Workshop 6: DNA Methylation Analysis using Bisulfite Sequencing

Fides D Lay
UCLA
QCB Fellow
lay.fides@gmail.com

Workshop 6 Outline

Day 1:

Introduction to DNA methylation & WGBS
Quick review of linux, Hoffman2 and high-throughput sequencing glossary.
Aligning WGBS reads using bwa-meth

Day 2:

DNA methylation calling using Bis-SNP
Analysis of differentially methylated regions (DMRs) using metilene

Day 3:

Visualization of DNA methylation data
WGBS analysis using BS-Seeker2

Day 3

Guo et al. BMC Genomics 2013, 14:774
http://www.biomedcentral.com/1471-2164/14/774

BMC Genomics

SOFTWARE Open Access

BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data

Weilong Guo[1,2], Petko Fiziev[3], Weihong Yan[4], Shawn Cokus[2], Xueguang Sun[5], Michael Q Zhang[1,6], Pao-Yang Chen[7*] and Matteo Pellegrini[2,8*]

Abstract

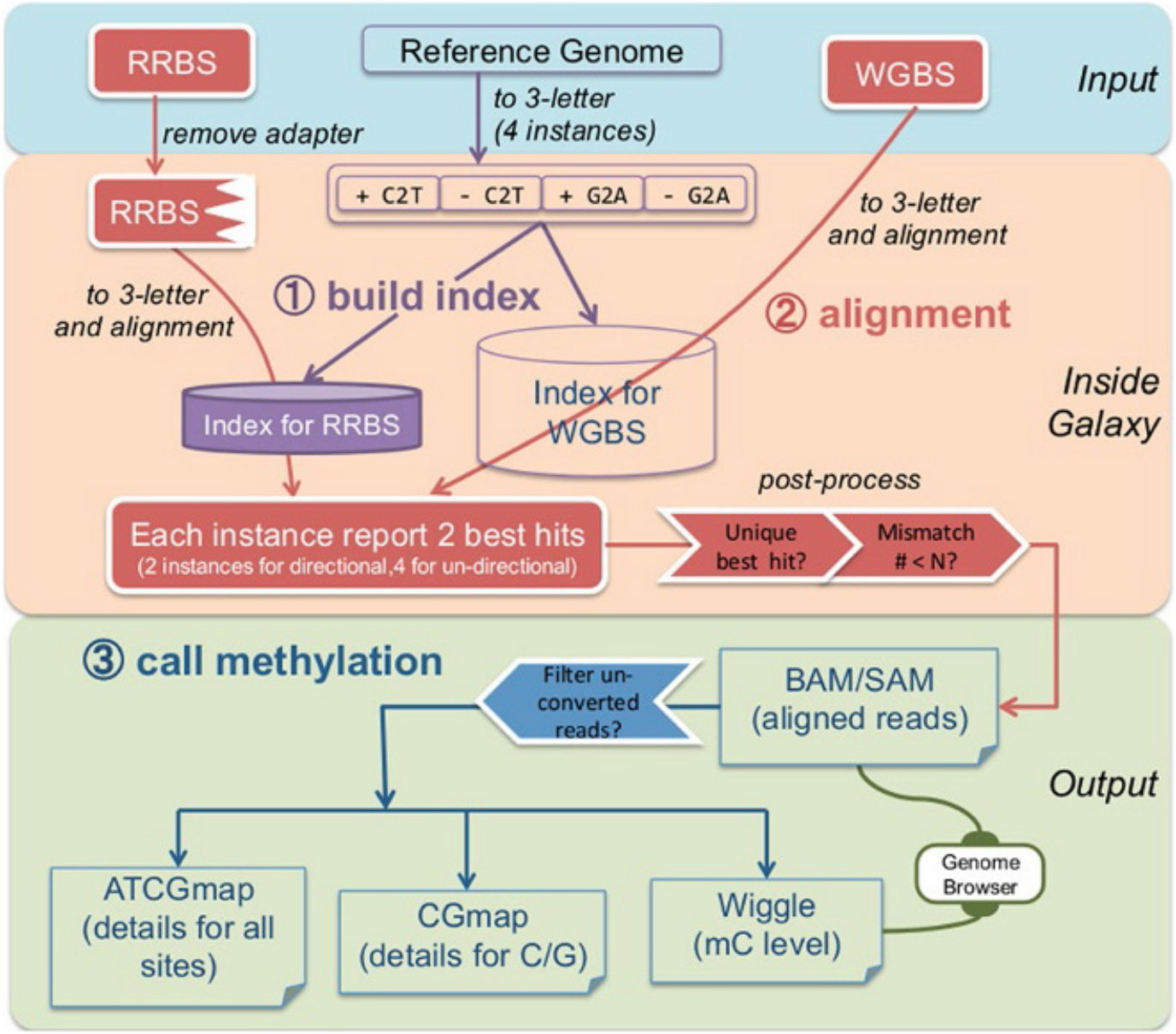
Background: DNA methylation is an important epigenetic modification involved in many biological processes. Bisulfite treatment coupled with high-throughput sequencing provides an effective approach for studying genome-wide DNA methylation at base resolution. Libraries such as whole genome bisulfite sequencing (WGBS) and reduced represented bisulfite sequencing (RRBS) are widely used for generating DNA methylomes, demanding efficient and versatile tools for aligning bisulfite sequencing data.

Results: We have developed BS-Seeker2, an updated version of BS Seeker, as a full pipeline for mapping bisulfite sequencing data and generating DNA methylomes. BS-Seeker2 improves mappability over existing aligners by using local alignment. It can also map reads from RRBS library by building special indexes with improved efficiency and accuracy. Moreover, BS-Seeker2 provides additional function for filtering out reads with incomplete bisulfite conversion, which is useful in minimizing the overestimation of DNA methylation levels. We also defined CGmap and ATCGmap file formats for full representations of DNA methylomes, as part of the outputs of BS-Seeker2 pipeline together with BAM and WIG files.

