

CDC_FastUniq

2022-03-04

20193852 문유빈

Before > bmtagger로 host로 판단된 시퀀스 제거

Now > FastUniq로 duplicate 제거

1. Script

FastUniq는 duplicate 제거 프로그램으로 A(animal), H(human), E(environment) 셋 다 진행해야 함

1) A script

```
inpath="/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/"
fastuniq="/home/bioware/FastUniq/bin/fastuniq"

for sample in ${inpath}*A*
do
    #echo ${sample}
    stub=${sample#${inpath}}
    echo $stub
    echo -e ${inpath}/${stub}/QcReads/QC.1.removed.fastq\n${inpath}/${stub}/
    QcReads/QC.2.removed.fastq' > ${inpath}/${stub}/inputlist.txt'

    ${fastuniq} -i ${inpath}/${stub}/inputlist.txt' -t q -o ${inpath}/${stub}
    '/'${stub}'.rmd.1.fastq' -p ${inpath}/${stub}/${stub}.rmd.2.fastq' -c 1
done
```

* \${inpath}*A* 경로상에 A가 있는 모든 파일

~2021_XXA_XXX

* echo -e를 통해 문자열 이스케이프를 읽을 수 있음(\n, \t 같은거)

** echo -e > newfile을 통해 간단하게 새로운 파일 생성 가능

* echo -e \${inpath}/\${stub}/QcReads/QC.1.removed.fastq\n'

줄바꾸기(이스케이프)

/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF1H_112/QcReads/QC.1.removed.fastq

** 즉 inputlist.txt란 파일에 2개의 QC.1.removed.fastq, QC.2.removed.fastq 파일의 절대 경로를 적는 것 ==> 이는 FastUniq에 inputlist로 들어감

inputlist.txt

```
/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF1H_112/QcReads/QC.1.re  
moved.fastq  
/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF1H_112/QcReads/QC.2.re  
moved.fastq
```

*** A, H는 bmtagger를 통해 host를 제거한 removed.fastq 파일을 사용하고
E는 제거되지 않은 trimmed.fastq 파일 사용

** output => rmd.1.fastq, rmd.2.fastq 파일

2) H script

```
inpath="/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/"  
fastuniq="/home/bioware/FastUniq/bin/fastuniq"  
  
for sample in ${inpath}*H*  
do  
    #echo ${sample}  
    stub=${sample#${inpath}}  
    echo $stub  
    echo -e ${inpath}/${stub}/QcReads/QC.1.removed.fastq\n${inpath}/${stub}/  
QcReads/QC.2.removed.fastq > ${inpath}/${stub}/inputlist.txt  
  
    ${fastuniq} -i ${inpath}/${stub}/inputlist.txt -t q -o ${inpath}/${stub}/  
/${stub}.rmd.1.fastq -p ${inpath}/${stub}/${stub}.rmd.2.fastq -c 1  
  
done
```

** \${inpath}*H*로 H 파일 전부
** removed.fastq 파일 사용

3) E script

```
inpath="/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/"  
fastuniq="/home/bioware/FastUniq/bin/fastuniq"  
  
for sample in ${inpath}*E*  
do  
    #echo ${sample}  
    stub=${sample#${inpath}}  
    echo $stub  
    echo -e ${inpath}/${stub}/QcReads/QC.1.trimmed.fastq\n${inpath}/${stub}/  
QcReads/QC.2.trimmed.fastq > ${inpath}/${stub}/inputlist.txt  
  
    ${fastuniq} -i ${inpath}/${stub}/inputlist.txt -t q -o ${inpath}/${stub}/  
/${stub}.rmd.1.fastq -p ${inpath}/${stub}/${stub}.rmd.2.fastq -c 1  
  
done
```

