

## CDC/Bmtagger

2022-02-22~

20193852 문유빈

▷ Bmtagger => host 제거를 위한 프로그램

▷ QC 완료 후 나온 파일 ...

2021_CF1H_112	→	CXXH
2021_CF2A_112	→	CXXA
2021_CF2E_111	→	CXXE

H => human
A => animal(pig)
E => environment

.	→	JXXH
.	→	JXXA
.	→	JXXE

C => 충청도
J => 전라도
K => 경기도

.	→	KXXH
.	→	KXXA
.	→	KXXE

paired
unpaired

### 1. 각 파일에 맞게 셸 스크립트 작성

```
[guest01@smel0:script]$ ll
total 36
drwxrwxr-x 2 guest01 guest01 154 Feb 22 14:45 01.QC
-rwxr-xr-x 1 guest01 guest01 1119 Feb 22 15:17 02.bmtagger_hostrm_C_A.sh
-rwxr-xr-x 1 guest01 guest01 1117 Feb 22 15:30 02.bmtagger_hostrm_C_H.sh
-rwxr-xr-x 1 guest01 guest01 1122 Feb 22 15:26 02.bmtagger_hostrm_human.sh
-rwxr-xr-x 1 guest01 guest01 1119 Feb 22 15:14 02.bmtagger_hostrm_J_A.sh
-rwxr-xr-x 1 guest01 guest01 1117 Feb 22 15:31 02.bmtagger_hostrm_J_H.sh
-rwxr-xr-x 1 guest01 guest01 1119 Feb 22 15:16 02.bmtagger_hostrm_K_A.sh
-rwxr-xr-x 1 guest01 guest01 1117 Feb 22 15:31 02.bmtagger_hostrm_K_H.sh
-rwxr-xr-x 1 guest01 guest01 1132 Feb 22 15:38 02.bmtagger_hostrm_unpaired_A.sh
-rwxr-xr-x 1 guest01 guest01 1130 Feb 22 15:38 02.bmtagger_hostrm_unpaired_H.sh
```

02.bmtagger\_hostrm\_C\_A.sh => paired CXXA

02.bmtagger\_hostrm\_K\_H.sh => paired KXXH

02.bmtagger\_hostrm\_unpaired\_A.sh => unpaired XXXA

\*\* unpaired는 양이 작아 C, J, K로 나누지 않고 A, H로만 나누기

### 1) 02.bmtagger\_hostrm\_C\_A.sh

```
inpath="/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/"
db="/home/s4645/build/bmtagger/DB/Sscrefall 1/GCF 000003025.6 Sscrefall.1 genomic"
bmtagger="/home/s4645/build/bmtagger/bmtagger.sh"

for sample in ${inpath}2021_C*A*
do
mkdir ${sample}/bmtagger_tmp
id=${sample#${inpath}}
echo $id
${bmtagger} -b ${db}.bitmask -x ${db}.srprism -T ${sample}/bmtagger_tmp -q1 -1 ${sample}/QcReads/QC.1.trimmed.fastq -2 ${sample}/QcReads/QC.2.trimmed.fastq -o ${sample}/QC.removed.bmtagger > ${sample}/bmtagger.log
done
```

- ① Database => 언니가 사전에 만들어 놓 Animal host database  
A or H에 따라 달라짐
- ② 파일 이름 => 2021\_CXXA인 모든 파일  
CXXA, CXXH, JXXA, JXXH, XXXH 등 필요한 파일에 따라 달라짐
- ③ bmtagger 옵션 => paired면 첫 번째, 두 번째 파일 넣기  
paired or unpaired에 따라 달라짐

### 2) 02.bmtagger\_hostrm\_K\_H.sh

```
inpath="/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/"
db="/home/s4645/build/bmtagger/DB/grch38p13/GCF 000001405.39 GRCh38.p13 genomic"
bmtagger="/home/s4645/build/bmtagger/bmtagger.sh"

for sample in ${inpath}2021_K*H*
do
mkdir ${sample}/bmtagger_tmp
id=${sample#${inpath}}
echo $id
${bmtagger} -b ${db}.bitmask -x ${db}.srprism -T ${sample}/bmtagger_tmp -q1 -1 ${sample}/QcReads/QC.1.trimmed.fastq -2 ${sample}/QcReads/QC.2.trimmed.fastq -o ${sample}/QC.removed.bmtagger > ${sample}/bmtagger.log
done
```

- ① Database => Human host database
- ② 파일 이름 => 2021\_KXXH인 모든 파일
- ③ bmtagger 옵션 => paired면 첫 번째, 두 번째 파일 넣기

