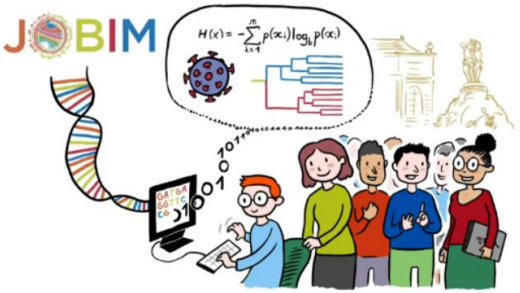
Report of the JOBIM conference

Club bioinfo - 13/10/2022

Audrey Gibert - Audrey Lapendry - Xavier Grand
ReGArDS team

Information on the JOBIM conference

JOBIM = Journées Ouvertes en Biologie Informatique et Mathématiques



SFBI video = https://www.youtube.com/watch?v=7iVD68aJuSU
- JOBIM déjà 20 ans

Dates = 5 to 8 July 2022

JOBIM 2022
05>08JUIL
Université de Rennes 1 | RENNES

Location = Campus de Beaulieu de l'Université de Rennes 1

→ Presentation of two talks or posters of interest per person

Talks and posters presented:

- Assessing conservation of alternative splicing with evolutionary splicing graphs

- Peer Community In: A free alternative to evaluate, validate (and publish) preprints


Audrey
LAPENDRY

- Semantic Web technologies are effective to remove redundancies from protein-protein interaction databases and define reproducible interactomes

- Green-BIM: a study to make young bioinformaticians aware of the carbon footprint of bioinformatics


Audrey
GIBERT

- TopoFun: improve the functional similarity of gene co-expression modules.

- Genome graphs detect human polymorphisms.

- nf-tower: make nextflow pipeline accessible to biologists.


Xavier
GRAND

Assessing conservation of alternative splicing with evolutionary splicing graphs

Diego Zea, Elodie Laine and Hugues Richard

Zea et al. Genome Research 2021

Goals = 1. Study the **transcript variability** of **different species** using a new type of splicing graph

2. Estimate **alternative splicing evolutionary conservation** and see how many **variations that are functionally relevant**

Tool available here: http://www.lcqb.upmc.fr/ThorAxe

Nodes = exons and edges = co-occurrences of exons in a set of transcripts observed for a gene

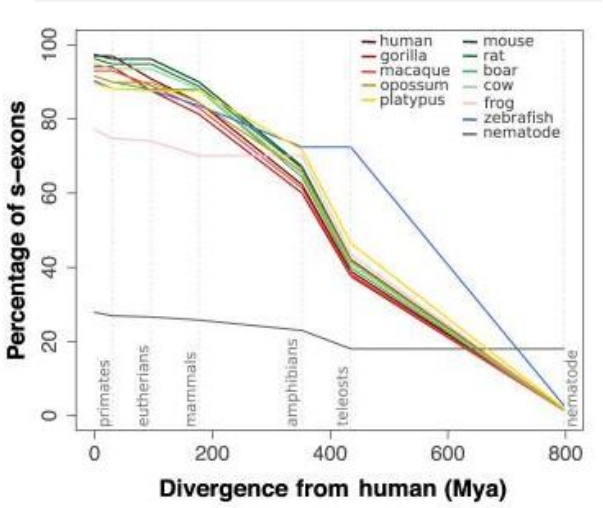
Example for a region of the gene CAMK2B:

→ 63 transcripts annotated in 10 species

Colors of nodes and edges = conservation levels, from yellow (low) to dark purple (high)

Assessing conservation of alternative splicing with evolutionary splicing graphs

ThorAxe summarizes across-species variations at the human proteome scale:



Percentages of s-exons conserved at different evolutionary distances from human

Peer Community In: A free alternative to evaluate, validate (and publish) preprints

Denis Bourguet

The Peer Community In project PCI: https://peercommunityin.org =

- **Alternative to the standard system of publication** & **non-profit** scientific organization
- Functioning =

  Recommenders evaluate preprints in their scientific fields based on peer-reviews.

