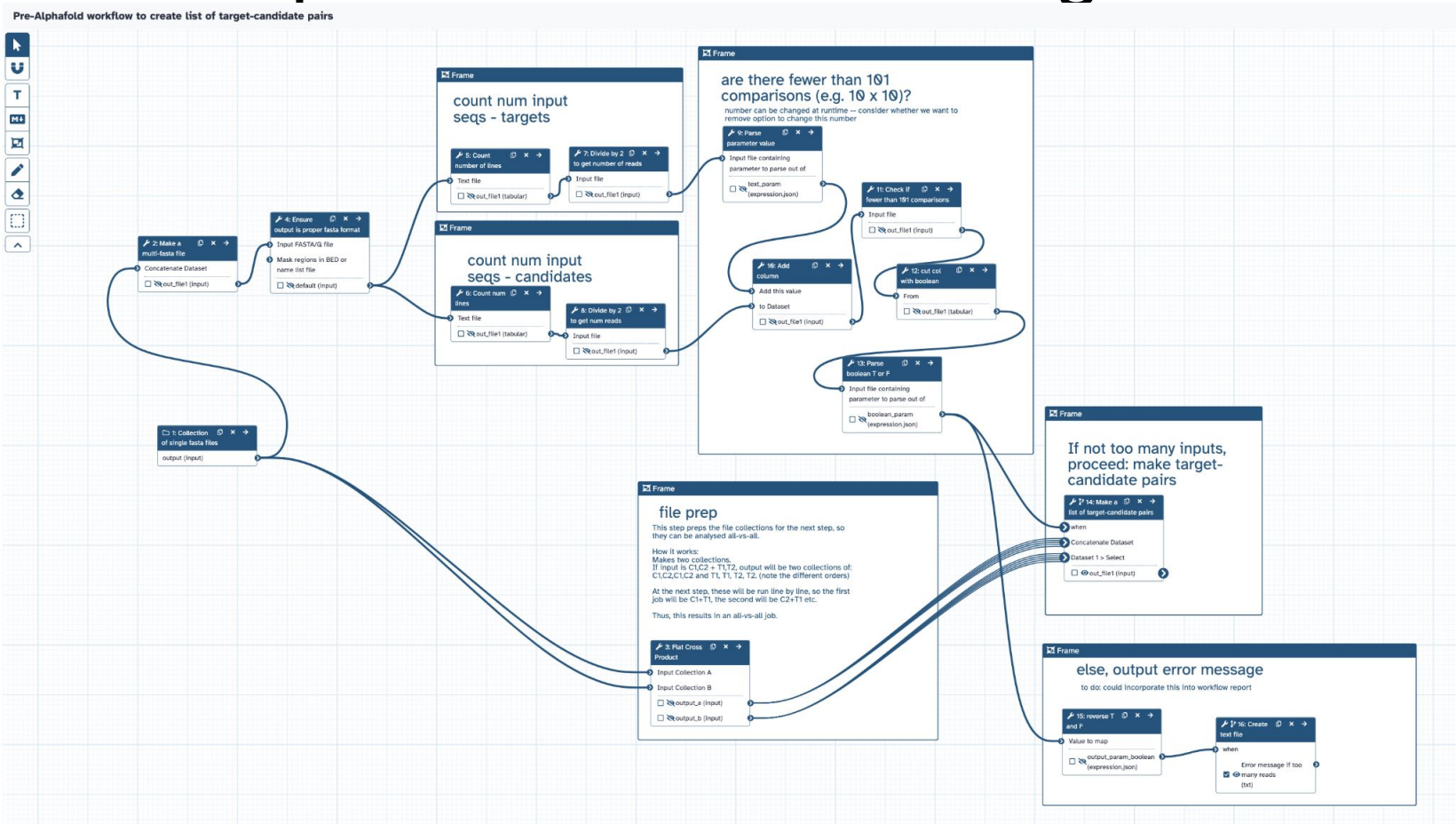
https://www.nature.com/articles/d41586-024-03214-7



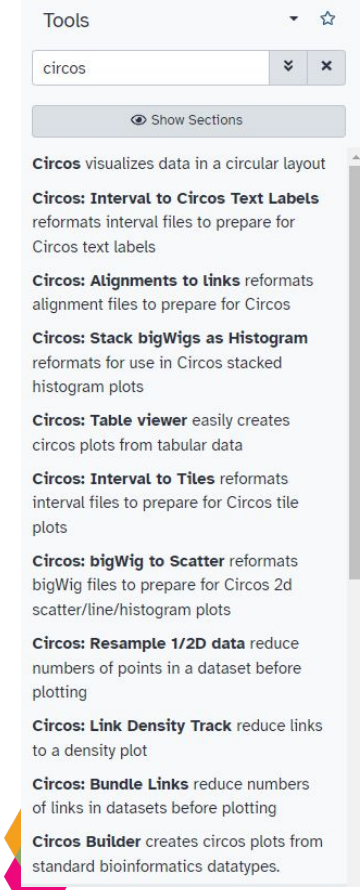
B.subtilis_FtsL_DivIC_DivIB_PBP2B_SpoVE_complex
aa: 117, 125, 263, 713, 366