Conclusions: Our evaluations on the performance show that BS-Seeker2 works efficiently and accurately for both WGBS data and RRBS data. BS-Seeker2 is freely available at http://pellegrini.mcdb.ucla.edu/BS_Seeker2/ and the Galaxy server.

Keywords: DNA methylation, Bisulfite sequencing aligner, WGBS, RRBS, BS Seeker, Bisulfite conversion failure, Galaxy toolshed

BS-Seeker2: Workflow



Guo et al, 2013

BS-Seeker2

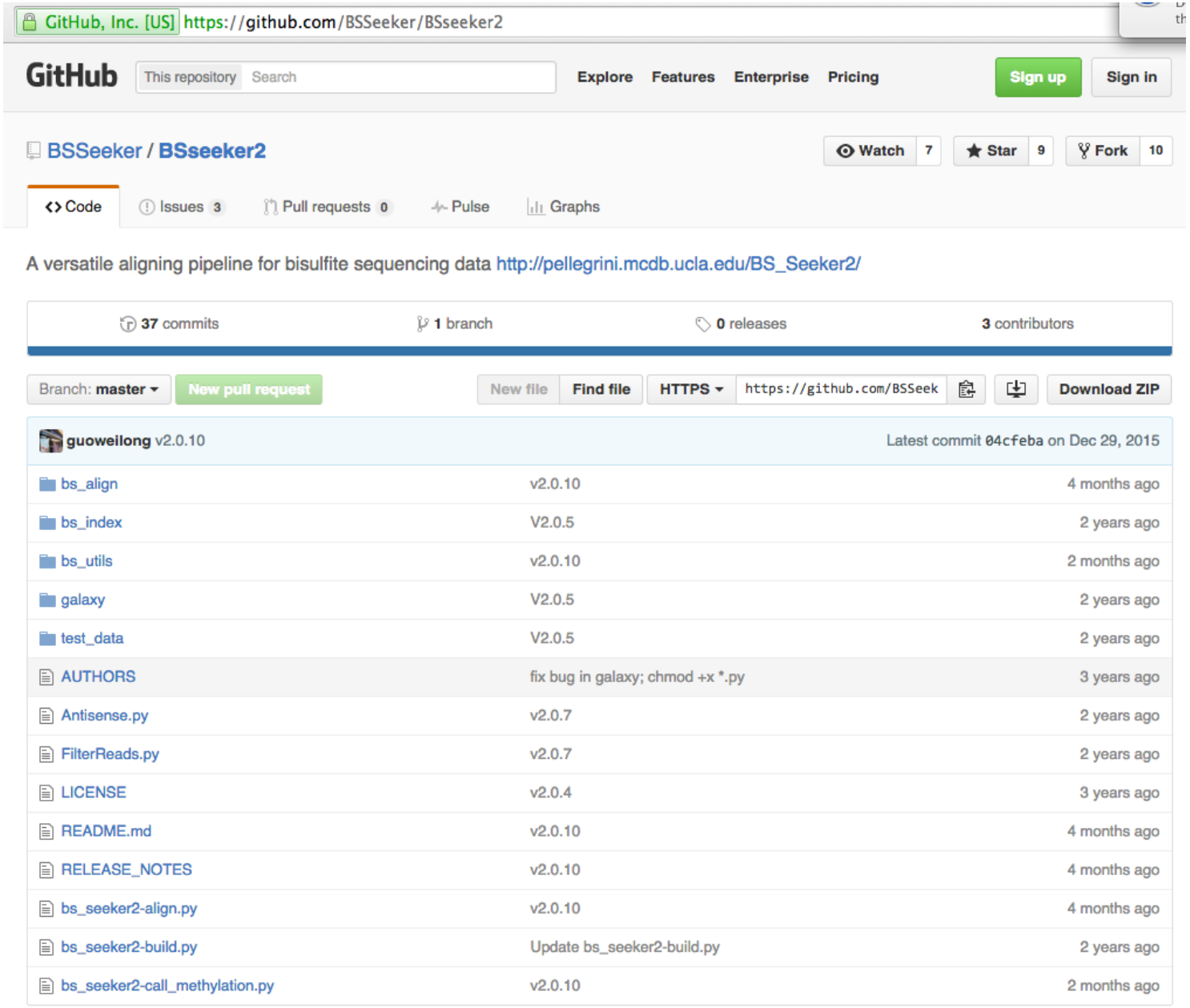
Advantages:

- Accept multiple aligners: bowtie, bowtie2, soap, rmap
- Accept multiple input files: qseq, fastq, fasta
- Can handle RRBS and non-directional libraries efficiently with minimal preprocessing
- Built-in trimming and methylation calling tool
- Built-in galaxy interface

Disadvantages:

- Slow speed
- Requires a lot of memory

Installing BS-Seeker2



Installing BS-Seeker2

#In your $HOME/software, git clone BS-Seeker2

git clone https://github.com/BSSeeker/BSseeker2.git

#enter directory

cd BSSeeker2

ls

```
[flay@login1 BSseeker2]$ ls -lh
total 140K
-rwxr-xr-x 1 flay matteop 6.8K Mar  9 16:50 Antisense.py
-rw-r--r-- 1 flay matteop  142 Mar  9 16:50 AUTHORS
drwxr-xr-x 2 flay matteop 4.0K Mar  9 16:50 bs_align
drwxr-xr-x 2 flay matteop 4.0K Mar  9 16:55 bs_index
-rwxr-xr-x 1 flay matteop  23K Mar  9 16:50 bs_seeker2-align.py
-rwxr-xr-x 1 flay matteop 4.8K Mar  9 16:50 bs_seeker2-build.py
-rwxr-xr-x 1 flay matteop  15K Mar  9 16:50 bs_seeker2-call_methylation.py
drwxr-xr-x 3 flay matteop 4.0K Mar  9 16:57 bs_utils
-rwxr-xr-x 1 flay matteop 7.7K Mar  9 16:50 FilterReads.py
drwxr-xr-x 2 flay matteop 4.0K Mar  9 16:50 galaxy
-rw-r--r-- 1 flay matteop 1.1K Mar  9 16:50 LICENSE
-rw-r--r-- 1 flay matteop  36K Mar  9 16:50 README.md
-rw-r--r-- 1 flay matteop 3.2K Mar  9 16:50 RELEASE_NOTES
drwxr-xr-x 2 flay matteop 4.0K Mar  9 16:50 test_data
```

Indexing Reference Genome

#Load up the modules

module load python

module load bowtie2

```
[flay@login1 BSseeker2]$ python bs_seeker2-build.py
Usage: bs_seeker2-build.py [options]

Options:
  -h, --help            show this help message and exit
  -f FILE, --file=FILE  Input your reference genome file (fasta)
  --aligner=ALIGNER     Aligner program to perform the analysis: bowtie,
                        bowtie2, soap, rmap [Default: bowtie]
  -p PATH, --path=PATH  Path to the aligner program. Detected:
                        bowtie: None
                        bowtie2: /u/local/apps/bowtie2/2.2.5
                        rmap: None
                        soap: None
  -d DBPATH, --db=DBPATH
                        Path to the reference genome library (generated in
                        preprocessing genome) [Default: /u/home/f/flay/softwar
                        e/BSseeker2/bs_utils/reference_genomes]
  -v, --version         show version of BS-Seeker2

  Reduced Representation Bisulfite Sequencing Options:
    Use this options with conjuction of -r [--rrbs]

    -r, --rrbs          Build index specially for Reduced Representation
                        Bisulfite Sequencing experiments. Genome other than
                        certain fragments will be masked. [Default: False]
    -l LOW_BOUND, --low=LOW_BOUND
                        lower bound of fragment length (excluding recognition
                        sequence such as C-CGG) [Default: 20]
    -u UP_BOUND, --up=UP_BOUND
                        upper bound of fragment length (excluding recognition
                        sequence such as C-CGG ends) [Default: 500]
    -c CUT_FORMAT, --cut-site=CUT_FORMAT
                        Cut sites of restriction enzyme. Ex: MspI(C-CGG),
                        Mael:(C-TAG), double-enzyme MspI&Mael:(C-CGG,C-TAG).
                        [Default: C-CGG]
```

Indexing Reference Genome

#Use chr1 as example:

cd /u/scratch/f/flay/workshop6/genome/chr1

python /u/home/f/flay/software/BSseeker2/bs_seeker2-build.py -f chr1.fa --aligner bowtie2 --db /u/scratch/f/flay/workshop6/

```
[flay@n2239 chr1]$ python /u/home/f/flay/software/BSseeker2/bs_seeker2-build.py -f chr1.fa --aligner bowtie2 --db /u/scratch/f/flay/workshop6/genome/chr1

     BS-Seeker2 v2.0.10 - Nov 5, 2015

Reference genome file: chr1.fa
Reduced Representation Bisulfite Sequencing: False
Short reads aligner you are using: bowtie2
Builder path: /u/local/apps/bowtie2/2.2.5/bowtie2-build
[Preprocessing chr1] Last: 0:00:27.668786       Total: 0:00:27.668879
[Genome preprocessing] Last: 0:00:00.319565     Total: 0:00:27.988525
[2016-07-05 18:07:35] Starting commands:
[2016-07-05 18:07:35] Launched: /u/local/apps/bowtie2/2.2.5/bowtie2-build -f /u/scratch/f/flay/workshop6/genome/chr1/chr1.fa_bowtie2/W_C2T.fa /u/scratch/f/f
lay/workshop6/genome/chr1/chr1.fa_bowtie2/W_C2T
[2016-07-05 18:07:35] Launched: /u/local/apps/bowtie2/2.2.5/bowtie2-build -f /u/scratch/f/flay/workshop6/genome/chr1/chr1.fa_bowtie2/W_G2A.fa /u/scratch/f/f
lay/workshop6/genome/chr1/chr1.fa_bowtie2/W_G2A
[2016-07-05 18:07:35] Launched: /u/local/apps/bowtie2/2.2.5/bowtie2-build -f /u/scratch/f/flay/workshop6/genome/chr1/chr1.fa_bowtie2/C_C2T.fa /u/scratch/f/f
lay/workshop6/genome/chr1/chr1.fa_bowtie2/C_C2T
[2016-07-05 18:07:35] Launched: /u/local/apps/bowtie2/2.2.5/bowtie2-build -f /u/scratch/f/flay/workshop6/genome/chr1/chr1.fa_bowtie2/C_G2A.fa /u/scratch/f/f
lay/workshop6/genome/chr1/chr1.fa_bowtie2/C_G2A
Building a SMALL index
Building a SMALL index
Building a SMALL index
Building a SMALL index
[2016-07-05 18:16:33] Finished: /u/local/apps/bowtie2/2.2.5/bowtie2-build -f /u/scratch/f/flay/workshop6/genome/chr1/chr1.fa_bowtie2/W_C2T.fa /u/scratch/f/f
lay/workshop6/genome/chr1/chr1.fa_bowtie2/W_C2T
[2016-07-05 18:16:33] Finished: /u/local/apps/bowtie2/2.2.5/bowtie2-build -f /u/scratch/f/flay/workshop6/genome/chr1/chr1.fa_bowtie2/W_G2A.fa /u/scratch/f/f
lay/workshop6/genome/chr1/chr1.fa_bowtie2/W_G2A
[2016-07-05 18:16:59] Finished: /u/local/apps/bowtie2/2.2.5/bowtie2-build -f /u/scratch/f/flay/workshop6/genome/chr1/chr1.fa_bowtie2/C_C2T.fa /u/scratch/f/f
lay/workshop6/genome/chr1/chr1.fa_bowtie2/C_C2T
[2016-07-05 18:16:59] Finished: /u/local/apps/bowtie2/2.2.5/bowtie2-build -f /u/scratch/f/flay/workshop6/genome/chr1/chr1.fa_bowtie2/C_G2A.fa /u/scratch/f/f
lay/workshop6/genome/chr1/chr1.fa_bowtie2/C_G2A
[Done] Last: 0:09:24.064488     Total: 0:09:52.053089
[flay@n2239 chr1]$
```

