

Reproducible bioinformatics workflows with Nextflow and nf-core

Phil Ewels

phil.ewels@scilifelab.se



NATIONAL
GENOMICS
INFRASTRUCTURE

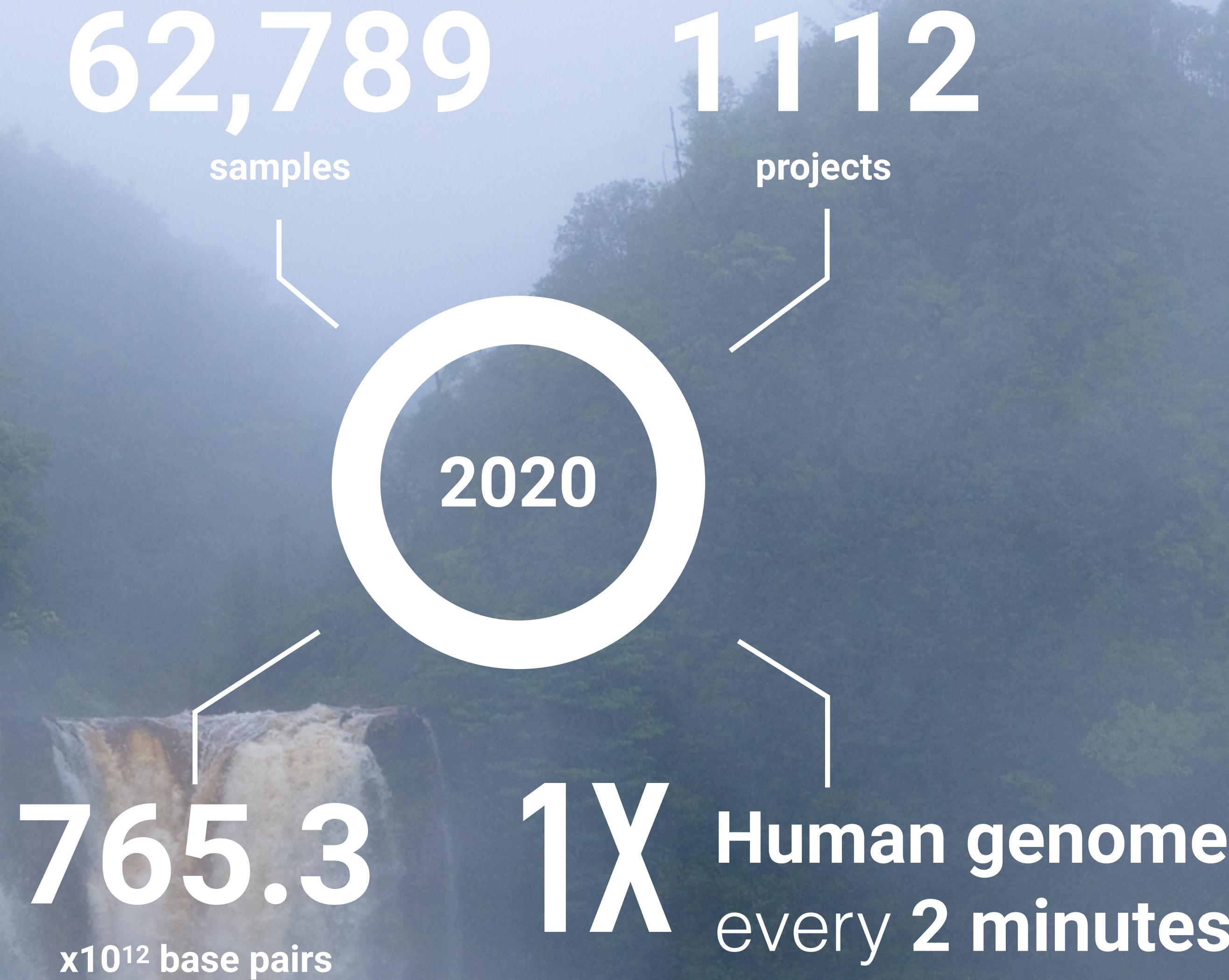


SciLifeLab

<https://scilifelab.se>
<https://ngisweden.scilifelab.se>



NATIONAL
GENOMICS
INFRASTRUCTURE



The background of the image is a photograph of a vast, misty mountain range with a dense forest of evergreen trees in the foreground. The sky is a clear, pale blue.

nextflow

<https://www.nextflow.io/>

nextflow

```
#!/usr/bin/env nextflow
input = Channel.fromFilePairs(params.reads)

process fastqc {
    input:
        file reads from input

    output:
        file "*_fastqc.{zip,html}" into results

    script:
    """
        fastqc -q $reads
    """

}
```