** \${inpath}*A*로 H 파일 전부
** trimmed.fastq 파일 사용

2. 결과물

1) A 폴더

```
[guest01@smel0:2021_KF2A_242]$ ll
total 10624048
-rw-rw-r-- 1 guest01 guest01 5101408448 Mar  4 22:33 2021_KF2A_242.rmd.1.fastq
-rw-rw-r-- 1 guest01 guest01 5099432570 Mar  4 22:33 2021_KF2A_242.rmd.2.fastq
-rw-rw-r-- 1 guest01 guest01      0 Feb 24 10:35 bmtagger.log
drwxrwxr-x 2 guest01 guest01     10 Feb 24 12:41 bmtagger_tmp
-rw-rw-r-- 1 guest01 guest01      0 Feb 22 17:17 bmtagger_unpaired.log
drwxrwxr-x 2 guest01 guest01     10 Feb 22 17:21 bmtagger_unpaired_tmp
-rw-rw-r-- 1 guest01 guest01     170 Mar  4 22:30 inputlist.txt
-rw-rw-r-- 1 guest01 guest01    52339 Feb 17 11:23 QC.log
drwxrwxr-x 2 guest01 guest01     4096 Mar  3 22:30 QcReads
-rw-rw-r-- 1 guest01 guest01   672858595 Feb 24 12:41 QC.removed.bmtagger
-rw-rw-r-- 1 guest01 guest01    5255066 Feb 22 17:21 QC.unpaired.removed.bmtagger
```

- ☒ inputlist.txt
- ☒ 2021_KF2A_242.rmd.1.fastq
- ☒ 2021_KF2A_242.rmd.2.fastq

2) H 폴더

```
[guest01@smel0:2021_KF2H_224]$ ll
total 18010020
-rw-rw-r-- 1 guest01 guest01 9222276416 Mar  4 21:55 2021_KF2H_224.rmd.1.fastq
-rw-rw-r-- 1 guest01 guest01 9218982260 Mar  4 21:55 2021_KF2H_224.rmd.2.fastq
-rw-rw-r-- 1 guest01 guest01      0 Feb 24 08:16 bmtagger.log
drwxrwxr-x 2 guest01 guest01     10 Feb 24 09:42 bmtagger_tmp
-rw-rw-r-- 1 guest01 guest01      0 Feb 22 16:50 bmtagger_unpaired.log
drwxrwxr-x 2 guest01 guest01     10 Feb 22 16:51 bmtagger_unpaired_tmp
-rw-rw-r-- 1 guest01 guest01     170 Mar  4 21:49 inputlist.txt
-rw-rw-r-- 1 guest01 guest01    46334 Feb 17 20:43 QC.log
drwxrwxr-x 2 guest01 guest01     4096 Mar  3 21:33 QcReads
-rw-rw-r-- 1 guest01 guest01    921886 Feb 24 09:42 QC.removed.bmtagger
-rw-rw-r-- 1 guest01 guest01    10133 Feb 22 16:51 QC.unpaired.removed.bmtagger
```

- ☒ inputlist.txt
- ☒ 2021_KF2H_224.rmd.1.fastq
- ☒ 2021_KF2H_224.rmd.2.fastq

3) E 폴더

```
[guest01@smel0:2021_KF2E_212P]$ ll
total 19816676
-rw-rw-r-- 1 guest01 guest01 10146745006 Mar  4 20:54 2021_KF2E_212P.rmd.1.fastq
-rw-rw-r-- 1 guest01 guest01 10145466398 Mar  4 20:54 2021_KF2E_212P.rmd.2.fastq
-rw-rw-r-- 1 guest01 guest01     172 Mar  4 20:47 inputlist.txt
-rw-rw-r-- 1 guest01 guest01    50802 Feb 17 15:12 QC.log
drwxrwxr-x 2 guest01 guest01     4096 Feb 17 15:13 QcReads
```

- ☒ inputlist.txt
- ☒ 2021_KF2H_224.rmd.1.fastq
- ☒ 2021_KF2H_224.rmd.2.fastq