  The recommendations are published in the thematic PCI websites with a DOI and can be sited.

  It can therefore be published in "Peer Community Journal" or submit in a PCI-friendly or other journal.

- Started in 2017 - now there are 15 thematic PCIs - 1,700 scientists as PCI recommenders

Semantic Web technologies are effective to remove redundancies from protein-protein interaction databases and define reproducible interactomes

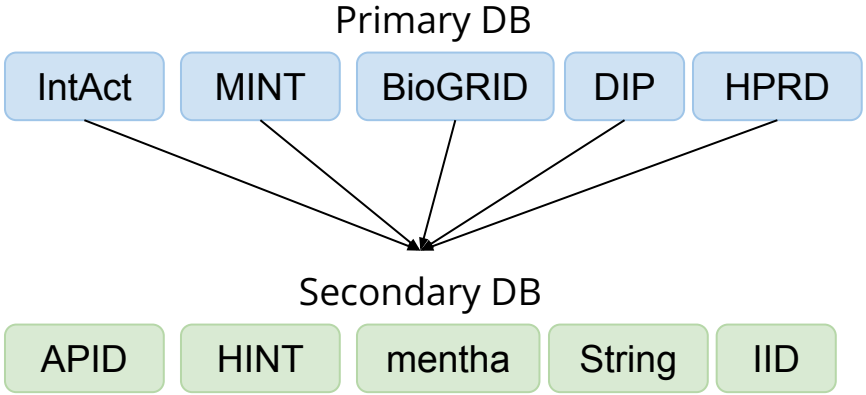
Talk by Olivier Dameron from INRIA Rennes

Based on **Melkonian, 2022** :
https://doi.org/10.1093/bioinformatics/btac013

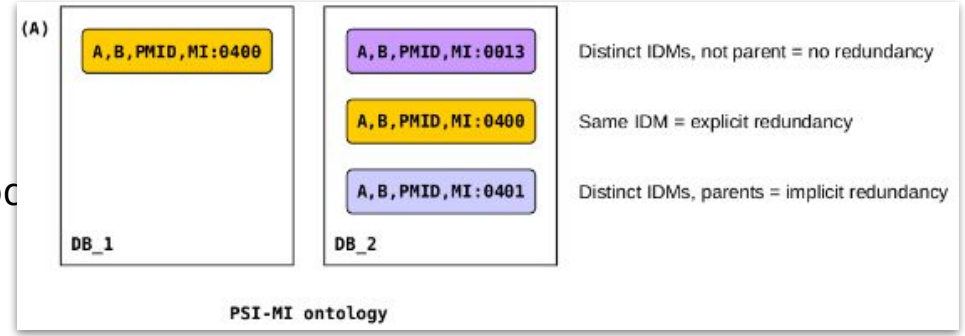
Protein-protein interaction (PPI) can be detected with numerous interaction detection methods. The confidence in the interaction grows with the number of paper proving it. 2 classes of databases records PPI :

Primary DB classifies the interaction detection method with vocabulary from PSI-MI (Proteomics Standard Initiative for Molecular Interaction). Chosen ontology term can be more or less specific according to the database.

Then, secondary DB query the primary ones to have more PPI. The information can be redundant because a same interaction proven by the same paper can be classified by different ontology term.



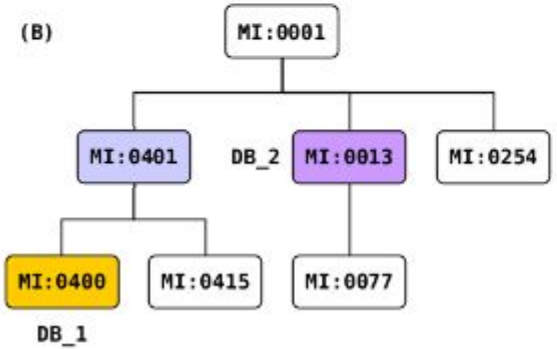
Semantic Web technologies are effective to remove redundancies from protein-protein interaction databases and define reproducible interactomes



2 redundancies :

- explicit : same PPI, PMID and detection metho
  DB are fine with it
- implicit : same PPI, same PMID but different detection method -> hard to say!

