

# Align reads

- Required software
  - bsmap2.9, fastx\_trimmer, samtools, bwa\_meth\_bias-plot.py
- Scripts
  - BSseq1\_bsmap\_mbias.sh, BSseq1.1\_fastx\_trimmer.sh
- Data
  - raw fastq files, genome fasta
- Output
  - sorted bam file for each subject

# Align reads

1. Align raw reads, plot methylation by read position
2. Trim reads based on m-bias plots
3. Align trimmed reads
4. Convert sam files to sorted bam files

BSseq1\_bsmap\_mbias.sh

Run 1

```
#!/bin/bash
#####
# Nov.20 2015, Austin Hilliard, Stanford University Biology, Fernald lab
# BSseq1_bsmap_mbias.sh
#####
#
# This script will run bsmap on raw WGBS read files
# output .sam files will be sorted and saved as .bam files
# m-bias plot will be made for each subject, to decide whether to trim bases and re-align
#
# When first called it will print relevant settings and reminders to the console,
# then wait for a user key press before continuing
#
#####
# IMPORTANT: This script is totally inflexible so follow these instructions exactly
#####
#
# Input data must be stored as follows:
# top directory contains one sub-directory for each subject
# subject directories contain the raw read files and ideally nothing else
#
# Command line args must be in the exact following order with spaces between
# make sure to put full pathnames for 1,3,4,6
# 1: full path to top directory with data
#     must contain only one sub-dir for each subject and nothing else
# 2: string to grep for raw fastq.gz files within subject dirs
#     format of filenames must be *1$2 and *2$2
# 3: full path to bsmap binary
# 4: full path to genome fasta file
# 5: base filename for bsmap output
#     make sure it reflects the hard-coded bsmap settings
# 6: full path to python script for making m-bias plot
#
# Bio-RDF14:abseq$ BSseq1_bsmap_mbias.sh data_dir raw_data_suffix bsmap_path genome bsmap_out_base mbias_path
#
#####
# Following the code below is an example use and the console output
#####
```

```
data_dir="$1"
raw_data_suffix="$2"
bsmap_path="$3"
genome="$4"
bsmap_out_base="$5"
mbias_path="$6"

cd "$data_dir"
subjects=$(ls)
echo -e "\n-----"
echo -e "Moved to:\n ${data_dir}"
echo -----
echo -e "Subject dirs are:\n${subjects}"
echo -----
echo -e "Name format of raw read files will be:\n *1${raw_data_suffix}\n *2${raw_data_suffix}"
echo -----
echo -e "\nWill use bsmap version:\n ${bsmap_path}"
echo -----
echo -e "bsmap will output files as:\n subject_dir/${bsmap_out_base}.sam"
echo "WARNING:"
echo " bsmap settings are hard-coded in this script, check them and make sure output name is smart"
echo -----
echo -e "Will use genome file:\n ${genome}"
echo -----
echo -e "Will convert output .sam into sorted .bam file, make index, then delete .sam file"
echo -----
echo -e "Will use ${mbias_path} to make m-bias plots"
echo -e "-----\n"
echo -e "-----\n"
read -rsp "If this is all good press any key to continue, or ctrl+c to quit..." -n1
```

```
for s in $subjects
do
echo -e "\n===="
echo "Working on ${s}"
echo "===="
reads1=$(ls "$s/"*1${raw_data_suffix})
reads2=$(ls "$s/"*2${raw_data_suffix})
echo -e "read files:\n ${reads1}\n ${reads2}\n"
echo -e "will write output file:\n ${s}/${bsmap_out_base}.sam\n"
"$bsmap_path" \
-a "$reads1" \
-b "$reads2" \
-d "$genome" \
-o "${s}/${bsmap_out_base}.sam" \
-A GAGCCGTAAGGACGACTTGG -A ACACCTTTCCCTACACGAC \
-q 30 -m 0 \
-S 1

echo -e "\n-----"
echo "Converting .sam to sorted .bam and creating index..."
echo -----
samtools view -bS "${s}/${bsmap_out_base}.sam" | \
samtools sort - "${s}/${bsmap_out_base}"
samtools index "${s}/${bsmap_out_base}.bam"
rm "${s}/${bsmap_out_base}.sam"

echo -----
echo "Making m-bias plot..."
echo -----
python "$mbias_path" "${s}/${bsmap_out_base}.bam" "$genome"
done
```

