# Parameters used for msGBS analysis:

NOTE:

Naming of samples : The script uses the names of the samples to understand the type of sample it is (Mono, standard or ratio sample). The samples need a -per sample type – uniform prefix. To use your own prefix in the Make\_Reference\_msGBS.py script replace ‘jenamono’ by the prefix you prefer. The mono prefix should be followed by a number starting by 1. Ratio samples should start by ‘ratio\_’ followed by a number starting by 1. Standard samples should start by ‘standaard\_’ followed by a number starting by 1. The script ‘Parse\_VCF.py’ was written to work with 2 species pools and should be adjusted to your own project (for example in the script a list exists were ‘standaard’ is be replaced to ‘STD’ for the standards you want to use as standard). In this way a standard that is clearly different from the others can simply be neglected by removing it from this list. Sorry if your programming /python skills are not enough to work with this file. Please contact me ([n.wagemaker@science.ru.nl](mailto:n.wagemaker@science.ru.nl)) if I can be of some help.

Good luck,

Niels Wagemaker

## Demultiplex msGBS

**The output of this script is adjusted to work with the NGmerge version of the make\_reference script. (as used in article)**

**IN folder**

/scratch/niels/raw/SGBS/

**IN files**

/scratch2/niels/raw/SGBS/barcodes\_SGBS\_pilot3.txt

/scratch2/niels/raw/SGBS/SGBS\_GBS\_pilot3\_KD17072297\_H53KHCCXY\_L6\_1.fq.gz

/scratch2/niels/raw/SGBS/SGBS\_GBS\_pilot3\_KD17072297\_H53KHCCXY\_L6\_2.fq.gz

UIT folder:

/scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge

**SCRIPT location:**

/scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/demultiplex\_msGBS.py

**Syntax:**

Screen –R demultiplex NGmerge

python demultiplex\_msGBS.py --r1 /scratch2/niels/raw/SGBS/SGBS\_GBS\_pilot3\_KD17072297\_H53KHCCXY\_L6\_1.fq.gz --r2 /scratch2/niels/raw/SGBS/SGBS\_GBS\_pilot3\_KD17072297\_H53KHCCXY\_L6\_2.fq.gz --mode pe -b /scratch2/niels/raw/SGBS/barcodes\_SGBS\_pilot3.txt --addRG -d --stat /scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/stat.txt /scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/NM2.fq.gz -o /scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/

**Adapter removal and trimming:**

Two files are needed in script folder :

adapters.txt

qual\_profile.txt

AdapterRemoval --file1 R1\_demultiplex\_NGmerge\_H53KHCCXY\_s\_6\_fastq.txt.gz --file2 R2\_demultiplex\_NGmerge\_H53KHCCXY\_s\_6\_fastq.txt.gz --basename output\_multi2 --trimns --trimqualities --adapter-list adapters.txt --gzip

## **Make\_reference parameters**

95% identity clustering

GIT NAME Script : Demultiplex\_msGBS.py

Original NAME script : NGMERGE\_make\_ref\_adj\_demultiplex\_PILOT3.py

**Uit folder**

/scratch2/niels/raw/SGBS

/scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/seqiroLML

**IN files**

/scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/seqiroLML/output\_multi2.pair1.truncated.gz

/scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/seqiroLML/output\_multi2.pair2.truncated.gz

/scratch2/niels/raw/SGBS/barcodes\_SGBS\_pilot3.txt

SCRIPT location:

Original NAME :NGMERGE\_make\_ref\_adj\_demultiplex.py

New GIT name : Demultiplex\_msGBS.py

**OUT folder**

/scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge

**Syntax:**

Screen -R PILOT3\_make\_ref

**Go to folder with script**

SCRATCH2

python NGMERGE\_make\_ref\_adj\_demultiplex\_PILOT3.py --forward /scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/seqiroLML/output\_multi2.pair1.truncated.gz --reverse /scratch2/niels/MERGETEST\_PILOT3/demultiplex\_NGmerge/seqiroLML/output\_multi2.pair2.truncated.gz -o /scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/merged.fastq.txt.gz --barcodes /scratch2/niels/raw/SGBS/barcodes\_SGBS\_pilot3.txt --cycles 150 -t /scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/tmp/ --threads 14 --outputdir /scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/ --log /scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/log.txt --samout SAMOUT --consensus CONSENSUS --consensus\_cluster CONSENSUS\_CLUSTER -q /scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/qual\_profile.txt -n 14 -u /scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/merged.unassembled

