



**The ELIXIR-Greece Galaxy server including best practices tools and workflows
for the analysis of SARS-CoV-2 data**

Alexandros Dimopoulos, Martin Reczko



BSRC Al. Fleming

www.elixir-greece.org

Galaxy-ELIXIR webinar series: FAIR data and Open Infrastructures to tackle the COVID-19 pandemic

30 April 2020 - 28 May 2020

- Session 1: Introduction to Galaxy and the Galaxy workflows for SARS-CoV-2 data analysis
- Session 2: Genomics/Variant Calling
- Session 3: Cheminformatics: Screening of the main protease
- Session 4: Evolution of the Virus
- Session 5: Behind the scenes: Global Open Infrastructures at work

<https://elixir-europe.org/events/webinar-galaxy-elixir-covid19>

Video + Slides available



Best practices for the analysis of SARS-CoV-2 data

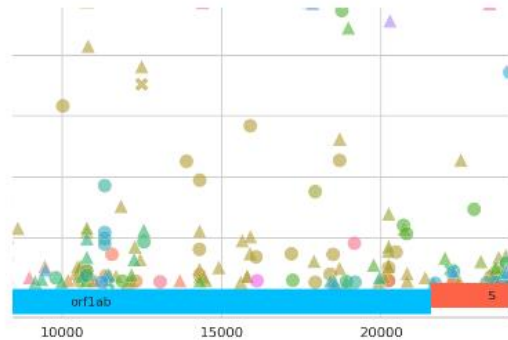
- The goal of this resource is to provide publicly accessible infrastructure and workflows for SARS-CoV-2 data analyses (<https://covid19.galaxyproject.org/>)



Three types of analyses

Genomics

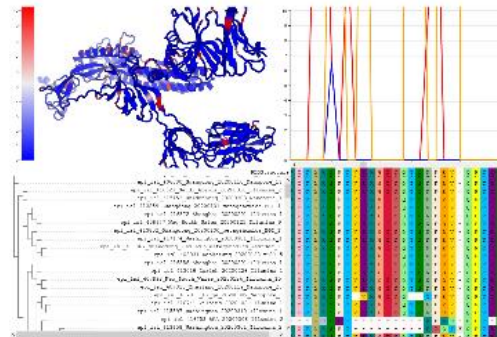
Assembly and intra-host variation



- Assembly
- MRCA timing
- Variation analysis
- Selection and recombination

Evolution

Sites under selection



- Natural Selection Analysis
- Analysis
- Visualizations
- Observable Notebooks

Cheminformatics

Screening of the main protease

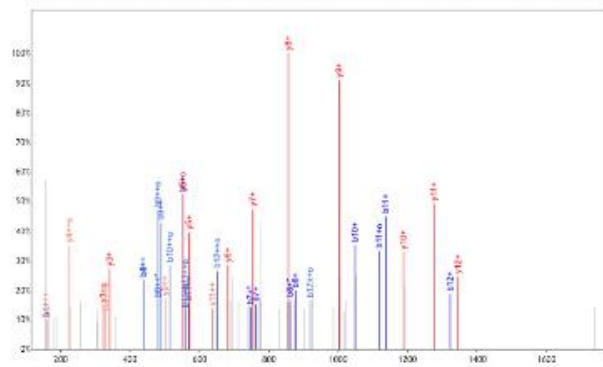


- Compound enumeration
- Generation of 3D conformations
- Docking
- Scoring
- Selection of compounds for synthesis

Additional types of analyses

Proteomics

Mass Spectrometry



- Reanalysis of PXD018804
- Reanalysis of PXD018682
- Reanalysis of PXD018117
- Metaproteomics of mPXD019423
- Metaproteomics of mPXD019119

