####Original Instructions by corey campbell Aug 2019. Modified by amejia, slozano, and ksaavedra Oct 2024

# Overview

The high throughput SNP analysis pipeline starts with fastq reads and ends with readcounts for each SNP according to reference genome SNP IDs. Readcounts are input into the R pipeline for population genetics calculations of LOD. The following workflow has been used in many exposed populations <citations>. This workflow uses as an example temephos (commercial name Abate) treated pools and Mexican natural population (Tapachula, Chiapas; 5 de Febrero neighborghood) datasets for exome-enriched sequencing libraries.

# Required programs

awk

cutadapt

gmap / gsnap (http://research-pub.gene.com/gmap)

pigz (optional but highly recommended)

samtools

# Required data

Reference genome

\*.fastq files

# Steps

1. Create the index file (Genome map) of the *Aedes aegypti* genome using the gmapdb program <citation>. This step is necessary if you are starting with a new reference genebuild. The index file must be generated prior to running GSNAP. The AaegL5 reference (2017) was used to index in idangs01, so if you are using it, you can skip to 2.
   1. First, copy the reference to /data/black\_lab/gmapdb
   2. Example command to de-compress a .gz file

$ pigz -d AaegL5.fasta.gz

* 1. Example command to build new reference (gmap\_build -d <genome> [-k <kmer size>] <fasta\_files...>)

$ gmap\_build -d AaegL5\_primaryDB -k 15 AaegL5\_primary.fsa

1. Align paired-end read libraries to the reference genome. The example command below uses the new AaegL5 reference file (/data/black\_lab/ReferenceSeqs/AGWG\_AaegL5\_primary.fsa). FYI- the example below is a single cmd and should be written on one line. It aligns raw fastq files to the AaegL4 index file, allowing a 10% mismatch and outputs a .sam file. In this example, our source files are in our current directory.
2. If starting with raw reads, the fastq files need to be quality trimmed prior to processing. We use cutadapt for this task. Migrate to fastq file directory.

Example cmd for Illumina lib adapters-

$ cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -g GCTCTTCCGATCT -G GCTCTTCCGATCT -a AGATGTGTATAAGAGACAG -A AGATGTGTATAAGAGACAG -g CTGTCTCTTATACACATCT -G CTGTCTCTTATACACATCT -q 30,30 --minimum-length 80 -o 5Feb\_Chl\_A2\_R1\_paired.fastq -p 5Feb\_Chl\_A2\_R2\_paired.fastq 5Feb\_Chlorpyr\_A2\_S52\_L004\_R1\_001.fastq 5Feb\_Chlorpyr\_A2\_S52\_L004\_R2\_001.fastq

1. Map reads to reference genome

$ gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam 5Feb\_Tem\_A1\_R1\_paired.fastq 5Feb\_Tem\_A1\_R2\_paired.fastq > 5Feb\_Tem\_A1\_L5.sam

1. Convert .sam to .bam file

$ samtools view -bS -o VCDA\_REP1\_L5.bam VCDA\_REP1\_L5.sam

1. Convert .bam to sorted .bam

$samtools sort VCDA\_REP1\_L5.bam -o VCDA\_REP1\_sorted.bam

1. The next command will generate an mpileup file. Be sure to have a genome reference file in the same folder for mpileup to use in the alignment. Mpileup will generate its own index when it runs the first time.