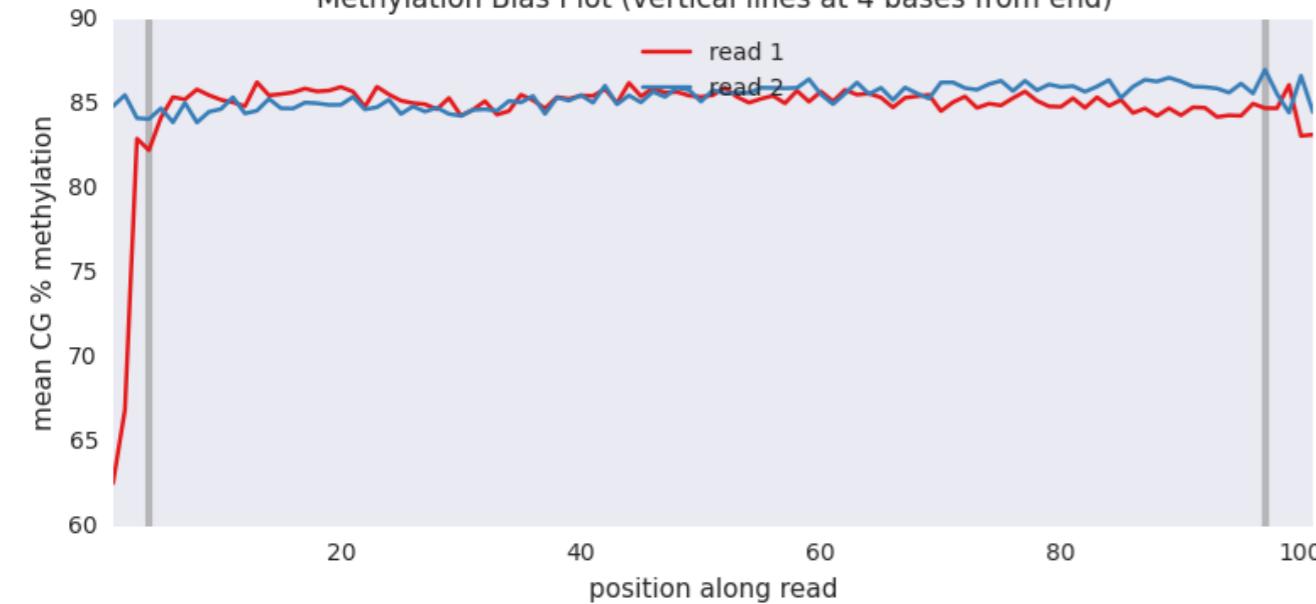
```
#####
# Example
#####
# Comments:
# After bsmap runs there's an error I don't understand:
# 40125 Abort trap: 6
# As far as I can tell bsmap writes a complete, valid .sam file
# The error may reflect some process trying to write to memory it doesn't own
#
# There's a samtools/matplotlib error after the m-bias plot script but the plots are fine
# The plotting script was taken from the bwa-meth github (https://github.com/brentp/bwa-meth)
# and I installed some other python libs to get it working, I'm surprised it works at all
#
# Based on the m-bias plots here I'll trim 3 bases from both ends of all reads then re-align
#####
#
# Bio-RDF14:Documents abseq$ /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh \
#                               /Users/abseq/Documents/_BS-seq_data/ _pf.fastq.gz /Users/abseq/Documents-bsmap-2.90/bsmap \
#                               /Users/abseq/Documents/H_burtoni_v1.assembly.fa \
#                               aligned.adapters.q30.m0_bsmap2.9 \
#                               /Volumes/fishstudies-1/_scripts/bwa-meth_bias-plot.py
#
# -----
# Moved to:
# /Users/abseq/Documents/_BS-seq_data/
# -----
# Subject dirs are:
# 3157_TENNISON 3165_BRISCOE 3581_LYNLEY 3677_MONK
# -----
# Name format of raw read files will be:
# *1_pf.fastq.gz
# *2_pf.fastq.gz
# -----
#
# Will use bsmap version:
# /Users/abseq/Documents/bsmap-2.90/bsmap
# -----
# bsmap will output files as:
# SUBJECT_DIR/aligned.adapters.q30.m0_bsmap2.9.sam
# WARNING:
# bsmap settings are hard-coded in this script, check them and make sure output name is smart
# -----
# Will use genome file:
# /Users/abseq/Documents/H_burtoni_v1.assembly.fa
# -----
# Will convert output .sam into sorted .bam file, make index, then delete .sam file
# -----
# Will use /Volumes/fishstudies-1/_scripts/bwa-meth_bias-plot.py to make m-bias plots
# -----
#
#
#
# If this is all good press any key to continue, or ctrl+c to quit...
# -----
```

```
# =====
# Working on 3157_TENNISON
# =====
# read files:
# 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_1_pf.fastq.gz
# 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_2_pf.fastq.gz
#
# will write output file:
# 3157_TENNISON/aligned.adapters.q30.m0_bsmap2.9.sam
#
# [bsmap] @Fri Nov 20 16:47:06 2015    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa      (format: FASTA)
# [bsmap] @Fri Nov 20 16:47:16 2015    8001 reference seqs loaded, total size 831411547 bp. 10 secs passed
# [bsmap] @Fri Nov 20 16:47:27 2015    create seed table. 21 secs passed
# [bsmap] @Fri Nov 20 16:47:27 2015    Pair-end alignment(8 threads),
#   Input read file #1: 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_1_pf.fastq.gz (format: gzipped FASTQ)
#   Input read file #2: 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_2_pf.fastq.gz (format: gzipped FASTQ)
#   Output file: 3157_TENNISON/aligned.adapters.q30.m0_bsmap2.9.sam (format: SAM)
# [bsmap] @Fri Nov 20 18:07:57 2015    total read pairs: 114891459    total time consumed: 4851 secs
#   aligned pairs: 66356318 (57.8%), unique pairs: 60551212 (52.7%), non-unique pairs: 5805106 (5.1%)
#   unpaired read #1: 12330042 (10.7%), unique reads: 9151551 (8.0%), non-unique reads: 3178491 (2.8%)
#   unpaired read #2: 12234190 (10.6%), unique reads: 8739382 (7.6%), non-unique reads: 3494808 (3.0%)
# /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh: line 56: 40125 Abort trap: 6          $bsmap_path -a $reads1 -b $reads2 -d $genome -o $s"/$bsmap_out_base".sam"
#
#
# Converting .sam to sorted .bam and creating index...
# -----
# [bam_sort_core] merging from 65 files...
#
# Making m-bias plot...
#
#
# wrote to 3157_TENNISON/aligned.adapters.q30.m0_bsmap2.9.bias.txt
# saving to 3157_TENNISON/aligned.adapters.q30.m0_bsmap2.9.bias.png
# /System/Library/Frameworks/Python.framework/Versions/2.7/Extras/lib/python/matplotlib/tight_layout.py:225: UserWarning: tight_layout : falling back to Agg renderer
#   warnings.warn("tight_layout : falling back to Agg renderer")
# samtools: writing to standard output failed: Broken pipe
# samtools: error closing standard output: -1
# =====
# Working on 3165_BRISCOE
# =====
# read files:
# 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_1_pf.fastq.gz
# 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_2_pf.fastq.gz
#
# will write output file:
# 3165_BRISCOE/aligned.adapters.q30.m0_bsmap2.9.sam
#
# [bsmap] @Fri Nov 20 19:15:37 2015    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa      (format: FASTA)
# [bsmap] @Fri Nov 20 19:15:47 2015    8001 reference seqs loaded, total size 831411547 bp. 10 secs passed
# [bsmap] @Fri Nov 20 19:15:58 2015    create seed table. 21 secs passed
# [bsmap] @Fri Nov 20 19:15:58 2015    Pair-end alignment(8 threads),
#   Input read file #1: 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_1_pf.fastq.gz (format: gzipped FASTQ)
#   Input read file #2: 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_2_pf.fastq.gz (format: gzipped FASTQ)
#   Output file: 3165_BRISCOE/aligned.adapters.q30.m0_bsmap2.9.sam (format: SAM)
# [bsmap] @Fri Nov 20 20:57:28 2015    total read pairs: 137675454    total time consumed: 6111 secs
#   aligned pairs: 73808493 (53.6%), unique pairs: 66568757 (48.4%), non-unique pairs: 7239736 (5.3%)
#   unpaired read #1: 16415368 (11.9%), unique reads: 12280698 (8.9%), non-unique reads: 4134670 (3.0%)
#   unpaired read #2: 16835165 (12.2%), unique reads: 11969447 (8.7%), non-unique reads: 4865718 (3.5%)
# /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh: line 56: 40242 Abort trap: 6          $bsmap_path -a $reads1 -b $reads2 -d $genome -o $s"/$bsmap_out_base".sam"
#
#
# Converting .sam to sorted .bam and creating index...
# -----
# [bam_sort_core] merging from 74 files...
#
# Making m-bias plot...
#
#
# wrote to 3165_BRISCOE/aligned.adapters.q30.m0_bsmap2.9.bias.txt
# saving to 3165_BRISCOE/aligned.adapters.q30.m0_bsmap2.9.bias.png
# /System/Library/Frameworks/Python.framework/Versions/2.7/Extras/lib/python/matplotlib/tight_layout.py:225: UserWarning: tight_layout : falling back to Agg renderer
#   warnings.warn("tight_layout : falling back to Agg renderer")
# samtools: writing to standard output failed: Broken pipe
# samtools: error closing standard output: -1
```

```
# =====
# Working on 3581_LYNLEY
# =====
# read files:
# 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_1_pf.fastq.gz
# 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_2_pf.fastq.gz
#
# will write output file:
# 3581_LYNLEY/aligned.adapters.q30.m0_bsmap2.9.sam
#
# [bsmap] @Fri Nov 20 22:27:34 2015    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa      (format: FASTA)
# [bsmap] @Fri Nov 20 22:27:43 2015    8001 reference seqs loaded, total size 831411547 bp. 9 secs passed
# [bsmap] @Fri Nov 20 22:27:55 2015    create seed table. 21 secs passed
# [bsmap] @Fri Nov 20 22:27:55 2015    Pair-end alignment(8 threads),
#   Input read file #1: 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_1_pf.fastq.gz (format: gzipped FASTQ)
#   Input read file #2: 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_2_pf.fastq.gz (format: gzipped FASTQ)
#   Output file: 3581_LYNLEY/aligned.adapters.q30.m0_bsmap2.9.sam (format: SAM)
# [bsmap] @Sat Nov 21 00:02:50 2015    total read pairs: 123614639    total time consumed: 5716 secs
# aligned pairs: 69852665 (56.5%), unique pairs: 63142731 (51.1%), non-unique pairs: 6709934 (5.4%)
# unpaired read #1: 15266437 (12.4%), unique reads: 11153983 (9.0%), non-unique reads: 4112454 (3.3%)
# unpaired read #2: 15069058 (12.2%), unique reads: 10608375 (8.6%), non-unique reads: 4460683 (3.6%)
# /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh: line 56: 40410 Abort trap: 6          $bsmap_path -a $reads1 -b $reads2 -d $genome -o $s"/$bsmap_out_base".sam"
#
# -----
# Converting .sam to sorted .bam and creating index...
# -----
#
# [bam_sort_core] merging from 70 files...
#
# Making m-bias plot...
#
#
# wrote to 3581_LYNLEY/aligned.adapters.q30.m0_bsmap2.9.bias.txt
# saving to 3581_LYNLEY/aligned.adapters.q30.m0_bsmap2.9.bias.png
# /System/Library/Frameworks/Python.framework/Versions/2.7/Extras/lib/python/matplotlib/tight_layout.py:225: UserWarning: tight_layout : falling back to Agg renderer
# warnings.warn("tight_layout : falling back to Agg renderer")
# samtools: writing to standard output failed: Broken pipe
# samtools: error closing standard output: -1
# =====
# Working on 3677_MONK
# =====
# read files:
# 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_1_pf.fastq.gz
# 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_2_pf.fastq.gz
#
# will write output file:
# 3677_MONK/aligned.adapters.q30.m0_bsmap2.9.sam
#
# [bsmap] @Sat Nov 21 01:22:05 2015    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa      (format: FASTA)
# [bsmap] @Sat Nov 21 01:22:15 2015    8001 reference seqs loaded, total size 831411547 bp. 10 secs passed
# [bsmap] @Sat Nov 21 01:22:27 2015    create seed table. 22 secs passed
# [bsmap] @Sat Nov 21 01:22:27 2015    Pair-end alignment(8 threads),
#   Input read file #1: 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_1_pf.fastq.gz (format: gzipped FASTQ)
#   Input read file #2: 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_2_pf.fastq.gz (format: gzipped FASTQ)
#   Output file: 3677_MONK/aligned.adapters.q30.m0_bsmap2.9.sam (format: SAM)
# [bsmap] @Sat Nov 21 02:54:04 2015    total read pairs: 122414964    total time consumed: 5519 secs
# aligned pairs: 68128108 (55.7%), unique pairs: 61988246 (50.6%), non-unique pairs: 6139862 (5.0%)
# unpaired read #1: 14551674 (11.9%), unique reads: 10758062 (8.8%), non-unique reads: 3793612 (3.1%)
# unpaired read #2: 14624569 (11.9%), unique reads: 10395751 (8.5%), non-unique reads: 4228818 (3.5%)
# /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh: line 56: 40595 Abort trap: 6          $bsmap_path -a $reads1 -b $reads2 -d $genome -o $s"/$bsmap_out_base".sam"
#
# -----
# Converting .sam to sorted .bam and creating index...
# -----
#
# [bam_sort_core] merging from 68 files...
#
# Making m-bias plot...
#
#
# wrote to 3677_MONK/aligned.adapters.q30.m0_bsmap2.9.bias.txt
# saving to 3677_MONK/aligned.adapters.q30.m0_bsmap2.9.bias.png
# /System/Library/Frameworks/Python.framework/Versions/2.7/Extras/lib/python/matplotlib/tight_layout.py:225: UserWarning: tight_layout : falling back to Agg renderer
# warnings.warn("tight_layout : falling back to Agg renderer")
# samtools: writing to standard output failed: Broken pipe
# samtools: error closing standard output: -1
#####
#####
```