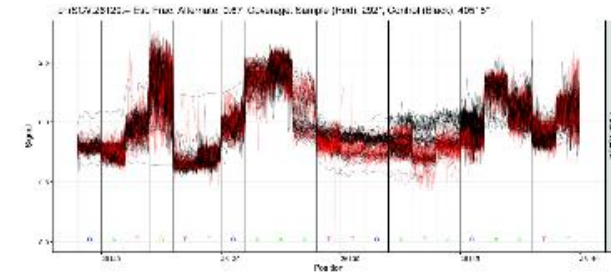
Artic

Amplicon based data analysis



direct RNA-seq

direct RNA-seq data analysis



- Pre-Processing
- RNA Epigenetics

Workflows

Galaxy Analyze Data Workflow Visualize Shared Data Admin Help User

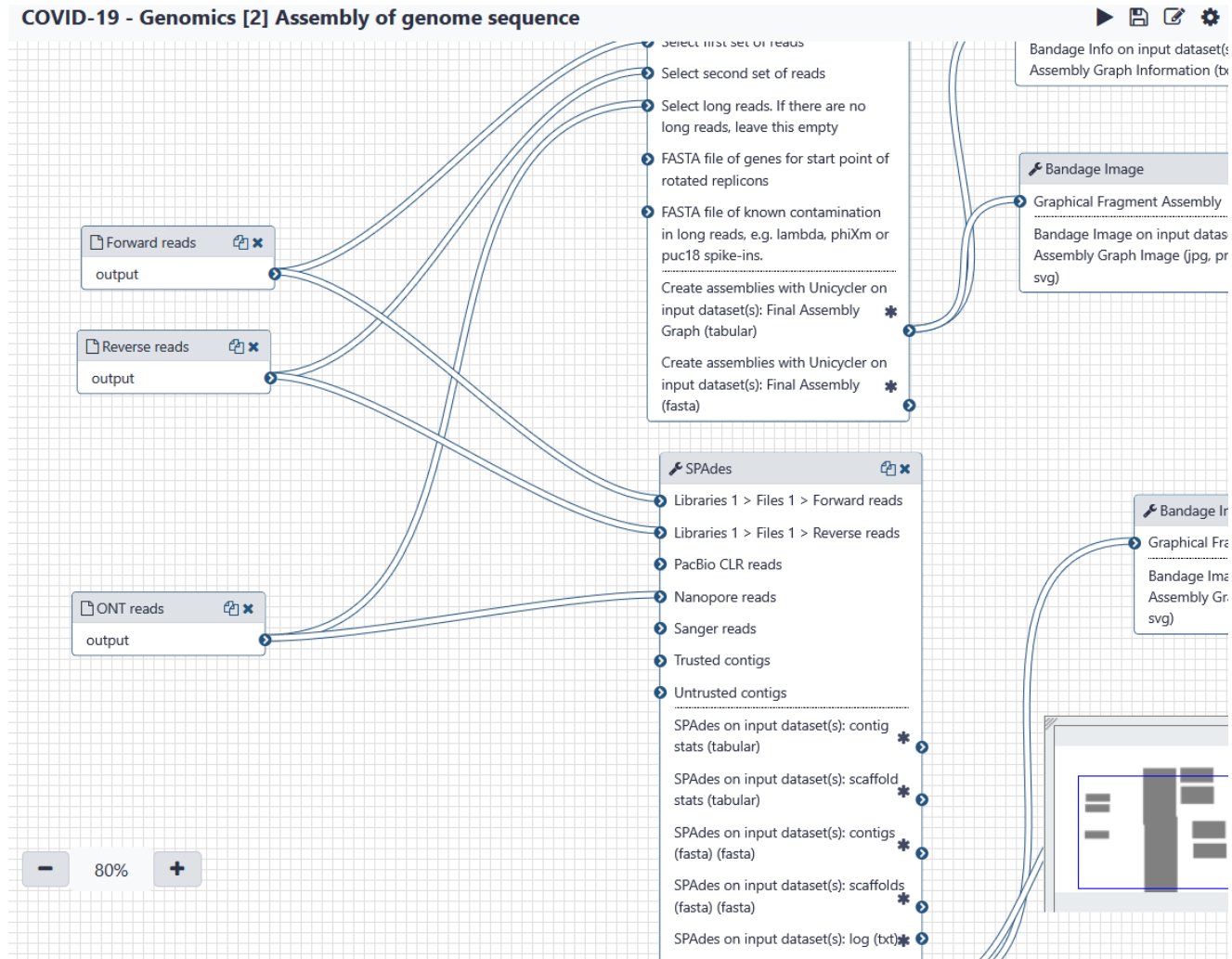
Tools ☆ ⬆

search tools ✖

- Get Data
- Send Data
- Collection Operations
- Expression Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- MaxQuant (using mqpar.xml)**
- FastQC** Read Quality reports
- Trim Galore!** Quality and adapter trimmer of reads
- Filter with SortMeRNA** of ribosomal

