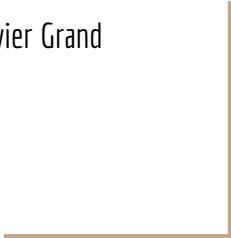


# Report of the JOBIM conference

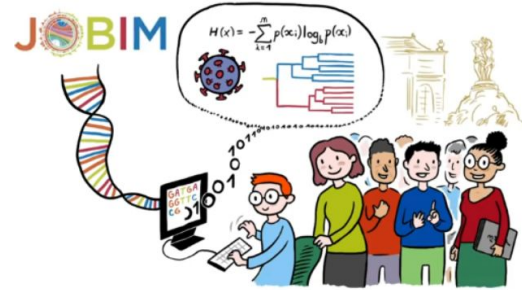
Club bioinfo - 13/10/2022  
Audrey Gibert - Audrey Lapendry - Xavier Grand  
ReGARDS team



# 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



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
- 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
- TopoFun: improve the functional similarity of gene co-expression modules.
- Genome graphs detect human polymorphisms.
- nf-tower: make nextflow pipeline accessible to biologists.



Audrey  
LAPENDRY



Audrey  
GIBERT



Xavier  
GRAND

# Assessing conservation of alternative splicing with evolutionary splicing graphs

Diego Zea, Elodie Laine and Hugues Richard

Zea et al. Genome Research 2021

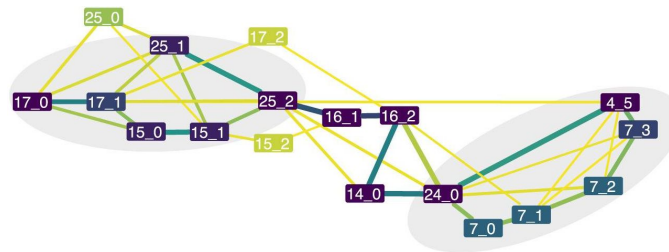
- Goals = 1. Study the **transcript variability** of **different species** using a new type of splicing graph
2. Estimate **alternative splicing evolutionary conservation** and see how many **variations that are functionally relevant**

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:

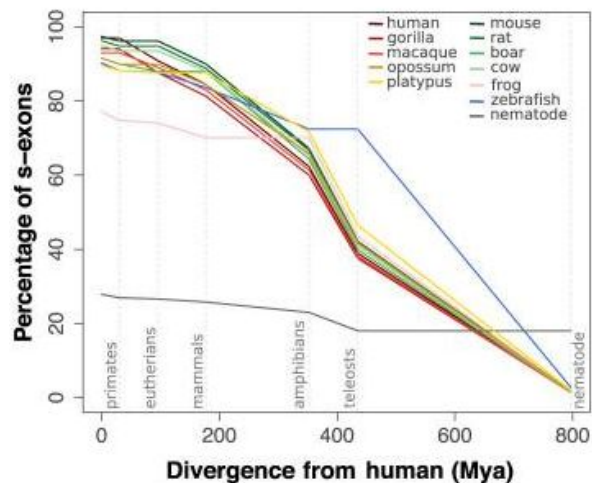
→ 63 transcripts annotated in 10 species



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

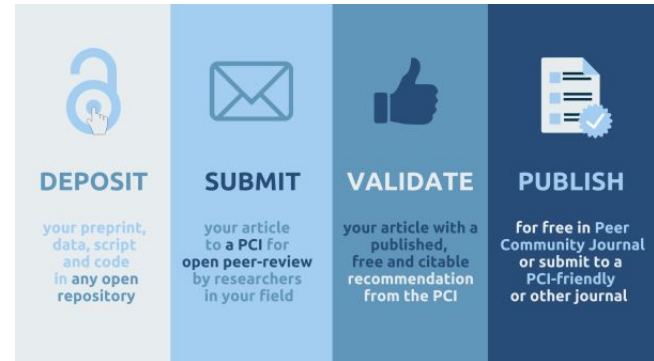
- Functioning =

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 cited.