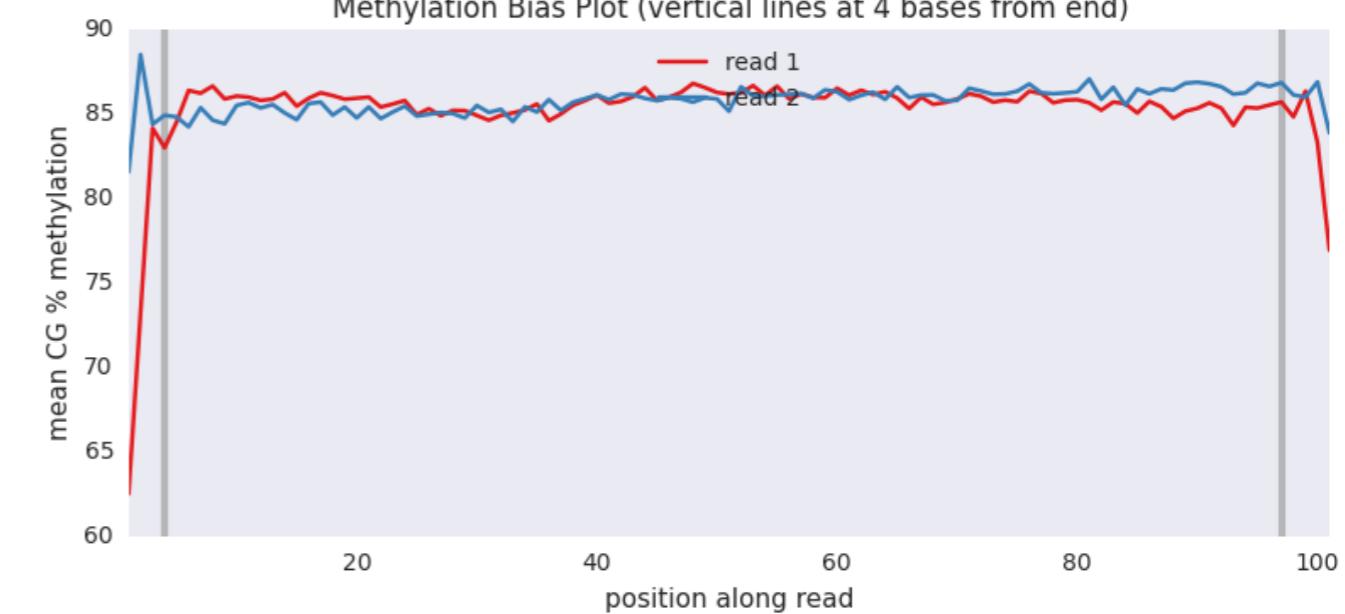
3157

Methylation Bias Plot (vertical lines at 4 bases from end)



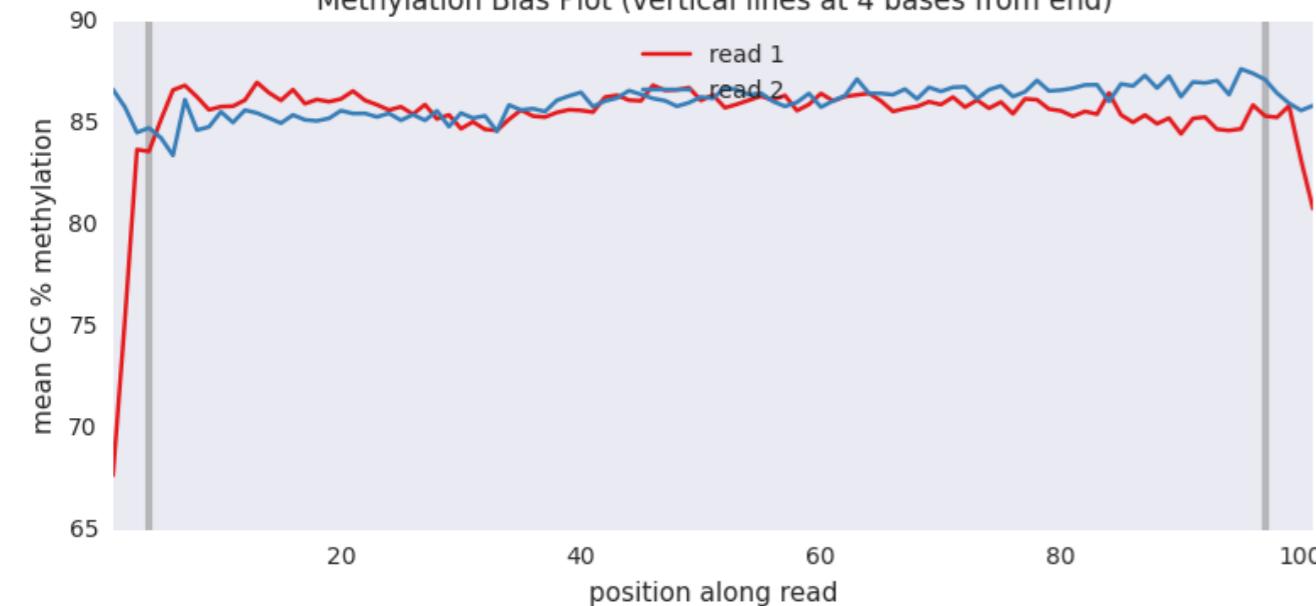
3165

Methylation Bias Plot (vertical lines at 4 bases from end)



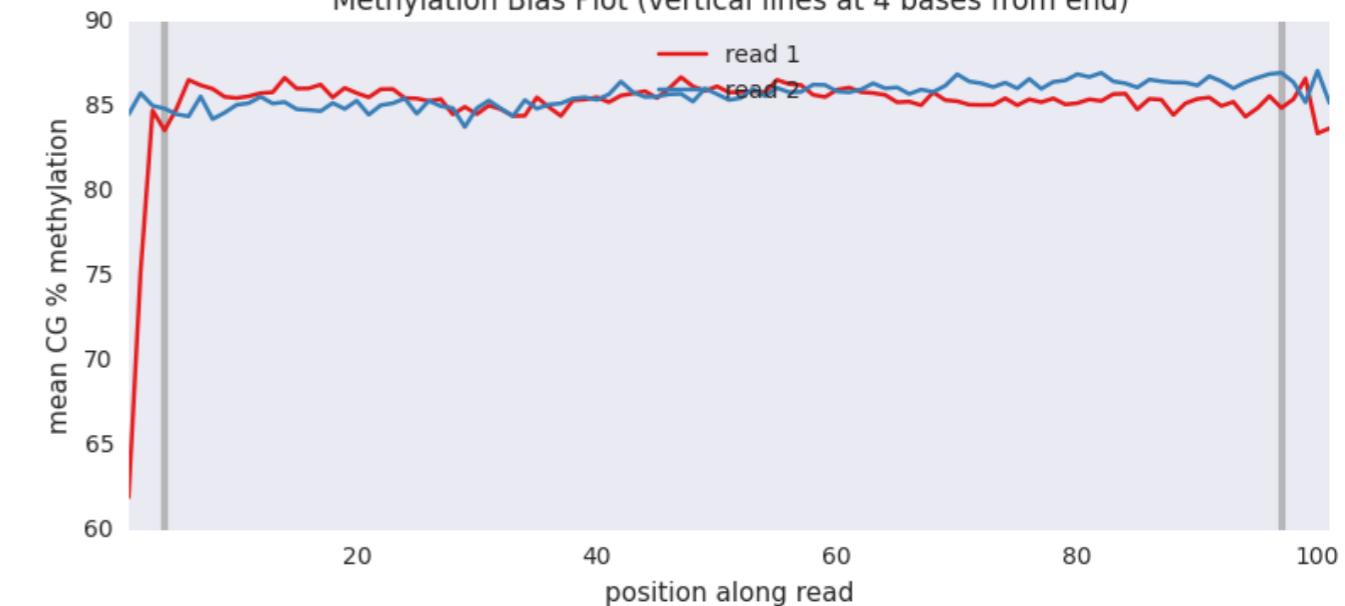
3581

Methylation Bias Plot (vertical lines at 4 bases from end)



3677

Methylation Bias Plot (vertical lines at 4 bases from end)



