BWA + SAMTOOLS_(1)

2022/02/09

20193852 문유빈

1. 프로그램 설명

① Bwa => 리드맵핑 해주는 프로그램, assembly 진행 후 리드들이 contig에 어떻게 붙어 있는지 맵핑, abundance 등을 알 수 있음

** 리드 맵핑에 bws, bowtie2 프로그램 두 개를 주로 사용

bwa 입력

```
[guest01@smel0:yb01]$ bwa

Program: bwa (alignment via Burrows-Wheeler transformation)
Version: 0.7.17-r1188
Contact: Heng Li <lh3@sanger.ac.uk>

Usage:   bwa <command> [options]

Command: index         index sequences in the FASTA format
         mem           BWA-MEM algorithm
         fastmap       identify super-maximal exact matches
         pemerge       merge overlapping paired ends (EXPERIMENTAL)
         aln           gapped/ungapped alignment
         samse         generate alignment (single ended)
         sampe         generate alignment (paired ended)
         bwasw         BWA-SW for long queries

         shm           manage indices in shared memory
         fa2pac        convert FASTA to PAC format
         pac2bwt       generate BWT from PAC
         pac2bwtgen    alternative algorithm for generating BWT
         bwtupdate     update .bwt to the new format
         bwt2sa        generate SA from BWT and Occ

Note: To use BWA, you need to first index the genome with `bwa index'.
      There are three alignment algorithms in BWA: `mem', `bwasw', and
      `aln/samse/sampe'. If you are not sure which to use, try `bwa mem'
      first. Please `man ./bwa.1' for the manual.
```

bwa mem 입력

```
[guest01@smel0:yb01]$ bwa mem

Usage: bwa mem [options] <idxbase> <in1.fq> [in2.fq]

Algorithm options:

       -t INT        number of threads [1]
       -k INT        minimum seed length [19]
       -w INT        band width for banded alignment [100]
       -d INT        off-diagonal X-dropoff [100]
       -r FLOAT      look for internal seeds inside a seed longer than {-k} * FLOAT [1.5]
       -y INT        seed occurrence for the 3rd round seeding [20]
       -c INT        skip seeds with more than INT occurrences [500]
       -D FLOAT      drop chains shorter than FLOAT fraction of the longest overlapping cha
in [0.50]
       -W INT        discard a chain if seeded bases shorter than INT [0]
       -m INT        perform at most INT rounds of mate rescues for each read [50]
       -S            skip mate rescue
       -P            skip pairing; mate rescue performed unless -S also in use

Scoring options:
```

** 해당 프로그램의 옵션까지 같이 넣어 검색하면 세부 정보 나옴

ex) bwa mem은 mem 옵션에대한 정보 알려줌

② samtools => BAM, SAM 형태의 파일을 읽고, 쓰고, 조작할 수 있게 해준다.

** samtools 관련 잘 설명된 사이트

https://hhj6212.github.io/biology/tech/2020/10/18/samtools.html

** 파일 형식 설명

https://hhj6212.github.io/biology/tech/2020/08/26/Bioinformatics-fileformats.html

2. 전체 셸 스크립트(진주 언니꺼)

```
input='/home/bbang9/Project/2021/KOPRI/shotgun/WWTP/Analysis/'
output='/home/bbang9/Project/2021/KOPRI/shotgun/WWTP/Analysis/'
bwa='/home/bioware/bwa-0.7.17/bwa'
samtools='/home/bioware/samtools-1.8/samtools'


for file in ${input}*
do
    stub=${file#$input}
    ${bwa} index ${file}/megahit/${stub}.contigs.fa
done

for file in ${input}*
do
    echo $file
    stub=${file#$input}
    echo $stub
    mkdir ${file}/megahit/readsMapping
    mkdir ${file}/megahit/readsMapping/bam
    mkdir ${file}/megahit/readsMapping/stat
    echo "bwa mem, samtools sorting >> bam"
    $bwa mem -t 24 ${file}/megahit/${stub}.contigs.fa ${file}/${stub}".rmd.1.fastq" ${file}
/${stub}".rmd.2.fastq" | $samtools view -@ 16 -ubhS -F4 -| $samtools sort -m 10G -@ 16 -o
${file}/megahit/readsMapping/bam/${stub}_paired.bam
    echo "indexing"
    $samtools index ${file}/megahit/readsMapping/bam/${stub}_paired.bam
    echo "contig stats"
    $samtools idxstats ${file}/megahit/readsMapping/bam/${stub}_paired.bam > ${file}/megahi
t/readsMapping/stat/${stub}_stat.contig.txt
    echo "mapping stat and genome mapped coverage calculation"
    $samtools flagstat ${file}/megahit/readsMapping/bam/${stub}_paired.bam > ${file}/megahi
t/readsMapping/stat/${stub}_stat.txt
done

"05.readmapping.sh" 30L, 1249C                                2,1           All
```

