



PEARC²⁴

PROVIDENCE, RI | HPC | HUMAN POWERED COMPUTING

HUMAN POWERED COMPUTING

nf-core on Open OnDemand: community-curated bioinformatics pipelines for everyone

Yucheng Zhang, Tufts University

Payas Bhutra, Purdue University

Shirley (Xue) Li, Tufts University

Bioinformatics workflow

- A pipeline/workflow is a collection of several analysis steps
- Steps are linked by input/output files
- One often needs to run the same workflow for several samples

Biological samples/Library preparation

Sequence reads

FASTQC

Adapter Trimming (Optional)

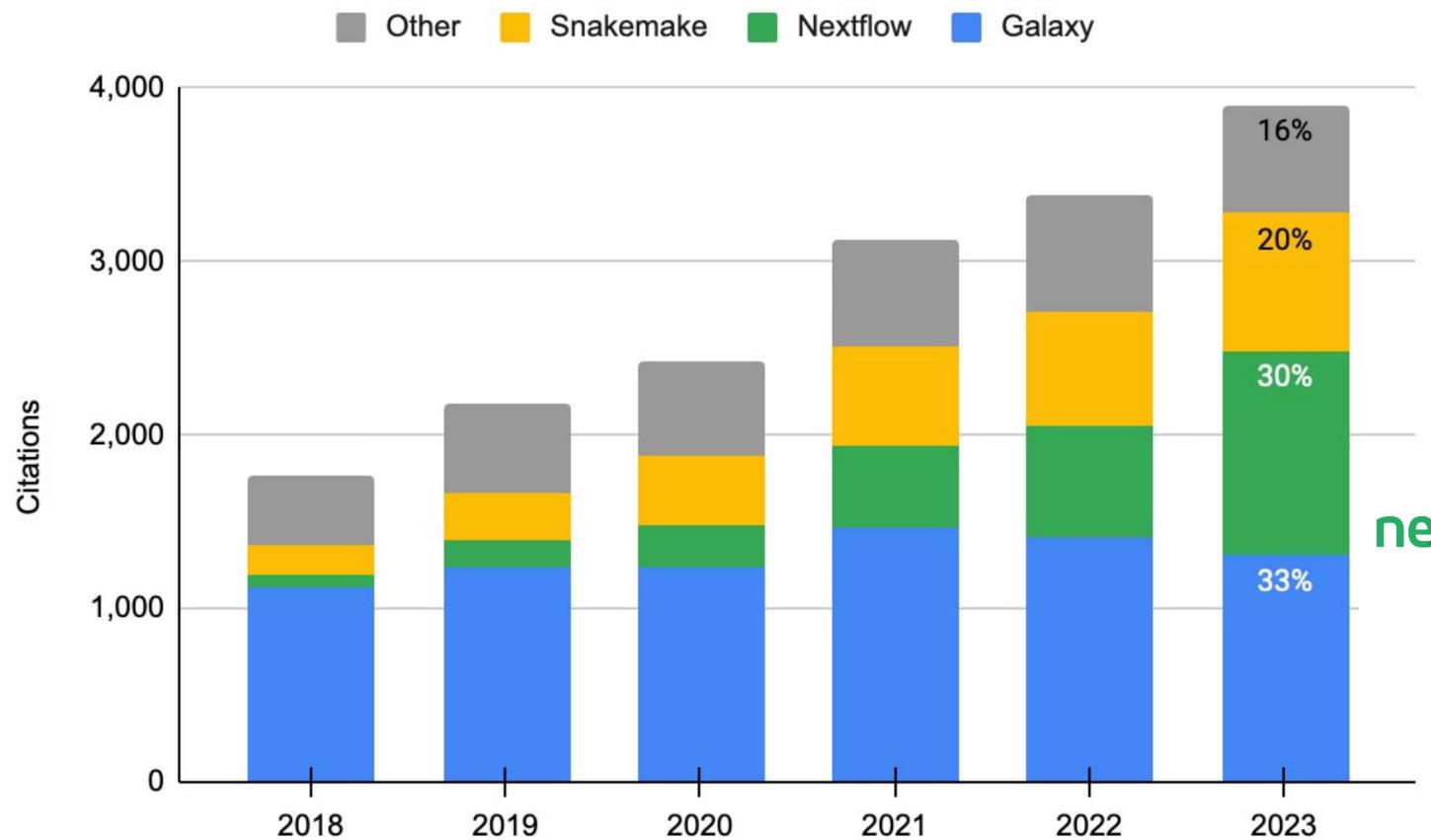
Splice-aware mapping to genome

Counting reads associated with genes

Statistical analysis to identify differentially expressed genes

An example workflow for gene expression analysis

Rising demand for bioinformatics workflows



nextflow

Langer, Bjorn E., et al.
"Empowering bioinformatics
communities with Nextflow and nf-
core." *bioRxiv* (2024): 2024-05.

FIGURE 1: Google Scholar citation counts for bioinformatics workflow management systems. Sum of citations of the major publications of Galaxy, Nextflow, and Snakemake between 2018 and 2023 (Data in Supplementary Table 1).

An overview of nextflow pipeline

nextflow pipeline

Write code
in any language



Orchestrate tasks with
dataflow programming



Define software
dependencies
via containers



Built-in version
control with Git



nextflow runtime

Task orchestration
and execution

Supported Platforms



nf-core



A community effort to collect a curated set of analysis pipelines built using Nextflow.

Pipelines

Browse the 109 pipelines that are currently available as part of nf-core.

Released 63

Under development 33

Archived 13

Last release ▾



funcscan ✓ ★ 62

(Meta-)genome screening for functional and natural product gene sequences

amp amr antibiotic-resistance
antimicrobial-peptides
antimicrobial-resistance-genes arg assembly
bgc biosynthetic-gene-clusters contigs
function metagenomics natural-products
screening secondary-metabolites

1.1.6 released 1 day ago

mag ✓ ★ 192

Assembly and binning of metagenomes