OUTPUT DIR IS :/scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge

Rename\_fast.py is used in script and can be found on github

**BLAST against NR database:**

blastn -query /scratch2/niels/ref/PILOT3/NGmerge/ref.fa -db nt -out /scratch2/niels/ref/PILOT3/outputblast\_kingdoms\_nt\_original\_tax.txt -num\_alignments 1 -num\_threads 12 -outfmt '6 qseqid sseqid pident evalue bitscore sskingdom sscinames length sstart send '

**separation of Bacteria and other contigs (vnl Eukaryote) contigs using script: github: NielsWagemaker/msGBS**/**scripts\_msGBS/**

Mapping 6 MRT : USE REF **FILTERED Flowering plant loci only** KINDOM OUT BY **NR BLAST**

script

Map\_STAR\_msGBS.py

Input dir:

/scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge

ref\_Eukaryota.txt = ref.fa (renamed file)

infiles:

/scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/merged.unassembled\_1\_correct\_header.fastq.gz

/scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/merged.unassembled\_2\_correct\_header.fastq.gz

/scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge/merged.assembled\_correct\_header.fastq.gz

/scratch2/niels/MERGETEST\_DINA/NGmerge\_reference/ref.fa

/scratch2/niels/raw/SGBS/barcodes\_SGBS\_pilot3.txt

output folder:

/niels/MERGETEST\_PILOT3/mapping\_NGmerge/radical

Syntax:

Screen –R pilot3\_ mapping

**Go to folder with script (op scratch2)**

Map\_STAR\_msGBS.py

python Map\_STAR\_msGBS.py --input\_dir /scratch2/niels/MERGETEST\_PILOT3/reference\_NGmerge --tmpdir /scratch2/niels/MERGETEST\_PILOT3/mapping\_NGmerge/radical\_flower/tmp --threads 24 --output\_dir /scratch2/niels/MERGETEST\_PILOT3/mapping\_NGmerge/radical\_flower --barcodes /scratch2/niels/raw/SGBS/barcodes\_SGBS\_pilot3.txt

**Mark\_PCR\_duplicates** 0.8 fail

Input files:

/scratch2/niels/raw/SGBS/barcodes\_SGBS\_pilot3.txt

/scratch2/niels/reference/SGBS3/ref.fa

/scratch2/niels/mapping/SGBS3/out.bam

outfile:

/scratch2/niels/mapping/SGBS3/out.dedup08.bam

Synthax:

python mark\_PCR\_duplicates\_08\_qual.py -i /scratch2/niels/mapping/SGBS3/out.bam -b /scratch2/niels/raw/SGBS/barcodes\_SGBS\_pilot3.txt -r /scratch2/niels/reference/SGBS3/ref.fa -o /scratch2/niels/mapping/SGBS3/out.dedup08.bam

**last step of script : Sort out.dedup.bam from pycharm did not work :**

**did it manualy afterwards: (works from server …)**

samtools index out.sorted.bam

takes 4 minuten

samtools sort -m 70G out.dedup.bam > out.dedup.sorted.bam

takes 19 minuten

samtools index out.dedup.sorted.bam

takes 4 minuten

# Parse BAM file to get raw read numbers:

# Parse CSV file to get final output files:

NAME : Parse\_VCF.py

Python Parse\_VCF.py -i /Volumes/Extreme900/SGBS/mapping/mapping\_Nov\_2018\_ref08\_map085\_noE/out\_dedup\_08.csv -op

/Volumes/Extreme900/SGBS/mapping/mapping\_Nov\_2018\_ref08\_map085\_noE/dedup\_08/nov\_2018\_testing\_ref80\_map085\_noBLAST\_dedup08\_ -os \_8\_15\_nov18.csv -f1 8 -f2 15 -f3

1000 --pool 1 --extra 'yes'

NOTE: You need to adjust this script to your own experimental conditions