



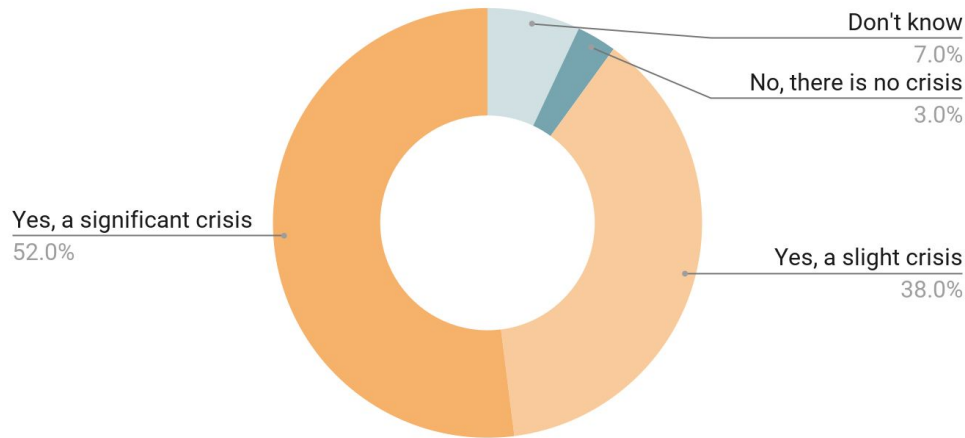
Reproducible NGS data analysis with Nextflow and nf-core

Gisela Gabernet
Team leader RDDS, QBiC

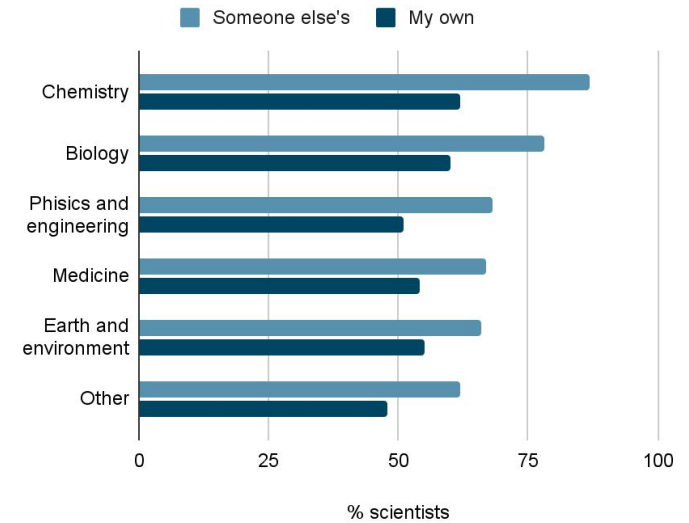


Reproducibility in Science

Is there a reproducibility crisis?