BSseq1.1\_fastx\_trimmer.sh

```
#!/bin/bash
#####
# Nov.20 2015, Austin Hilliard, Stanford University Biology, Fernald lab
# BSseq1.1_fastx_trimmer.sh
#####
#
# This script runs fastx_trimmer on compressed raw reads
# The data should be organized in the same way as required for BSseq1_bsmap_mbias.sh
#
#####
# IMPORTANT: This script is totally inflexible so follow these instructions exactly
#####
#
# Input data must be stored as follows:
# top directory contains one sub-directory for each subject
# subject directories contain the raw read files and ideally nothing else
#
# There are 4 command line arguments to this script:
# They must be in the exact following order with spaces between
# 1: full path to top directory with data
#     must contain only one sub-dir for each subject and nothing else
# 2: string to grep for raw fastq.gz files within subject dirs
#     format of filenames must be *1$2 and *2$2
# 3: -f option to fastx_trimmer (first base to keep)
# 4: -l option to fastx_trimmer (last base to keep)
#
# It's assumed the input raw read files and output trimmed files are compressed
#
#####
# Following the code below is an example use and the console output
#####
```

```
data_dir="$1"
raw_data_suffix="$2"
first_to_keep="$3"
last_to_keep="$4"

cd "$data_dir"
subjects=$(ls)
echo -e "\n===="
echo -e "Moved to:\n ${data_dir}"
echo -----
echo -e "Subject dirs are:\n${subjects}"
echo -----
echo -e "Name format of raw read files will be:\n *1${raw_data_suffix}\n *2${raw_data_suffix}"
echo -----
echo =====
echo -e "Trimming... \n"
for s in $subjects
do
    echo -----
    echo "$s"
    echo -----
    date
    reads1=$(ls "$s/*1${raw_data_suffix}")
    reads2=$(ls "$s/*2${raw_data_suffix}")
    stripped1=$(echo "$reads1" | awk '{gsub(/.fastq.gz/, "");print}')
    stripped2=$(echo "$reads2" | awk '{gsub(/.fastq.gz/, "");print}')
    out1="${stripped1}_trimmed${first_to_keep}-${last_to_keep}.fastq.gz"
    out2="${stripped2}_trimmed${first_to_keep}-${last_to_keep}.fastq.gz"
    echo -e "input read files:\n ${reads1}\n ${reads2}\n"
    echo -e "output files:\n ${out1}\n ${out2}\n"
    gunzip -c "$reads1" | \
    fastx_trimmer -f "$first_to_keep" -l "$last_to_keep" -z -o "$out1"
    gunzip -c "$reads2" | \
    fastx_trimmer -f "$first_to_keep" -l "$last_to_keep" -z -o "$out2"
done
```

```
#####
# Example
#####
#
# Bio-RDF14:Documents abseq$ /Volumes/fishstudies-1/_scripts/BSseq1.1_fastx_trimmer.sh \
#                               /Users/abseq/Documents/_BS-seq_data \
#                               _pf.fastq.gz \
#                               4 \
#                               98
#
# -----
#
# Moved to:
#   /Users/abseq/Documents/_BS-seq_data
# -----
#
# Subject dirs are:
#   3157_TENNISON 3165_BRISCOE 3581_LYNLEY 3677_MONK
# -----
#
# Name format of raw read files will be:
#   *1_pf.fastq.gz
#   *2_pf.fastq.gz
# -----
# -----
#
# Trimming...
#
```

```
#  
# 3157_TENNISON  
# -----  
# Mon Nov 23 14:27:46 PST 2015  
# input read files:  
# 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_1_pf.fastq.gz  
# 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_2_pf.fastq.gz  
#  
# output files:  
# 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_1_pf_trimmed4-98.fastq.gz  
# 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_2_pf_trimmed4-98.fastq.gz  
#  
# -----  
# 3165_BRISCOE  
# -----  
# Mon Nov 23 15:42:55 PST 2015  
# input read files:  
# 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_1_pf.fastq.gz  
# 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_2_pf.fastq.gz  
#  
# output files:  
# 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_1_pf_trimmed4-98.fastq.gz  
# 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_2_pf_trimmed4-98.fastq.gz  
#  
# -----  
# 3581_LYNLEY  
# -----  
# Mon Nov 23 17:15:00 PST 2015  
# input read files:  
# 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_1_pf.fastq.gz  
# 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_2_pf.fastq.gz  
#  
# output files:  
# 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_1_pf_trimmed4-98.fastq.gz  
# 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_2_pf_trimmed4-98.fastq.gz  
#  
# -----  
# 3677_MONK  
# -----  
# Mon Nov 23 18:36:24 PST 2015  
# input read files:  
# 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_1_pf.fastq.gz  
# 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_2_pf.fastq.gz  
#  
# output files:  
# 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_1_pf_trimmed4-98.fastq.gz  
# 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_2_pf_trimmed4-98.fastq.gz
```

