

FaQC 실습

2022-01-18(화)
20193852 문유빈

1) 목표 : FaQCs 프로그램으로 Quality Control

2) FaQCs 정리

- FaQCs : Quality Control of Next Generation Sequencing Data
 - > Quality control에선 시퀀스 데이터의 퀄리티를 검사하고 정리한다.
- 사용한 시퀀싱 기구 : Illumina Hiseq
 - forward 약 150bp, backward 약 150bp 씩 읽음
- 시퀀싱 된 파일 => fastq format

@ -> sequence identifier and optional description

-> low sequence letters

+

-> line 2 시퀀스에 대한 quality values(scores)

** 참고) fastq format 위키피디아 : https://en.wikipedia.org/wiki/FASTQ_format

=> Quality score for each sequence is a string of characters, one for each base of the nucleic sequence, used to characterize the probability of mis-identification of each base. The score is encoded using the ASCII character table

- xshell /home/bioware/FaQCs/FaQCs.pl -h 입력

Usage: perl /home/bioware/FaQCs/FaQCs.pl [options] [-u unpaired.fastq] -p reads1.fastq reads2.fastq -d out_directory		
Version 1.36(*1)		
Input File: (can use more than once)	-p	<Files> Paired reads in two files and separate by space
Trim:	-q	<INT> Targets # as quality level (default 5) for trimming
	-adapter	<bool> Trim reads with illumina adapter/primers (default: no)
Output:	-d	<PATH> Output directory.

Options:	-t	<INT > # of CPUs to run the script (default:2)
----------	----	---

3) QC 진행 -1 : xshell 입력어

[형식]

1 2 3 4
 /home/bioware/FaQCs/FaQCs.pl -p sample.1.fastq samples.2.fastq -d QcReads q 20
5 -t 8 -adapter > QC.log 6 7

원래 -q 20 인데 잘못 한거

```
[quest01@smel0:test01]$ /home/bioware/FaQCs/FaQCs.pl -p /data/Original_data/KOPRI/WWTP/KP-WWTP-1512_1.fastq.gz /data/Original_data/KOPRI/WWTP/KP-WWTP-1512_2.fastq.gz -d output q 20 -t 8 -adapter > output/QC.log
```

① /home/bioware/FaQCs/FaQCs.pl => 프로그램이 있는 절대 위치

** 서버마다 프로그램의 위치 다름, 각각의 버전도 다를 수 있기 때문에 절대 위치를 써주는 습관 들이기

② -p /data/Original_data/KOPRI/WWTP/KP-WWTP-1512_1.fastq.gz /data/Original_data/KOPRI/WWTP/KP-WWTP-1512_2.fastq.gz => QC 진행할 파일 2개

** 파일의 절대 경로/파일명 형식으로 작성

** 파일이 있는지 앞에 철자 친 후 tab키를 눌러 확인하기

③ -d output => 결과 파일이 들어갈 디렉토리

** 내가 생성한 output 디렉토리에 결과 파일이 들어감

④ q 20 => Quality score(Phred quality)

Phred Quality Score	Error	Accuracy (1 - Error)
10	1/10 = 10%	90%
20	1/100 = 1%	99%
30	1/1000 = 0.1%	99.9%

** q 20 이라면 bp 100개 중 1개의 오류 발생 확률

** 보통 quality score는 20 정도로 조정

⑤ -t 8 => 스크립트 돌릴 CPU 수, CPU thread

** 8개의 CPU 사용

** CPU 사용할 때는 xshell에 htop or top 입력해 현재 사용되는 CPU 보고 돌리기

⑥ -adapter => 어댑터를 부쳐 실험 시 어댑터는 합성 시퀀스이므로 제거해야 함

** 실제 시퀀싱된 영역(amplicon)만 남도록 제거

⑦ > output/QC.log => 진행 사항 로그로 남기기

** 어떤 명령을 시행할 때도 마지막에 '> 파일명' 입력 시 로그 남음

** 백그라운드에서 실행 시 왜 에러가 발생했는지, 어떤 문제가 있는지 알기 위해 로그를 남겨야 한다.

4) QC 진행 -2 : output 디렉토리에서 진행 상황 보기

진행 (1)

```
[guest01@smel0:test01]$ cd output
[guest01@smel0:output]$ ll
total 3816728
-rw-rw-r-- 1 guest01 guest01 374637580 Jan 18 15:00 KP-WWTP-1512_1_00000.fastq
-rw-rw-r-- 1 guest01 guest01 374602261 Jan 18 15:00 KP-WWTP-1512_1_00001.fastq
-rw-rw-r-- 1 guest01 guest01 374595513 Jan 18 15:00 KP-WWTP-1512_1_00002.fastq
-rw-rw-r-- 1 guest01 guest01 374602626 Jan 18 15:00 KP-WWTP-1512_1_00003.fastq
-rw-rw-r-- 1 guest01 guest01 374623520 Jan 18 15:00 KP-WWTP-1512_1_00004.fastq
-rw-rw-r-- 1 guest01 guest01 374592022 Jan 18 15:00 KP-WWTP-1512_1_00005.fastq
-rw-rw-r-- 1 guest01 guest01 374614612 Jan 18 15:01 KP-WWTP-1512_1_00006.fastq
-rw-rw-r-- 1 guest01 guest01 374634429 Jan 18 15:01 KP-WWTP-1512_1_00007.fastq
-rw-rw-r-- 1 guest01 guest01 374601863 Jan 18 15:01 KP-WWTP-1512_1_00008.fastq
-rw-rw-r-- 1 guest01 guest01 374599889 Jan 18 15:01 KP-WWTP-1512_1_00009.fastq
-rw-rw-r-- 1 guest01 guest01 0 Jan 18 15:00 QC.fastqCount.txt
-rw-rw-r-- 1 guest01 guest01 169 Jan 18 15:00 QC.log
```