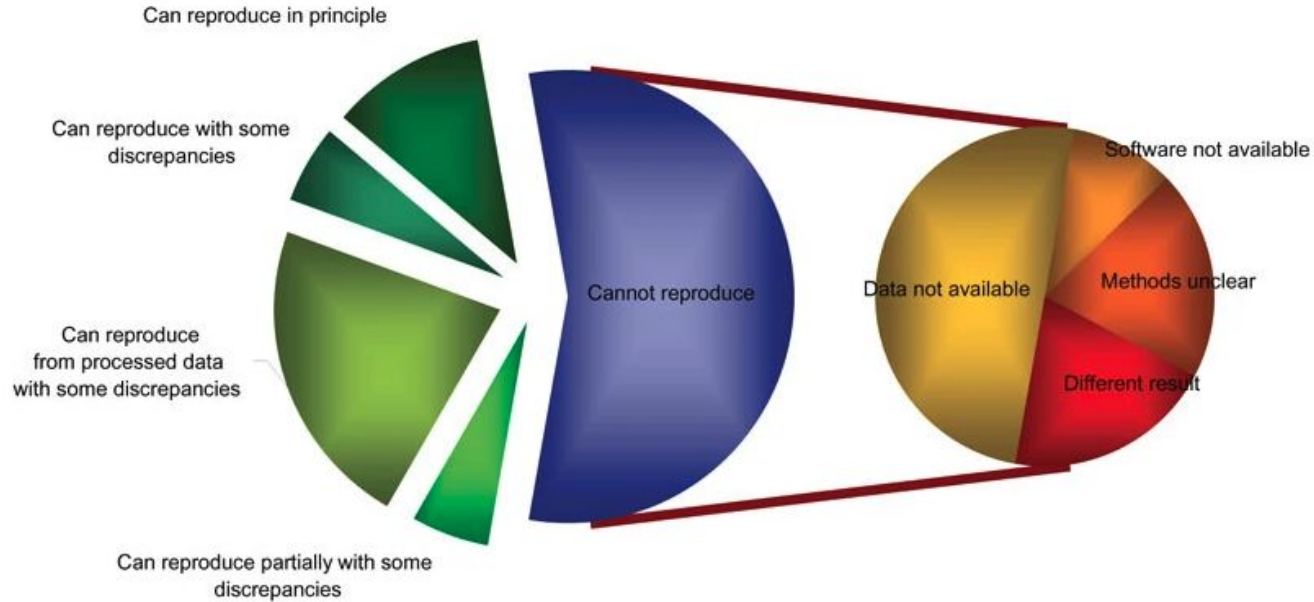


Have you failed to reproduce an experiment?