Methods from semantic web allow to call the difference.
Tests were conducted on human and yeast.



Results obtained for the Human interactome :
size ↘ 56% from 159,192 PPI to 70,554
reproducible

Green-BIM: a study to make young bioinformaticians aware of the carbon footprint of bioinformatics

Hélène DAUCHEL and her students from M2 Bioinformatics of Rouen

Make (young) professionals in bioinformatics aware of their work carbon footprint

Based on 2 papers :

- The Carbon Footprint of Bioinformatics
- Green Algorithms: Quantifying the Carbon Footprint of Computation

Carbon footprint = energy needed x carbon intensity

7 students calculated their carbon footprint according to the distance to their workplace, and work in bioinformatics

Green-BIM: a study to make young bioinformaticians aware of the carbon footprint of bioinformatics

Students	Runtime	GPU	CPU	Memory	Location
A	+	-	+	-	-
B	+	-	-	+	++
C	-	-	++	-	-
D	-	-	-	-	-
E	+	++	+	-	-
F	++	-	++	++	-
G	-	-	+	++	-

On what subject was D working? We are consumming carbon for our work

Advices :

- Selecting a datacenter based on its location
- Sometimes losing time allows to save energy comparing to unrestrained parallelization
- Updating softwares : ↘73% CF
- Carbon offsetting

TopoFun: improve the functional similarity of gene co-expression modules

Laurent Journot, Institut de Génomique Fonctionnelle, Univ. Montpellier, CNRS, INSERM, Montpellier

Janbain et al. (2021) TopoFun: a machine learning method to improve the functional similarity of gene co-expression modules, NAR Genom Bioinform, 3:lqab103. https://doi.org/10.1093/nargab/lqab103

https://github.com/ljournot/TopoFun

Goal: improve gene co-expression analysis, and add new genes related to functional modules.

> Functional modules in co-expression analysis are, usually, only defined by expression correlations. It doesn't take into account Functional annotation such as Gene Ontology.
>
> TopoFun propose to:

- Learn (machine learning) the links between genes from Gene Ontology Biological Processes Modules: $Score_{topo}$. Generate Random and Curated functional modules to determine discriminant descriptors.
- Designe a functional similarity score based on the distance in the GO tree of the annotations of the genes that constitute a module: $Score_{fun}$.
- Combine the two scores to evaluate if a co-expressed module is made of functionally related genes.

Genome graphs detect human polymorphisms

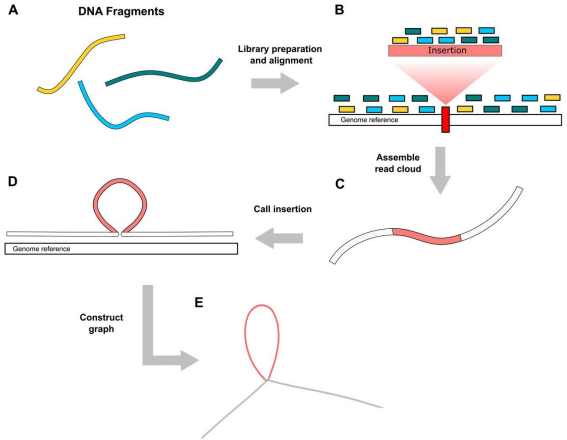
Guillaume Bourque, Canadian Center for Computational Genomics, McGill University, Montréal, Québec.