BSseq1\_bsmap\_mbias.sh

Run 2

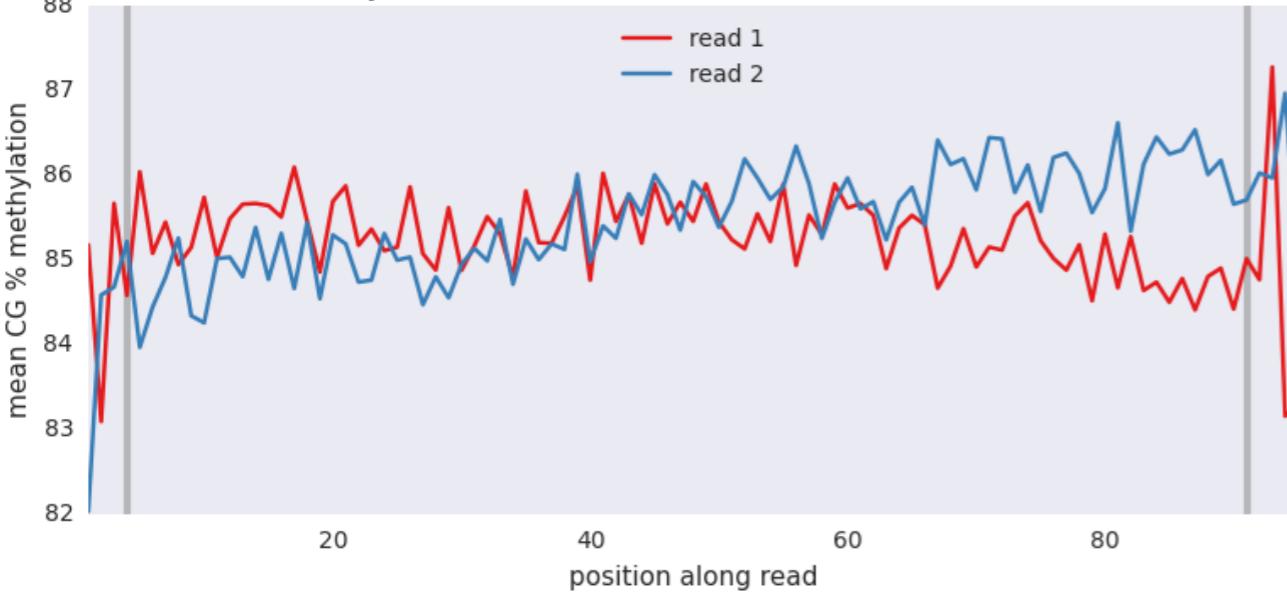
```
#####
# Another example run
#####
# Comments:
# After trimming 3bp from both ends of reads
# Mapping percentages were slightly better than they were before trimming
#####
#
# Bio-RDF14:Documents abseq$ /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh \
#                               /Users/abseq/Documents/_BS-seq_data/ \
#                               _pf_trimmed4-98.fastq.gz \
#                               /Users/abseq/Documents/bsmap-2.90/bsmap \
#                               /Users/abseq/Documents/H_burtoni_v1.assembly.fa \
#                               aligned_trimmed4-98.adapters.q30.m0_bsmap2.9 \
#                               /Volumes/fishstudies-1/_scripts/bwa-meth_bias-plot.py
#
# -----
# Moved to:
# /Users/abseq/Documents/_BS-seq_data/
# -----
# Subject dirs are:
# 3157_TENNISON 3165_BRISCOE 3581_LYNLEY 3677_MONK
# -----
# Name format of raw read files will be:
# *1_pf_trimmed4-98.fastq.gz
# *2_pf_trimmed4-98.fastq.gz
# -----
#
# Will use bsmap version:
# /Users/abseq/Documents/bsmap-2.90/bsmap
# -----
# bsmap will output files as:
# SUBJECT_DIR/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam
# WARNING:
# bsmap settings are hard-coded in this script, check them and make sure output name is smart
# -----
# Will use genome file:
# /Users/abseq/Documents/H_burtoni_v1.assembly.fa
# -----
# Will convert output .sam into sorted .bam file, make index, then delete .sam file
# -----
# Will use /Volumes/fishstudies-1/_scripts/bwa-meth_bias-plot.py to make m-bias plots
# -----
#
# -----
#
# If this is all good press any key to continue, or ctrl+c to quit...
```

```
# =====
# Working on 3157_TENNISON
# =====
# read files:
# 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_1_pf_trimmed4-98.fastq.gz
# 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_2_pf_trimmed4-98.fastq.gz
#
# will write output file:
# 3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam
#
# [bsmap] @Tue Nov 24 11:52:35 2015    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa      (format: FASTA)
# [bsmap] @Tue Nov 24 11:52:44 2015    8001 reference seqs loaded, total size 831411547 bp. 9 secs passed
# [bsmap] @Tue Nov 24 11:52:56 2015    create seed table. 21 secs passed
# [bsmap] @Tue Nov 24 11:52:56 2015    Pair-end alignment(8 threads),
#   Input read file #1: 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_1_pf_trimmed4-98.fastq.gz (format: gzipped FASTQ)
#   Input read file #2: 3157_TENNISON/130917_TENNISON_0250_AD2H9VACXX_L4_2_pf_trimmed4-98.fastq.gz (format: gzipped FASTQ)
#   Output file: 3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam (format: SAM)
# [bsmap] @Tue Nov 24 13:05:03 2015    total read pairs: 114891459    total time consumed: 4348 secs
#   aligned pairs: 67412600 (58.7%), unique pairs: 61112215 (53.2%), non-unique pairs: 6300385 (5.5%)
#   unpaired read #1: 12589712 (11.0%), unique reads: 9237250 (8.0%), non-unique reads: 3352462 (2.9%)
#   unpaired read #2: 12458311 (10.8%), unique reads: 8812416 (7.7%), non-unique reads: 3645895 (3.2%)
# /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh: line 69: 3104 Abort trap: 6          $bsmap_path -a $reads1 -b $reads2 -d $genome -o $s"/$bsmap_out_base".sam"
#
# -----
# Converting .sam to sorted .bam and creating index...
# -----
# [bam_sort_core] merging from 66 files...
#
# Making m-bias plot...
#
# 
# wrote to 3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bias.txt
# saving to 3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bias.png
# /System/Library/Frameworks/Python.framework/Versions/2.7/Extras/lib/python/matplotlib/tight_layout.py:225: UserWarning: tight_layout : falling back to Agg renderer
#   warnings.warn("tight_layout : falling back to Agg renderer")
# samtools: writing to standard output failed: Broken pipe
# samtools: error closing standard output: -1
# =====
# Working on 3165_BRISCOE
# =====
# read files:
# 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_1_pf_trimmed4-98.fastq.gz
# 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_2_pf_trimmed4-98.fastq.gz
#
# will write output file:
# 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam
#
# [bsmap] @Tue Nov 24 14:12:22 2015    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa      (format: FASTA)
# [bsmap] @Tue Nov 24 14:12:32 2015    8001 reference seqs loaded, total size 831411547 bp. 10 secs passed
# [bsmap] @Tue Nov 24 14:12:43 2015    create seed table. 21 secs passed
# [bsmap] @Tue Nov 24 14:12:43 2015    Pair-end alignment(8 threads),
#   Input read file #1: 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_1_pf_trimmed4-98.fastq.gz (format: gzipped FASTQ)
#   Input read file #2: 3165_BRISCOE/130920_BRISCOE_0120_BC2HPBACXX_L2_2_pf_trimmed4-98.fastq.gz (format: gzipped FASTQ)
#   Output file: 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam (format: SAM)
# [bsmap] @Tue Nov 24 15:48:40 2015    total read pairs: 137675454    total time consumed: 5778 secs
#   aligned pairs: 75136885 (54.6%), unique pairs: 67270054 (48.9%), non-unique pairs: 7866831 (5.7%)
#   unpaired read #1: 16963090 (12.3%), unique reads: 12567031 (9.1%), non-unique reads: 4396059 (3.2%)
#   unpaired read #2: 17207508 (12.5%), unique reads: 12119535 (8.8%), non-unique reads: 5087973 (3.7%)
# /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh: line 69: 3353 Abort trap: 6          $bsmap_path -a $reads1 -b $reads2 -d $genome -o $s"/$bsmap_out_base".sam"
#
# -----
# Converting .sam to sorted .bam and creating index...
# -----
# [bam_sort_core] merging from 76 files...
#
# Making m-bias plot...
#
# 
# wrote to 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bias.txt
# saving to 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bias.png
# /System/Library/Frameworks/Python.framework/Versions/2.7/Extras/lib/python/matplotlib/tight_layout.py:225: UserWarning: tight_layout : falling back to Agg renderer
#   warnings.warn("tight_layout : falling back to Agg renderer")
# samtools: writing to standard output failed: Broken pipe
# samtools: error closing standard output: -1
```

```
# =====
# Working on 3581_LYNLEY
# =====
# read files:
# 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_1_pf_trimmed4-98.fastq.gz
# 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_2_pf_trimmed4-98.fastq.gz
#
# will write output file:
# 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam
#
# [bsmap] @Tue Nov 24 17:13:32 2015    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa      (format: FASTA)
# [bsmap] @Tue Nov 24 17:13:42 2015    8001 reference seqs loaded, total size 831411547 bp. 10 secs passed
# [bsmap] @Tue Nov 24 17:13:54 2015    create seed table. 22 secs passed
# [bsmap] @Tue Nov 24 17:13:54 2015    Pair-end alignment(8 threads),
#   Input read file #1: 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_1_pf_trimmed4-98.fastq.gz (format: gzipped FASTQ)
#   Input read file #2: 3581_LYNLEY/131004_LYNLEY_0370_AD2HMEACXX_L6_2_pf_trimmed4-98.fastq.gz (format: gzipped FASTQ)
#   Output file: 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam      (format: SAM)
# [bsmap] @Tue Nov 24 18:36:37 2015    total read pairs: 123614639    total time consumed: 4985 secs
#   aligned pairs: 71192785 (57.6%), unique pairs: 63872941 (51.7%), non-unique pairs: 7319844 (5.9%)
#   unpaired read #1: 15587990 (12.6%), unique reads: 11262188 (9.1%), non-unique reads: 4325802 (3.5%)
#   unpaired read #2: 15398786 (12.5%), unique reads: 10706652 (8.7%), non-unique reads: 4692134 (3.8%)
# /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh: line 69: 3984 Abort trap: 6          $bsmap_path -a $reads1 -b $reads2 -d $genome -o $s"/$bsmap_out_base".sam"
#
# -----
# Converting .sam to sorted .bam and creating index...
# -----
#
# [bam_sort_core] merging from 71 files...
#
# Making m-bias plot...
#
# 
# wrote to 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bias.txt
# saving to 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bias.png
# /System/Library/Frameworks/Python.framework/Versions/2.7/Extras/lib/python/matplotlib/tight_layout.py:225: UserWarning: tight_layout : falling back to Agg renderer
#   warnings.warn("tight_layout : falling back to Agg renderer")
# samtools: writing to standard output failed: Broken pipe
# samtools: error closing standard output: -1
# =====
# Working on 3677_MONK
# =====
# read files:
# 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_1_pf_trimmed4-98.fastq.gz
# 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_2_pf_trimmed4-98.fastq.gz
#
# will write output file:
# 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam
#
# [bsmap] @Tue Nov 24 19:53:05 2015    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa      (format: FASTA)
# [bsmap] @Tue Nov 24 19:53:15 2015    8001 reference seqs loaded, total size 831411547 bp. 10 secs passed
# [bsmap] @Tue Nov 24 19:53:27 2015    create seed table. 22 secs passed
# [bsmap] @Tue Nov 24 19:53:27 2015    Pair-end alignment(8 threads),
#   Input read file #1: 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_1_pf_trimmed4-98.fastq.gz (format: gzipped FASTQ)
#   Input read file #2: 3677_MONK/131023_MONK_0319_AC2YY8ACXX_L3_2_pf_trimmed4-98.fastq.gz (format: gzipped FASTQ)
#   Output file: 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.sam      (format: SAM)
# [bsmap] @Tue Nov 24 21:14:28 2015    total read pairs: 122414964    total time consumed: 4883 secs
#   aligned pairs: 69427255 (56.7%), unique pairs: 62709140 (51.2%), non-unique pairs: 6718115 (5.5%)
#   unpaired read #1: 14974490 (12.2%), unique reads: 10944191 (8.9%), non-unique reads: 4030299 (3.3%)
#   unpaired read #2: 15007540 (12.3%), unique reads: 10531881 (8.6%), non-unique reads: 4475659 (3.7%)
# /Volumes/fishstudies-1/_scripts/BSseq1_bsmap_mbias.sh: line 69: 4517 Abort trap: 6          $bsmap_path -a $reads1 -b $reads2 -d $genome -o $s"/$bsmap_out_base".sam"
#
# -----
# Converting .sam to sorted .bam and creating index...
# -----
#
# [bam_sort_core] merging from 69 files...
#
# Making m-bias plot...
#
# 
# wrote to 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bias.txt
# saving to 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bias.png
# /System/Library/Frameworks/Python.framework/Versions/2.7/Extras/lib/python/matplotlib/tight_layout.py:225: UserWarning: tight_layout : falling back to Agg renderer
#   warnings.warn("tight_layout : falling back to Agg renderer")
# samtools: writing to standard output failed: Broken pipe
# samtools: error closing standard output: -1
```

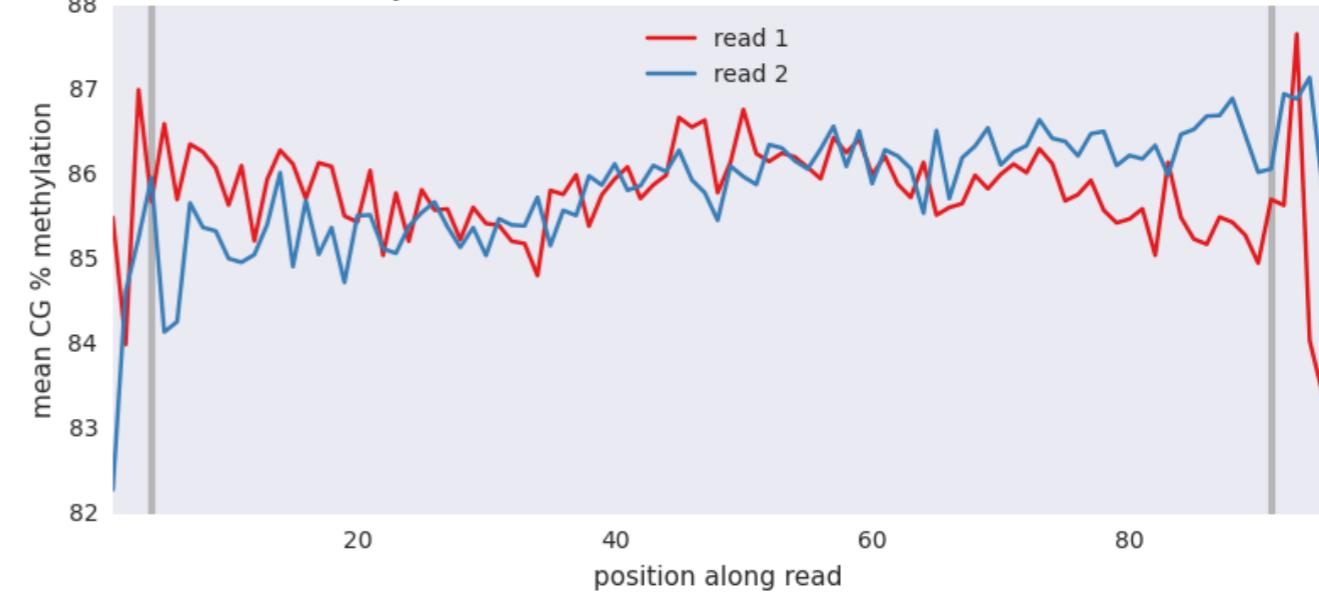
3157

Methylation Bias Plot (vertical lines at 4 bases from end)