SGE



Google Cloud



LSF

PBS

<https://www.nextflow.io/>



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



Guidelines



Tools



Pipelines

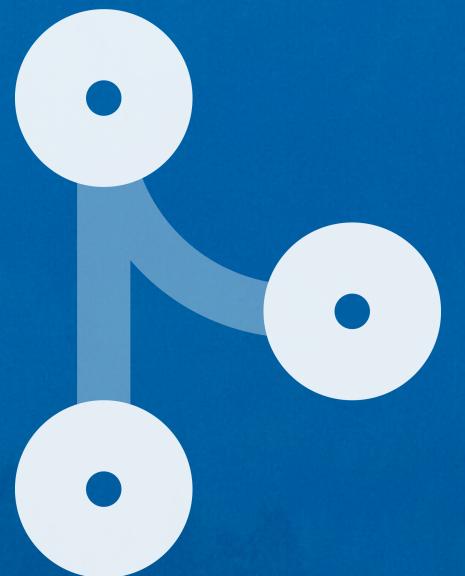
nf-core



Develop with
the community



Start from the
template



Collaborate,
don't duplicate

nf-core



33

RELEASED



15

UNDER DEVELOPMENT



5

ARCHIVED



Available Pipelines

Can you think of another pipeline that would fit in well? [Let us know!](#)

Search keywords

Filter: Released 33 Under development 15 Archived 5

Sort: Last Release Alphabetical Stars

Display:

[nf-core/rnaseq_](#) ✓

[rna](#) [rna-seq](#)

RNA sequencing analysis pipeline using STAR, RSEM, HISAT2 or Salmon with gene/isoform counts and extensive quality control.

Version 3.4

Published 1 month ago

★ 395

[nf-core/chipseq_](#) ✓

[chip](#) [chip-seq](#) [chromatin-immunoprecipitation](#) [macs2](#) [peak-calling](#)

ChIP-seq peak-calling, QC and differential analysis pipeline.

Version 1.2.2

Published 7 months ago

★ 92

[nf-core/sarek](#) ✓

[annotation](#) [cancer](#) [gatk4](#) [genomics](#) [germline](#) [pre-processing](#) [somatic](#)
[target-panels](#) [variant-calling](#) [whole-exome-sequencing](#) [whole-genome-sequencing](#)

Analysis pipeline to detect germline or somatic variants (pre-processing, variant calling and annotation) from WGS / targeted sequencing

Version 2.7.1

Published 5 months ago

★ 125

[nf-core/atacseq_](#) ✓

[atac-seq](#) [chromatin-accessibility](#)

ATAC-seq peak-calling, QC and differential analysis pipeline

Version 1.2.1

Published 1 year ago

★ 91

nf-core



2030

Slack users

337

GitHub organisation
members

934

GitHub contributors

2272

Twitter followers

70

Repositories

6.8K

Pull Requests

32.35K

Commits

3.07K

Issues

nf-core



<https://nf-co.re/community>

nf-core



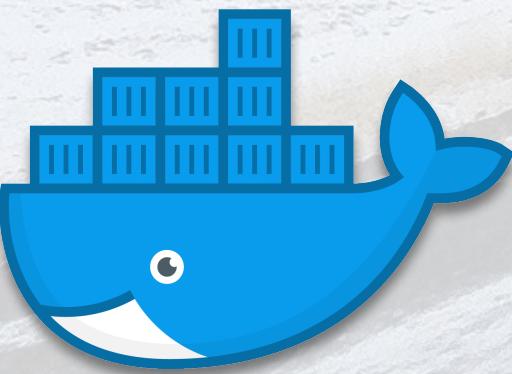
<https://nf-co.re/join>

Requirements

1

nextflow

2



docker®

or



or

CONDA

3

nf-core/tools



Running a pipeline

```
nextflow run nf-core/<pipeline> -r <version>
```

GitHub pipeline name
Clones to `~/.nextflow`

Pipeline release tag
Repository branch

Running a pipeline

```
nextflow run nf-core/<pipeline> -r <version>  
--input my_samples.csv --genome GRCh38
```

Parsed as params.genome = 'GRCh38'

Can be supplied in a file (config / YAML / JSON)

Running a pipeline

```
nextflow run nf-core/<pipeline> -r <version>  
--input my_samples.csv --genome GRCh38  
-profile docker
```

Specifies a configuration profile

Common setups bundled with pipelines, also shared institutional profiles available

Introduction  Results Usage docs **Parameter docs** Output docs Releases & Statistics  3.4 

Parameters

> Input/output options

Define where the pipeline should find input data and save output data.