annotation assembly binning
long-read-sequencing metagenomes
metagenomics nanopore
nanopore-sequencing

3.0.2 released 5 days ago

eager ✓ ★ 128

A fully reproducible and state-of-the-art ancient DNA analysis pipeline

adna ancient-dna-analysis ancientdna
genome metagenomics pathogen-genomics
population-genetics

2.5.2 released 11 days ago

ampliseq ✓ ★ 163

Amplicon sequencing analysis workflow using DADA2 and QIIME2

16s 18s amplicon-sequencing edna
illumina iontorrent its metabarcoding
metagenomics microbiome pacbio qiime2
rrna taxonomic-classification
taxonomic-profiling

2.10.0 released 12 days ago

scrnaseq ✓ ★ 179

A single-cell RNAseq pipeline for 10X genomics data

proteinfold ✓ ★ 41

Protein 3D structure prediction pipeline

bacass ✓ ★ 55

Simple bacterial assembly and annotation pipeline

assembly bacterial-genomes denovo

demo ✓ ★ 1

nf-core/demo is a simple nf-core style bioinformatics pipeline for workshops and demos.



109 Pipelines Developed
as of July 2024



See what nf-core is up to

Slack is where work happens for companies of all sizes.



Chris Hakkaart, Phil Ewels and 9,224 others have already joined.

<https://nfcore.slack.com/>

nf-core config files

nf-core pipelines leverage nexflow's configuration files to specify the following:

- How the pipelines run,
- Custom parameters,
- Software management system to use
(e.g., docker, singularity or conda).



Default 'base' config (always loaded)



Core profiles (e.g. singularity, conda, test)



Institutional profiles (nf-core/configs)



Your local config files (-c flag)

Institutional profiles

```
nextflow run nf-core/<pipeline> -profile mycluster
```

- Specifies job submission
- 📦 Specify software packaging

Works for:

- ↳ For all pipelines
- 👥 For all users on your system
- ⟳ Single point to update

Benefits of institutional profiles

Resource Allocation and Improved Efficiency

- Institutional best practices and preferred settings can be embedded into the profile.
- Saving users time and effort by eliminating the need to configure individual pipelines from scratch.