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

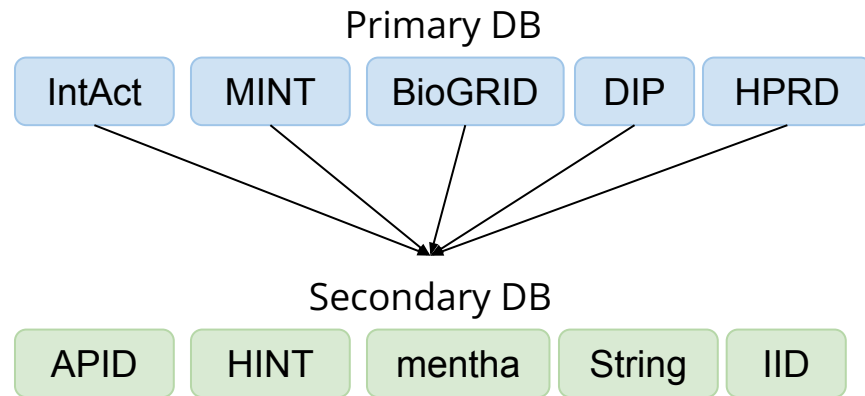
Based on **Melkonian, 2022** :

<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.

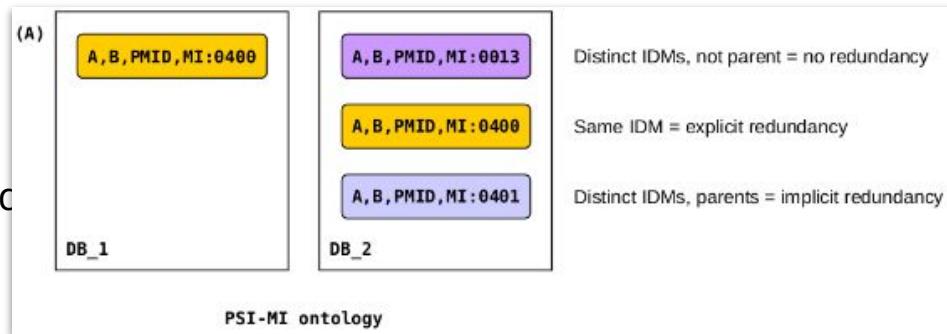
Then, secondary DB query the primary ones to have more PPI. The information can be redundant because a same interaction proven by the same paper can be classified by different ontology term.



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

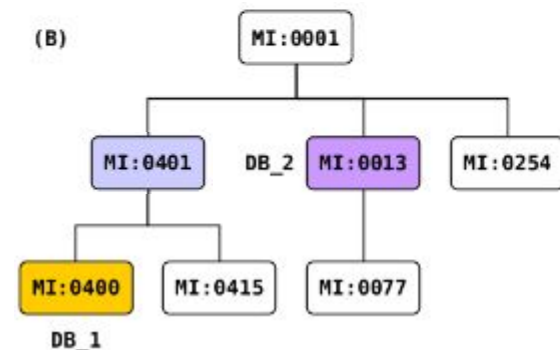
2 redundancies :

- **explicit** : same PPI, PMID and detection method  
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 consuming 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

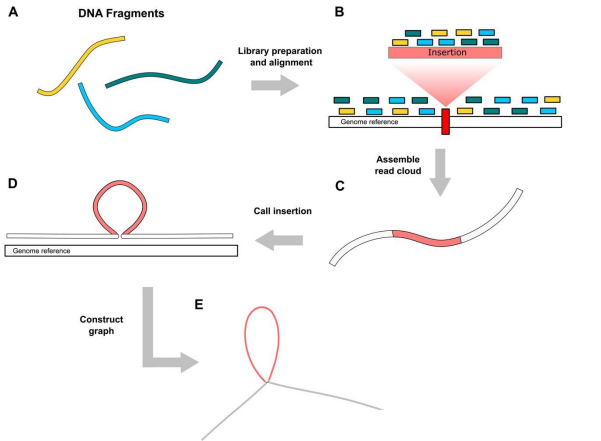
Guillaume Bourque, Canadian Center for Computational Genomics, McGill University, Montréal, Québec.