 <code>--input</code>	Path to comma-separated file containing information about the samples in the experiment.	
 <code>--outdir</code>	Path to the output directory where the results will be saved.	default: './results'
 <code>--email</code>	Email address for completion summary.	
 <code>--multiqc_title</code>	MultiQC report title. Printed as page header, used for filename if not otherwise specified.	
 <code>--save_merged_fastq</code>	Save FastQ files after merging re-sequenced libraries in the results directory.	

On this page

Parameters

- > Input/output options
- UMI options
- Read filtering options
- Reference genome options
- Read trimming options
- Alignment options
- Process skipping options

 Show all help  Show hidden params

 [Back to top](#)

UMI options

Options for processing reads with unique molecular identifiers

Running a pipeline

<https://nf-co.re/launch>

Read trimming options

`--save_trimmed` True False

Save the trimmed FastQ files in the results directory.

Alignment options

Options to adjust parameters and filtering criteria for read alignments.

Aligner `--aligner` star_salmon

Specifies the alignment algorithm to use - available options are 'star_salmon', 'star_rsem' and 'hisat2'.

Pseudo aligner `--pseudo_aligner` [Select an option]

Specifies the pseudo aligner to use - available options are 'salmon'. Runs in addition to '--aligner'.

BAM CSI index `--bam_csi_index` True False

Create a CSI index for BAM files instead of the traditional BAI index. This will be required for genomes with larger chromosome sizes.

STAR ignore sjdbgtf `--star_ignore_sjdbgtf` True False

When using pre-built STAR indices do not re-extract and use splice junctions from the GTF file.

Salmon quant libtype `--salmon_quant_libtype`

On this page

Nextflow command-line flags

> Input/output options

UMI options

Read filtering options

Reference genome options

Read trimming options

Alignment options

Process skipping options

Show hidden params

[Back to top](#)

Running a pipeline

<https://nf-co.re/launch>

```
nf-core launch --id 1637063024_d92e0f1632c2
```

```
nextflow run nf-core/rnaseq -params-file nf-params.json
```

Running in the cloud

nextflow



Google Cloud



Microsoft
Azure

Running in the cloud

<https://github.com/nf-core/viralrecon>

 nf-core / **viralrecon** Public

Watch ▾ 16 Star 49 Fork 42

Code Issues 9 Pull requests 1 Discussions Actions Security Insights Settings

 **nf-core/viralrecon v2.1 - Lead Mink nf-core AWS full size tests #3** ...

Triggered via release 5 months ago	Status	Total duration	Artifacts
 drpatelh published 2.1	Success	2m 23s	—

Summary

Jobs

-  Run AWS full tests (illumina)
-  Run AWS full tests (nanopore)

awsfulltest.yml
on: release

Matrix: Run AWS full tests

 **2 jobs completed**

Show all jobs

Running in the cloud

<https://nf-co.re/viralrecon>

aws nf-core-awsmegatests / viralrecon / [results-2ebae61442598302c64916bd5127cf23c8ab5611](#) / platform_illumina

[Copy Bucket S3 URL](#)

Name	Last Modified	Size
..		
assembly/		
fastp/		
fastqc/		
kraken2/		
multiqc/		
pipeline_info/		
variants/		