COVID-19 - Genomics [1] Read pre-processing without downloading from SRA Preprocessing of raw SARS-CoV-2 reads	covid-19 ✖	5 months ago	<input type="checkbox"/>	▶
COVID-19 - Genomics [5] Analysis of S-protein polymorphism Analysis of S-protein polymorphism	covid-19 ✖	5 months ago	<input type="checkbox"/>	▶
COVID-19 - Genomics [1] Read pre-processing with download Preprocessing of raw SARS-CoV-2 reads	covid-19 ✖	5 months ago	<input type="checkbox"/>	▶
COVID-19 - Genomics [4] PE Variation Analysis of variation within individual COVID-19 samples	covid-19 ✖	5 months ago	<input type="checkbox"/>	▶
COVID-19 - Genomics [2] Assembly of genome sequence Assembly of SARS-CoV-2 from pre-processed reads	covid-19 ✖	5 months ago	<input type="checkbox"/>	▶
COVID-19 - Genomics [4] SE Variation Analysis of variation within individual COVID-19 samples	covid-19 ✖	5 months ago	<input type="checkbox"/>	▶
COVID-19 - Genomics [6] Recombination and selection analysis Evolutionary Analysis	covid-19 ✖	5 months ago	<input type="checkbox"/>	▶
COVID-19 - Genomics [3] MRCA analysis Dating the most recent common ancestor (MRCA) of SARS-CoV-2	covid-19 ✖	5 months ago	<input type="checkbox"/>	▶

Workflow design



Workflow execution

Galaxy Analyze Data Workflow Visualize Shared Data Admin Help User Using 3.1 GB

Tools search tools

- Get Data
- Send Data
- Collection Operations
- Expression Tools
- Lift-Over
- Text Manipulation
- Convert Formats
- Filter and Sort
- Join, Subtract and Group
- Fetch Alignments/Sequences
- Operate on Genomic Intervals
- Statistics
- Graph/Display Data
- Phenotype Association
- MaxQuant (using mqpar.xml)**
- FastQC** Read Quality reports
- Trim Galore!** Quality and adapter trimmer of reads
- Filter with SortMeRNA** of ribosomal

Workflow: COVID-19 - Genomics [2] Assembly of genome sequence Run workflow

History Options

Send results to a new history
Yes No

1: Forward reads

34: Galaxy1-[2.fasta].fasta

2: Reverse reads

34: Galaxy1-[2.fasta].fasta

3: ONT reads

34: Galaxy1-[2.fasta].fasta

4: Create assemblies with Unicycler (Galaxy Version 0.4.6.0)

Paired or Single end data?
Paired

Select first set of reads
Output dataset 'output' from step 1 (-1)

History search datasets

Unnamed history
13 shown, 21 deleted
3.11 GB

- 34: Galaxy1-[2.fasta].fasta
- 33: Bowtie2 index
- 32: Bowtie2 index
- 31: Create DBKey and Reference Genome
- 30: Create DBKey and Reference Genome
- 29: FastQC on data 27: RawData
- 28: FastQC on data 27: Webpage
- 27: X161.s10q15150.fastq.gz
- 26: X161.s10q15150.head.fastq

ELIXIR Greece Galaxy instance



<https://usegalaxy.elixir-greece.org/>

- Small cluster of 3 (VM) nodes
 - 48 CPUs
 - 192 GB RAM
 - 3 TB of storage (no quota yet)
- Connected over a private network
- Common filesystem (nfs)
- Slurm Workload Manager

Depending on user needs, more resources (CPU, RAM, storage) and tools can be installed

Available to **everyone**

<https://usegalaxy.elixir-greece.org>

Welcome to Galaxy, please log in

Public name or Email Address

Password

Forgot password? [Click here to reset your password.](#)

Login

Don't have an account? [Register here.](#)



Select your identity provider

your previous selection

★ Alexander Fleming Biomedical Sciences Research Center

or

Sign in using another institute or account



Usage Stats

Galaxy Reports

Reports



Jobs

- Today's jobs
- Jobs per day this month
- Jobs in error per day this month
- All unfinished jobs
- Jobs per month
- Jobs in error per month
- Jobs per user
- Jobs per tool
- Errors per tool

Histories

- Histories and Datasets per User
- States of Datasets per History

Tools

- States of Jobs per Tool
- Execution Time per Tool

Workflows

- Runs per Workflows
- Workflows per month
- Workflows per user

Users

- Registered users
- Date of last login

Jobs Per Month

← 1 →

Click Month to view details. Graph goes from the 1st to the last of the month.

Max items:

Month ↑	User and Monitor Jobs
September 2020	23
July 2020	106
June 2020	47
April 2020	33



Thank you for your attention!



www.elixir-greece.org