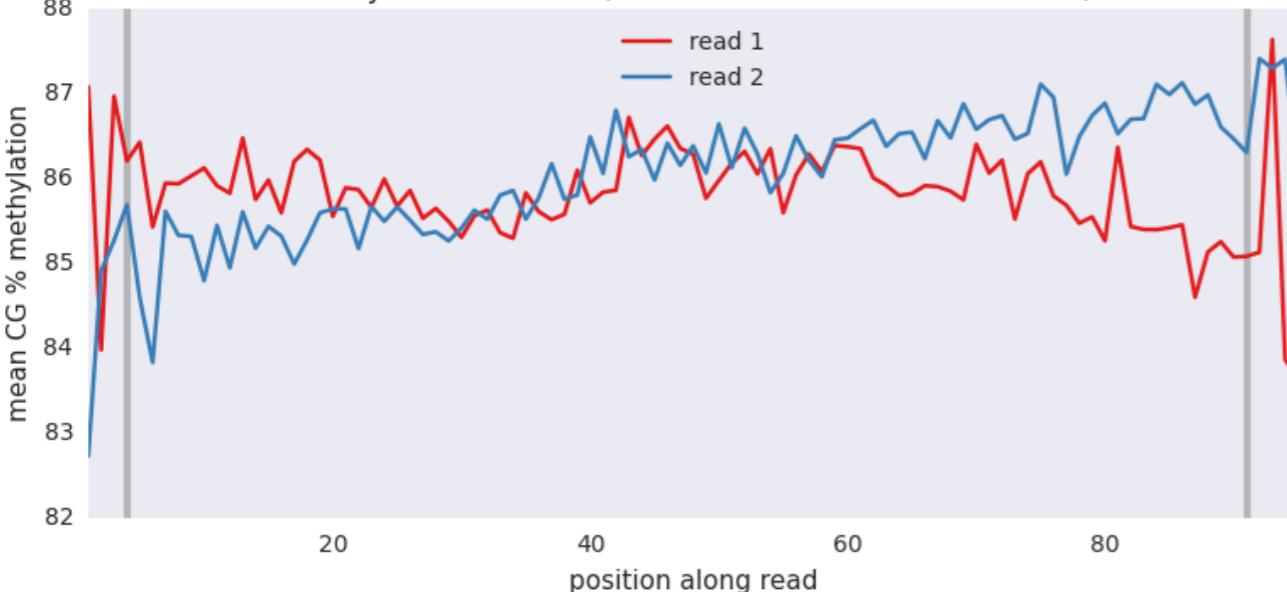
3165

Methylation Bias Plot (vertical lines at 4 bases from end)



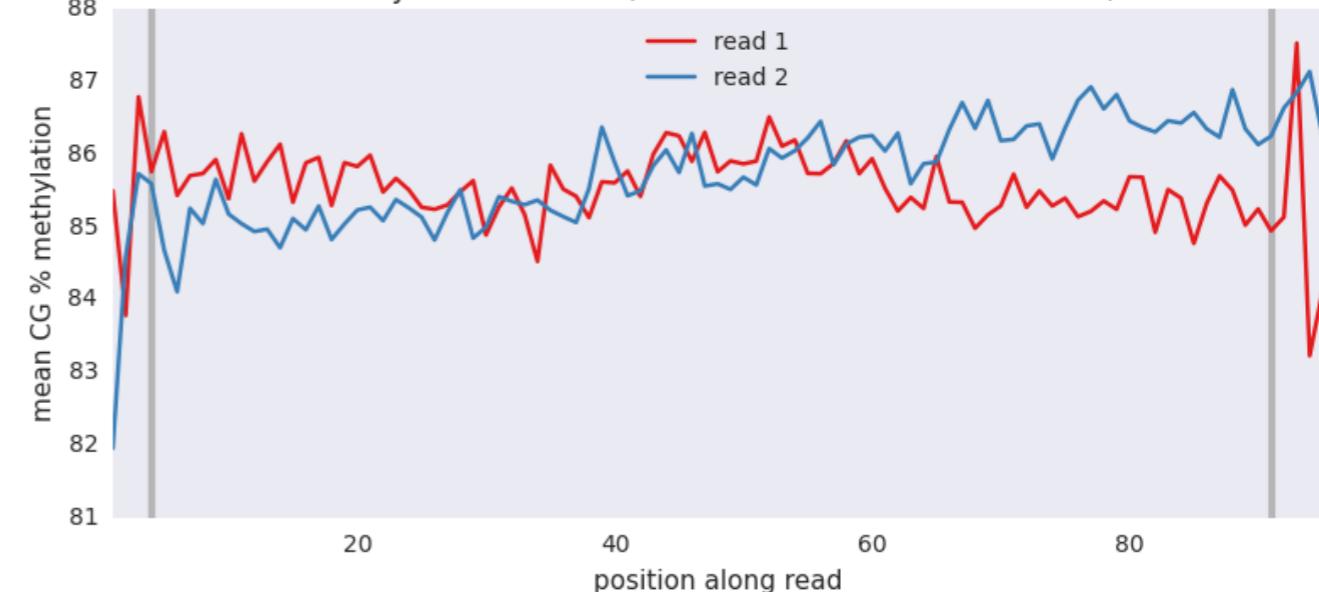
3581

Methylation Bias Plot (vertical lines at 4 bases from end)



3677

Methylation Bias Plot (vertical lines at 4 bases from end)



# Compute methylation ratios

- Required software
  - methratio.py, samtools-0.1.19
- Scripts
  - BSseq2\_methratio.sh
- Data
  - sorted bam files, genome fasta
- Output
  - text files with methylation ratios for each subject

# Compute methylation ratios

1. Compute ratios with methratio.py
  - 1.1. Strand specific (no -g)
  - 1.2. Combine CpGs across strands
2. Filter output files to only CpG loci

BSseq2\_methratio.sh

```
#!/bin/bash
#####
# Dec.09 2015, Austin Hilliard, Stanford University Biology, Fernald lab
# BSseq2_methratio.sh
#####
#
# This script will run methratio.py (from bsmap) on .bam files
#
# When first called it will print relevant settings and reminders to the console,
# then wait for a user key press before continuing
#
#####
# IMPORTANT: This script is totally inflexible so follow these instructions exactly
#####
#
# Input data must be stored as follows:
# top directory contains one sub-directory for each subject
# subject directories contain the .bam files
#
# methratio.py will run twice for each subject
# the second run will combine CpGs across strands (use -g option)
# output files will be filtered down to CpGs only, creating a total of 4 output files
#
# Command line args must be in the exact following order with spaces between
# make sure to put full pathnames for 1,3,4
# 1: full path to top directory with data
#     must contain only one sub-dir for each subject and nothing else
# 2: string to grep for suffixes of .bam files within subject dirs
#     should match only a single .bam file in the subject directory
# 3: full path to methratio.py script
# 4: full path to genome fasta file
# 5: base filename for output files
#     will be appended to .bam file name and followed by value for -m option
# 6: required coverage (-m option)
#
# Bio-RDF14:abseq$ BSseq2_methratio.sh data_dir bam_suffix script_path genome out_base req_cov
#
#####
# Following the code below is an example use and the console output
#####
```

```
data_dir="$1"
bam_suffix="$2"
script_path="$3"
genome="$4"
out_base="$5"
req_cov="$6"

# need to use old samtools because of methratio.py reference to deprecated -X flag for samtools view
sam_path "~/Documents/samtools-0.1.19"

cd "$data_dir"
subjects=$(ls)
echo -e "\n===="
echo -e "Moved to:\n ${data_dir}"
echo -----
echo -e "Subject dirs are:\n ${subjects}"
echo -----
echo -e "Will analyze .bam files ending in:\n ${bam_suffix}"
echo -----
echo -e "Will use methratio.py at:\n ${script_path}"
echo -----
echo -e "Will use genome file:\n ${genome}"
echo -----
read -rsp "If this is all good press any key to continue, or ctrl+c to quit..." -n1
echo
```

```

for s in $subjects
do
    echo -e "\n===="
    echo "Working on ${s}"
    echo "===="
    bam=$(ls "${s}/*${bam_suffix}")
    echo -e "Analyzing:\n ${bam}"
    echo -e "Will write output files:\n ${bam}_$out_base-m$req_cov"
    echo " ${bam}_$out_base-m$req_cov.CG"
    echo " ${bam}_$out_base-m$req_cov-CpGcombined"
    echo " ${bam}_$out_base-m$req_cov-CpGcombined.CG"

    echo -e "\n---"
    echo "Running, no -g option..."
    echo "----"
    "$script_path" \
-o "${bam}_$out_base-m$req_cov" \
-d "$genome" \
-s "$sam_path" \
-u -p -z -r -m "$req_cov" \
"$bam"

    echo -e "\n---"
    echo "Running, yes -g option..."
    echo "----"
    "$script_path" \
-o "${bam}_$out_base-m$req_cov-CpGcombined" \
-d "$genome" \
-s "$sam_path" \
-u -p -z -r -g -m "$req_cov" \
"$bam"

    echo -e "\n---"
    echo "Filtering both output files for CpGs..."
    echo "----"

#####
## Old regexes required for older versions of methratio.py script that reported context as
## e.g. GGCGTT instead of CG, ATCTGG instead of CHG, GTCAAA instead of CHH,
## and did not convert nucleotides for minus strand hits

# awk '($3=="-" && $4~/^.{1}CG/ ) || ($3== "+" && $4~/^.{2}CG/)' \
# $bam"_$out_base-m$req_cov > $bam"_$out_base-m$req_cov.CG"
# awk '($3=="-" && $4~/^.{1}CG/ ) || ($3== "+" && $4~/^.{2}CG/)' \
# $bam"_$out_base-m$req_cov-CpGcombined" > $bam"_$out_base-m$req_cov-CpGcombined.CG"
#
#####

echo $bam"_$out_base-m$req_cov..."
awk '$4=="CG"' $bam"_$out_base-m$req_cov \
> $bam"_$out_base-m$req_cov.CG"

echo $bam"_$out_base-m$req_cov-CpGcombined..."
awk '$4=="CG"' $bam"_$out_base-m$req_cov-CpGcombined" \
> $bam"_$out_base-m$req_cov-CpGcombined.CG"
done