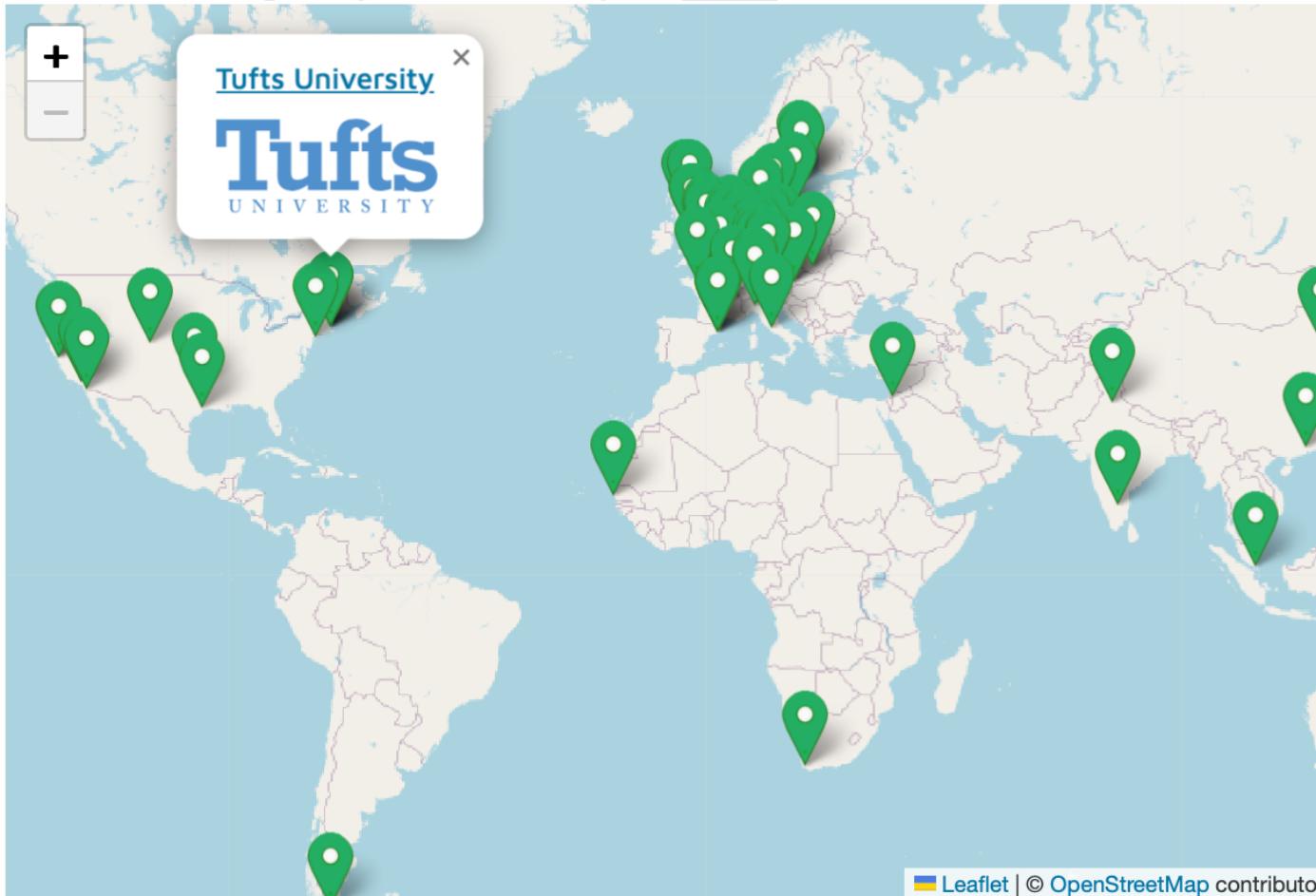
Performance Degradation Avoidance

- The default behavior of nextflow is to submit jobs as quickly as possible, which can be disruptive.
- We can incorporate specific limitations in institutional profiles to control job submission.

Visibility and Recognition

- An effective way for organizations to showcase their expertise, capabilities, and contributions to the bioinformatics community.