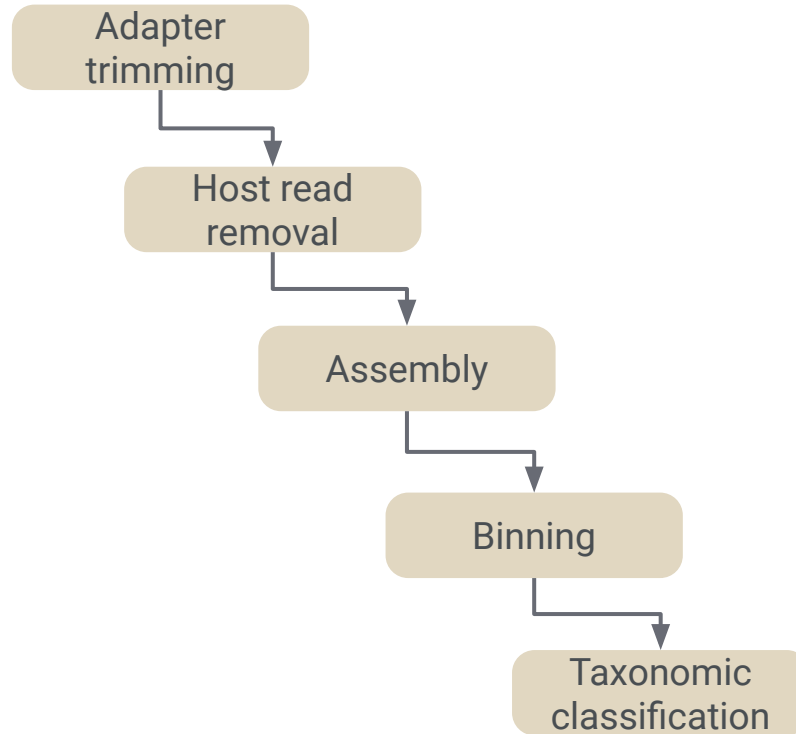
Reproducibility in Science





Reproducibility in NGS data analysis

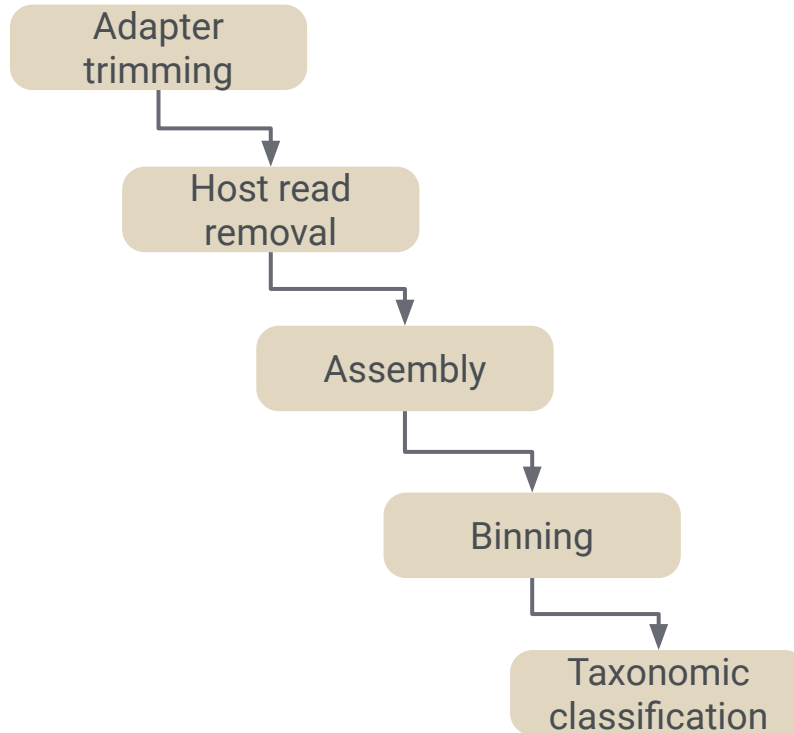
Metagenomics analysis





Reproducibility in NGS data analysis

Metagenomics analysis



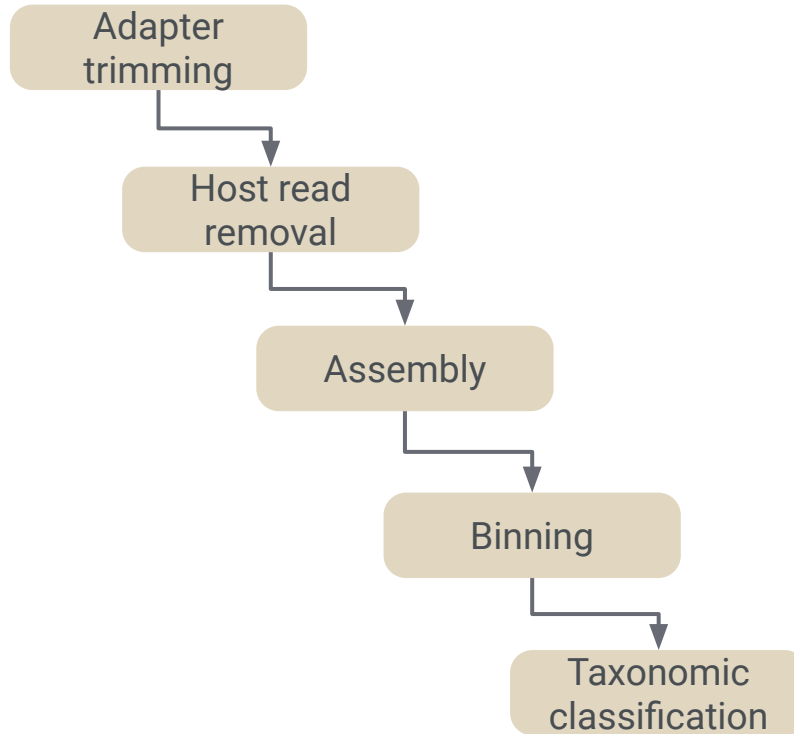
What affects reproducibility?

