



Nextflow Introduction

JeBiF@JOBIM June/2024

Workflows

Using computers to collect, store, analyze, and disseminate data and information



Large files

> 100 GB for one raw
human genome...



Many languages

Bash, Python, R, PERL...



Complex interactions

Networks of software
and their dependencies...



Reproducibility

Hidden reproducibility issues are like an iceberg



“

First, we tried to re-run the analysis with the code and data provided by the authors.

Second, we reimplemented the whole method in a Python package...

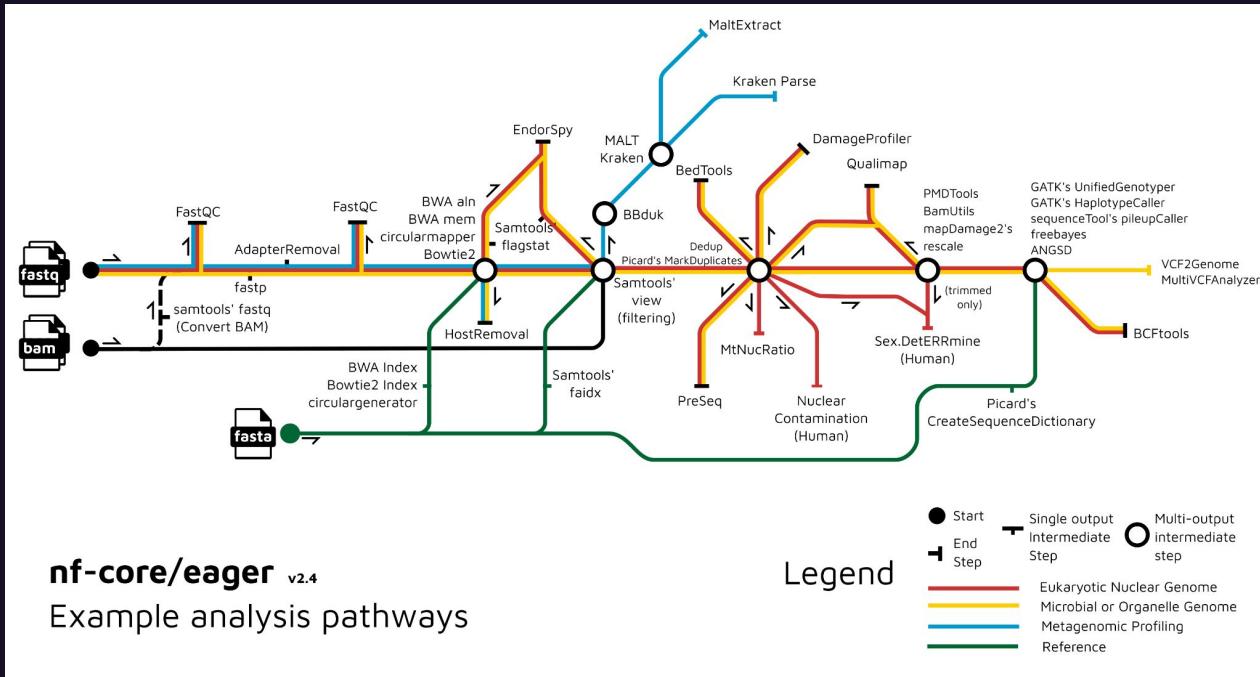
”

Experimenting with reproducibility:
a case study of robustness in bioinformatics
Kim et al., GigaScience (2018).
<https://doi.org/10.1093/gigascience/giy077>