Indexing Reference Genome: Output

```
[flay@n46 bsseeker2]$ cd chr1.fa_bowtie2/
[flay@n46 chr1.fa_bowtie2]$ ls -lh
total 1.4G
-rw-r--r-- 1 flay matteop  65M Mar  9 17:16 C_C2T.1.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:16 C_C2T.2.bt2
-rw-r--r-- 1 flay matteop  332 Mar  9 17:12 C_C2T.3.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:11 C_C2T.4.bt2
-rw-r--r-- 1 flay matteop  12K Mar  9 17:22 C_C2T.log
-rw-r--r-- 1 flay matteop  65M Mar  9 17:22 C_C2T.rev.1.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:22 C_C2T.rev.2.bt2
-rw-r--r-- 1 flay matteop  65M Mar  9 17:16 C_G2A.1.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:16 C_G2A.2.bt2
-rw-r--r-- 1 flay matteop  332 Mar  9 17:12 C_G2A.3.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:11 C_G2A.4.bt2
-rw-r--r-- 1 flay matteop  11K Mar  9 17:21 C_G2A.log
-rw-r--r-- 1 flay matteop  65M Mar  9 17:21 C_G2A.rev.1.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:21 C_G2A.rev.2.bt2
-rw-r--r-- 1 flay matteop 189M Mar  9 17:11 chr1.data
-rw-r--r-- 1 flay matteop 1.5K Mar  9 17:22 log
-rw-r--r-- 1 flay matteop   16 Mar  9 17:11 refname.data
-rw-r--r-- 1 flay matteop  65M Mar  9 17:17 W_C2T.1.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:17 W_C2T.2.bt2
-rw-r--r-- 1 flay matteop  323 Mar  9 17:12 W_C2T.3.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:11 W_C2T.4.bt2
-rw-r--r-- 1 flay matteop  12K Mar  9 17:22 W_C2T.log
-rw-r--r-- 1 flay matteop  65M Mar  9 17:22 W_C2T.rev.1.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:22 W_C2T.rev.2.bt2
-rw-r--r-- 1 flay matteop  65M Mar  9 17:17 W_G2A.1.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:17 W_G2A.2.bt2
-rw-r--r-- 1 flay matteop  323 Mar  9 17:12 W_G2A.3.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:11 W_G2A.4.bt2
-rw-r--r-- 1 flay matteop  12K Mar  9 17:21 W_G2A.log
-rw-r--r-- 1 flay matteop  65M Mar  9 17:21 W_G2A.rev.1.bt2
-rw-r--r-- 1 flay matteop  46M Mar  9 17:21 W_G2A.rev.2.bt2
```

BS-Seeker2 Alignment Options

```
[flay@login1 BSseeker2]$ python bs_seeker2-align.py
Usage: bs_seeker2-align.py {-i <single> | -1 <mate1> -2 <mate2>} -g <genome.fa> [options]

Options:
  -h, --help            show this help message and exit

  For single end reads:
    -i INFILE, --input=INFILE
                        Input read file (FORMAT: sequences, qseq, fasta,
                        fastq). Ex: read.fa or read.fa.gz

  For pair end reads:
    -1 FILE, --input_1=FILE
                        Input read file, mate 1 (FORMAT: sequences, qseq,
                        fasta, fastq)
    -2 FILE, --input_2=FILE
                        Input read file, mate 2 (FORMAT: sequences, qseq,
                        fasta, fastq)
    -I MIN_INSERT_SIZE, --minins=MIN_INSERT_SIZE
                        The minimum insert size for valid paired-end
                        alignments [Default: 0]
    -X MAX_INSERT_SIZE, --maxins=MAX_INSERT_SIZE
                        The maximum insert size for valid paired-end
                        alignments [Default: 500]

  Reduced Representation Bisulfite Sequencing Options:
    -r, --rrbs          Map reads to the Reduced Representation genome
    -c pattern, --cut-site=pattern
                        Cutting sites of restriction enzyme. Ex: MspI(C-CGG),
                        MaeI:(C-TAG), double-enzyme MspI&MaeI:(C-CGG,C-TAG).
                        [Default: C-CGG]
    -L RRBS_LOW_BOUND, --low=RRBS_LOW_BOUND
                        Lower bound of fragment length (excluding C-CGG ends)
                        [Default: 20]
    -U RRBS_UP_BOUND, --up=RRBS_UP_BOUND
                        Upper bound of fragment length (excluding C-CGG ends)
                        [Default: 500]
```