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>

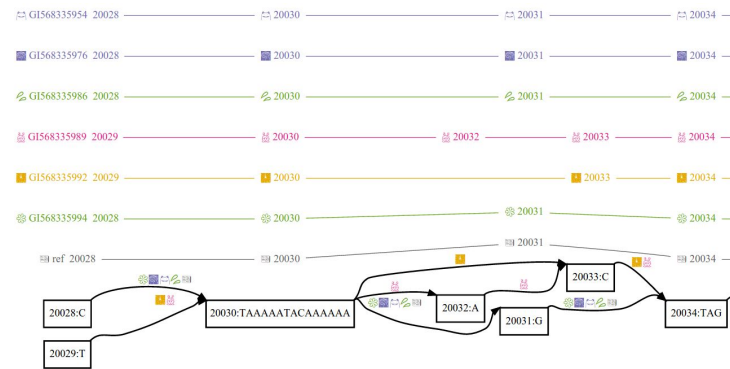
Reference genomes do not take into account individual variation, such as SNPs, Indels and other variations such as Mobile Element Insertions (MEIs). Whole genome sequencing is going to be easy and cheap, then it is possible to capture these variations from individuals.

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>





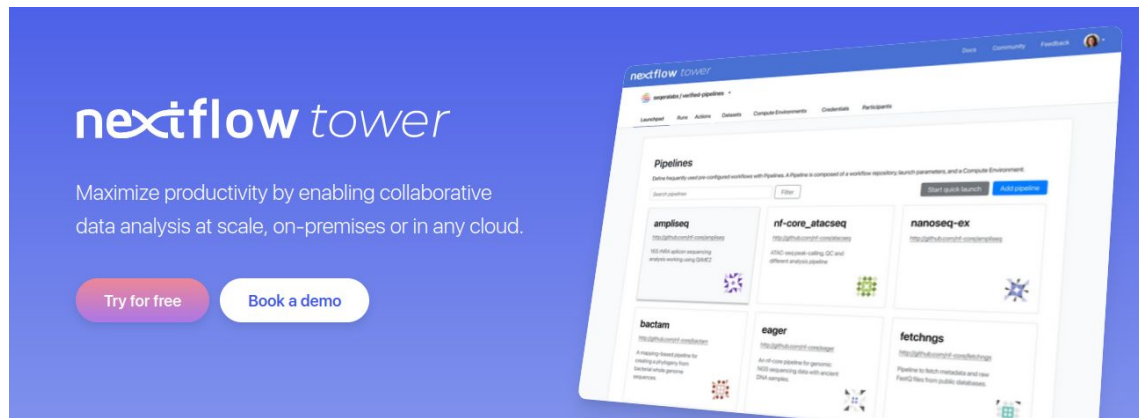
# nf-tower: make nextflow pipelines accessible to biologists.

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

Open discussion during coffee break:

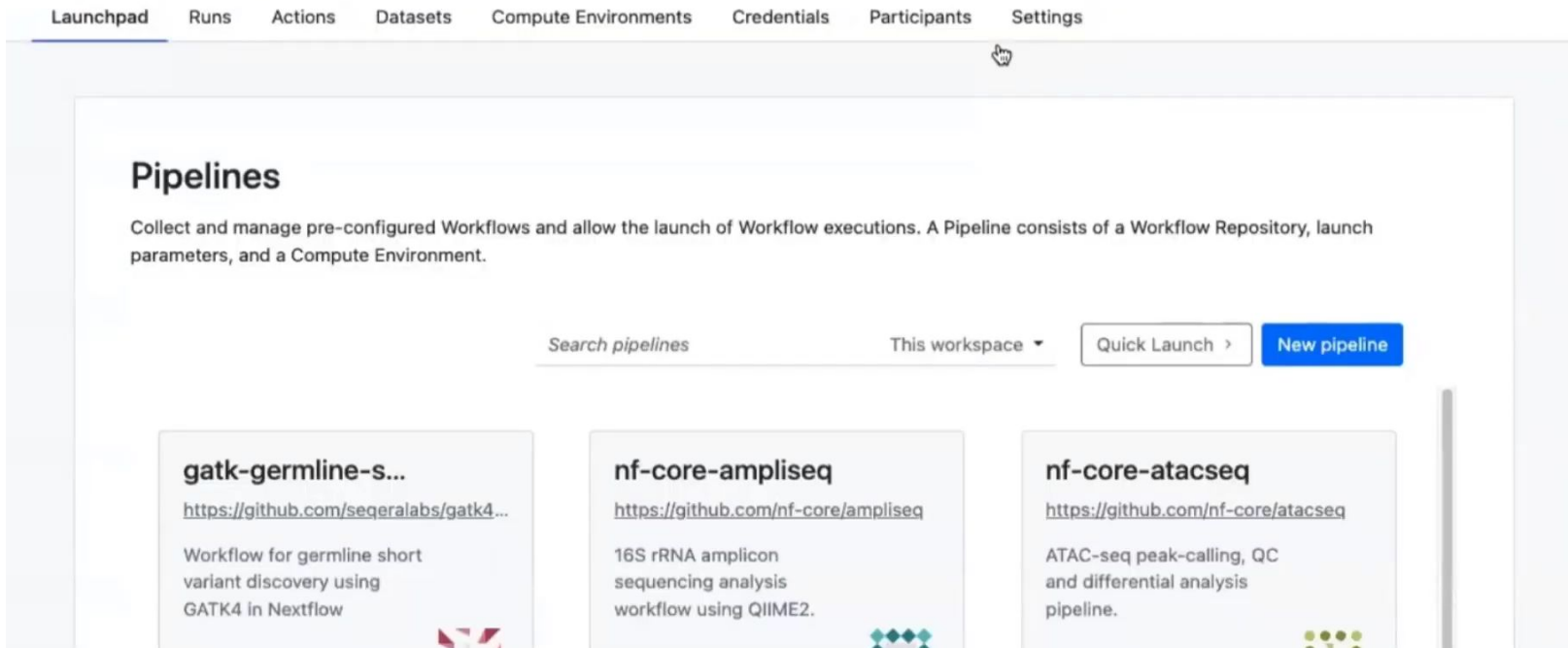
Question: Is there a GUI to nextflow and nf-core pipelines ? To facilitate the use of bioinformatics pipelines to biologists...

Answer: Yes, nf-tower.



The image shows a promotional banner for 'nextflow tower'. On the left, the text reads 'nextflow tower' in a white font on a blue background. Below this, it says 'Maximize productivity by enabling collaborative data analysis at scale, on-premises or in any cloud.' At the bottom left of this section are two buttons: 'Try for free' (pink) and 'Book a demo' (white with blue border). On the right, there is a screenshot of the 'nextflow tower' web interface. The interface has a blue header with the 'nextflow tower' logo and navigation links like 'Home', 'Community', and 'Feedback'. Below the header, there are tabs for 'Learn More', 'Run Actions', 'Details', 'Compute Environments', 'Credentials', and 'Participants'. The main content area is titled 'Pipelines' and contains a grid of pipeline cards. Each card includes a title, a brief description, a URL, and a small icon. The visible cards are: 'ampliseq' (RNA-seq analysis), 'nf-core\_atacseq' (ATAC-seq analysis), 'nanoseq-ex' (nanopore sequencing), 'bactan' (bacterial analysis), 'eager' (genomic analysis), and 'fetchngs' (genomic data fetching).