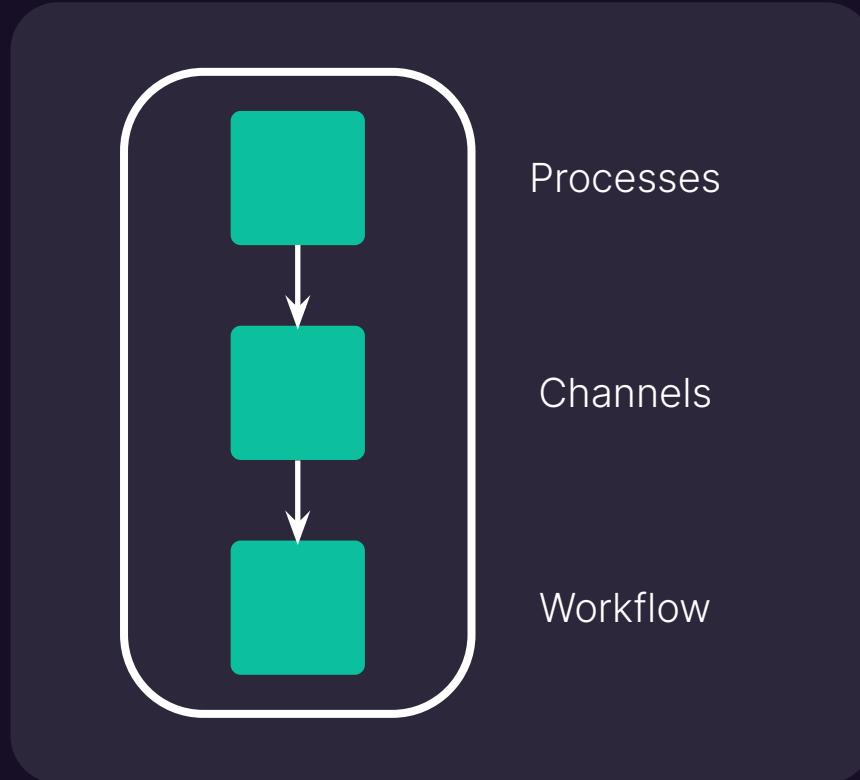
Nextflow

Managing modern workflows is complicated



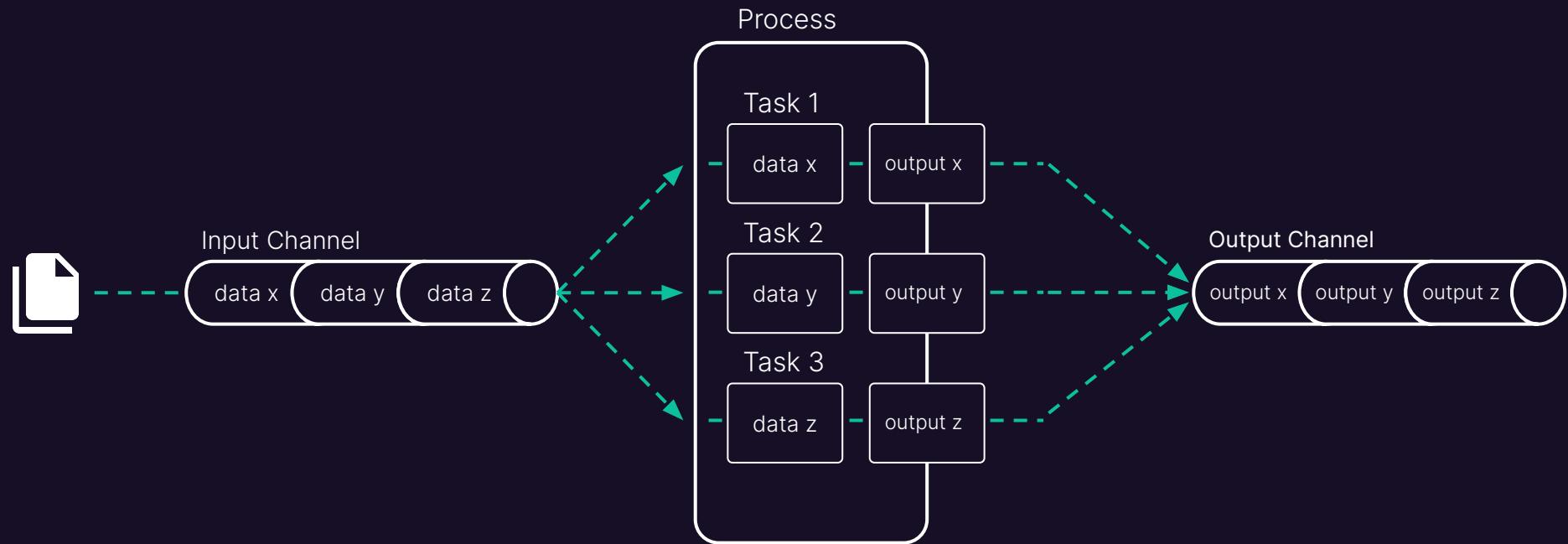
Nextflow

A reactive workflow framework and a programming DSL



Nextflow

A reactive workflow framework and a programming DSL



Nextflow

A reactive workflow framework and a programming DSL

```
process fastqc {
    input:
    path input

    output:
    path "*_fastqc.{zip,html}"

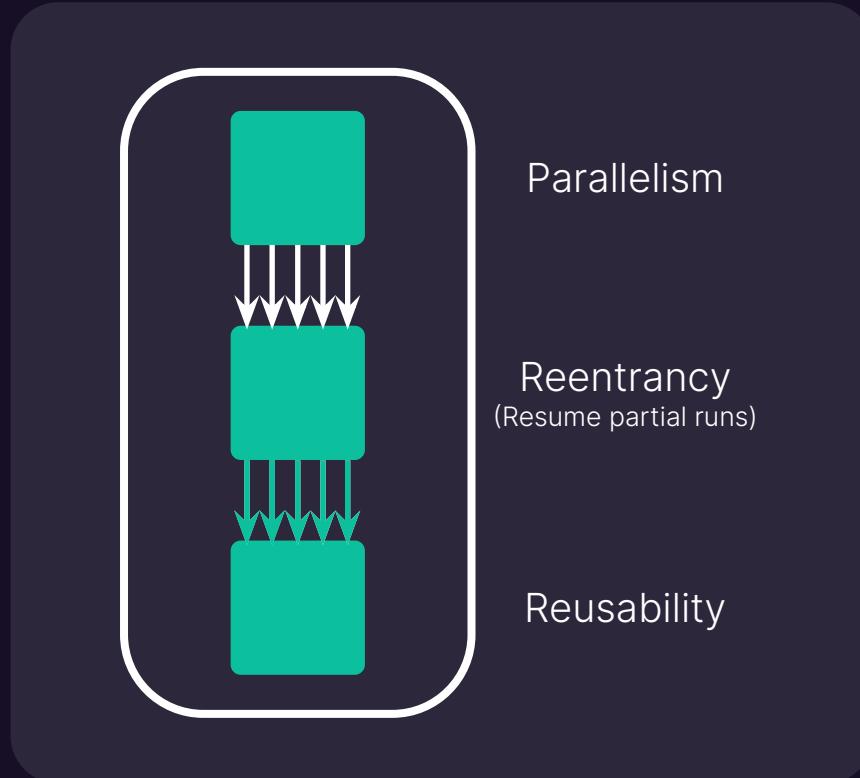
    script:
    .....
    fastqc -q $input
    .....
}

workflow {
    Channel.fromPath("*.fastq.gz") | fastqc
}
```