### 3) 02.bmtagger\_hostrm\_unpaired\_A.sh

```

inpath="/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/"
db="/home/$4645/build/bmtagger/DB/Sscrofall_1/GCF_000003025.6_Sstrefall1.1_genomic"
bmtagger="/home/$4645/build/bmtagger/bmtagger.sh"

for sample in ${inpath}*A*
do
    mkdir ${sample}/bmtagger_unpaired_tmp
    id=${sample#${inpath}}
    echo $id
    ${bmtagger} -b ${db}.bitmask -x ${db}.srprism -T ${sample}/bmtagger_unpaired_tmp -q1 -1 ${sample}/QcReads/QC.unpaired.trimmed.fastq -o ${sample}/QC.unpaired.removed.bmtagger > ${sample}/bmtagger_unpaired.log
done

```

① Database => Animal host database

② 파일 이름 => XXXA인 모든 파일

③ bmtagger 옵션 => unpaired면 unpaired.trimmed.fastq 파일 한 개만 넣기

## 2. Unpaired부터 돌려보기

▷ unpaired A

```

[guest01@smel0:script]$ nohup ./02.bmtagger_hostrm_unpaired_A.sh > /home/guest01/2021/yb/yb01/CDC_JJ/script/02.bmtagger_hostrm_unpaired_A.nhlog
nohup: ignoring input and redirecting stderr to stdout

```

▷ unpaired H

```

[guest01@smel0:script]$ nohup ./02.bmtagger_hostrm_unpaired_H.sh > /home/guest01/2021/yb/yb01/CDC_JJ/script/02.bmtagger_hostrm_unpaired_H.nhlog
nohup: ignoring input and redirecting stderr to stdout

```

▷ unpaired 실행 시 샘플 폴더

```

[guest01@smel-cluster:2021_CF1H_112]$ ll
total 56
-rw-rw-r-- 1 guest01 guest01    0 Feb 22 15:43 bmtagger_unpaired.log
drwxrwxr-x 2 guest01 guest01 4096 Feb 22 15:44 bmtagger_unpaired_tmp
-rw-rw-r-- 1 guest01 guest01 47720 Feb 16 13:01 QC.log
drwxrwxr-x 2 guest01 guest01 4096 Feb 16 13:02 QcReads
[guest01@smel-cluster:2021_CF1H_112]$

```

## ▷ unpaired.nhlog

```
mkdir: cannot create directory '/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF2A_112/bmtagger_unpaired_tmp': File exists
2021_CF2A_112
Info: no ./bmtagger.conf found
Using following programs:
/home/s4645/build/bmtagger/bmfilter
/home/s4645/build/bmtagger/srprism
/home/bioware/ncbi-blast-2.7.1+/bin/blastn
/home/bioware/bin/extract_fullseq
MAIN SCRIPT IS /home/s4645/build/bmtagger/bmtagger.sh (PID=45486)
RUNNING bmfilter
* Attaching /home/s4645/build/bmtagger/DB/Scrofall_1/GCF_000003025.6_Scrofall_1_genomic.bitmask.
(pattern = 0b1111111111111111 of len 18 using 36 bits)
* Notice: creating CFastqFileReader( 1, 0, /home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF2A_112/QcReads/QC.unpaired.trimmed.fastq, )
* Info: Created /home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF2A_112/bmtagger_unpaired_tmp/bmtagger.1645512129.smel0.45486.tag
* Info: Created /home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF2A_112/bmtagger_unpaired_tmp/bmtagger.1645512129.smel0.45486.short.fa
Completed: [15:42:35,2022-02-22] 0b Processing files: 259.7k op in 18s (14.4k op/sec)

real    0m25.761s
user    0m13.840s
sys      0m9.481s
RUNNING align_short for '/home/guest01/2021/yb/yb01/CDC_JJ/Analysis/2021_CF2A_112/bmtagger_unpaired_tmp/bmtagger.1645512129.smel0.45486.short'
reading queries....done
Mpass 1: searching...0% Mpass 1: searching...0.100192% Mpass 1: searching...0.200271% Mpass 1: searching...0.300351% Mpass 1: searching...0.40043% Mpass 1: searching...0.500509% Mpass 1: searching...0.600589% Mpass 1: searching...0.700668% Mpass 1: searching...0.800748% Mpass 1: searching...0.900827% Mpass 1: searching...1.00091% Mpass 1: searching...1.10099% Mpass 1: searching...1.20107% Mpass 1: searching...1.30114% Mpass 1: searching...1.40122% Mpass 1: searching...1.5013% Mpass 1: searching...1.60138% Mpass 1: searching...1.70146% Mpass 1: searching...1.80154% Mpass 1: searching...1.90162% Mpass 1: searching...2.0017% Mpass 1: searching...2.10178% Mpass 1: searching...2.20186% Mpass 1: searching...2.30194% Mpass 1: searching...2.40202% Mpass 1: searching...2.5021% Mpass 1: searching...2.60218% Mpass 1: searching...2.70226% Mpass 1: searching...2.80234% Mpass 1: searching...2.90242% Mpass 1: searching...3.0025% Mpass 1: searching...3.10257% Mpass 1: searching...3.20265% Mpass 1: searching...3.30273% Mpass 1: searching...3.40281% Mpass 1: searching...3.50289% Mpass 1: searching...3.6029% Mpass 1: searching...3.70305% Mpass 1: searching...3.80313% Mpass 1: searching...3.90321% Mpass 1: searching...4.00329% Mpass 1:
```

