Galaxy

A LONG TIME AGO IN A GALAXY FAR, FAR AWAY,
BIOINFORMATIC TOOLS ARE EASY TO USE,
INTERPRETATION OF THE RESULT IS SIMPLE,
EVERYTHING IS JUST ONE CLICK AWAY ...

12/02/2021

Important link

Documentation and training material:

- https://training.galaxyproject.org/training-material/

Galaxy platform:

- https://galaxyproject.org/use/ ⇒ https://usegalaxy.eu/

Workflow repository:

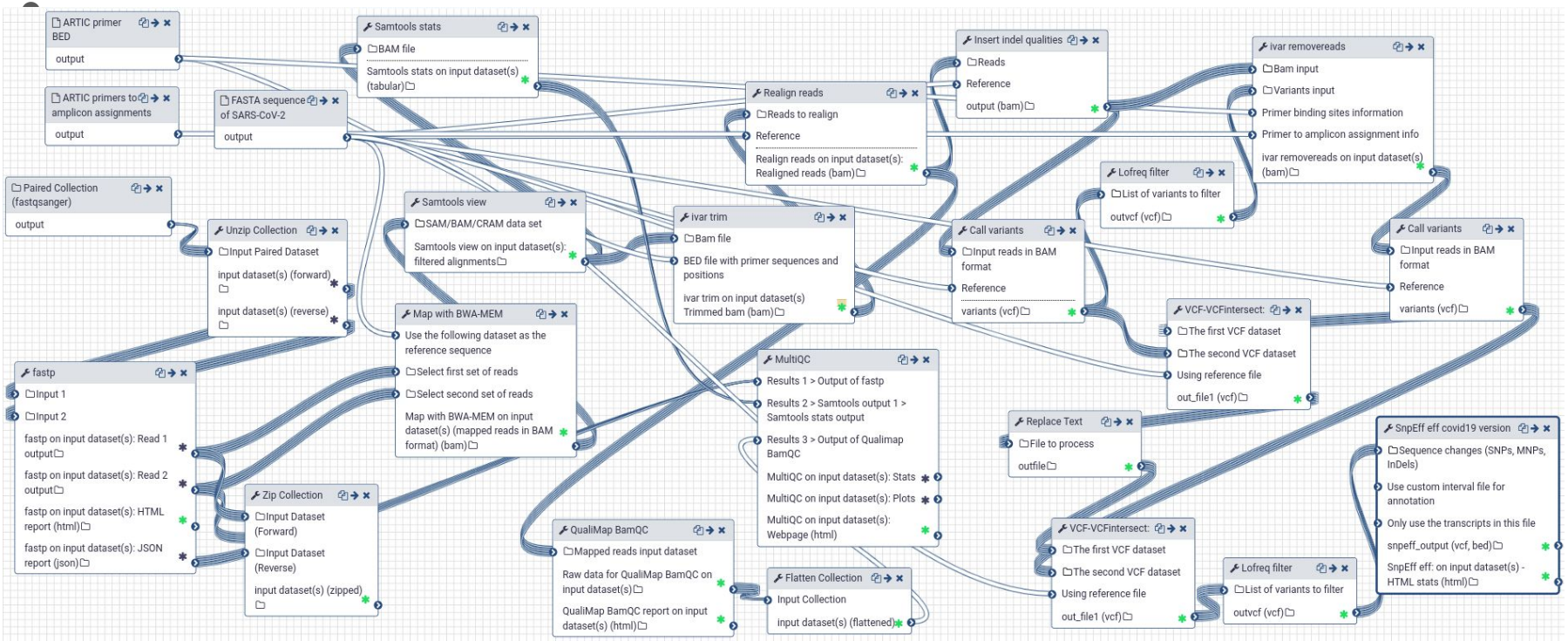
- https://workflowhub.eu/ ⇒ https://workflowhub.eu/projects/27#workflows

Download prepared référence:

- https://usegalaxy.eu/u/phytopat/h/reference-data

May the force be with us

- Workflow in galaxy https://usegalaxy.eu/workflows/list

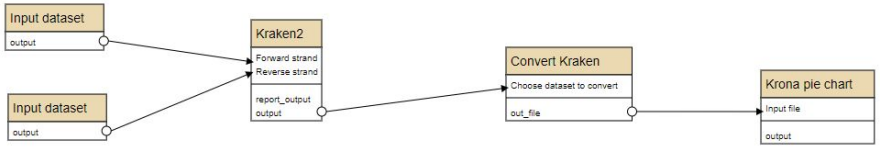


Find variant in covid19: https://covid19.galaxyproject.org/artic/

Kraken: a new hope

Workflow proposal: **1: Plant virus detection with kraken2**

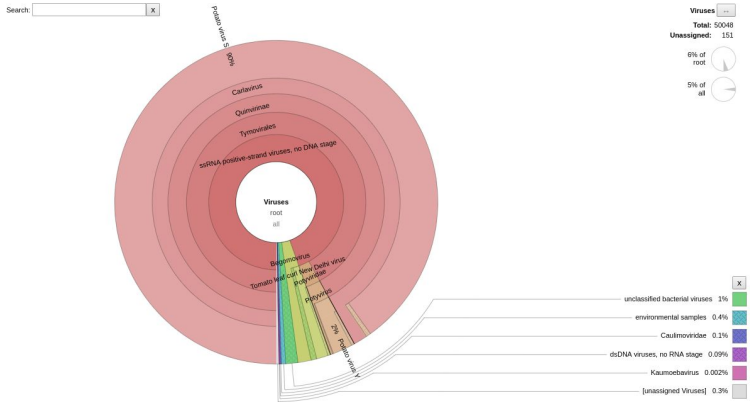
Make a list of candidate virus:



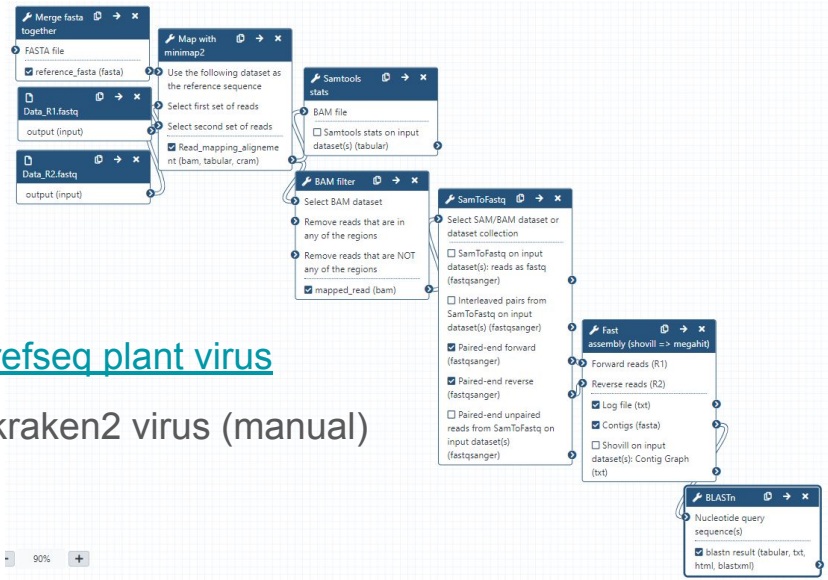
⇒ A lot of false hits

⇒ Don’t miss any present known virus (so far)

You can change the database !



BLAST: the old king
Mapping: the young Jedi



Workflow proposal: **2: Plant virus confirmation**

I) Map the candidate virus (on reads)

⇒ List of refseq plant virus

⇒ List of kraken2 virus (manual)

⇒ Reference may not be the closest

⇒ A lot of non viral read will map

⇒ Mapping profils can confirm presence of viruses

II) DeNovo contig of the mapped read

⇒ A lot of non viral contigs

⇒ Intentionally favorise numerus short contig

III) Blastn these contigs

⇒ Confirm presence of virus

⇒ Find the closest reference

⇒ Hard to read the result

Tools combination: Infinity and beyond

Workflow proposal: **3: Plant virus exploration**

I) Reads cleaning (trimming, merging, host removing, deduplicating …)

⇒ May vary according to your objective

II) DeNovo assembly

⇒ Favorise short or long contig according to your objective

III) Blast or mapping to flag contig

⇒ Lead to further exploration

Use case: find new viruses

To boldly go where no human has gone before

Clues about new viruses in your datasets:

- Kraken2 reads repartition between specie, genus and family can indicate new viruses.
- A tblastx on global DeNovo contig can find distant hit leading to new viruses.
- HMM analysis on protein from global DeNovo contig can find distant hit leading to new viruses.

Use case: Bacterial 16S

The dark side of Qiime underestimate you must not

Qiime is available on Galaxy, but fully dedicated team have developed specialized workflow/tools for bacterial analysis like:

FROGS on MIGALE plateform

FROGS article: *Escudie F., et al. Bioinformatics, 2018. FROGS: Find, Rapidly, OTUs with Galaxy Solution.*

- 16S analysis using migale Galaxy instance
- Tutorial and training

Same for pasteur: Galaxy instance (several original tools)

Same for South Green: Galaxy instance (plant analysis)

Same for phylogeny: https://ngphylogeny.fr/

To try to find some resources(tool/platform) for France, Germany or worldwide

Conclusion:

Advantage:

- Reproductible
- Easy to use
- Easy to explore intermediary analysis
- Goes well with networking party (and dataset sharing)
- User have to understand BioIT step !

Drawback:

- Fixed tool and database
- Difficult visualization (compared to Geneious)
- User have to understand BioIT step !

0: View complete virus identification ⇒ Non-functional workflow to get a global view of possibilities for plant virus classification.