Groza et al. (2021) Genome graphs detect human polymorphisms in active epigenomic states during influenza infection, bioRxiv 2021.09.29.462206; doi: https://doi.org/10.1101/2021.09.29.462206

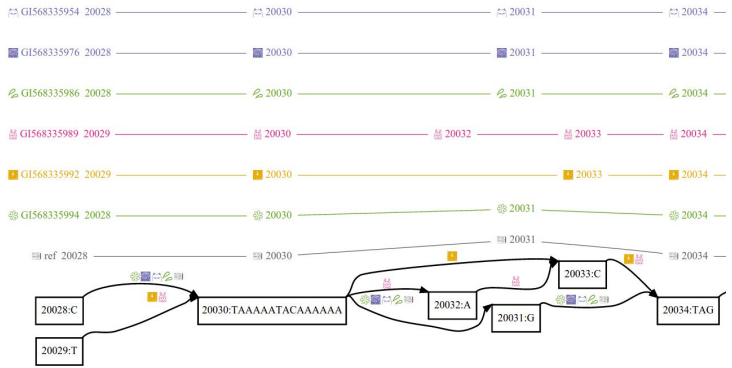
Reference genomes do not take into account individual variation, such as SNPs, Indels and other variations such as Mobile Element Insertions (MEIs). Whole genome sequencing is going to be easy and cheap, then it is possible to capture these variations from individuals.

Goal: Compile these information into a Graph Reference Genome instead of a linear reference sequence.

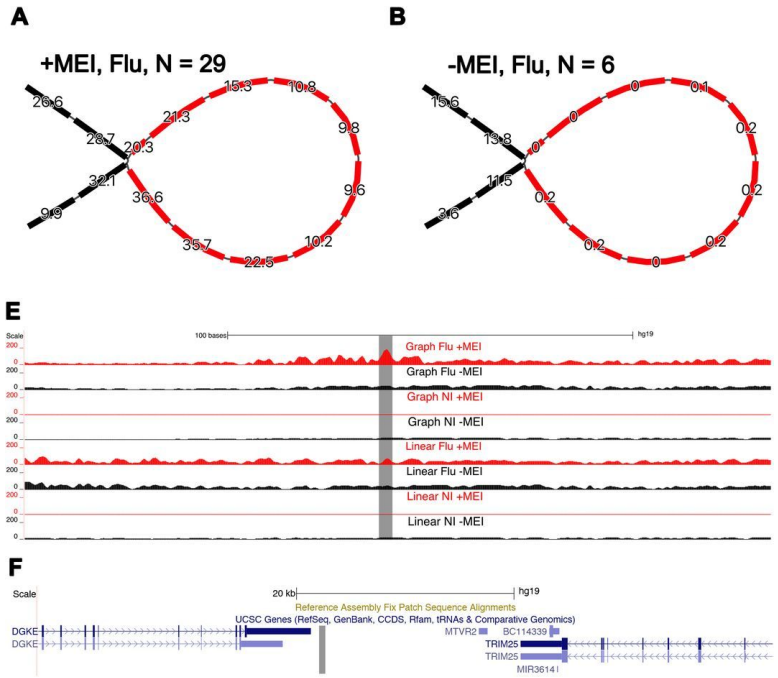
Using variation graph data structures to construct the graph reference from sequencing data and assembled reference genome.



https://github.com/vgteam/vg



Genome graphs detect human polymorphisms



As a results:

ATAC-seq results mapping on Reference sequence and Reference graph. A pic is highlighted when reads are mapped on Graph.

In this case, a regulation locus is impaired by MEI, specifically associated with the expression of TRIM25, a gene that restricts influenza RNA synthesis.

Graph reference genome can aggregate species genetic variations.

It's a way to store and explore pangenomes.