$ samtools mpileup -f /data/black\_lab/ReferenceSeqs/ AGWG\_AaegL5\_primary.fsa VCDA\_REP1\_sorted.bam > VCDA\_REP1\_L5.pileup

1. Remove lines with zero reads from mpileup prior to Varscan.

$ awk '$4 != 0 {print $0}' VCDA\_REP1\_L5.pileup > VCDA\_REP1\_no\_zeros.pileup

1. Convert mpileup file to readcounts file using Varscan.

$ java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts VCDA\_REP1\_no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file VCDA\_REP1.readcounts

# EXAMPLE output from Varscan

VCD\_1D.readcounts

Min coverage: 15

Min base qual: 30

Reading input from VCD\_1D\_no\_zeros.pileup

446914553 positions in pileup file

446914553 included in readcount analysis

34150107 met minimum coverage

1. Modifying Varscan readcount files. Replace header labels with sorter label

#replace in place exact string (e.g., AGWG\_AaegL5\_hic\_chr\_1\_PBJ\_arrow) with another string (  
1”)

$ awk -i inplace '{ gsub("AGWG\_AaegL5\_hic\_chr\_1\_PBJ\_arrow", "1"); print }' 5feb\_tem\_a1.readcounts

$ awk -i inplace '{ gsub("AGWG\_AaegL5\_hic\_chr\_2\_PBJ\_arrow", "2"); print }' 5feb\_tem\_a1.readcounts

$ awk -i inplace '{ gsub("AGWG\_AaegL5\_hic\_chr\_3\_PBJ\_arrow", "3"); print }' 5feb\_tem\_a1.readcounts

#replace AGWG\_AaegL5\_hic\_scaff\_

$ awk -i inplace '{ gsub("AGWG\_AaegL5\_hic\_scaff\_", ""); print }' 5feb\_tem\_a1.readcounts

$ awk -i inplace '{ gsub("\_PBJ\_arrow", ""); print }' 5feb\_tem\_a1.readcounts

#replace tabs with a single space

$ awk -i inplace '{ gsub(/\t/, " "); print }' col\_chlorpyr\_d1.readcounts

1. Send readcounts file to R pipeline to continue analysis

# Examples

####9/4/19 I will run the step to align to AaelL5\_ Colinas Temephos\_A1, A2, D1, D2, each library has two ends....two files each

##Colinas Tem A1

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam Col\_Tem\_A1\_R1\_paired.fastq Col\_Tem\_A1\_R2\_paired.fastq > Col\_Tem\_A1\_L5.sam

##Colinas Tem A2

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam Col\_Tem\_A2\_R1\_paired.fastq Col\_Tem\_A2\_R2\_paired.fastq > Col\_Tem\_A2\_L5.sam

##Colinas Tem D1

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam Col\_Tem\_D1\_R1\_paired.fastq Col\_Tem\_D1\_R2\_paired.fastq > Col\_Tem\_D1\_L5.sam

##Colinas Tem D2

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam Col\_Tem\_D2\_R1\_paired.fastq Col\_Tem\_D2\_R2\_paired.fastq > Col\_Tem\_D2\_L5.sam

#########9/4/19 transforming sam to bam

samtools view -bS -o Col\_Tem\_A1\_L5.bam Col\_Tem\_A1\_L5.sam

samtools view -bS -o Col\_Tem\_A2\_L5.bam Col\_Tem\_A2\_L5.sam

samtools view -bS -o Col\_Tem\_D1\_L5.bam Col\_Tem\_D1\_L5.sam

samtools view -bS -o Col\_Tem\_D2\_L5.bam Col\_Tem\_D2\_L5.sam

#########9/4/19 transforming bam to sorted.bam

samtools sort Col\_Tem\_A1\_L5.bam -o Col\_Tem\_A1\_L5.sorted.bam

samtools sort Col\_Tem\_A2\_L5.bam -o Col\_Tem\_A2\_L5.sorted.bam

samtools sort Col\_Tem\_D1\_L5.bam -o Col\_Tem\_D1\_L5.sorted.bam

samtools sort Col\_Tem\_D2\_L5.bam -o Col\_Tem\_D2\_L5.sorted.bam

####Starting from compressed fastq.files

###############Col-Perm 11/20/2019

####1) decompress gz files (fasqfiles folder)

cd /data/black\_lab/Col-Perm

gzip --decompress \*.gz

####2) 11/21/19 remove adapter sequence. Change the name of the file

####Alive1

cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -g GCTCTTCCGATCT -G GCTCTTCCGATCT -a AGATGTGTATAAGAGACAG -A AGATGTGTATAAGAGACAG -g CTGTCTCTTATACACATCT -G CTGTCTCTTATACACATCT -q 30,30 --minimum-length 80 -o col\_perm\_A1\_R1\_paired.fastq -p col\_perm\_A1\_R2\_paired.fastq Col\_Perm\_A1\_S37\_L004\_R1\_001.fastq Col\_Perm\_A1\_S37\_L004\_R2\_001.fastq