```

```
#####
#
# Bio-RDF14:_BS-seq_data abseq$ /Volumes/fishstudies-1/_scripts/BSseq2_methratio.sh \
# ~/Documents/_BS-seq_data/BROAD_genome/ \
# trimmed4-98.adapters.q30.m0_bsmap2.9.bam \
# ~/Documents-bsmap-2.90/methratio.py \
# ~/Documents/H_burtoni_v1.assembly.fa \
# methratio_samtools0.1.19 \
# 4
#
# =====
#
# Moved to:
# /Users/abseq/Documents/_BS-seq_data/BROAD_genome/
# -----
#
# Subject dirs are:
# 3157_TENNISON
# 3165_BRISCOE
# 3581_LYNLEY
# 3677_MONK
# -----
#
# Will analyze .bam files ending in:
# trimmed4-98.adapters.q30.m0_bsmap2.9.bam
# -----
#
# Will use methratio.py at:
# /Users/abseq/Documents-bsmap-2.90/methratio.py
# -----
#
# Will use genome file:
# /Users/abseq/Documents/H_burtoni_v1.assembly.fa
# =====
#
# If this is all good press any key to continue, or ctrl+c to quit...
#
```

```
# =====
# Working on 3157_TENNISON
#
# Analyzing:
#   3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam
# Will write output files:
#   3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4
#   3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4.CG
#   3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined
#   3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined.CG
#
#
# -----
# Running, no -g option...
#
# [methratio] @Mon Mar 14 17:04:48 2016      loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa ...
# [methratio] @Mon Mar 14 17:07:41 2016      read 10000000 lines
# [methratio] @Mon Mar 14 17:10:07 2016      read 20000000 lines
# [methratio] @Mon Mar 14 17:12:26 2016      read 30000000 lines
# [methratio] @Mon Mar 14 17:14:54 2016      read 40000000 lines
# [methratio] @Mon Mar 14 17:17:18 2016      read 50000000 lines
# [methratio] @Mon Mar 14 17:19:36 2016      read 60000000 lines
# [methratio] @Mon Mar 14 17:21:50 2016      read 70000000 lines
# [methratio] @Mon Mar 14 17:24:06 2016      read 80000000 lines
# [methratio] @Mon Mar 14 17:26:16 2016      read 90000000 lines
# [methratio] @Mon Mar 14 17:28:24 2016      read 100000000 lines
# [methratio] @Mon Mar 14 17:30:26 2016      read 110000000 lines
# [methratio] @Mon Mar 14 17:32:34 2016      read 120000000 lines
# [methratio] @Mon Mar 14 17:34:30 2016      read 130000000 lines
# [methratio] @Mon Mar 14 17:36:14 2016      read 140000000 lines
# [methratio] @Mon Mar 14 17:38:03 2016      read 150000000 lines
# [methratio] @Mon Mar 14 17:39:39 2016      read 159873223 lines
# [methratio] @Mon Mar 14 17:39:39 2016      writing 3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4 ...
# [methratio] @Mon Mar 14 17:45:55 2016      total 49698939 valid mappings, 50195186 covered cytosines, average coverage: 17.07 fold.
#
#
# -----
# Running, yes -g option...
#
# [methratio] @Mon Mar 14 17:45:56 2016      loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa ...
# [methratio] @Mon Mar 14 17:49:40 2016      read 10000000 lines
# [methratio] @Mon Mar 14 17:52:46 2016      read 20000000 lines
# [methratio] @Mon Mar 14 17:55:39 2016      read 30000000 lines
# [methratio] @Mon Mar 14 17:58:05 2016      read 40000000 lines
# [methratio] @Mon Mar 14 18:00:23 2016      read 50000000 lines
# [methratio] @Mon Mar 14 18:02:40 2016      read 60000000 lines
# [methratio] @Mon Mar 14 18:04:52 2016      read 70000000 lines
# [methratio] @Mon Mar 14 18:07:07 2016      read 80000000 lines
# [methratio] @Mon Mar 14 18:09:16 2016      read 90000000 lines
# [methratio] @Mon Mar 14 18:11:22 2016      read 100000000 lines
# [methratio] @Mon Mar 14 18:13:25 2016      read 110000000 lines
# [methratio] @Mon Mar 14 18:15:46 2016      read 120000000 lines
# [methratio] @Mon Mar 14 18:17:57 2016      read 130000000 lines
# [methratio] @Mon Mar 14 18:19:41 2016      read 140000000 lines
# [methratio] @Mon Mar 14 18:21:08 2016      read 150000000 lines
# [methratio] @Mon Mar 14 18:22:17 2016      read 159873223 lines
# [methratio] @Mon Mar 14 18:22:17 2016      combining CpG methylation from both strands ...
# [methratio] @Mon Mar 14 18:22:33 2016      writing 3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined ...
# [methratio] @Mon Mar 14 18:27:42 2016      total 49698939 valid mappings, 49127032 covered cytosines, average coverage: 17.49 fold.
#
#
# -----
# Filtering both output files for CpGs...
#
# 3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4...
# 3157_TENNISON/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined...
```

```
# =====
# Working on 3165_BRISCOE
#
# Analyzing:
# 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam
# Will write output files:
# 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio_samtools0.1.19-m4
# 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio_samtools0.1.19-m4.CG
# 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio_samtools0.1.19-m4-CpGcombined
# 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio_samtools0.1.19-m4-CpGcombined.CG
#
#
# Running, no -g option...
#
# [methratio] @Mon Mar 14 18:32:41 2016    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa ...
# [methratio] @Mon Mar 14 18:35:30 2016    read 10000000 lines
# [methratio] @Mon Mar 14 18:37:47 2016    read 20000000 lines
# [methratio] @Mon Mar 14 18:39:57 2016    read 30000000 lines
# [methratio] @Mon Mar 14 18:42:12 2016    read 40000000 lines
# [methratio] @Mon Mar 14 18:44:31 2016    read 50000000 lines
# [methratio] @Mon Mar 14 18:46:41 2016    read 60000000 lines
# [methratio] @Mon Mar 14 18:48:51 2016    read 70000000 lines
# [methratio] @Mon Mar 14 18:51:00 2016    read 80000000 lines
# [methratio] @Mon Mar 14 18:53:06 2016    read 90000000 lines
# [methratio] @Mon Mar 14 18:55:07 2016    read 100000000 lines
# [methratio] @Mon Mar 14 18:57:08 2016    read 110000000 lines
# [methratio] @Mon Mar 14 18:59:02 2016    read 120000000 lines
# [methratio] @Mon Mar 14 19:00:58 2016    read 130000000 lines
# [methratio] @Mon Mar 14 19:02:49 2016    read 140000000 lines
# [methratio] @Mon Mar 14 19:04:36 2016    read 150000000 lines
# [methratio] @Mon Mar 14 19:06:09 2016    read 160000000 lines
# [methratio] @Mon Mar 14 19:07:28 2016    read 170000000 lines
# [methratio] @Mon Mar 14 19:08:30 2016    read 180000000 lines
# [methratio] @Mon Mar 14 19:09:00 2016    read 184444368 lines
# [methratio] @Mon Mar 14 19:09:00 2016    writing 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio_samtools0.1.19-m4 ...
# [methratio] @Mon Mar 14 19:13:32 2016    total 52514031 valid mappings, 41704613 covered cytosines, average coverage: 22.20 fold.
#
#
# Running, yes -g option...
#
# [methratio] @Mon Mar 14 19:13:33 2016    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa ...
# [methratio] @Mon Mar 14 19:16:26 2016    read 10000000 lines
# [methratio] @Mon Mar 14 19:18:46 2016    read 20000000 lines
# [methratio] @Mon Mar 14 19:20:59 2016    read 30000000 lines
# [methratio] @Mon Mar 14 19:23:16 2016    read 40000000 lines
# [methratio] @Mon Mar 14 19:25:36 2016    read 50000000 lines
# [methratio] @Mon Mar 14 19:27:49 2016    read 60000000 lines
# [methratio] @Mon Mar 14 19:30:01 2016    read 70000000 lines
# [methratio] @Mon Mar 14 19:32:12 2016    read 80000000 lines
# [methratio] @Mon Mar 14 19:34:19 2016    read 90000000 lines
# [methratio] @Mon Mar 14 19:36:22 2016    read 100000000 lines
# [methratio] @Mon Mar 14 19:38:25 2016    read 110000000 lines
# [methratio] @Mon Mar 14 19:40:18 2016    read 120000000 lines
# [methratio] @Mon Mar 14 19:42:16 2016    read 130000000 lines
# [methratio] @Mon Mar 14 19:44:10 2016    read 140000000 lines
# [methratio] @Mon Mar 14 19:45:58 2016    read 150000000 lines
# [methratio] @Mon Mar 14 19:47:33 2016    read 160000000 lines
# [methratio] @Mon Mar 14 19:48:54 2016    read 170000000 lines
# [methratio] @Mon Mar 14 19:49:57 2016    read 180000000 lines
# [methratio] @Mon Mar 14 19:50:28 2016    read 184444368 lines
# [methratio] @Mon Mar 14 19:50:28 2016    combining CpG methylation from both strands ...
# [methratio] @Mon Mar 14 19:50:44 2016    writing 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio_samtools0.1.19-m4-CpGcombined ...
# [methratio] @Mon Mar 14 19:55:14 2016    total 52514031 valid mappings, 40840159 covered cytosines, average coverage: 22.70 fold.
#
#
# Filtering both output files for CpGs...
#
# 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio_samtools0.1.19-m4...
# 3165_BRISCOE/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio_samtools0.1.19-m4-CpGcombined...
```