→ command
`nextflow run nf-core/viralrecon`

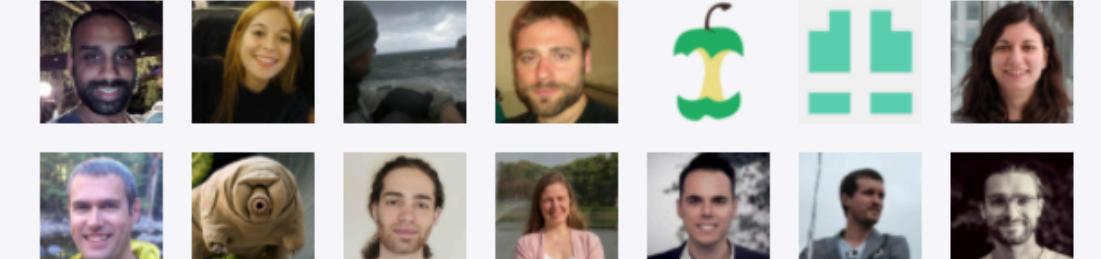
↓ clones in last 2 years
13054

stars 49 watchers 16

last release last updated
1 month ago

open issues pull requests
9 168

collaborators



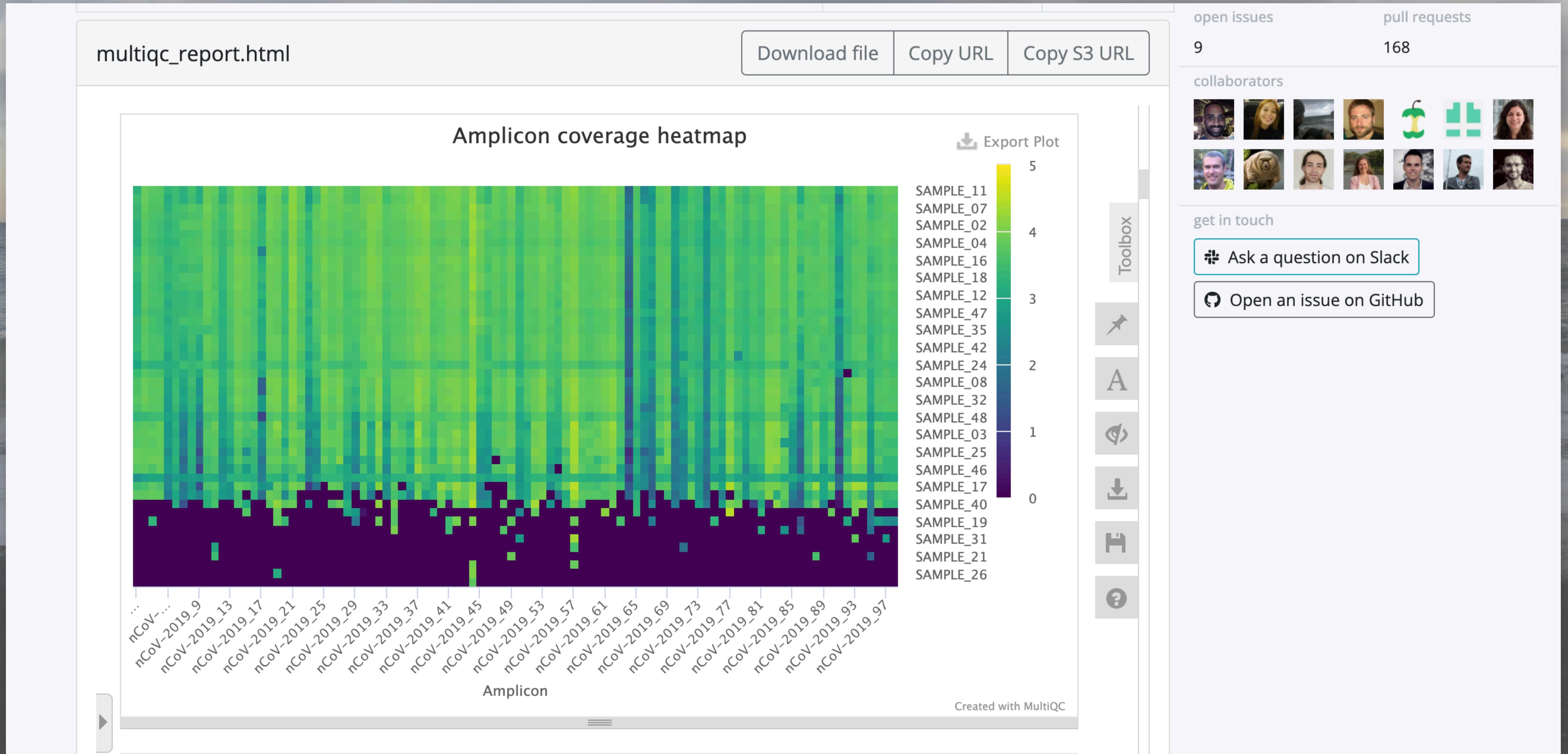
get in touch

[Ask a question on Slack](#)

[Open an issue on GitHub](#)

Running in the cloud

<https://nf-co.re/viralrecon>



Phil Ewels

<https://phil.ewels.co.uk>

 phil.ewels@scilifelab.se

 @ewels

 @tallphil



<https://nf-co.re>



<https://www.nextflow.io>



<https://tower.nf>

Ewels, P.A., Peltzer, A., *et al.*
Nat Biotechnol **38**, 276–278 (2020).
<https://doi.org/10.1038/s41587-020-0439-x>



<https://ngisweden.se>

<https://scilifelab.se>



<https://multiqc.info>



MegaQC <https://megaqc.info>



<https://sra-explorer.info/>