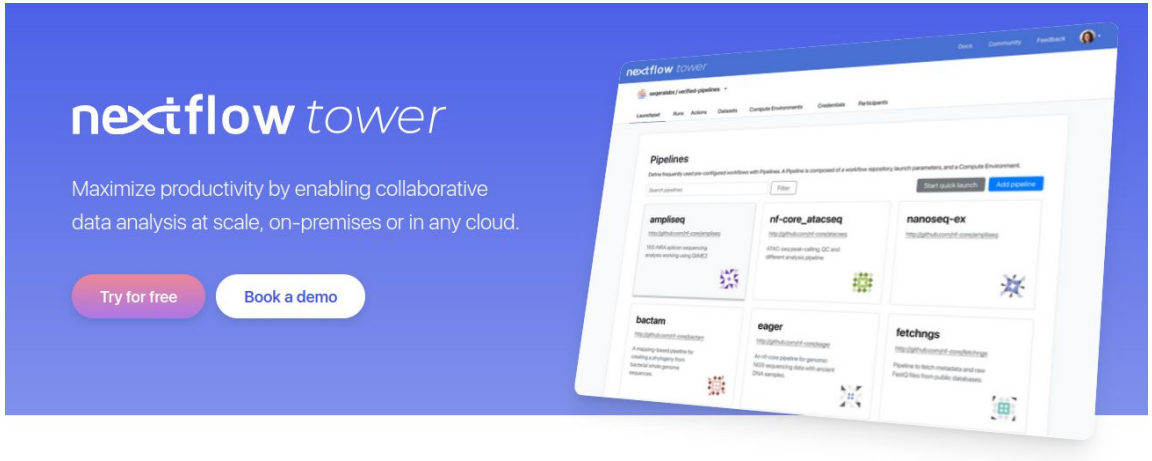
nf-tower: make nextflow pipelines accessible to biologists.

Cédric Notredame, Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Barcelona, Spain.

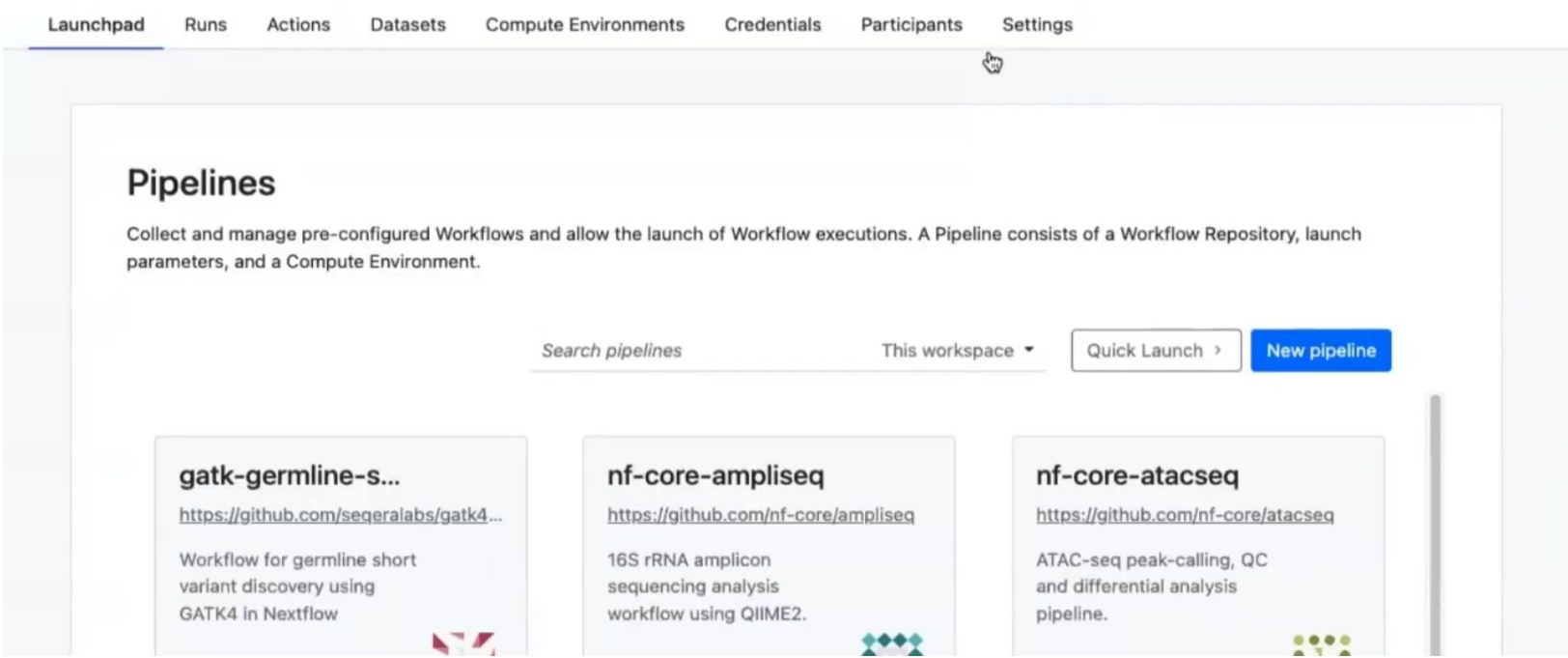
Open discussion during coffee break:

Question: Is there a GUI to nextflow and nf-core pipelines ? To facilitate the use of bioinformatics pipelines to biologists...

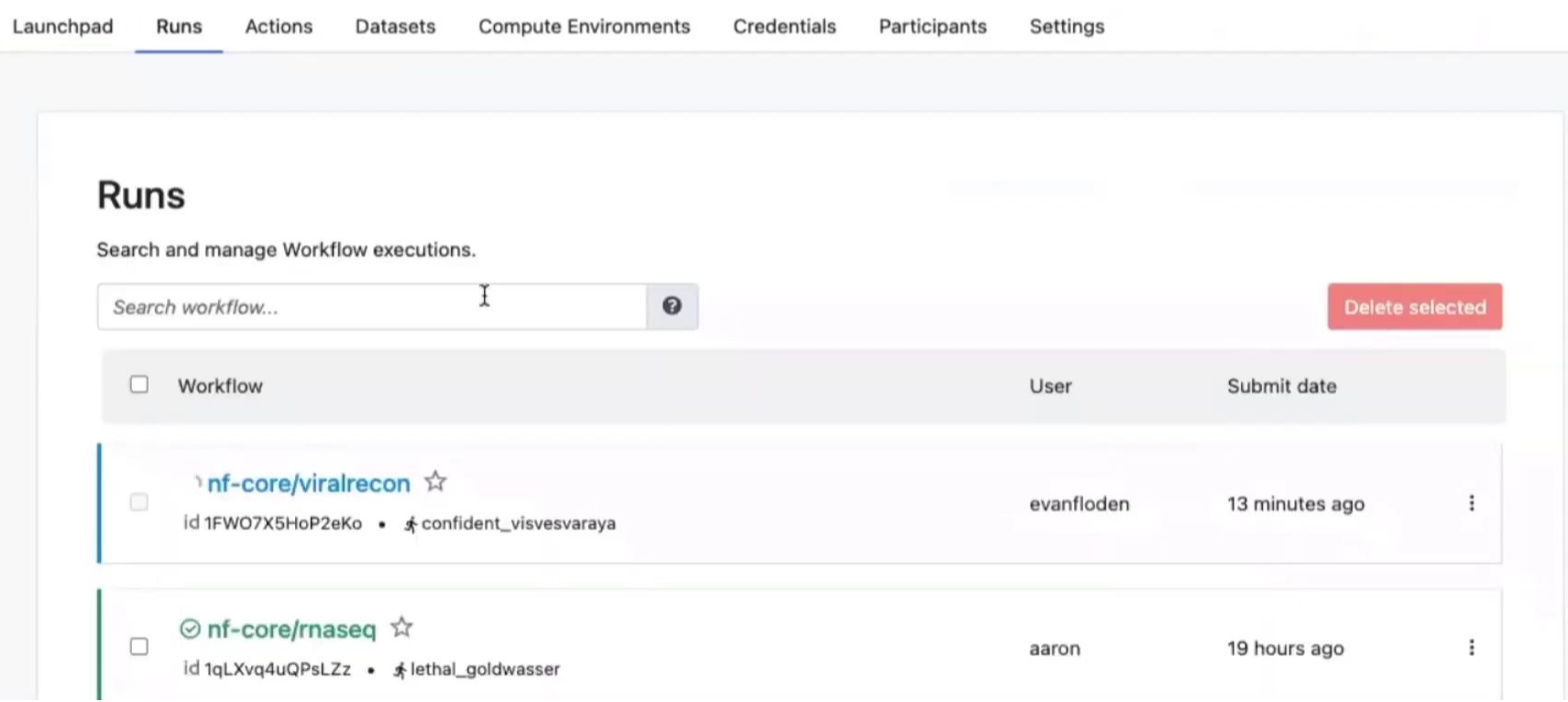
Answer: Yes, nf-tower.



