

NGS workflows with Nextflow

NGS analysis for gene regulation and epigenomics

Physalia 2021

What is Nextflow

SCALABILITY

PARALLELISATION

- Nextflow allows to write scalable and highly parallelizable scientific workflows.

What is Nextflow

- It uses software containers for reproducibility and portability.

REPRODUCIBILITY

PORTABILITY ₃

What is Nextflow

- Pipelines can be written in the most common scripting languages.

UNIVERSAL

What is Nextflow

- Continuous checkpoints are automatically tracked, and workflows can be resumed seamlessly.

PROGRESS TRACKING

What is Nextflow

SCALABILITY

PARALLELISATION

REPRODUCIBILITY

PROGRESS TRACKING

UNIVERSAL PORTABILITY

6

Why would I use Nextflow

- For all the reasons mentioned just before
- In plain words, for example:
 - To split fastq/bam/... files in smaller files
 - To map hundreds of ChIP-seq datasets in parallel, in 1 keystroke
 - To work efficiently in an HPC (High Performance Cluster) environment (e.g. AWS, Slurm, ...)
 - To reproduce a published analysis without having to wonder what parameter was used here or there
 - ...

How it works

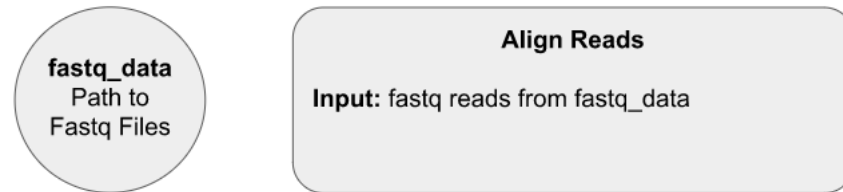
Align Reads

How it works

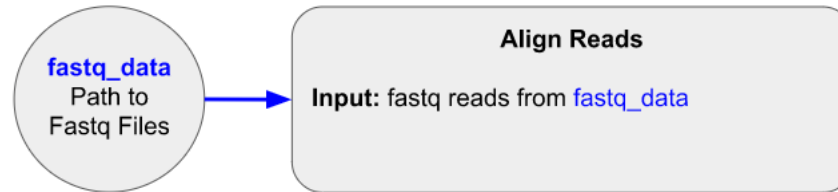
Align Reads

Input: fastq reads from fastq_data

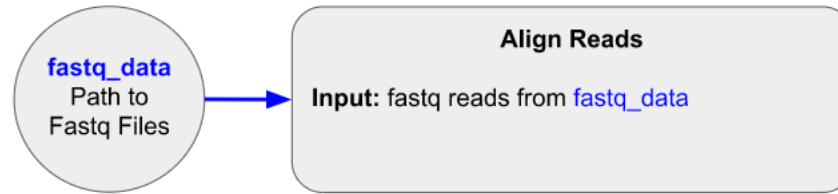
How it works



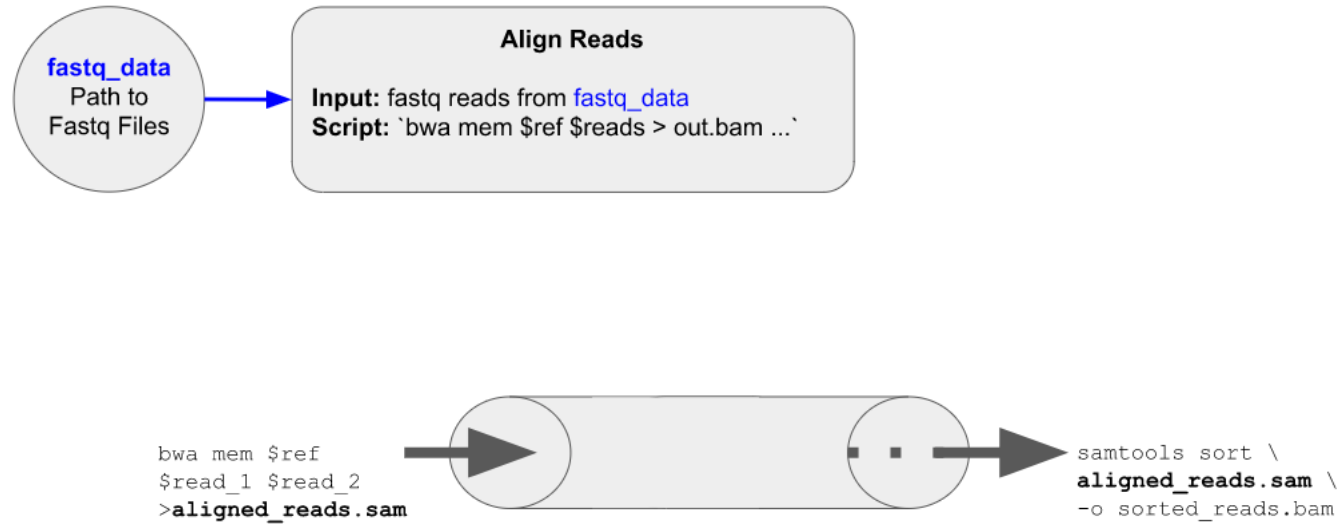
How it works



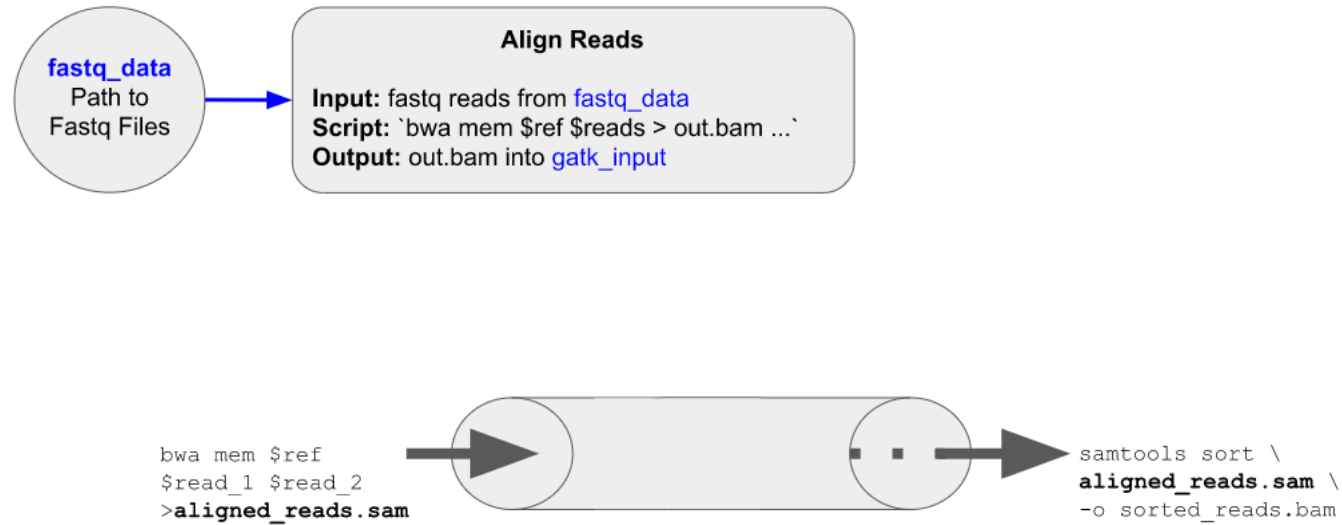
How it works



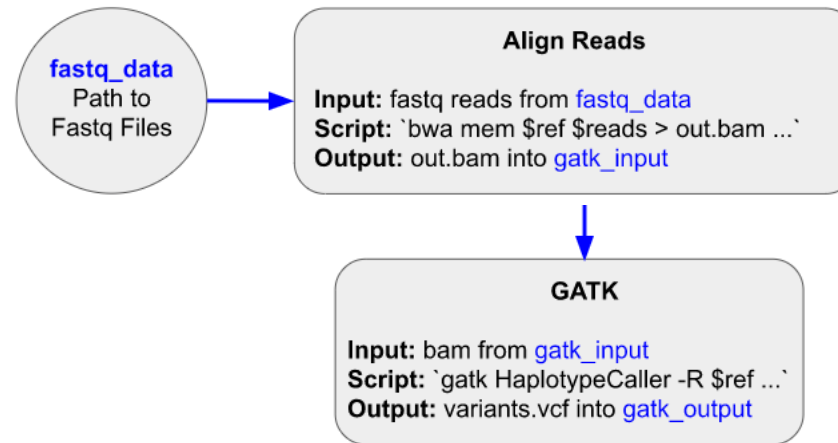
How it works



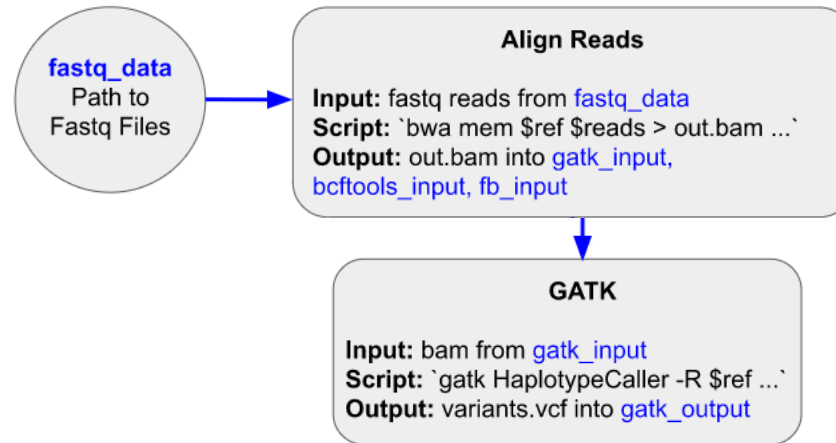