####Alive2

cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -g GCTCTTCCGATCT -G GCTCTTCCGATCT -a AGATGTGTATAAGAGACAG -A AGATGTGTATAAGAGACAG -g CTGTCTCTTATACACATCT -G CTGTCTCTTATACACATCT -q 30,30 --minimum-length 80 -o col\_perm\_A2\_R1\_paired.fastq -p col\_perm\_A2\_R2\_paired.fastq Col\_Perm\_A2\_S38\_L004\_R1\_001.fastq Col\_Perm\_A2\_S38\_L004\_R2\_001.fastq

####Dead1

cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -g GCTCTTCCGATCT -G GCTCTTCCGATCT -a AGATGTGTATAAGAGACAG -A AGATGTGTATAAGAGACAG -g CTGTCTCTTATACACATCT -G CTGTCTCTTATACACATCT -q 30,30 --minimum-length 80 -o col\_perm\_D1\_R1\_paired.fastq -p col\_perm\_D1\_R2\_paired.fastq Col\_Perm\_D1\_S41\_L004\_R1\_001.fastq Col\_Perm\_D1\_S41\_L004\_R2\_001.fastq

####Dead2

cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -g GCTCTTCCGATCT -G GCTCTTCCGATCT -a AGATGTGTATAAGAGACAG -A AGATGTGTATAAGAGACAG -g CTGTCTCTTATACACATCT -G CTGTCTCTTATACACATCT -q 30,30 --minimum-length 80 -o col\_perm\_D2\_R1\_paired.fastq -p col\_perm\_D2\_R2\_paired.fastq Col\_Perm\_D2\_S42\_L004\_R1\_001.fastq Col\_Perm\_D2\_S42\_L004\_R2\_001.fastq

###Rec1

cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -g GCTCTTCCGATCT -G GCTCTTCCGATCT -a AGATGTGTATAAGAGACAG -A AGATGTGTATAAGAGACAG -g CTGTCTCTTATACACATCT -G CTGTCTCTTATACACATCT -q 30,30 --minimum-length 80 -o col\_perm\_R1\_R1\_paired.fastq -p col\_perm\_R1\_R2\_paired.fastq Col\_Perm\_R1\_S39\_L004\_R1\_001.fastq Col\_Perm\_R1\_S39\_L004\_R2\_001.fastq

###Rec2

cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -g GCTCTTCCGATCT -G GCTCTTCCGATCT -a AGATGTGTATAAGAGACAG -A AGATGTGTATAAGAGACAG -g CTGTCTCTTATACACATCT -G CTGTCTCTTATACACATCT -q 30,30 --minimum-length 80 -o col\_perm\_R2\_R1\_paired.fastq -p col\_perm\_R2\_R2\_paired.fastq Col\_Perm\_R2\_S40\_L004\_R1\_001.fastq Col\_Perm\_R2\_S40\_L004\_R2\_001.fastq

#####3)Align to AaeL5 reference genome

###3a)Convert end-pairs of fastq files to .sam file

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam col\_perm\_A1\_R1\_paired.fastq col\_perm\_A1\_R2\_paired.fastq > col\_perm\_A1\_R1\_L5.sam

####

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam col\_perm\_A2\_R1\_paired.fastq col\_perm\_A2\_R2\_paired.fastq > col\_perm\_A2\_L5.sam