→ Tools & versions



Reproducibility in NGS data analysis

Metagenomics analysis

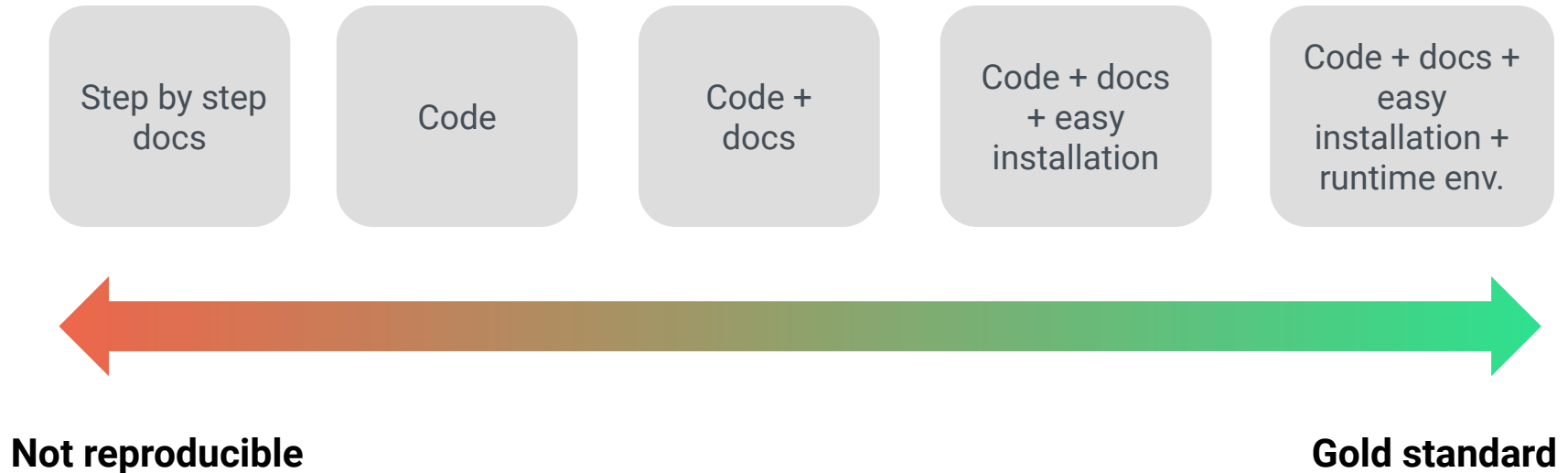


What affects reproducibility?

- Tools & versions
- Tool parameters
- Dependency versions
- Runtime environment



Reproducibility in NGS data analysis





Reproducible pipelines / workflows

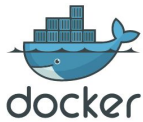
- **Code comprises all steps** – processes call all necessary analysis tools
- **Specialized programming languages** – e.g. Common Workflow Language, Snakemake, Nextflow
- **Tools and dependencies are packaged with the pipeline** – container(s) ships all necessary tool dependencies, fixing their versions and the runtime environment





nextflow workflow language

Reproducibility – supports multiple container engines



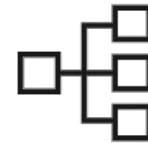
Portable – supports a variety of computing infrastructures



Continuous checkpoints – allow pipeline resuming



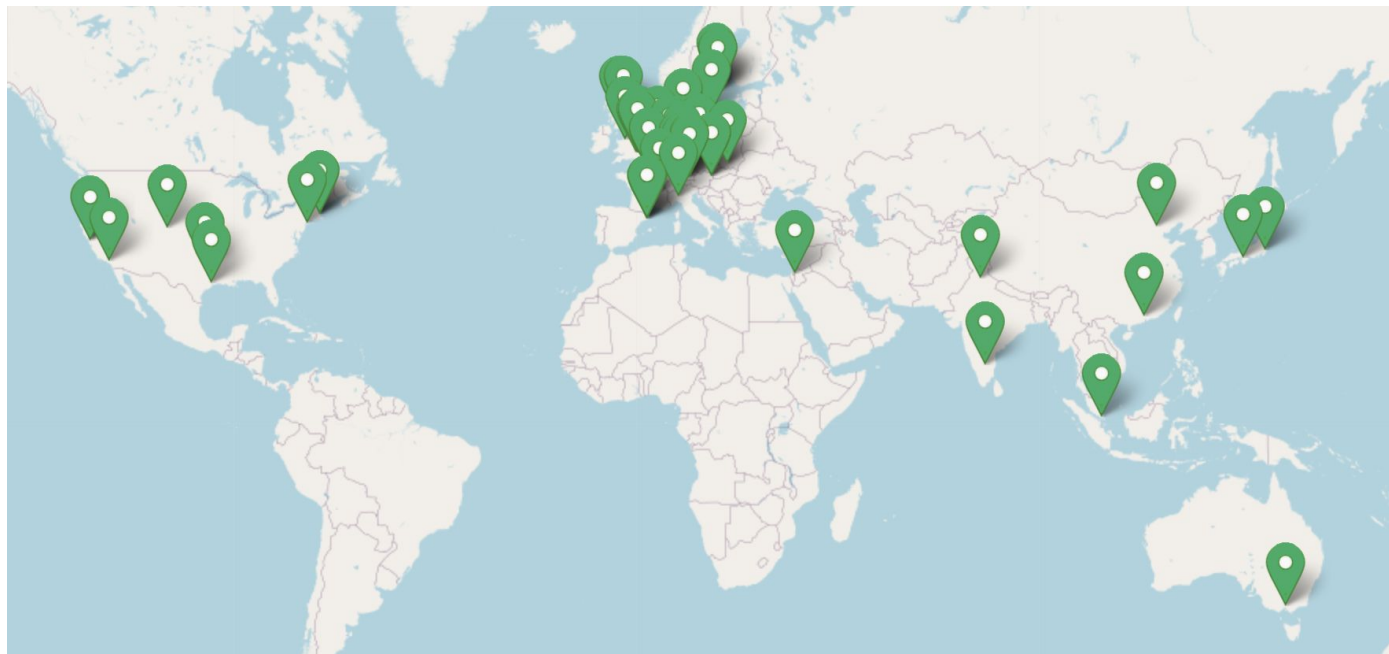
Implicit parallelism – processes samples in parallel