진행 (2)

```
[guest01@smel0:output]$ ll
total 8069944
-rw-rw-r-- 1 guest01 guest01 374637580 Jan 18 15:00 KP-WWTP-1512_1_00000.fastq
-rw-rw-r-- 1 guest01 guest01 374602261 Jan 18 15:00 KP-WWTP-1512_1_00001.fastq
-rw-rw-r-- 1 guest01 guest01 374595513 Jan 18 15:00 KP-WWTP-1512_1_00002.fastq
-rw-rw-r-- 1 guest01 guest01 374602626 Jan 18 15:00 KP-WWTP-1512_1_00003.fastq
-rw-rw-r-- 1 guest01 guest01 374623520 Jan 18 15:00 KP-WWTP-1512_1_00004.fastq
-rw-rw-r-- 1 guest01 guest01 374592022 Jan 18 15:00 KP-WWTP-1512_1_00005.fastq
-rw-rw-r-- 1 guest01 guest01 374614612 Jan 18 15:01 KP-WWTP-1512_1_00006.fastq
-rw-rw-r-- 1 guest01 guest01 374634429 Jan 18 15:01 KP-WWTP-1512_1_00007.fastq
-rw-rw-r-- 1 guest01 guest01 374601863 Jan 18 15:01 KP-WWTP-1512_1_00008.fastq
-rw-rw-r-- 1 guest01 guest01 374599889 Jan 18 15:01 KP-WWTP-1512_1_00009.fastq
-rw-rw-r-- 1 guest01 guest01 374600378 Jan 18 15:01 KP-WWTP-1512_1_00010.fastq
-rw-rw-r-- 1 guest01 guest01 374621521 Jan 18 15:01 KP-WWTP-1512_1_00011.fastq
-rw-rw-r-- 1 guest01 guest01 374605847 Jan 18 15:01 KP-WWTP-1512_1_00012.fastq
-rw-rw-r-- 1 guest01 guest01 374637013 Jan 18 15:01 KP-WWTP-1512_1_00013.fastq
-rw-rw-r-- 1 guest01 guest01 374620333 Jan 18 15:01 KP-WWTP-1512_1_00014.fastq
-rw-rw-r-- 1 guest01 guest01 374613056 Jan 18 15:01 KP-WWTP-1512_1_00015.fastq
-rw-rw-r-- 1 guest01 guest01 374632354 Jan 18 15:01 KP-WWTP-1512_1_00016.fastq
-rw-rw-r-- 1 guest01 guest01 374626361 Jan 18 15:01 KP-WWTP-1512_1_00017.fastq
-rw-rw-r-- 1 guest01 guest01 374627354 Jan 18 15:02 KP-WWTP-1512_1_00018.fastq
-rw-rw-r-- 1 guest01 guest01 374619464 Jan 18 15:02 KP-WWTP-1512_1_00019.fastq
-rw-rw-r-- 1 guest01 guest01 374603537 Jan 18 15:02 KP-WWTP-1512_1_00020.fastq
-rw-rw-r-- 1 guest01 guest01 374610944 Jan 18 15:02 KP-WWTP-1512_1_00021.fastq
-rw-rw-r-- 1 guest01 guest01 22061056 Jan 18 15:02 KP-WWTP-1512_1_00022.fastq
-rw-rw-r-- 1 guest01 guest01 0 Jan 18 15:00 QC.fastqCount.txt
-rw-rw-r-- 1 guest01 guest01 169 Jan 18 15:00 QC.log
```

** 진행되는데 몇 시간 소요

** output 파일 계속 업데이트 됨

5) QC 종료

QC output

```
[guest01@smel-cluster:output]$ ll
total 17980536
-rw-rw-r-- 1 guest01 guest01 9208044907 Jan 18 16:10 QC.1.trimmed.fastq
-rw-rw-r-- 1 guest01 guest01 9203088729 Jan 18 16:10 QC.2.trimmed.fastq
-rw-rw-r-- 1 guest01 guest01      102 Jan 18 15:05 QC.fastqCount.txt
-rw-rw-r-- 1 guest01 guest01     4040 Jan 18 16:10 QC.log
-rw-rw-r-- 1 guest01 guest01    266994 Jan 18 16:11 QC_qc_report.pdf
-rw-rw-r-- 1 guest01 guest01     1093 Jan 18 16:11 QC.stats.txt
-rw-rw-r-- 1 guest01 guest01    650612 Jan 18 16:11 QC.unpaired.trimmed.fastq
```