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam col\_perm\_D1\_R1\_paired.fastq col\_perm\_D1\_R2\_paired.fastq > col\_perm\_D1\_L5.sam

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam col\_perm\_D2\_R1\_paired.fastq col\_perm\_D2\_R2\_paired.fastq > col\_perm\_D2\_L5.sam

####

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam col\_perm\_R1\_R1\_paired.fastq col\_perm\_R1\_R2\_paired.fastq > col\_perm\_R1\_L5.sam

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam col\_perm\_R2\_R1\_paired.fastq col\_perm\_R2\_R2\_paired.fastq > col\_perm\_R2\_L5.sam

####4) Convert Sam to Bam

samtools view -bS -o col\_perm\_A1\_L5.bam col\_perm\_A1\_L5.sam

samtools view -bS -o col\_perm\_A2\_L5.bam col\_perm\_A2\_L5.sam

samtools view -bS -o col\_perm\_D1\_L5.bam col\_perm\_D1\_L5.sam

samtools view -bS -o col\_perm\_D2\_L5.bam col\_perm\_D2\_L5.sam

samtools view -bS -o col\_perm\_R1\_L5.bam col\_perm\_R1\_L5.sam

samtools view -bS -o col\_perm\_R2\_L5.bam col\_perm\_R2\_L5.sam

###########5) Convert to sorted.bam

samtools sort col\_perm\_A1\_L5.bam -o col\_perm\_A1\_L5\_sorted.bam

samtools sort col\_perm\_A2\_L5.bam -o col\_perm\_A2\_L5\_sorted.bam

samtools sort col\_perm\_D1\_L5.bam -o col\_perm\_D1\_L5\_sorted.bam

samtools sort col\_perm\_D2\_L5.bam -o col\_perm\_D2\_L5\_sorted.bam

samtools sort col\_perm\_R1\_L5.bam -o col\_perm\_R1\_L5\_sorted.bam

samtools sort col\_perm\_R2\_L5.bam -o col\_perm\_R2\_L5\_sorted.bam

##############6) align to aael5

###This works 11/25/19 Copy the AGWG\_AaegL5\_primary.fsa in the folder with the sorted.bam files

samtools mpileup -f /data/black\_lab/fastqfiles/Tem-Chlor-Perm/Col-Perm/AGWG\_AaegL5\_primary.fsa col\_perm\_A1\_sorted.bam > col\_perm\_A1.pileup

###works 11/26/19

samtools mpileup -f AGWG\_AaegL5\_primary.fsa col\_perm\_A2\_L5\_sorted.bam > col\_perm\_A2.pileup

samtools mpileup -f AGWG\_AaegL5\_primary.fsa col\_perm\_D1\_L5\_sorted.bam > col\_perm\_D1.pileup

samtools mpileup -f AGWG\_AaegL5\_primary.fsa col\_perm\_D2\_L5\_sorted.bam > col\_perm\_D2.pileup

samtools mpileup -f AGWG\_AaegL5\_primary.fsa col\_perm\_R1\_L5\_sorted.bam > col\_perm\_R1.pileup

samtools mpileup -f AGWG\_AaegL5\_primary.fsa col\_perm\_R2\_L5\_sorted.bam > col\_perm\_R2.pileup

#####7)remove zeros from pileup 11/26/19

awk '$4 != 0 {print $0}' col\_perm\_A1.pileup > col\_perm\_A1\_no\_zeros.pileup

awk '$4 != 0 {print $0}' col\_perm\_A2.pileup > col\_perm\_A2\_no\_zeros.pileup

awk '$4 != 0 {print $0}' col\_perm\_D1.pileup > col\_perm\_D1\_no\_zeros.pileup

awk '$4 != 0 {print $0}' col\_perm\_D2.pileup > col\_perm\_D2\_no\_zeros.pileup

awk '$4 != 0 {print $0}' col\_perm\_R1.pileup > col\_perm\_R1\_no\_zeros.pileup

awk '$4 != 0 {print $0}' col\_perm\_R2.pileup > col\_perm\_R2\_no\_zeros.pileup

##########8) convert mpileup file to readcounts file using Varscan 11/26/19

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts col\_perm\_A1\_no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file col\_perm\_A1.readcounts