nf-core

A community effort to collect a curated set of Nextflow pipelines





nf-core

A community effort to collect a curated set of Nextflow pipelines



Stable releases



Packaged software



Documentation



Portable



Continuous integration

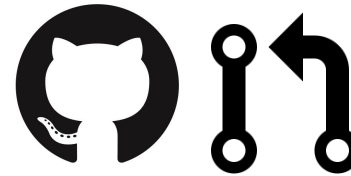
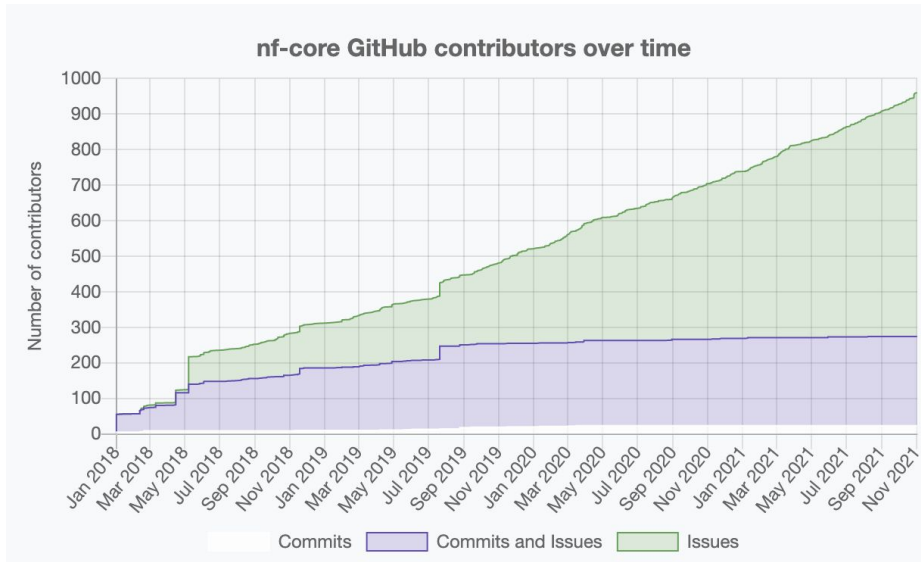