BS-Seeker2 Alignment Options

```
General options:
  -t TAG, --tag=TAG     [Y]es for undirectional lib, [N]o for directional
                        [Default: N]
  -s CUTNUMBER1, --start_base=CUTNUMBER1
                        The first cycle of the read to be mapped [Default: 1]
  -e CUTNUMBER2, --end_base=CUTNUMBER2
                        The last cycle of the read to be mapped [Default: 200]
  -a FILE, --adapter=FILE
                        Input text file of your adaptor sequences (to be
                        trimmed from the 3'end of the reads, ). Input one seq
                        for dir. lib., twon seqs for undir. lib. One line per
                        sequence. Only the first 10bp will be used
  --am=ADAPTER_MISMATCH
                        Number of mismatches allowed in adapter [Default: 0]
  -g GENOME, --genome=GENOME
                        Name of the reference genome (should be the same as
                        "-f" in bs_seeker2-build.py ) [ex. chr21_hg18.fa]
  -m NO_MISMATCHES, --mismatches=NO_MISMATCHES
                        Number(>=1)/Percentage([0, 1)) of mismatches in one
                        read. Ex: 4 (allow 4 mismatches) or 0.04 (allow 4%
                        mismatches) [Default: 4]
  --aligner=ALIGNER     Aligner program for short reads mapping: bowtie,
                        bowtie2, soap, rmap [Default: bowtie]
  -p PATH, --path=PATH
                        Path to the aligner program. Detected:
                        bowtie: None
                        bowtie2: /u/local/apps/bowtie2/2.2.5
                        rmap: None
                        soap: None
  -d DBPATH, --db=DBPATH
                        Path to the reference genome library (generated in
                        preprocessing genome) [Default: /u/home/f/flay/softwar
                        e/BSseeker2/bs_utils/reference_genomes]
  -l INT, --split_line=INT
                        Number of lines per split (the read file will be split
                        into small files for mapping. The result will be
                        merged. [Default: 4000000]
  -o OUTFILE, --output=OUTFILE
                        The name of output file [INFILE.bs(se|pe|rrbs)]
  -f FORMAT, --output-format=FORMAT
                        Output format: bam, sam, bs_seeker1 [Default: bam]
  --no-header           Suppress SAM header lines [Default: False]
  --temp_dir=PATH       The path to your temporary directory [Detected: /tmp]
  --XS=XS_FILTER        Filter definition for tag XS, format X,Y. X=0.8 and
                        y=5 indicate that for one read, if #(mCH sites)/#(all
                        CH sites)>0.8 and #(mCH sites)>5, then tag XS=1; or
                        else tag XS=0. [Default: 0.5,5]
  -M FileName, --multiple-hit=FileName
                        File to store reads with multiple-hits
  -u FileName, --unmapped=FileName
                        File to store unmapped reads
  -v, --version         show version of BS-Seeker2
```

BS-Seeker2 Alignment

#copy hg19 index files for BS-Seeker2

cd /u/scratch/f/flay/workshop6/genome/
cp -r /u/scratch/f/flay/workshop6/genome/hg19_rCRSchrm.fa_bowtie2/ .

#change directory and make a new directory for BS-Seeker2
cd /u/scratch/f/flay/workshop6/data/
mkdir bs-seeker2
cd bs-seeker2

#to check that you are in the right directory,
    pwd

BS-Seeker2 Alignment

#Alignment using the same fastq files we used previously:

python /u/home/f/flay/software/BSseeker2/bs_seeker2-align.py -1 /u/scratch/f/flay/workshop6/data/raw/N25_R1.fastq -2 /u/scratch/f/flay/workshop6/data/raw/N25_R2.fastq -g hg19_rCRSchrm.fa --aligner bowtie2 -d /u/scratch/f/flay/workshop6/genome/ -f bam -o N25_bsseeker2.bam