# nf-tower: make nextflow pipelines accessible to biologists.



The screenshot displays the 'nf-tower' web interface. At the top, a navigation bar includes links for 'Launchpad', 'Runs', 'Actions', 'Datasets', 'Compute Environments', 'Credentials', 'Participants', and 'Settings'. The 'Launchpad' link is underlined. Below the navigation bar, the main content area is titled 'Pipelines' and contains a descriptive paragraph: 'Collect and manage pre-configured Workflows and allow the launch of Workflow executions. A Pipeline consists of a Workflow Repository, launch parameters, and a Compute Environment.' Below this text, there is a search bar labeled 'Search pipelines', a dropdown menu set to 'This workspace', a 'Quick Launch >' button, and a prominent blue 'New pipeline' button. Three pipeline cards are visible below the search area. The first card is titled 'gatk-germline-s...' with a GitHub link and a description: 'Workflow for germline short variant discovery using GATK4 in Nextflow'. The second card is titled 'nf-core-ampliseq' with a GitHub link and a description: '16S rRNA amplicon sequencing analysis workflow using QIIME2.'. The third card is titled 'nf-core-atacseq' with a GitHub link and a description: 'ATAC-seq peak-calling, QC and differential analysis pipeline.' Each card features a small decorative icon in the bottom right corner.

Launchpad Runs Actions Datasets Compute Environments Credentials Participants Settings

## Pipelines

Collect and manage pre-configured Workflows and allow the launch of Workflow executions. A Pipeline consists of a Workflow Repository, launch parameters, and a Compute Environment.

Search pipelines This workspace ▾ Quick Launch > **New pipeline**

**gatk-germline-s...**  
<https://github.com/seqeralabs/gatk4...>  
Workflow for germline short variant discovery using GATK4 in Nextflow

**nf-core-ampliseq**  
<https://github.com/nf-core/ampliseq>  
16S rRNA amplicon sequencing analysis workflow using QIIME2.

**nf-core-atacseq**  
<https://github.com/nf-core/atacseq>  
ATAC-seq peak-calling, QC and differential analysis pipeline.

# nf-tower: make nextflow pipelines accessible to biologists.

Launchpad **Runs** Actions Datasets Compute Environments Credentials Participants Settings

## Runs

Search and manage Workflow executions.

Search workflow...



Delete selected



Workflow

User

Submit date



[nf-core/viralrecon](#) ☆

id 1FWO7X5HoP2eKo • [confident\\_visvesvaraya](#)

evanfloden

13 minutes ago



[nf-core/rnaseq](#) ☆

id 1qLXvq4uQPsLZz • [lethal\\_goldwasser](#)

aaron

19 hours ago





# nf-tower: make nextflow pipelines accessible to biologists.

The screenshot displays the nf-tower web interface. At the top, there are navigation tabs: "Command line", "Parameters", "Configuration", "Execution log", and "Reports". The "Execution log" tab is active, showing a terminal window with the following command:

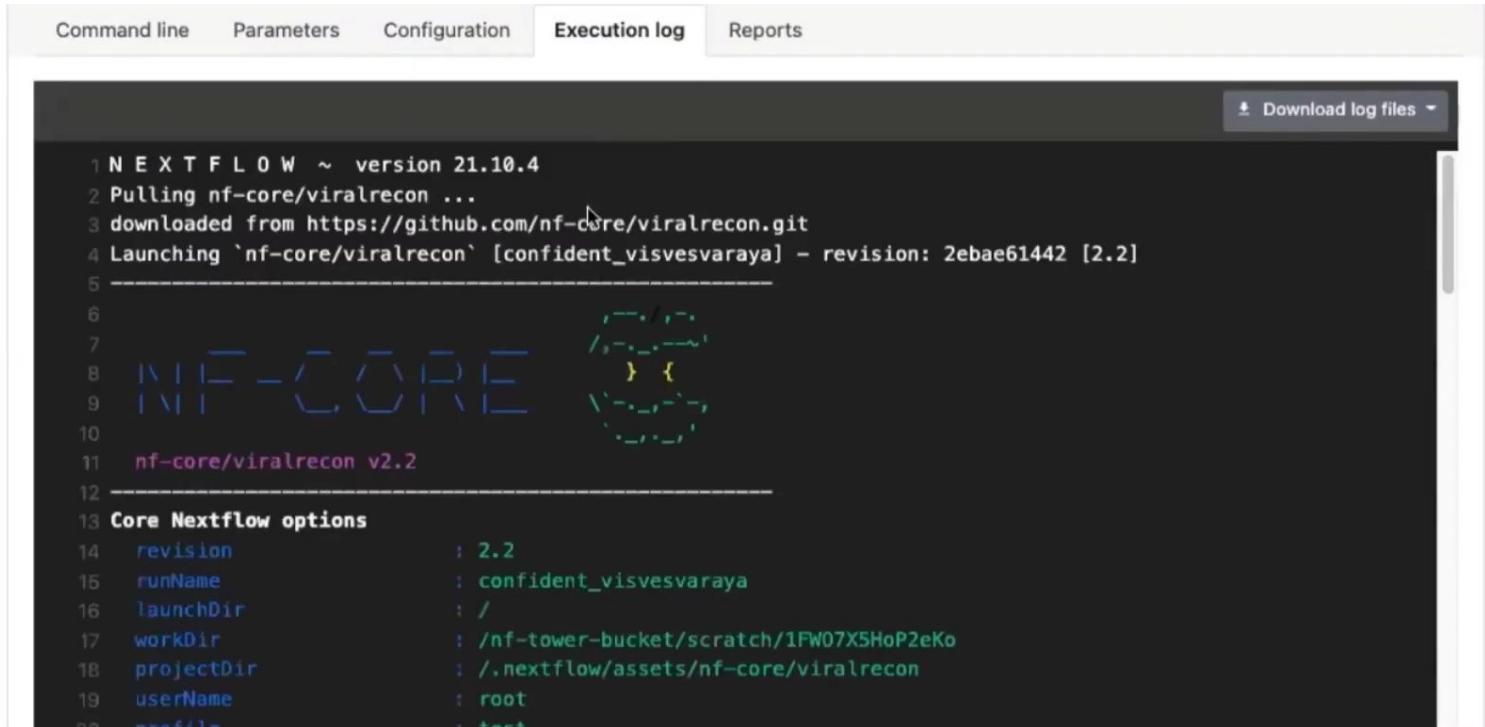
```
nextflow run 'https://github.com/nf-core/viralrecon'  
  -name confident_visvesvaraya  
  -params-file nf-1FW07X5HoP2eKo.params.json  
  -with-tower  
  -r 2.2  
  -profile test
```

Below the terminal window, there are two main panels:


- General:** A list of pipeline details:
  - id: 1FW07X5HoP2eKo
  - confident\_visvesvaraya
  - 2022-02-01 16:27:45
  - 2ebae61442598302c64916bd5127cf23c8ab5611 (2.2)
  - 46263280-eb1d-4719-a782-82bb3bf02f1f
  - evanfloden
- Status:** A grid of status indicators for different pipeline components:

0	4	5
pending	submitted	running
0	13	0

nf-tower: make nextflow pipelines accessible to biologists.



The screenshot displays the 'Execution log' tab of the nf-tower interface. The log content is as follows:

```
1 N E X T F L O W ~ version 21.10.4
2 Pulling nf-core/viralrecon ...
3 downloaded from https://github.com/nf-core/viralrecon.git
4 Launching `nf-core/viralrecon` [confident_visvesvaraya] - revision: 2ebae61442 [2.2]
5 -----
6
7
8 NF-CORE 
9
10 nf-core/viralrecon v2.2
11 -----
12 Core Nextflow options
13
14 revision           : 2.2
15 runName            : confident_visvesvaraya
16 launchDir          : /
17 workDir            : /nf-tower-bucket/scratch/1FW07X5HoP2eKo
18 projectDir         : /.nextflow/assets/nf-core/viralrecon
19 userName           : root
20 profile            : test
```

The interface includes navigation tabs for 'Command line', 'Parameters', 'Configuration', 'Execution log', and 'Reports'. A 'Download log files' button is located in the top right corner of the log area.

# 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>