Nextflow

A reactive workflow framework and a programming DSL



Nextflow

Nextflow is a language, a runtime, and a community



Write code
in any language



Orchestrate tasks with
dataflow programming



Define software
dependencies via containers



Built-in version control
with Git



Task orchestration
and execution



Nextflow

Nextflow is a language, a runtime, and a community



Reproducible

Integration with code management tools, with versioned releases.



Portable

Docker, Singularity, Conda, works with most compute environments.



Scalable

5 samples on your laptop, 5k on an HPC or 5 million in the cloud.



Nextflow plugins

nf-validation: a Nextflow plugin to natively handle schema files



Parameter validation

Functions to validate pipeline parameters and print CLI help text



Samplesheet Validation

Validate input sample sheet - formatting and contents



Samplesheet channels

fromSamplesheet()
channel factor,
including meta maps



nextflow-io/nf-validation



<https://github.com/nextflow-io/nf-validation>



nf-core

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

8k+

Slack
users

2k+

GitHub
contributors

120+

GitHub
repositories

40k

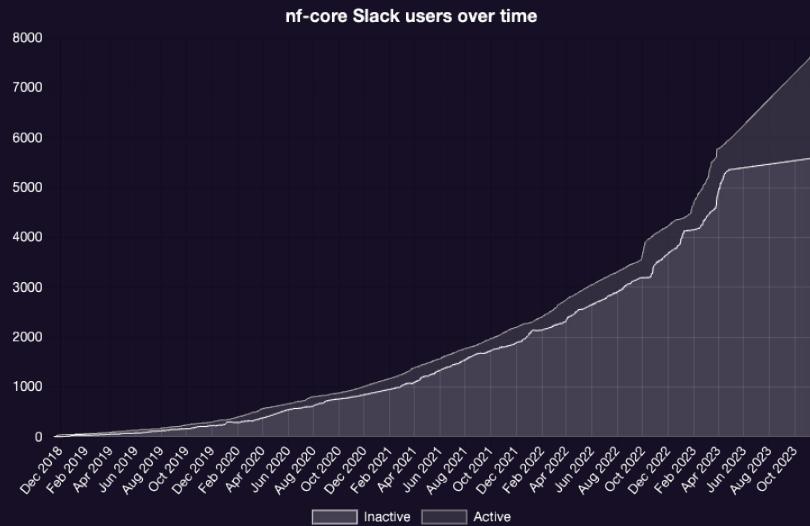
GitHub
commits

16k+

Pull
requests

7k+

GitHub
issues



nf-core principles

Principles that guide the nf-core community



Cooperation

Develop with the community



Standards

Use a common template



Collaboration

No duplicate pipelines within nf-core



Helper Tools

Tools built for everyone



Compatibility

Tools work for any Nextflow pipeline



Components

Collaborate on component-level



nf-core components

Pick and choose which component you need



Pipelines

>95 pipelines and a base template



Subworkflows

>55 subworkflows



Modules

>1150 modules



Linting

Choose conventions to test for consistency



Schema

Validation, channels and user interface



Tooling

Development and deployment



nf-core/tools

Command line tools to help you build your pipeline with ease



Pipelines

Create from template,
sync to get updates



Subworkflows

Create, install and
update



Modules

Create, install, update,
patch, test



Schema

Build your pipeline
schema with a GUI



Linting

Test nf-core standards
and best practices



Download

Fetch with singularity
images for offline use



Participate

Seminars, training, hackathons, and more

- Bytesize seminars
- Training sessions
- Hackathons
- Social media
- Blogs
- Community Forum and Slack
- Documentation
- Mentorships





Maxime U Garcia, Ph.D.

Bioinformatics Engineer
maxime.garcia@seqera.io
Barcelona | Stockholm

Thank you