BS-Seeker2 Alignment

```
[flay@n6281 bs-seeker2]$ python /u/home/f/flay/software/BSseeker2/bs_seeker2-align.py -1 /u/scratch/f/flay/workshop6/data/raw/N25_R1.fastq -2 /u/scratch/f/flay/workshop6/data/raw/
N25_R2.fastq -g hg19_rCRSchrm.fa --aligner bowtie2 -d /u/scratch/f/flay/workshop6/genome/ -f bam -o N25_bsseeker2.bam

     BS-Seeker2 v2.0.10 - Nov 5, 2015

[2016-07-05 21:16:05] Mode: Bowtie2, local alignment
[2016-07-05 21:16:05] Filter for tag XS: #(mCH)/#(all CH)>50.00% and #(mCH)>5
[2016-07-05 21:16:05] Temporary directory: /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc
[2016-07-05 21:16:05] Reduced Representation Bisulfite Sequencing: False
[2016-07-05 21:16:05] Pair end
[2016-07-05 21:16:05] Aligner command: /u/local/apps/bowtie2/2.2.5/bowtie2 --local --quiet -D 50 --no-mixed --norc --sam-nohead --no-discordant -k 2 -p 2 -X 500 --fr  -x %(referen
ce_genome)s  -f -1 %(input_file_1)s -2 %(input_file_2)s -S %(output_file)s
[2016-07-05 21:16:05] ----------------------------------------------------
[2016-07-05 21:16:05] Filename for 1st mate: /u/scratch/f/flay/workshop6/data/raw/N25_R1.fastq
[2016-07-05 21:16:05] Filename for 2nd mate: /u/scratch/f/flay/workshop6/data/raw/N25_R2.fastq
[2016-07-05 21:16:05] The first base (for mapping): 1
[2016-07-05 21:16:05] The  last base (for mapping): 200
[2016-07-05 21:16:05] Path for short reads aligner: /u/local/apps/bowtie2/2.2.5/bowtie2 --local --quiet -D 50 --no-mixed --norc --sam-nohead --no-discordant -k 2 -p 2 -X 500 --fr
 -x %(reference_genome)s  -f -1 %(input_file_1)s -2 %(input_file_2)s -S %(output_file)s

[2016-07-05 21:16:05] Reference genome library path: /u/scratch/f/flay/workshop6/genome/hg19_rCRSchrm.fa_bowtie2
[2016-07-05 21:16:05] Directional library
[2016-07-05 21:16:05] Number of mismatches allowed: 4
[2016-07-05 21:16:05] ----------------------------------
[2016-07-05 21:16:05] Start reading and trimming the input sequences
Detected data format: fastq
[2016-07-05 21:16:05] Start mapping
[2016-07-05 21:16:05] Starting commands:
[2016-07-05 21:16:05] Launched: /u/local/apps/bowtie2/2.2.5/bowtie2 --local --quiet -D 50 --no-mixed --norc --sam-nohead --no-discordant -k 2 -p 2 -X 500 --fr  -x /u/scratch/f/fla
y/workshop6/genome/hg19_rCRSchrm.fa_bowtie2/W_C2T  -f -1 /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc/Trimed_FCT_1.fa.tmp-9410179 -2 /tmp/bs_seeker2_N25_bsseeker2.b
am_-bowtie2-local-TMP-4Zu4Lc/Trimed_RGA_2.fa.tmp-9410179 -S /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc/W_C2T_fr_m4.mapping.tmp-9410179
[2016-07-05 21:16:05] Launched: /u/local/apps/bowtie2/2.2.5/bowtie2 --local --quiet -D 50 --no-mixed --norc --sam-nohead --no-discordant -k 2 -p 2 -X 500 --fr  -x /u/scratch/f/fla
y/workshop6/genome/hg19_rCRSchrm.fa_bowtie2/C_C2T  -f -1 /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc/Trimed_FCT_1.fa.tmp-9410179 -2 /tmp/bs_seeker2_N25_bsseeker2.b
am_-bowtie2-local-TMP-4Zu4Lc/Trimed_RGA_2.fa.tmp-9410179 -S /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc/C_C2T_fr_m4.mapping.tmp-9410179
[2016-07-05 21:17:44] Finished: /u/local/apps/bowtie2/2.2.5/bowtie2 --local --quiet -D 50 --no-mixed --norc --sam-nohead --no-discordant -k 2 -p 2 -X 500 --fr  -x /u/scratch/f/fla
y/workshop6/genome/hg19_rCRSchrm.fa_bowtie2/W_C2T  -f -1 /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc/Trimed_FCT_1.fa.tmp-9410179 -2 /tmp/bs_seeker2_N25_bsseeker2.b
am_-bowtie2-local-TMP-4Zu4Lc/Trimed_RGA_2.fa.tmp-9410179 -S /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc/W_C2T_fr_m4.mapping.tmp-9410179
[2016-07-05 21:17:44] Finished: /u/local/apps/bowtie2/2.2.5/bowtie2 --local --quiet -D 50 --no-mixed --norc --sam-nohead --no-discordant -k 2 -p 2 -X 500 --fr  -x /u/scratch/f/fla
y/workshop6/genome/hg19_rCRSchrm.fa_bowtie2/C_C2T  -f -1 /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc/Trimed_FCT_1.fa.tmp-9410179 -2 /tmp/bs_seeker2_N25_bsseeker2.b
am_-bowtie2-local-TMP-4Zu4Lc/Trimed_RGA_2.fa.tmp-9410179 -S /tmp/bs_seeker2_N25_bsseeker2.bam_-bowtie2-local-TMP-4Zu4Lc/C_C2T_fr_m4.mapping.tmp-9410179
--> N25_R1.fastq-E1-1 N25_R2.fastq-E2-1 (1/1)
```

BS-Seeker2 Alignment

```
--> N25_R1.fastq-E1-1 N25_R2.fastq-E2-1 (1/1)
[2016-07-05 21:18:39] ----------------------------------
[2016-07-05 21:18:39] Number of raw BS-read pairs: 6250
[2016-07-05 21:18:39] Number of bases in total: 950000
[2016-07-05 21:18:39] Number of reads rejected because of multiple hits: 142
[2016-07-05 21:18:39] Number of unique-hits reads (before post-filtering): 5425

[2016-07-05 21:18:39]      2669 FW-RC pairs mapped to Watson strand (before post-filtering)
[2016-07-05 21:18:39]      2756 FW-RC pairs mapped to Crick strand (before post-filtering)
[2016-07-05 21:18:39]   4777 uniquely aligned pairs, where each end has mismatches <= 4
[2016-07-05 21:18:39]      2336 FW-RC pairs mapped to Watson strand
[2016-07-05 21:18:39]      2441 FW-RC pairs mapped to Crick strand
[2016-07-05 21:18:39] Mappability = 76.4320%
[2016-07-05 21:18:39] Total bases of uniquely mapped reads :  726104
[2016-07-05 21:18:39] Unmapped read pairs: 1473

[2016-07-05 21:18:39] ----------------------------------
[2016-07-05 21:18:39] Methylated C in mapped reads
[2016-07-05 21:18:39]  mCG  74.136%
[2016-07-05 21:18:39]  mCHG 0.850%
[2016-07-05 21:18:39]  mCHH 0.844%
[2016-07-05 21:18:39] ----------------------------------
[2016-07-05 21:18:39] Files : /u/scratch/f/flay/workshop6/data/raw/N25_R1.fastq and /u/scratch/f/flay/workshop6/data/raw/N25_R2.fastq
[Resource / CPU time] Last: 0:02:34.597460      Total: 0:02:34.597481
[2016-07-05 21:18:39] ------------------- END ----------------------
[flay@n6281 bs-seeker2]$
```

