

After Bmtagger

2022-3-03

20193852 문유빈

1. Bmtagger가 완료되면 host 시퀀스가 나온다.

=> 해당 host를 제거하는 작업이 필요하다.

```
A00721:393:H3NNLDSX3:3:1101:30852:9706
A00721:393:H3NNLDSX3:3:1101:30888:9706
A00721:393:H3NNLDSX3:3:1101:25156:10019
A00721:393:H3NNLDSX3:3:1101:25500:10050
A00721:393:H3NNLDSX3:3:1101:25066:10363
A00721:393:H3NNLDSX3:3:1101:1253:10551
QC.unpaired.removed.bmtagger
```

-> 여기에 기록된 시퀀스가 host이다. 이것들은 제거해줘야 한다.

2. After bmtagger

- 1) 전체 코드

```
#!/opt/ohpc/pub/apps/python3/3.5.4/bin/python3

import sys,os
from Bio.SeqIO.QualityIO import FastqGeneralIterator

Marked_P=set()
Marked_unP=set()
bmt_path=sys.argv[1] # /home/bbang9/Project/2020/CDC/20_03/HiSeq/EDGE/KF1H-154
fq_path=sys.argv[2] # /home/bbang9/Project/2020/CDC/20_03/HiSeq/EDGE/KF1H-154/QcReads

with open(bmt_path+'QC.removed.bmtagger','r') as M1, open(bmt_path+'QC.unpaired.removed.bmtagger',
'r') as M2:
    for l in M1:
        l=l.rstrip('\n')
        Marked_P.add(l)
    for l in M2:
        l=l.rstrip('\n')
        Marked_unP.add(l)

def New_fastq(o_read,n_read,Mark):
    with open(o_read,'r') as F, open(n_read,'w') as N:
        for t,s,q in FastqGeneralIterator(F):
            tmp=t.split()[0]
            if not tmp in Mark:
                N.write('@'+t+'\n'+s+'\n'+q+'\n')

New_fastq(fq_path+'QC.1.trimmed.fastq',fq_path+'QC.1.removed.fastq',Marked_P)
New_fastq(fq_path+'QC.2.trimmed.fastq',fq_path+'QC.2.removed.fastq',Marked_P)
New_fastq(fq_path+'QC.unpaired.trimmed.fastq',fq_path+'QC.un.removed.fastq',Marked_unP)
```

2) 코드 설명 -1

```
Marked_P=set()
Marked_unP=set()
bmt_path=sys.argv[1] # /home/bbang9/Project/2020/CDC/20_03/HiSeq/EDGE/KF1H-154
fq_path=sys.argv[2] # /home/bbang9/Project/2020/CDC/20_03/HiSeq/EDGE/KF1H-154/QcReads

with open(bmt_path+'/QC.removed.bmtagger','r') as M1, open(bmt_path+'/QC.unpaired.removed.bmtagger',
'r') as M2:
    for l in M1:
        l=l.rstrip('\n')
        Marked_P.add(l)
    for l in M2:
        l=l.rstrip('\n')
        Marked_unP.add(l)
```

- * Marked_P -> paired된 host set에 append
- * Marked_unP -> unpaired된 host set에 append

3) 코드 설명 -2

```
def New_fastq(o_read,n_read,Mark):
    with open(o_read,'r') as F, open(n_read,'w') as N:
        for t,s,q in FastqGeneralIterator(F):
            tmp=t.split()[0]
            if not tmp in Mark:
                N.write('@'+t+'\n'+s+'\n'+q+'\n')

New_fastq(fq_path+'/QC.1.trimmed.fastq',fq_path+'/QC.1.removed.fastq',Marked_P)
New_fastq(fq_path+'/QC.2.trimmed.fastq',fq_path+'/QC.2.removed.fastq',Marked_P)
New_fastq(fq_path+'/QC.unpaired.trimmed.fastq',fq_path+'/QC.un.removed.fastq',Marked_unP)
```

- * def New_fastq(인자) -> New_fastq라는 함수 생성
- ** r = 읽기 모드, w = 쓰기 모드

** QC.1.trimmed.fastq(bmtagger 안한 파일, 'r' mode)에서 t, s, p를 읽어오고 t를 split한 첫 번째를 tmp로 지정한다. 그 후 Marked_P(paired된 host set)에 tmp가 없다면(=host가 아니라면) QC.1.removed.fastq(bmtagger 후 host를 제거한 파일)를 생성해 그곳에 쓴다('w' mode)

*** 즉, bmtagger 하기 전 전체에서 bmtagger 후 host만 모아놓은 파일을 제외하는 과정

3. after_bmtagger_A, after_bmtagger_H

1) after_A

```
Path="/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/"

for i in ${Path}*A*
do
    #echo ${i}
    python After_bmtagger.py ${i} ${i}/QcReads/
done
```

=> 전체 파일에서 ~ XXXA ~ 인것들 모두

ex > /home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF2A_112

즉 python After_bmtagger.py \${i} \${i}/QcReads/

/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF2A_112

/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF2A_112/QcReads/

```
import sys,os
from Bio.SeqIO.QualityIO import FastqGeneralIterator

Marked_P=set()
Marked_unP=set()
bmt_path=sys.argv[1] # /home/bbang9/Project/2020/CDC/20_03/HiSeq/EDGE/KF1H-154
fq_path=sys.argv[2] # /home/bbang9/Project/2020/CDC/20_03/HiSeq/EDGE/KF1H-154/QcReads
```

위에서 받은 I, I/QcReads는 각각 After_bmtagger.py 파일에서 bmt_path, fq_path가 됨

2) after_H

```
Path="/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/"

for i in ${Path}*H*
do
    #echo ${i}
    python After_bmtagger.py ${i} ${i}/QcReads/
done
```

=> 전체 파일에서 ~ XXXH ~ 인것들 모두

ex > /home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF1H_112

언니가 알려준 팁

```
[guest01@smel0:script]$ ./03.after_bmtagger_A.sh > 03.after_bmtagger_A.log 1
2 ^Z
[1]+  Stopped 3 ./03.after_bmtagger_A.sh > 03.after_bmtagger_A.log
[guest01@smel0:script]$ bg 1
[1]+ ./03.after_bmtagger_A.sh > 03.after_bmtagger_A.log &
[guest01@smel0:script]$ jobs
[1]+  Running ./03.after_bmtagger_A.sh > 03.after_bmtagger_A.log &
[guest01@smel0:script]$ disown 4
[guest01@smel0:script]$ jobs
[guest01@smel0:script]$ htop
[guest01@smel0:script]$
```

- (1) 처음에 ./03.after_bmtagger_A.sh로 nohup이 아닌 일반적으로 돌려보기
- (2) 잘 돌아가는지 확인 후 이상이 없다면 ctrl+Z 눌러 중단
 - ** ctrl+Z는 실행 프로세스를 잠시 멈춤
 - ** ctrl+C는 실행 프로세스를 아예 중단
- (3) bg1 입력시 background로 돌아감 (끝에 붙은 &)
- (4) disown 입력시 nohup으로 돌아감 (jobs해도 안보임)