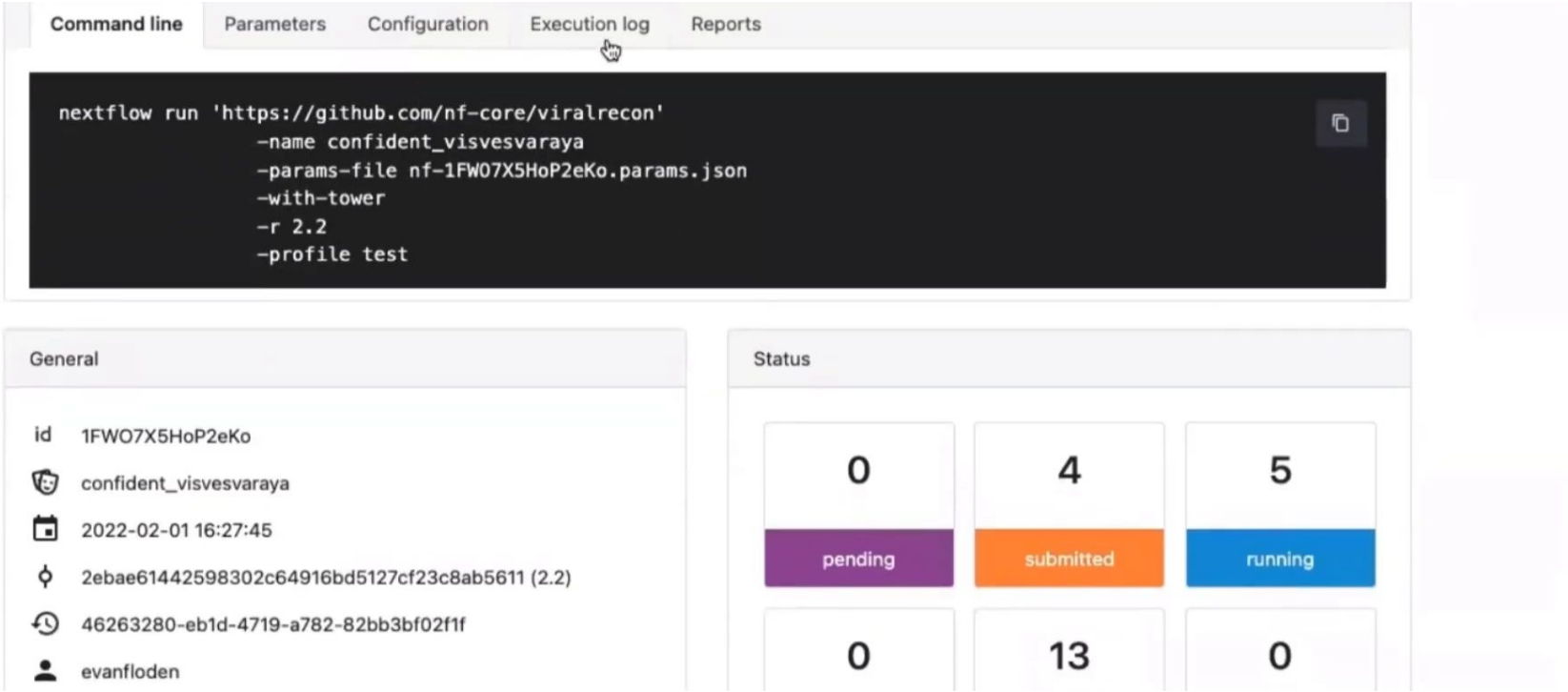
nf-tower: make nextflow pipelines accessible to biologists.