BS-Seeker2 Alignment: Output Files

```
[flay@n6281 bs-seeker2]$ ls -lh
total 1.2M
-rw-r--r-- 1 flay matteop 866K Jul  5 21:18 N25_bsseeker2.bam
-rw-r--r-- 1 flay matteop 5.0K Jul  5 21:18 N25_bsseeker2.bam.bs_seeker2_log
```

BS-Seeker2 Methylation Calling

```
[flay@n2175 bs-seeker2]$ python /u/home/f/flay/software/BSseeker2/bs_seeker2-call_methylation.py
Usage: bs_seeker2-call_methylation.py [options]

Options:
  -h, --help            show this help message and exit
  -i INFILE, --input=INFILE
                        BAM output from bs_seeker2-align.py
  -d DBPATH, --db=DBPATH
                        Path to the reference genome library (generated in
                        preprocessing genome) [Default: /u/home/f/flay/softwar
                        e/BSseeker2/bs_utils/reference_genomes]
  -o OUTFILE, --output-prefix=OUTFILE
                        The output prefix to create ATCGmap and wiggle files
                        [INFILE]
  --sorted              Specify when the input bam file is already sorted, the
                        sorting step will be skipped [Default: False]
  --wig=OUTFILE         The output .wig file [INFILE.wig]
  --CGmap=OUTFILE       The output .CGmap file [INFILE.CGmap]
  --ATCGmap=OUTFILE     The output .ATCGmap file [INFILE.ATCGmap]
  -x, --rm-SX           Removed reads with tag 'XS:i:1', which would be
                        considered as not fully converted by bisulfite
                        treatment [Default: False]
  --rm-CCGG             Removed sites located in CCGG, avoiding the bias
                        introduced by artificial DNA methylation status
                        'XS:i:1', which would be considered as not fully
                        converted by bisulfite treatment [Default: False]
  --rm-overlap          Removed one mate if two mates are overlapped, for
                        paired-end data [Default: False]
  --txt                 Show CGmap and ATCGmap in .gz [Default: False]
  -r READ_NO, --read-no=READ_NO
                        The least number of reads covering one site to be
                        shown in wig file [Default: 1]
  -v, --version         show version of BS-Seeker2
```

BS-Seeker2 Methylation Calling

python /u/home/f/flay/software/BSseeker2/bs_seeker2-call_methylation.py -i N25_bsseeker2.bam -o N25_bsseeker2_meth -d /u/scratch/f/flay/workshop6/genome/hg19_rCRSchrm.fa_bowtie2/

BS-Seeker2 Methylation Calling

```
[flay@n6281 bs-seeker2]$
[flay@n6281 bs-seeker2]$ python /u/home/f/flay/software/BSseeker2/bs_seeker2-call_methylation.py -i N25_bsseeker2.bam -o N25_bsseeker2_meth -d /u/scratch/f/flay/workshop6/genome/hg19_rCRSchrm.fa_bowtie2/

    BS-Seeker2 v2.0.10 - Nov 5, 2015

[2016-07-05 21:38:19] sorting BS-Seeker alignments
[2016-07-05 21:38:20] indexing sorted alignments
[2016-07-05 21:38:20] calculating methylation levels
[2016-07-05 21:38:20] Processing chromosome: chr1
[2016-07-05 21:38:23] Processing chromosome: chr10
[2016-07-05 21:38:24] Processing chromosome: chr11
[2016-07-05 21:38:26] Processing chromosome: chr12
[2016-07-05 21:38:28] Processing chromosome: chr13
[2016-07-05 21:38:29] Processing chromosome: chr14
[2016-07-05 21:38:30] Processing chromosome: chr15
[2016-07-05 21:38:31] Processing chromosome: chr16
[2016-07-05 21:38:32] Processing chromosome: chr17
[2016-07-05 21:38:32] Processing chromosome: chr17_gl000205_random
[2016-07-05 21:38:32] Processing chromosome: chr18
[2016-07-05 21:38:33] Processing chromosome: chr19
[2016-07-05 21:38:34] Processing chromosome: chr1_gl000192_random
[2016-07-05 21:38:34] Processing chromosome: chr2
[2016-07-05 21:38:36] Processing chromosome: chr20
[2016-07-05 21:38:37] Processing chromosome: chr21
[2016-07-05 21:38:38] Processing chromosome: chr22
[2016-07-05 21:38:38] Processing chromosome: chr3
[2016-07-05 21:38:41] Processing chromosome: chr4
[2016-07-05 21:38:43] Processing chromosome: chr4_gl000193_random
[2016-07-05 21:38:43] Processing chromosome: chr5
[2016-07-05 21:38:45] Processing chromosome: chr6
[2016-07-05 21:38:47] Processing chromosome: chr7
[2016-07-05 21:38:49] Processing chromosome: chr8
[2016-07-05 21:38:51] Processing chromosome: chr9
[2016-07-05 21:38:52] Processing chromosome: chr9_gl000198_random
[2016-07-05 21:38:52] Processing chromosome: chr9_gl000199_random
[2016-07-05 21:38:52] Processing chromosome: chrM
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000212
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000214
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000216
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000218
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000220
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000224
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000225
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000226
[2016-07-05 21:38:52] Processing chromosome: chrUn_gl000232
[2016-07-05 21:38:52] Processing chromosome: chrX
[2016-07-05 21:38:54] Processing chromosome: chrY
[2016-07-05 21:38:54] Call methylation is finished.
[2016-07-05 21:38:54] ==============================
[2016-07-05 21:38:54] Files are saved as:
[2016-07-05 21:38:54]   Wiggle: N25_bsseeker2_meth.wig
[2016-07-05 21:38:54]   ATCGMap: N25_bsseeker2_meth.ATCGmap.gz
[2016-07-05 21:38:54]   CGmap: N25_bsseeker2_meth.CGmap.gz
[flay@n6281 bs-seeker2]$
```

