codingene/nextflow-base (v1.0)

Developed by Codingene.

Last modified: 19 Jun 2020

Contents

1 About

Currently on this pipeline three most based steps on any Sequence based analysis (starts from fastq files)

- Quality Check (using fastqc)
- Filtering (Using fastp)
- Sequence Read Quantification (Using kallisto)

This can be used as a base to add other `process`.

1.1 Outlines

- Adapter removal and Filtering of RAW reads using fastp

2 Setup

Pipeline dependency.

- Nextflow on which this workflow framework is based.
- Docker or Conda for tools environment. (It is recommended to use Docker for this workflow.)

2.1 Getting and Installing Nextflow

This is required only once per system. Check if your system already have it by typing `nextflow` from any terminal location. If not follow there steps -

```
curl -s https://get.nextflow.io | bash
mv nextflow usr/bin/
```

2.2 Getting and Installing Docker

Follow this - How to install and use docker on ubuntu

2.3 Installing Conda

We will use miniconda for this purpose.

```
wget https://repo.continuum.io/miniconda/Miniconda3-latest-Linux-x86_64.sh -O miniconda.sh
bash miniconda.sh -b -p $HOME/miniconda3
export PATH="$HOME/miniconda3/bin:$PATH"
rm miniconda.sh
```

2.4 Getting This Workflow

```
git clone https://github.com/codingene/nextflow-base.git
```

3 Test the Workflow

Test is to check if basic components of a workflow is able to run in a system with everything setup properly.

Supposing you are in workflow directory, run following -

```
nextflow run mian.nf -profile test,docker
```

> Note: Test run may take some time on a first time, because it will download all the tools environment (docker-images/conda-env) automatically in background.

If this success you are good to go on running with your own datasets.

4 Running the workflow

4.1 Basic Usage

Check help menu

```
nextflow run path-to/nextflow-base/main.nf --help
```

The typical command for running the pipeline is as follows

```
nextflow run path-to/raw-read-qc/main.nf [arguments] -profile docker
```

4.2 Arguments

4.2.1 Workflow Arguments

4.2.1.1 `--reads` [mandatory]

A fasta file directory where all the paired-end reads present.

They must follow this naming convention of `*_{1,2}.fastq.gz` or `*_{1,2}.fq.gz`

4.2.1.2 `--cdna` [mandatory]

Path to a cDNA fasta file.

4.2.1.3 `--outdir` [optional]

Output folder name. If not given it will create a `results` named directory on working location. This is where you can find all the results post pipeline run.

4.2.1.4 Individual Tool parameters [optional]

For details of individual tool parameters check respective documentation. All are optional with default values (please check bellow)

- `--fastp.length_required` (default: 75)
- `--fastp.length_limit` (default: 151)
- `--fastp.qualified_quality_phred` (default: 30)

4.2.2 System Arguments

This arguments are optional but recommended to provided with higher numbers as per system configuration and data need.

–max_cpus : [Recommended] Number of threads/CPU to assign (default = 1)
–max_memory : [Recommended] Maximum Memory in GB (default = '2 GB')
–max_time : [Optional] Maximum time for a single step (default = '1h')

5 Output Directory Structure

```
|- Sample-Name/ID
    |- fastp_filtred_reads
  |- fastqc_report_post_filter
  |- fastqc_report_pre_filter
```

6 Changelog

More information about Changelog (version updates) can be found in NEWS.md

7 FAQs