## ▷ QC.unpaired.removed.bmtagger

```
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
```

\*\* 샘플에서 host seq인 부분을 알려줌

## ▷ QC.unpaired.trimmed.fastq

```
@A00721:393:H3NNLDSX3:3:1101:30888:9706 1:N:0:CGGCTATG+AGGATAGG
CTCTCAGTTCATACTGTAAGTTCATGCATCAGACTATATTGAAATTGCCTCTTTGTTGGCAAGACACTTTTTTTGTATTTTTTAAATTAATATATAG
TTGATTACAATACTGTGACACTTTGTGATAATAAGGTCTTCGAAGGCAGGGT
+
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF
```

\*\* 실제 샘플에서 /A00721:393:H3NNLDSX3:3:1101:30888:9706 검색 하면 해당 시퀀스를 볼 수 있다.



### 3. Paired 돌려보기

expanding queries...done

real 0m35.021s  
user 0m23.471s  
sys 0m11.530s

real 0m0.003s  
user 0m0.000s  
sys 0m0.002s

real 0m0.002s  
user 0m0.000s  
sys 0m0.001s

ONE /home/s4645/build/bmtagger/bmtagger.sh (PID=66959)

다 끝나면 해당 nhlog에 done

품

```
[guest01@smel0:script]$ ll
total 16304
drwxrwxr-x 3 guest01 guest01 104 Feb 23 10:29 01.QC
-rw-rw-r-- 1 guest01 guest01 3425831 Feb 25 05:04 02.bmtagger_hostrm_C_A.nhlog
-rwxr-xr-x 1 guest01 guest01 1119 Feb 22 15:17 02.bmtagger_hostrm_C_A.sh
-rw-rw-r-- 1 guest01 guest01 1869265 Feb 24 11:00 02.bmtagger_hostrm_C_H.nhlog
-rwxr-xr-x 1 guest01 guest01 1117 Feb 22 15:30 02.bmtagger_hostrm_C_H.sh
-rwxr-xr-x 1 guest01 guest01 1122 Feb 22 15:26 02.bmtagger_hostrm_human.sh
-rwxr-xr-x 1 guest01 guest01 1119 Feb 22 15:14 02.bmtagger_hostrm_J_A.sh
-rw-rw-r-- 1 guest01 guest01 1403926 Feb 24 01:28 02.bmtagger_hostrm_J_H.nhlog
-rwxr-xr-x 1 guest01 guest01 1117 Feb 22 15:31 02.bmtagger_hostrm_J_H.sh
-rw-rw-r-- 1 guest01 guest01 1880433 Feb 24 12:41 02.bmtagger_hostrm_K_A.nhlog
-rwxr-xr-x 1 guest01 guest01 1119 Feb 22 15:16 02.bmtagger_hostrm_K_A.sh
-rw-rw-r-- 1 guest01 guest01 2022078 Feb 24 09:42 02.bmtagger_hostrm_K_H.nhlog
-rwxr-xr-x 1 guest01 guest01 1117 Feb 22 15:31 02.bmtagger_hostrm_K_H.sh
-rw-rw-r-- 1 guest01 guest01 3463286 Feb 22 17:21 02.bmtagger_hostrm_unpaired_A.nhlog
-rwxr-xr-x 1 guest01 guest01 1132 Feb 22 15:38 02.bmtagger_hostrm_unpaired_A.sh
-rw-rw-r-- 1 guest01 guest01 2556575 Feb 22 16:51 02.bmtagger_hostrm_unpaired_H.nhlog
-rwxr-xr-x 1 guest01 guest01 1130 Feb 22 15:38 02.bmtagger_hostrm_unpaired_H.sh
```

언니 꺼(예시)

J\_A.nhlog

J\_H.nhlog

따로 적어야 하는데  
실수로 똑같은 거에  
적음

끝난 후 Analysis에 샘플 폴더에 들어가면

```
[guest01@smel0:2021_KF2H_224]$ ll
total 968
-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 46334 Feb 17 20:43 QC.log
drwxrwxr-x 2 guest01 guest01 4096 Feb 17 20:44 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
```

이렇게 표시됨