BS-Seeker2 Methylation Calling: Output

```
[flay@n6281 bs-seeker2]$ ls -lh
total 9.8M
-rw-r--r-- 1 flay matteop 866K Jul  5 21:18 N25_bsseeker2.bam
-rw-r--r-- 1 flay matteop 5.0K Jul  5 21:18 N25_bsseeker2.bam.bs_seeker2_log
-rw-r--r-- 1 flay matteop 2.6K Jul  5 21:38 N25_bsseeker2.bam.call_methylation_log
-rw-r--r-- 1 flay matteop 875K Jul  5 21:38 N25_bsseeker2.bam_sorted.bam
-rw-r--r-- 1 flay matteop 1.6M Jul  5 21:38 N25_bsseeker2.bam_sorted.bam.bai
-rw-r--r-- 1 flay matteop 2.4M Jul  5 21:38 N25_bsseeker2_meth.ATCGmap.gz
-rw-r--r-- 1 flay matteop 421K Jul  5 21:38 N25_bsseeker2_meth.CGmap.gz
-rw-r--r-- 1 flay matteop 1.8M Jul  5 21:38 N25_bsseeker2_meth.wig
```

```
[flay@n6281 bs-seeker2]$ head N25_bsseeker2_meth.wig
type wiggle_0
variableStep chrom=chr1
1256559 -0.00
1256561 -0.00
1256562 -0.00
1256568 -1.00
1256569 -0.00
1256575 -0.00
1256579 -0.00
1256582 -0.00
```

BS-Seeker2 Methylation Calling: Output

```
[flay@n6281 bs-seeker2]$ zcat N25_bsseeker2_meth.CGmap.gz | head -10
chr1    G       1256559 CHG     CA      0.0     0       1
chr1    G       1256561 CHH     CT      0.0     0       1
chr1    G       1256562 CHH     CC      0.0     0       1
chr1    G       1256568 CG      CG      1.0     1       1
chr1    G       1256569 CHG     CC      0.0     0       1
chr1    G       1256575 CHH     CT      0.0     0       1
chr1    G       1256579 CHG     CA      0.0     0       1
chr1    G       1256582 CHG     CA      0.0     0       1
chr1    G       1256584 CHH     CA      0.0     0       1
chr1    G       1256586 CHH     CA      0.0     0       1
```

Format descriptions:

```
(1) chromosome
(2) nucleotide on Watson (+) strand
(3) position
(4) context (CG/CHG/CHH)
(5) dinucleotide-context (CA/CC/CG/CT)
(6) methylation-level = #_of_C / (#_of_C + #_of_T).
(7) #_of_C (methylated C, the count of reads showing C here)
(8) = #_of_C + #_of_T (all Cytosines, the count of reads showing C or T here)
```

BS-Seeker2 Methylation Calling: Output

```
[flay@n2175 bs-seeker2]$ zcat K562_bsseeker2_meth.ATCGmap.gz |head -10
chr1    C       847682  CHH     CC      0       0       0       0       0       0       0       1       0       0       na
chr1    C       847683  CHH     CA      0       0       0       0       0       0       0       1       0       0       na
chr1    A       847684  --      --      0       0       0       0       0       1       0       0       0       0       na
chr1    C       847685  CHH     CC      0       0       0       0       0       0       0       1       0       0       na
chr1    C       847686  CHH     CA      0       0       0       0       0       0       0       1       0       0       na
chr1    A       847687  --      --      0       0       0       0       0       1       0       0       0       0       na
chr1    C       847688  CHH     CC      0       0       0       0       0       0       0       1       0       0       na
chr1    C       847689  CHG     CT      0       0       0       0       0       0       0       1       0       0       na
chr1    T       847690  --      --      0       0       0       0       0       0       1       0       0       0       na
chr1    G       847691  CHG     CA      0       0       0       0       0       1       0       0       0       0       0.0
```

Format descriptions:

```
(1) chromosome
(2) nucleotide on Watson (+) strand
(3) position
(4) context (CG/CHG/CHH)
(5) dinucleotide-context (CA/CC/CG/CT)

(6) - (10) plus strand
(6) # of reads from Watson strand mapped here, support A on Watson strand
(7) # of reads from Watson strand mapped here, support T on Watson strand
(8) # of reads from Watson strand mapped here, support C on Watson strand
(9) # of reads from Watson strand mapped here, support G on Watson strand
(10) # of reads from Watson strand mapped here, support N

(11) - (15) minus strand
(11) # of reads from Crick strand mapped here, support A on Watson strand and T on Crick strand
(12) # of reads from Crick strand mapped here, support T on Watson strand and A on Crick strand
(13) # of reads from Crick strand mapped here, support C on Watson strand and G on Crick strand
(14) # of reads from Crick strand mapped here, support G on Watson strand and C on Crick strand
(15) # of reads from Crick strand mapped here, support N

(16) methylation_level = #C/(#C+#T) = C8/(C7+C8) for Watson strand, =C14/(C11+C14) for Crick strand;
"nan" means none reads support C/T at this position.
```

Converting BS-Seeker Output Format

#.wig format file can be used for visualization in IGV. However, if the file is too large, IGV can crash. It is recommended to convert wig to bigwig file for a more compressed version.

wigToBigWig

#bigwig file can be converted to bedgraph format to run metilene

bigWigToBedGraph