Australian BioCommons

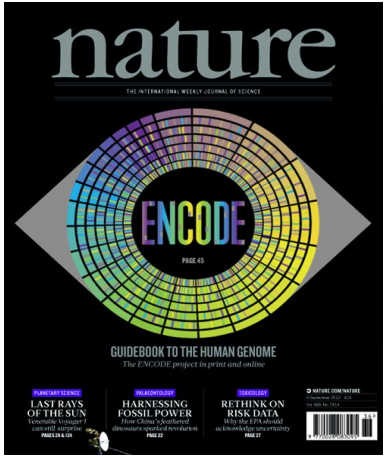
Alphafold as a screening tool

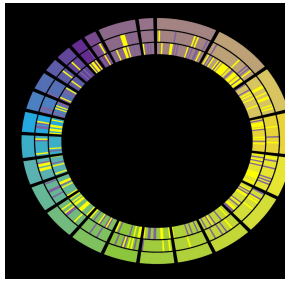


Circos in Galaxy



https://www.youtube.com/watch?v=j2R10doH5_4





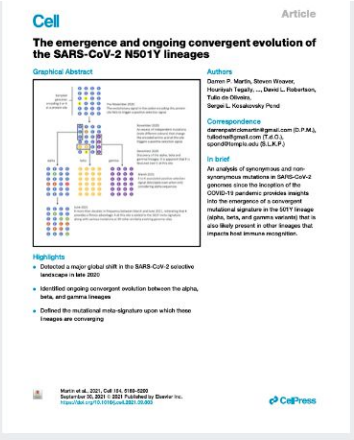
Visualisation with Circos

https://training.galaxyproject.org/training-material/topics/visualisation/tutorials/circos/tutorial.html

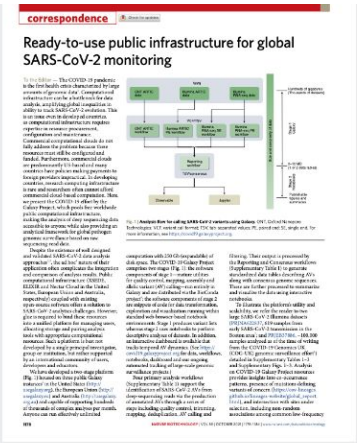
Galaxy's response to that annoying virus:
https://galaxyproject.org/projects/covid19/


PLOS PATHOGENS

No more business as usual: Agile and effective responses to emerging pathogen threats require open data and open analytics


Cell

The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages


correspondence

Ready-to-use public infrastructure for global SARS-CoV-2 monitoring


Brief Communications

Evolution and Exceptional Conservation of ORF1a/b Overlap in Coronaviruses

Vertebrate Genome Project (VGP) on Galaxy

nature biotechnology

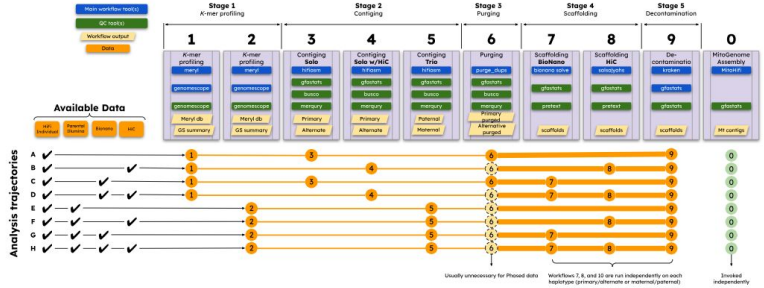
Correspondence | Published: 26 January 2024