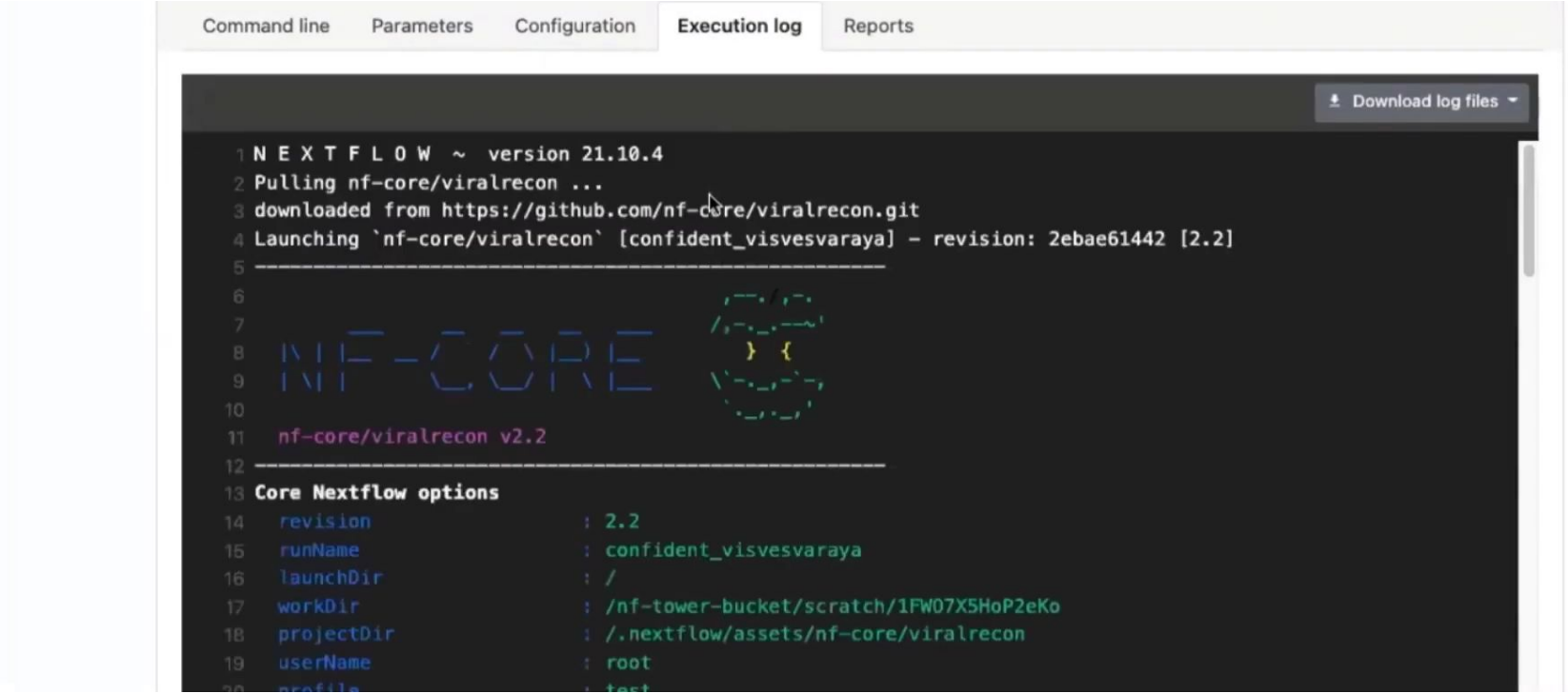
nf-tower: make nextflow pipelines accessible to biologists.



nf-tower: make nextflow pipelines accessible to biologists.



nf-tower: make nextflow pipelines accessible to biologists.



nf-tower: make nextflow pipelines accessible to biologists.

Nextflow Tower is an open source monitoring and management platform for Nextflow workflows developed by Seqera Labs.

Website:

https://cloud.tower.nf/

Github:

https://github.com/seqeralabs/nf-tower