1

만들어 줘야 하는 디렉토리

2

3

4

5

** ①→②→③→④→⑤ 순서대로 진행해야 함

2-1>

① /home/bioware/bwa-0.7.17/bwa index ② /home/guest01/2021/yb/yb01/ybMegahit/PG-16-sps-01_megahit/PG-16-sps-01.contigs.fa

① bwa 프로그램 경로(버전)
② index assembly 후의 contig 파일 경로

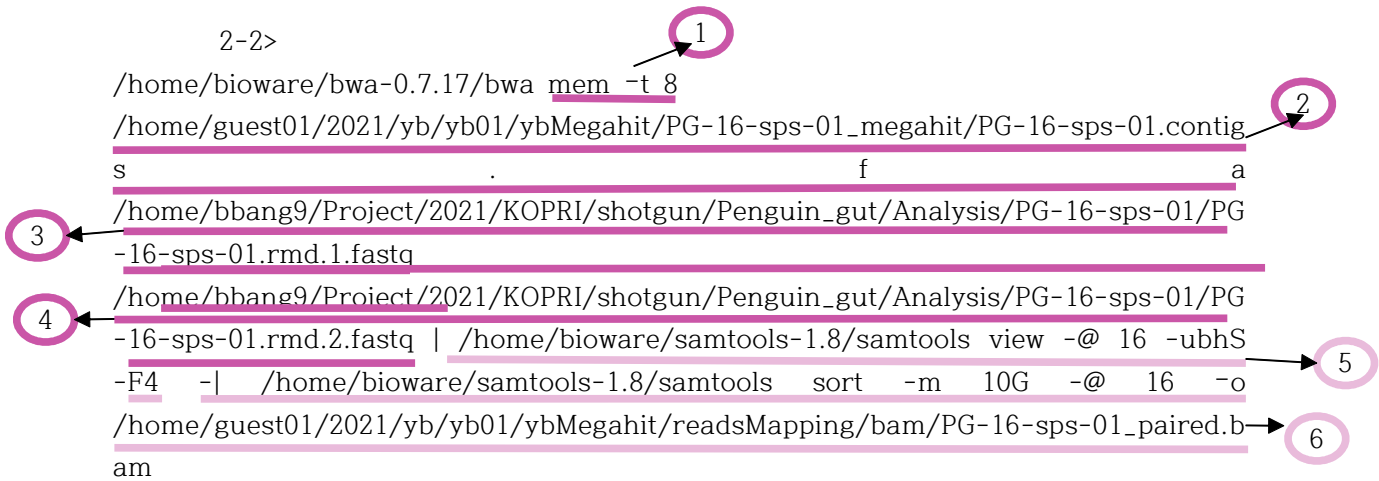
실행 중

```
[guest01@smel0:ybBwa]$ /home/bioware/bwa-0.7.17/bwa index /home/guest01/2021/yb/yb01/ybMegahit/PG-16-sps-01_megahit/PG-16-sps-01.contigs.fa
[bwa_index] Pack FASTA... 7.14 sec
[bwa_index] Construct BWT for the packed sequence...
[BWTIncCreate] textLength=1519770024, availableWord=118936152
[BWTIncConstructFromPacked] 10 iterations done. 99999992 characters processed.
[BWTIncConstructFromPacked] 20 iterations done. 199999992 characters processed.
[BWTIncConstructFromPacked] 30 iterations done. 299999992 characters processed.
[BWTIncConstructFromPacked] 40 iterations done. 399999992 characters processed.
[BWTIncConstructFromPacked] 50 iterations done. 499999992 characters processed.
[BWTIncConstructFromPacked] 60 iterations done. 599999992 characters processed.
[BWTIncConstructFromPacked] 70 iterations done. 699999992 characters processed.
[BWTIncConstructFromPacked] 80 iterations done. 799999992 characters processed.
[BWTIncConstructFromPacked] 90 iterations done. 898125368 characters processed.
[BWTIncConstructFromPacked] 100 iterations done. 986312232 characters processed.
[BWTIncConstructFromPacked] 110 iterations done. 1064688936 characters processed.
[BWTIncConstructFromPacked] 120 iterations done. 1134346328 characters processed.
[BWTIncConstructFromPacked] 130 iterations done. 1196253992 characters processed.
[BWTIncConstructFromPacked] 140 iterations done. 1251273688 characters processed.
[BWTIncConstructFromPacked] 150 iterations done. 1300171320 characters processed.
[BWTIncConstructFromPacked] 160 iterations done. 1343627608 characters processed.
[BWTIncConstructFromPacked] 170 iterations done. 1382247656 characters processed.
[BWTIncConstructFromPacked] 180 iterations done. 1416569192 characters processed.
[BWTIncConstructFromPacked] 190 iterations done. 1447070264 characters processed.
[BWTIncConstructFromPacked] 200 iterations done. 1474175656 characters processed.
[BWTIncConstructFromPacked] 210 iterations done. 1498262936 characters processed.
```