How it works



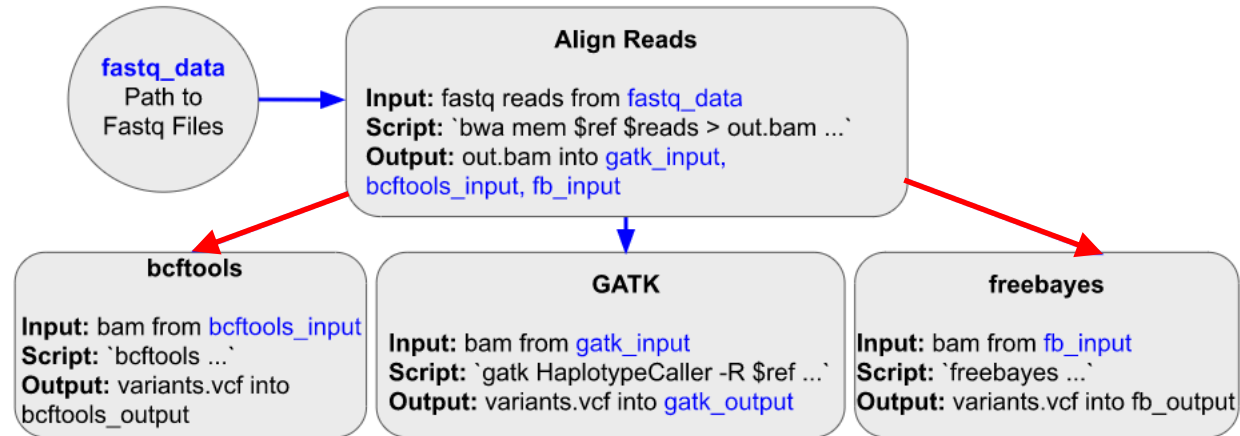
How it works



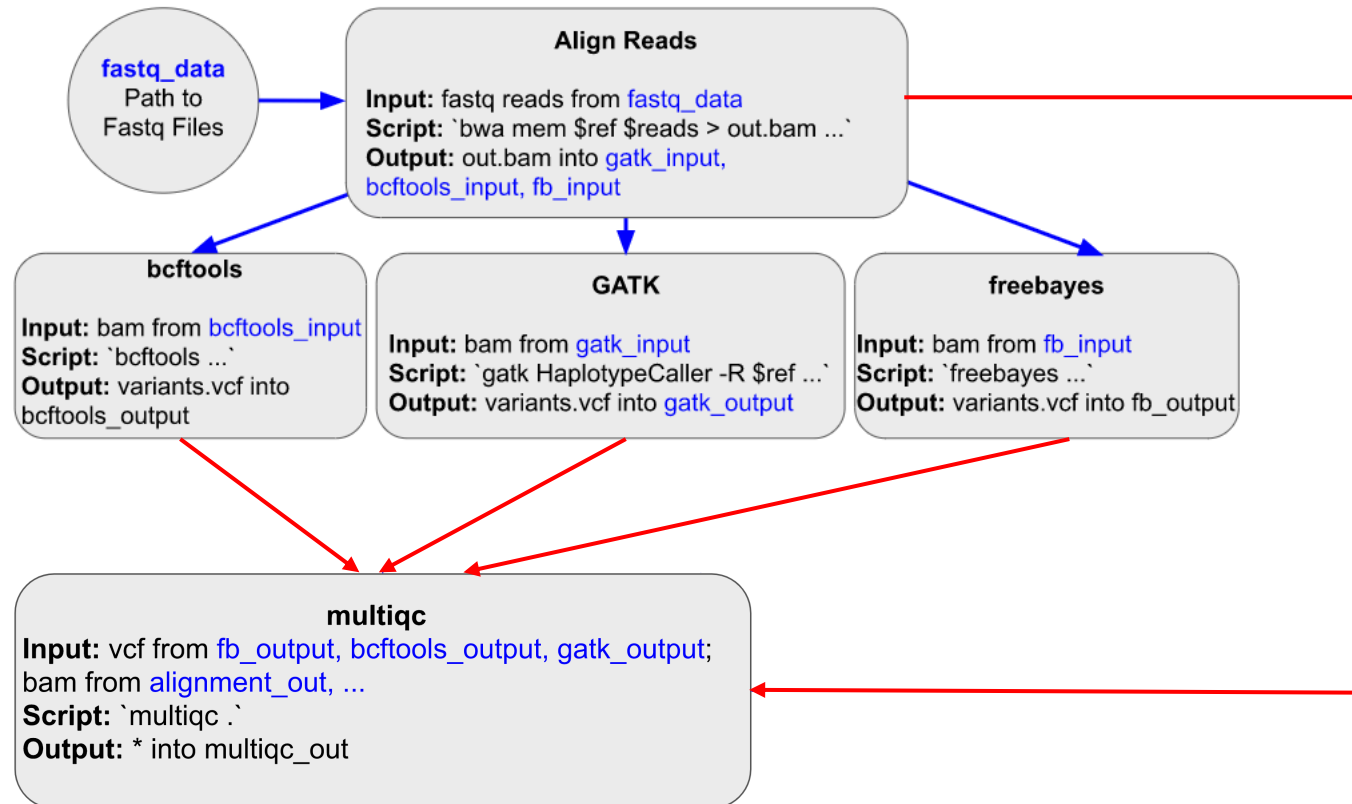
How it works



How it works



How it works



Resources

- Latest documentation: <https://www.nextflow.io/docs/latest/index.html>
- Working examples: <https://github.com/nextflow-io/awesome-nextflow>
- Workflows for standard protocols: <https://github.com/nf-core>
- Recurrent implementation patterns used in Nextflow applications:
<https://nextflow-io.github.io/patterns/index.html>