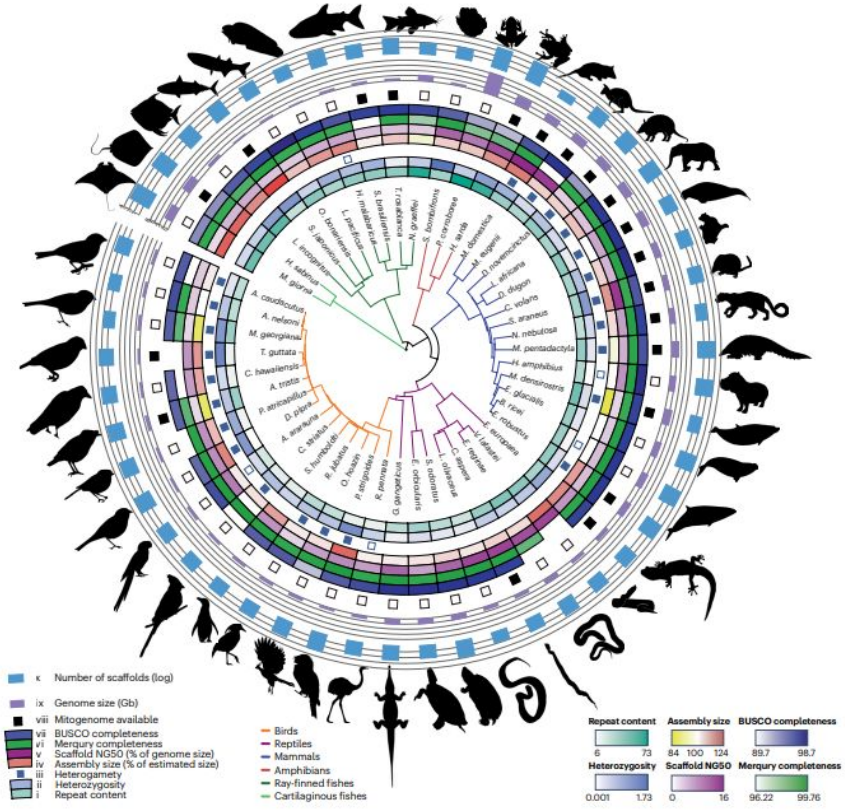
Scalable, accessible and reproducible reference genome assembly and evaluation in Galaxy

Delphine Larivière, Linelle Abueg, Nadolina Brajuka, Cristóbal Gallardo-Alba, Bjorn Grüning, Byung June Ko, Alex Ostrovsky, Marc Palmada-Flores, Brandon D. Pickett, Keon Rabbani, Agostinho Antunes, Jennifer R. Balacco, Mark J. P. Chaisson, Haoyu Cheng, Joanna Collins, Melanie Couture, Alexandra Denisova, Olivier Fedrigo, Guido Roberto Gallo, Alice Maria Giani, Grenville MacDonald Gooder, Kathleen Horan, Nivesh Jain, Cassidy Johnson, … Giulio Formenti ✉ + Show authors

https://doi.org/10.1038/s41587-023-02100-3

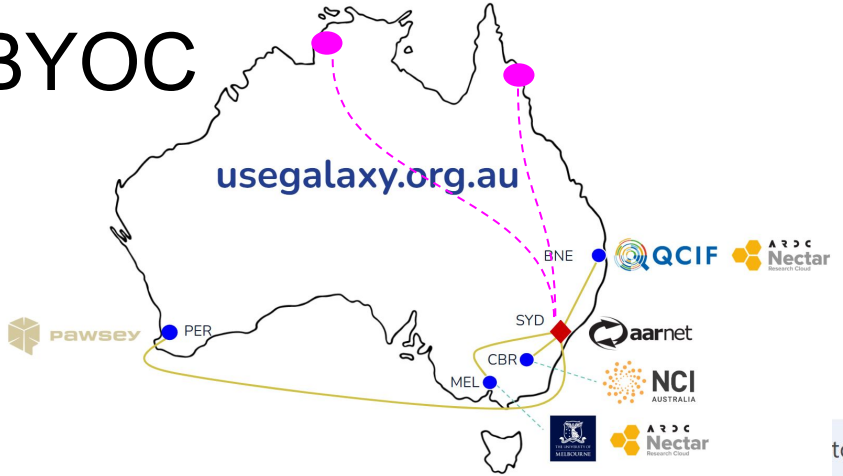
https://vgp.usegalaxy.org/





Bring Your Own Compute / Storage (BYOC / BYOS)

BYOC



BYOS

tore_instances/create

Europe | Workflow | Visualize | Data | Help | User

- Azure Blob Storage
- Amazon Web Services S3 Storage
- Any S3 Compatible Storage
- Google Cloud Storage

Select storage location template to create new storage location with. These templates are configured by your Galaxy administrator.

Australian BioCommons

Galaxy Australia Team

Gareth Price

Catherine Bromhead

Justin Lee

Nuwan Goonasekera

Michael D'Silva

Igor Makunin

Michael Thang

Cameron Hyde

Tom Harrop

Anna Syme

Australian BioCommons

QCIF

Melbourne Bioinformatics

aarnet

pawsey

NCI AUSTRALIA

BIOPLATFORMS AUSTRALIA

ARDC Australian Research Data Commons

Galaxy PROJECT

A final word

Galaxy Australia Acknowledgment Statement

When you use Galaxy Australia to support your publication or project, please acknowledge its use with the following statement:

"This work is supported by Galaxy Australia, a service provided by Australian BioCommons and its partners. The service receives NCRIS funding through Bioplatforms Australia, as well as The University of Melbourne and Queensland Government RICF funding."

Please also cite the Galaxy publication:

The Galaxy Community, The Galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update, *Nucleic Acids Research*, Volume 52, Issue W1, 5 July 2024, Pages W83-W94, https://doi.org/10.1093/nar/gkae410/

and **Thanks for having us**

Australian BioCommons