실행 완료시 megahit 폴더에 contigs.fa 파일 외에 새로운 파일들이 생성 됨

```
[guest01@smel1:PG-16-sps-01_megahit]$ ll
total 2110720
-rw-rw-r-- 1 guest01 guest01         0 Jan 20 20:18 done
drwxrwxr-x 2 guest01 guest01      4096 Jan 21 10:04 intermediate_contigs
-rw-rw-r-- 1 guest01 guest01       469 Jan 20 15:22 opts.txt
-rw-rw-r-- 1 guest01 guest01 789891711 Jan 20 20:18 PG-16-sps-01.contigs.fa
-rw-rw-r-- 1 guest01 guest01        19 Feb  9 17:03 PG-16-sps-01.contigs.fa.amb
-rw-rw-r-- 1 guest01 guest01  41518137 Feb  9 17:03 PG-16-sps-01.contigs.fa.ann
-rw-rw-r-- 1 guest01 guest01 759885108 Feb  9 17:03 PG-16-sps-01.contigs.fa.bwt
-rw-rw-r-- 1 guest01 guest01 189971255 Feb  9 17:03 PG-16-sps-01.contigs.fa.pac
-rw-rw-r-- 1 guest01 guest01 379942560 Feb  9 17:07 PG-16-sps-01.contigs.fa.sa
-rw-rw-r-- 1 guest01 guest01    145715 Jan 20 20:18 PG-16-sps-01.log
```

2-2>




```
/home/bioware/bwa-0.7.17/bwa mem -t 8
/home/guest01/2021/yb/yb01/ybMegahit/PG-16-sps-01_megahit/PG-16-sps-01.contigs.fa
/home/bbang9/Project/2021/KOPRI/shotgun/Penguin_gut/Analysis/PG-16-sps-01/PG-16-sps-01.rmd.1.fastq
/home/bbang9/Project/2021/KOPRI/shotgun/Penguin_gut/Analysis/PG-16-sps-01/PG-16-sps-01.rmd.2.fastq | /home/bioware/samtools-1.8/samtools view -@ 16 -ubhS -F4 -| /home/bioware/samtools-1.8/samtools sort -m 10G -@ 16 -o /home/guest01/2021/yb/yb01/ybMegahit/readsMapping/bam/PG-16-sps-01_paired.bam
```

** 찐한 보라 - bwa / 연한 보라 - samtools

① mem -t 8 => CPU thread 8로 돌린다.

②
/home/guest01/2021/yb/yb01/ybMegahit/PG-16-sps-01_megahit/PG-16-sps-01.contigs.fa => mapping 시도할 contig 파일 ** 내 파일에 있는 것

③
/home/bbang9/Project/2021/KOPRI/shotgun/Penguin_gut/Analysis/PG-16-sps-01/PG-16-sps-01.rmd.1.fastq => 리드 1(forward)

**언니 파일에 있는 것

④
/home/bbang9/Project/2021/KOPRI/shotgun/Penguin_gut/Analysis/PG-16-sps-01/PG-16-sps-01.rmd.2.fastq => 리드 2(backward)

⑤ | /home/bioware/samtools-1.8/samtools view -@ 16 -ubhS -F4
=> samtools의 view 설정 (samtools view 입력시 상세히 나옴)
=> -F4는 unmapped 된 거 포함 x

⑥ -| /home/bioware/samtools-1.8/samtools sort -m 10G -@ 16 -o /home/guest01/2021/yb/yb01/ybMegahit/readsMapping/bam/PG-16-sps-01_paired.bam => samtools의 sort 설정
=> -o로 output 디렉토리 설정(미리 만들어준 readsMapping 폴더에 넣기)

끝

```
[main] CMD: /home/bioware/bwa-0.7.17/bwa mem -t 8 /home/guest01/
lysis/PG-16-sps-01/PG-16-sps-01.rmd.1.fastq /home/bbang9/Project
[main] Real time: 2380.181 sec; CPU: 19078.432 sec
[bam_sort_core] merging from 0 files and 16 in-memory blocks...
[guest01@smel1:yb01]$ ll
```

cd bam

```
[guest01@smel1:bam]$ ll
total 2681848
-rw-rw-r-- 1 guest01 guest01 2746200573 Feb  9 19:10 PG-16-sps-01_paired.bam
```