ontpipeline Documentation

Yan Zhou

Jun 17, 2019

Contents

CHAPTER 1

Get Started

1.1 Installation

1.1.1 Anaconda Installation

Installing on Linux https://docs.anaconda.com/anaconda/install/linux/

1.1.2 Guppy Installation

```
sudo apt-get update
sudo apt-get install wget lsb-release
export PLATFORM=$(lsb_release -cs)
wget -O- https://mirror.oxfordnanoportal.com/apt/ont-repo.pub | sudo apt-key add -
echo "deb http://mirror.oxfordnanoportal.com/apt ${PLATFORM}-stable non-free" | sudo␣
↪tee /etc/apt/sources.list.d/nanoporetech.sources.list
sudo apt-get update
apt-get install ont-guppy[-cpu]
```

1.1.3 Porechop Installation

```
/opt/anaconda3/bin/conda create -n porechop python=3.7
source activate /opt/anaconda3/envs/porechop
/opt/anaconda3/bin/conda install -c bioconda porechop
source /opt/anaconda3/bin/deactivate
```

1.1.4 NanoStat Installation

```
/opt/anaconda3/bin/conda create -n nanostat python=3.7
source activate /opt/anaconda3/envs/nanostat
/opt/anaconda3/bin/conda install -c bioconda nanostat
source /opt/anaconda3/bin/deactivate
```

1.1.5 NanoFilt Installation

```
/opt/anaconda3/bin/conda create -n nanofilt python=3.7
source activate /opt/anaconda3/envs/nanofilt
/opt/anaconda3/bin/conda install -c bioconda nanofilt
source /opt/anaconda3/bin/deactivate
```

1.1.6 Canu Installation

```
/opt/anaconda3/bin/conda create -n canu python=3.7
source activate /opt/anaconda3/envs/canu
/opt/anaconda3/bin/conda install -c bioconda canu
source /opt/anaconda3/bin/deactivate
```

1.1.7 Flye Installation

```
/opt/anaconda3/bin/conda create -n flye python=2.7
source activate /opt/anaconda3/envs/flye
/opt/anaconda3/bin/conda install -c bioconda flye
source /opt/anaconda3/bin/deactivate
```

1.1.8 Unicycler Installation

```
/opt/anaconda3/bin/conda create -n unicycler python=3.7
source activate /opt/anaconda3/envs/unicylcer
/opt/anaconda3/bin/conda install -c bioconda unicycler
source /opt/anaconda3/bin/deactivate
```

CHAPTER 2

General Settings

2.1 Job Name (Optional)

Set the name of your job.

Note:

- Up to 15 characters, no blank spaces, start with alphanumeric character.
- If you want to use the default value, leave it blank.

2.2 Workspace (Required)

Set the path to the folder where contains the .fast5 files.

Note:

- **Example:** `/home/ynzhou/NCCTData/20181109_1435_20181109_VRE_I/fast5`

2.3 Barcodes (Optional)

Set which barcodes should be included into the analysis. Put in the numbers and separate them with a comma.

Note:

- **Example:** 1,2,3,4
- If you want to include all barcodes in your analysis, leave it blank.

2.4 Threads (Required)

Set the number of threads to run the analysis.

Note:

- The program will automatically find the max number of threads available on the system and set it to the default value.

CHAPTER 3

Basecalling Settings

3.1 Flowcell ID (Required)

Choose the Flowcell ID from the select list.

3.2 Kit Number (Requied)

Choose the kit number from the select list.

3.3 Barcode kit (Optional)

Choose the barcode kit(s) from the list if used.

Note:

- If no barcode kit was used, leave it blank.
- Choose by press `Ctrl` to select multiple barcode kits.

3.4 cpu_threads_per_caller[1] (Default)

Note:

- Set value: 1.

[1] Guppy v2.3.1 Release https://community.nanoporetech.com/posts/guppy-2-3-1-release

3.5 records_per_fastq[2] (Default)

Note:

- select value: 0.
- Use a single file (per worker, per run id).

3.6 recursive[2] (Default)

Note:

- Set value: search for input files recursively.

3.7 enable_trimming[2] (Default)

Note:

- Set value: enable trimming of the adapter from the beginning of the read before basecalling.

[2] How to configure Guppy parameters https://community.nanoporetech.com/protocols/Guppy-protocol-preRev/v/gpb_2003_v1_revg_14dec2018/how-to-configure-guppy-parameters

CHAPTER 4

Reads Filter Settings

4.1 Read Score (Required)

Set a minimum average read quality score to filter the reads.

Note:

- Default value: 9.
- If you want to keep all the reads, set the value to 0, or only a postive integer is acceptable.

4.2 Read Length (Required)

Set a minimum read length to filter reads.

Note:

- Default value: 500.
- If you want to keep all the reads, set the value to 0, or only a postive integer is acceptable.

4.3 Head Crop (Required)

Note:

- Default value: 50.
- If you do not want to crop any reads, set the value to 0, or only a postive integer is acceptable.

4.4 no_split[1] (Default)

Note:

- Set value: Split reads when an adapter is found in the middle.

[1] Porechop https://github.com/rrwick/Porechop

CHAPTER 5

Assembly Settings

5.1 Assembler (Required)

Choose an assembler tool from the select list.

5.1.1 Canu

Genome Size (Required)

Set the estimated genome size.

Note:

- e.g. 5m or 2.6g, which is used for solid k-mers selection.
- The estimate could be rough (e.g. within 0.5x-2x range) and does not affect the other assembly stages.

assembly-prefix[1] (Default)

Note:

- Set value: asm.
- Set the file name prefix with "asm" of intermediate and output files.

[1] Canu Tutorial https://canu.readthedocs.io/en/latest/tutorial.html

5.1.2 Flye

Genome Size (Required)

Set the estimated genome size.

Note:

- e.g. 5m or 2.6g, which is used for solid k-mers selection.
- The estimate could be rough (e.g. within 0.5x-2x range) and does not affect the other assembly stages.

5.1.3 Unicycler

Short 1 (Optional)

Set the path to the fastq file of first short reads in each pair.

Short 2 (Optional)

Set the path to the fastq file of second short reads in each pair

Note:

- These two options are required in **hybrid assembly** mode, or leave it blank for **long-read-only assembly** mode.
- These two fastq files should be obtained from Illumina sequencing.

CHAPTER 6

FAQ

6.1 What bioinformatics tools are used?

- Guppy v2.3.1 https://community.nanoporetech.com
- Porechop 0.2.3 https://github.com/rrwick/Porechop
- NanoStat 1.1.2 https://github.com/wdecoster/nanostat
- NanoFilt 2.2.0 https://github.com/wdecoster/nanofilt
- Canu v1.8 https://github.com/marbl/canu
- Flye 2.4-release https://github.com/fenderglass/Flye
- Unicycler v0.4.8-beta https://github.com/rrwick/Unicycler

6.2 Where can I find output files?

- `\Basecalled` (reads after basecalling)
- `\Barcodes` (reads after demultiplexing)
- `\Analysis\AdapterTrimmedFiles` (reads after adapter trimming)
- `\Analysis\FiltedFiles` (reads after filtering)
- `\Analysis\StatFiles` (statitstical data before and after filtering)
- `\Analysis\Logs` (log files)
- `\Assembly` (reads after assembly)