Cloud ready



nf-core

Best practice pipelines, peer reviewed by the community



Collaboration over
GitHub



Code review



Pipelines with active QBIC contributors

Metagenomics / microbiology

nf-core / mag
metagenomics
assembly and binning

nf-core / ampliseq
16S amplicon

nf-core / bacass
bacterial assembly

Immunoinformatics

**nf-core /
epitopeprediction**
epitope prediction

nf-core / hlatyping
HLA typing

nf-core / mhcquant
MHC peptides MS data
analysis

nf-core / bcellmagic
TCR and BCR analysis

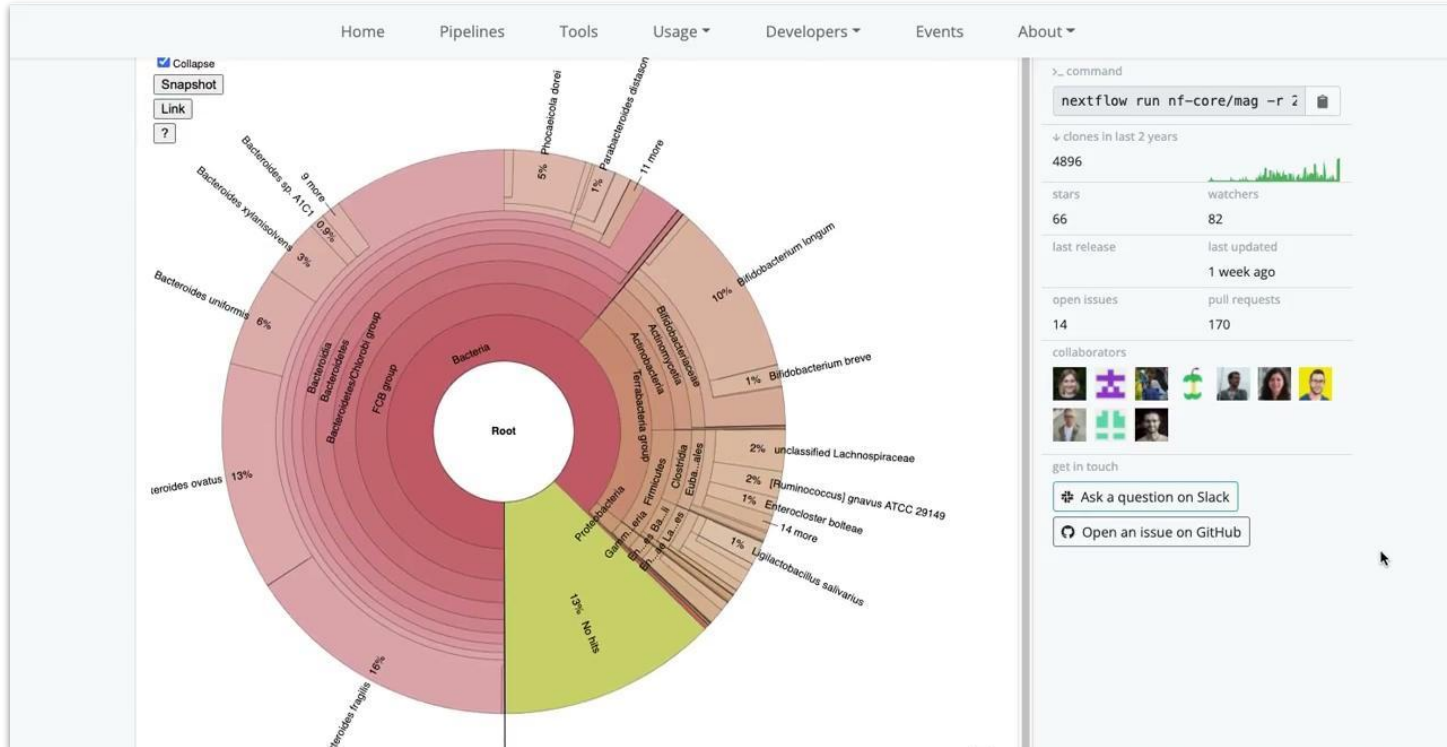
Genomics / transcriptomics

nf-core / sarek
somatic & germline
variant calling

nf-core / scrnaseq
single-cell RNAseq
analysis



nf-core





nf-core

Prerequisites:

- Java
- Nextflow
- Container engine
(Singularity / Docker)
- Unix system
(MacOS / Linux)


```

[-] process > NFCORE_MAG:MAG:NANOPLT_RAW -
[-] process > NFCORE_MAG:MAG:PORECHOP -
[-] process > NFCORE_MAG:MAG:NANOLYSE -
[-] process > NFCORE_MAG:MAG:FILTLONG -
[-] process > NFCORE_MAG:MAG:NANOPLT_FILTERED -
[92/a5095a] process > NFCORE_MAG:MAG:CENTRIFUGE_DB_PREPARATION [100%] 1 of 1 ✓
[79/e6a7e6] process > NFCORE_MAG:MAG:CENTRIFUGE (test_minigut-m... [100%] 2 of 2 ✓
[24/7aa435] process > NFCORE_MAG:MAG:KRAKEN2_DB_PREPARATION [100%] 1 of 1 ✓
[76/d8debc] process > NFCORE_MAG:MAG:KRAKEN2 (test_minigut-mini... [100%] 2 of 2 ✓
[2a/239fc4] process > NFCORE_MAG:MAG:MEGAHIT (test_minigut) [100%] 2 of 2 ✓
[1f/70e216] process > NFCORE_MAG:MAG:SPADES (test_minigut) [100%] 2 of 2 ✓
[-] process > NFCORE_MAG:MAG:SPADESHYBRID -
[39/2faf4f] process > NFCORE_MAG:MAG:QUAST (SPAdes-test_minigut) [100%] 4 of 4 ✓
[be/5c1a09] process > NFCORE_MAG:MAG:METABAT2_BINNING:BOWTIE2_A... [100%] 4 of 4 ✓
[ad/076c73] process > NFCORE_MAG:MAG:METABAT2_BINNING:BOWTIE2_A... [100%] 8 of 8 ✓
[43/ce06ce] process > NFCORE_MAG:MAG:METABAT2_BINNING:METABAT2 ... [100%] 4 of 4 ✓
[a3/25f8ed] process > NFCORE_MAG:MAG:METABAT2_BINNING:MAG_DEPTH... [100%] 4 of 4 ✓
[31/8b7d26] process > NFCORE_MAG:MAG:METABAT2_BINNING:MAG_DEPTH... [100%] 4 of 4 ✓
[ec/8ab7ef] process > NFCORE_MAG:MAG:METABAT2_BINNING:MAG_DEPTH... [100%] 1 of 1 ✓
[95/8bd6fb] process > NFCORE_MAG:MAG:BUSCO_QC:BUSCO_DB_PREPARAT... [100%] 1 of 1 ✓
[43/e86378] process > NFCORE_MAG:MAG:BUSCO_QC:BUSCO (SPAdes-tes... [100%] 12 of 12 ✓
[0c/674a6e] process > NFCORE_MAG:MAG:BUSCO_QC:BUSCO_PLOT (MEGAH... [100%] 4 of 4 ✓
[6c/504576] process > NFCORE_MAG:MAG:BUSCO_QC:BUSCO_SUMMARY [100%] 1 of 1 ✓
[37/adf4d5] process > NFCORE_MAG:MAG:QUAST_BINS (MEGAHIT-test_m... [100%] 4 of 4 ✓
[6a/2164f7] process > NFCORE_MAG:MAG:QUAST_BINS_SUMMARY [100%] 1 of 1 ✓
[-] process > NFCORE_MAG:MAG:CAT -
[35/932dac] process > NFCORE_MAG:MAG:BIN_SUMMARY (1) [100%] 1 of 1 ✓
[4b/3f61b0] process > NFCORE_MAG:MAG:GET_SOFTWARE_VERSIONS [100%] 1 of 1 ✓
[51/6b9d11] process > NFCORE_MAG:MAG:MULTIQC [ 0%] 0 of 1

```



nf-core

🏠 **nextflow tower** Docs 

QBiC / cfc ▾

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

⏪

↶ nf-core/mag
🔑 mad_newton
🕒 Duration: -
📅 Started: 2021-11-02 13:38:33 🗑

🕒 qbic-pipelines/cellranger
👤 qeajl01
🔑 backstabbing_swanson
🕒 Duration: 7 h 57 m 12 s
📅 Started: 2021-10-26 13:36:14 🗑

🕒 snpEff.nf
👤 qealk01
🔑 kickass_swanson
🕒 Duration: 3 h 30 m 42 s
📅 Started: 2021-10-26 13:14:42 🗑

🕒 snpEff.nf

qbic-pipelines/cellranger

backstabbing_swanson

✔

Command line Parameters Configuration

```
nextflow run qbic-pipelines/cellranger
-r 1.0.1
-profile cfc
--input /sfs/7/workspace/ws/qeajl01-QUBFL_scrna-0/data/sample_sheet_cellranger.tsv
--genome mm10
--outdir /sfs/7/workspace/ws/qeajl01-QUBFL_scrna-0/results/cellranger_1
--resume
```

📄

General

Status



Data analysis pipelines at QBiC

- Emphasis on reproducibility
- Scalable and parallelizable for large datasets
- Benefit from nf-core community development and peer review



Acknowledgements



Quantitative Biology Center
RDDS team