HPC centers with institutional profiles



Step-by-step guide to writing an institutional profile

Walkthrough on what you need to set up an nf-core institutional profile

https://nf-co.re/docs/tutorials/use_nf-core_pipelines/config_institutional_profile

tufts profile →



SINGULARITYCE

```
params {  
    max_memory = 120.GB  
    max_cpus = 72  
    max_time = 168.h  
    igenomes_base = '/cluster/tufts/biocontainers/datasets/igenomes/'  
}  
  
process {  
    executor = 'slurm'  
    clusterOptions = '-N 1 -n 1 -p batch'  
}  
  
executor {  
    queueSize = 16  
    pollInterval = '1 min'  
    queueStatInterval = '5 min'  
    submitRateLimit = '10 sec'  
}  
  
// Set $NXF_SINGULARITY_CACHEDIR in your ~/.bashrc  
// to stop downloading the same image for every run  
singularity {  
    enabled = true  
    autoMounts = true  
}
```

<https://github.com/nf-core/configs/blob/master/conf/tufts.config>

Complexity of nf-core workflows for biologists

```
#!/bin/bash

#SBATCH --time=00-48:00:00
#SBATCH -p batch
#SBATCH -N 1
#SBATCH -n 2 ## This is the parent script used for submitting children slurm jobs, 2 cores are enough
#SBATCH --job-name nf-core
#SBATCH --output=%x-%J-%u.out
#SBATCH --error=%x-%J-%u.err
#SBATCH --mail-type=ALL
#SBATCH --mail-user=XXX@tufts.edu

module load nf-core

export NXF_SINGULARITY_CACHEDIR=/cluster/tufts/biocontainers/nf-core/singularity-images

nextflow run /cluster/tufts/biocontainers/nf-core/pipelines/nf-core-rnaseq/3.14.0/3_14_0/ \
    --input samplesheet.csv --outdir output \
    --fasta ref.fasta --gtf ref.gtf \
    --aligner star_salmon \
    -profile tufts --partition preempt
```

nf-core



OPEN
OnDemand

NOTIFICATIONS and SUPPORT

- Request Assistance: Email tts-rese
- Upload/Download: Via OnDemand
- Acknowledging Usage of NSF CC*:
- Acknowledging Usage of Tufts HP Cluster

OPEN
OnDemand

OnDemand provides an integrated

Apps

-  AlphaFold
-  CellProfiler
-  CellProfiler GPU(beta)
-  CellProfiler-Analyst
-  FastQC
-  Jupyter Bioinfo
-  QualiMap
-  RELiON
-  RStudio for bioinformatics
-  RStudio for scRNA-Seq
-  Shinyngs

ing Tufts HPC Cluster.
which will be increased in the future.
Cluster - [Click Here](#)

for all of your HPC resources.

**User-friendly
interactive
bioinformatics apps**

<https://ondemand.pax.tufts.edu/>

Bioinformatics Apps

- Apps
- AlphaFold
- CellProfiler
- CellProfiler GPU(beta)
- FastQC
- Jupyter Bioinfo
- QualiMap
- RELION
- RStudio for bioinformatics
- RStudio for scRNA-Seq
- Shinyngs
- nf-core pipelines
- ampliseq
- atacseq
- bacass
- bamtofastq
- chipseq
- detaxizer
- differentialabundance
- eager
- fetchngs**
- funcscan
- hic
- mag
- metatdenovo
- methylseq
- nanoseq
- nanostring

fetchngs

This app will launch the [fetchngs](#) pipeline developed by nf-core community.

Number of hours

2

Which nextflow executor to use?

slurm

With slurm, tasks will be distributed to different nodes, local means all tasks will run on a single node.

Partition

batch

NOTE: Please do not choose specific lab partitions if you are not a member of that lab.

Reservation for class, training, workshop

Default

If you don't know about specific reservation, select default.

Version

1.12.0

Working Directory

/cluster/tufts/workshop/yzhang85/fetchngs

Select your project directory; defaults to \$HOME

outdir

fetchngsOut

The output directory where the results will be saved. You have to use absolute paths to storage on Cloud infrastructure.

input

samplesheet.csv

File containing SRA/ENA/GEO/DDBJ identifiers one per line to download their associated metadata and FastQ files.

ena_metadata_fields

Comma-separated list of ENA metadata fields to fetch before downloading data.

Which nextflow executor to use?

local

With slurm, tasks will be distributed to different nodes, local means all tasks will run on a single node.

Partition

batch

NOTE: Please do not choose specific lab partitions if you are not a member of that lab.

Cores

24

Number of cores (up to 128) for a shared job. Non-shared jobs will have exclusive nodes and be charged at 128 cores per node requested

Amount of memory

64GB

Reservation for class, training, workshop

Default

If you don't know about specific reservation, select default.