```
# =====
# Working on 3581_LYNLEY
#
# Analyzing:
# 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam
# Will write output files:
# 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4
# 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4.CG
# 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined
# 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined.CG
#
#
# Running, no -g option...
#
# [methratio] @Mon Mar 14 19:59:31 2016    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa ...
# [methratio] @Mon Mar 14 20:02:30 2016    read 10000000 lines
# [methratio] @Mon Mar 14 20:04:59 2016    read 20000000 lines
# [methratio] @Mon Mar 14 20:07:20 2016    read 30000000 lines
# [methratio] @Mon Mar 14 20:09:45 2016    read 40000000 lines
# [methratio] @Mon Mar 14 20:12:13 2016    read 50000000 lines
# [methratio] @Mon Mar 14 20:14:24 2016    read 60000000 lines
# [methratio] @Mon Mar 14 20:16:44 2016    read 70000000 lines
# [methratio] @Mon Mar 14 20:19:05 2016    read 80000000 lines
# [methratio] @Mon Mar 14 20:21:12 2016    read 90000000 lines
# [methratio] @Mon Mar 14 20:23:26 2016    read 100000000 lines
# [methratio] @Mon Mar 14 20:25:22 2016    read 110000000 lines
# [methratio] @Mon Mar 14 20:27:26 2016    read 120000000 lines
# [methratio] @Mon Mar 14 20:29:27 2016    read 130000000 lines
# [methratio] @Mon Mar 14 20:31:26 2016    read 140000000 lines
# [methratio] @Mon Mar 14 20:33:07 2016    read 150000000 lines
# [methratio] @Mon Mar 14 20:34:33 2016    read 160000000 lines
# [methratio] @Mon Mar 14 20:35:40 2016    read 170000000 lines
# [methratio] @Mon Mar 14 20:36:04 2016    read 173372346 lines
# [methratio] @Mon Mar 14 20:36:04 2016    writing 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4 ...
# [methratio] @Mon Mar 14 20:40:30 2016    total 49890822 valid mappings, 39656095 covered cytosines, average coverage: 22.14 fold.
#
#
# Running, yes -g option...
#
# [methratio] @Mon Mar 14 20:40:31 2016    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa ...
# [methratio] @Mon Mar 14 20:43:23 2016    read 10000000 lines
# [methratio] @Mon Mar 14 20:45:44 2016    read 20000000 lines
# [methratio] @Mon Mar 14 20:47:58 2016    read 30000000 lines
# [methratio] @Mon Mar 14 20:50:16 2016    read 40000000 lines
# [methratio] @Mon Mar 14 20:52:37 2016    read 50000000 lines
# [methratio] @Mon Mar 14 20:54:42 2016    read 60000000 lines
# [methratio] @Mon Mar 14 20:56:54 2016    read 70000000 lines
# [methratio] @Mon Mar 14 20:59:10 2016    read 80000000 lines
# [methratio] @Mon Mar 14 21:01:12 2016    read 90000000 lines
# [methratio] @Mon Mar 14 21:03:20 2016    read 100000000 lines
# [methratio] @Mon Mar 14 21:05:10 2016    read 110000000 lines
# [methratio] @Mon Mar 14 21:07:08 2016    read 120000000 lines
# [methratio] @Mon Mar 14 21:09:03 2016    read 130000000 lines
# [methratio] @Mon Mar 14 21:10:56 2016    read 140000000 lines
# [methratio] @Mon Mar 14 21:12:32 2016    read 150000000 lines
# [methratio] @Mon Mar 14 21:13:53 2016    read 160000000 lines
# [methratio] @Mon Mar 14 21:14:58 2016    read 170000000 lines
# [methratio] @Mon Mar 14 21:15:21 2016    read 173372346 lines
# [methratio] @Mon Mar 14 21:15:21 2016    combining CpG methylation from both strands ...
# [methratio] @Mon Mar 14 21:15:37 2016    writing 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined ...
# [methratio] @Mon Mar 14 21:19:55 2016    total 49890822 valid mappings, 38858234 covered cytosines, average coverage: 22.63 fold.
#
#
# Filtering both output files for CpGs...
#
# 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4...
# 3581_LYNLEY/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined...
```

```
# =====
# Working on 3677_MONK
#
# Analyzing:
# 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam
# Will write output files:
# 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4
# 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4.CG
# 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined
# 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined.CG
#
# -----
# Running, no -g option...
#
# [methratio] @Mon Mar 14 21:24:06 2016    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa ...
# [methratio] @Mon Mar 14 21:27:00 2016    read 10000000 lines
# [methratio] @Mon Mar 14 21:29:25 2016    read 20000000 lines
# [methratio] @Mon Mar 14 21:31:43 2016    read 30000000 lines
# [methratio] @Mon Mar 14 21:34:04 2016    read 40000000 lines
# [methratio] @Mon Mar 14 21:36:26 2016    read 50000000 lines
# [methratio] @Mon Mar 14 21:38:36 2016    read 60000000 lines
# [methratio] @Mon Mar 14 21:40:53 2016    read 70000000 lines
# [methratio] @Mon Mar 14 21:43:07 2016    read 80000000 lines
# [methratio] @Mon Mar 14 21:45:11 2016    read 90000000 lines
# [methratio] @Mon Mar 14 21:47:23 2016    read 100000000 lines
# [methratio] @Mon Mar 14 21:49:16 2016    read 110000000 lines
# [methratio] @Mon Mar 14 21:51:20 2016    read 120000000 lines
# [methratio] @Mon Mar 14 21:53:20 2016    read 130000000 lines
# [methratio] @Mon Mar 14 21:55:08 2016    read 140000000 lines
# [methratio] @Mon Mar 14 21:56:44 2016    read 150000000 lines
# [methratio] @Mon Mar 14 21:58:03 2016    read 160000000 lines
# [methratio] @Mon Mar 14 21:59:01 2016    read 168836540 lines
# [methratio] @Mon Mar 14 21:59:01 2016    writing 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4 ...
# [methratio] @Mon Mar 14 22:03:33 2016    total 50337454 valid mappings, 39975402 covered cytosines, average coverage: 22.06 fold.
#
# -----
# Running, yes -g option...
#
# [methratio] @Mon Mar 14 22:03:34 2016    loading reference file: /Users/abseq/Documents/H_burtoni_v1.assembly.fa ...
# [methratio] @Mon Mar 14 22:06:28 2016    read 10000000 lines
# [methratio] @Mon Mar 14 22:08:52 2016    read 20000000 lines
# [methratio] @Mon Mar 14 22:11:09 2016    read 30000000 lines
# [methratio] @Mon Mar 14 22:13:30 2016    read 40000000 lines
# [methratio] @Mon Mar 14 22:15:52 2016    read 50000000 lines
# [methratio] @Mon Mar 14 22:18:01 2016    read 60000000 lines
# [methratio] @Mon Mar 14 22:20:16 2016    read 70000000 lines
# [methratio] @Mon Mar 14 22:22:30 2016    read 80000000 lines
# [methratio] @Mon Mar 14 22:24:34 2016    read 90000000 lines
# [methratio] @Mon Mar 14 22:26:45 2016    read 100000000 lines
# [methratio] @Mon Mar 14 22:28:37 2016    read 110000000 lines
# [methratio] @Mon Mar 14 22:30:39 2016    read 120000000 lines
# [methratio] @Mon Mar 14 22:32:38 2016    read 130000000 lines
# [methratio] @Mon Mar 14 22:34:26 2016    read 140000000 lines
# [methratio] @Mon Mar 14 22:36:02 2016    read 150000000 lines
# [methratio] @Mon Mar 14 22:37:20 2016    read 160000000 lines
# [methratio] @Mon Mar 14 22:38:19 2016    read 168836540 lines
# [methratio] @Mon Mar 14 22:38:19 2016    combining CpG methylation from both strands ...
# [methratio] @Mon Mar 14 22:38:35 2016    writing 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined ...
# [methratio] @Mon Mar 14 22:42:56 2016    total 50337454 valid mappings, 39184975 covered cytosines, average coverage: 22.54 fold.
#
# -----
# Filtering both output files for CpGs...
#
# 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4...
# 3677_MONK/aligned_trimmed4-98.adapters.q30.m0_bsmap2.9.bam_methratio.samtools0.1.19-m4-CpGcombined...
```