2578439	OnDemand/+	batch	default	2	COMPLETED	0:0
2578451	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578452	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578453	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578454	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578455	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578456	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578457	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578458	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578459	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578460	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578461	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2578462	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578477	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578630	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578692	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578693	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578710	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578711	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2578712	nf-NFCORE+	batch	default	2	COMPLETED	0:0
2578786	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2578909	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2579166	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2579310	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2580524	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2580697	nf-NFCORE+	batch	default	1	COMPLETED	0:0
2580704	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583415	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583416	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583417	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583418	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583421	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583422	nf-NFCORE+	batch	default	6	COMPLETED	0:0
2583439	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2583453	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2583470	nf-NFCORE+	batch	default	12	COMPLETED	0:0
2583483	nf-NFCORE+	batch	default	12	COMPLETED	0:0

←master job

Research Technology at Tufts Technology Services (TTS) presents...

RNA-SEQ WORKSHOP SERIES

Master bioinformatics workflows and analytical techniques to perform RNA sequencing (RNA-seq) analyses, from working with pre-processed data to creating publication-quality figures and results!

ONLINE event!



WEDS
APRIL 3
1PM–3PM

Introduction to **nextflow**
nf-core  and
bioinformatics pipelines



THUR
APRIL 4
1PM–3PM

Running **nf-core**  RNA-seq pipelines



THUR
APRIL 11
1PM–3PM

RNA-seq downstream analysis
with **nf-core**  differentialabundance pipelines

Expert-led office hours on April 5 & April 12!

A unique opportunity for personalized
guidance & error troubleshooting!

Register at:
[GO.TUFTS.EDU/
WORKSHOPS](http://GO.TUFTS.EDU/WORKSHOPS)



Tufts
Bioinformatics
Workshops

Nextflow
Summit
2024 Boston



Download nf-core pipelines to HPC

```
module load nf-core singularity  
export NXF_SINGULARITY_CACHEDIR=/XX/singularity-images  
nf-core download $pipeline_name -r $version --outdir $version -d -s singularity -u amend -x none
```

```
[yzhang85@d1cmp004 pipelines]$ pwd  
/cluster/tufts/biocontainers/nf-core/pipelines  
[yzhang85@d1cmp004 pipelines]$ tree -L 2  
.  
├── nf-core-ampliseq  
│   ├── 2.10.0  
│   ├── 2.8.0  
│   └── 2.9.0  
├── nf-core-atacseq  
│   └── 2.1.2  
├── nf-core-bacass  
│   ├── 2.2.0  
│   └── 2.3.1  
├── nf-core-bamtofastq  
│   └── 2.1.1  
├── nf-core-chipseq  
│   └── 2.0.0  
├── nf-core-detaxizer  
│   └── 1.0.0  
├── nf-core-differentialabundance  
│   ├── 1.4.0  
│   └── 1.5.0  
├── nf-core-eager  
│   └── 2.5.1  
├── nf-core-fetchngs  
│   ├── 1.11.0  
│   └── 1.12.0  
└── nf-core-funcscan  
    └── 1.1.4
```

Convert nf-core pipelines to OOD apps

```
git clone https://github.com/TuftsRT/nf-core_open-on-demand.git
```

```
cd nf-core_open-on-demand
```

```
## Edit nf-core_template according to your cluster's set up.
```

```
bash nfcore2ood.sh -i /cluster/tufts/biocontainers/nf-core/pipelines -o ood_apps
```

What's next?

- Integrate new features of OOD v3.1.
 - The workflow was developed for our center's OOD v1.8.
- Implement functionality to easily switch forms depending on the versions of pipelines.
- Integrate the OOD dynamic form widgets to hide some parameters based on certain selections.
 - Some pipelines have an excessive number of parameters.
 - Many pipelines have different routes.
 - Hiding parameters specific to certain routes will make it clearer for users.

Thank you!

PEARC24
Conference Committee

Tufts Research Technology

Patrick Florance

Delilah Maloney

Tom Phimmassen

Christina Divoll

Lillian Kelting

Kyle Monahan

Carolyn Talmadge



Fangping Mu (Univ. of Pittsburgh)