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts col\_perm\_A2\_no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file col\_perm\_A2.readcounts

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts col\_perm\_D1\_no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file col\_perm\_D1.readcounts

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts col\_perm\_D2\_no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file col\_perm\_D2.readcounts

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts col\_perm\_R1\_no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file col\_perm\_R1.readcounts

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts col\_perm\_R2\_no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file col\_perm\_R2.readcounts

###Next step is to open the readcount files using ultra edit and remove the header, tabs and description. Save as .rc just to differentiate from the

###non edited file

####11/27/2019 Start with 5 de Febrero Temephos

####Starting from compressed fastq.files move to /data/black\_lab/fastqfiles..../5Feb-Tem

####1) decompress gz files (fasqfiles folder)

gzip --decompress \*.gz

####2) 11/27/19 remove adapter sequence: I think file swere already procesed. skipto step 3

cutadapt -a AGATCGGAAGAGC -A AGATCGGAAGAGC -g GCTCTTCCGATCT -G GCTCTTCCGATCT -a AGATGTGTATAAGAGACAG -A AGATGTGTATAAGAGACAG -g CTGTCTCTTATACACATCT -G CTGTCTCTTATACACATCT -q 30,30 --minimum-length 80 -o col\_perm\_A1\_R1\_paired.fastq -p col\_perm\_A1\_R2\_paired.fastq Col\_Perm\_A1\_S37\_L004\_R1\_001.fastq Col\_Perm\_A1\_S37\_L004\_R2\_001.fastq

###3) 11/27/19 Convert end-pairs of fastq files to .sam file

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam 5Feb\_Tem\_A1\_R1\_paired.fastq 5Feb\_Tem\_A1\_R2\_paired.fastq > 5feb\_tem\_A1.sam

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam 5Feb\_Tem\_A2\_R1\_paired.fastq 5Feb\_Tem\_A2\_R2\_paired.fastq > 5feb\_tem\_A2.sam

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam 5Feb\_Tem\_D1\_R1\_paired.fastq 5Feb\_Tem\_D1\_R2\_paired.fastq > 5feb\_tem\_D1.sam

/data/black\_lab/gmap-2019-06-10/src/gsnap -d AaegL5\_primaryDB -D /data/black\_lab/gmapdb/AaegL5\_primaryDB --max-mismatches=0.10 --trim-mismatch-score 0 -i 2 -n 10 -Q --nofails -B 4 -t 8 --allow-pe-name-mismatch --format=sam 5Feb\_Tem\_D2\_R1\_paired.fastq 5Feb\_Tem\_D2\_R2\_paired.fastq > 5feb\_tem\_D2.sam

###4) 11/27/19 Convert Sam to Bam

samtools view -bS -o 5feb\_tem\_A1.bam 5feb\_tem\_A1.sam

samtools view -bS -o 5feb\_tem\_A2.bam 5feb\_tem\_A2.sam

samtools view -bS -o 5feb\_tem\_D1.bam 5feb\_tem\_D1.sam

samtools view -bS -o 5feb\_tem\_D2.bam 5feb\_tem\_D2.sam

###########5) Convert to sorted.bam

samtools sort 5feb\_tem\_A1.bam -o 5feb\_tem\_A1\_sorted.bam

samtools sort 5feb\_tem\_A2.bam -o 5feb\_tem\_A2\_sorted.bam

samtools sort 5feb\_tem\_D1.bam -o 5feb\_tem\_D1\_sorted.bam

samtools sort 5feb\_tem\_D2.bam -o 5feb\_tem\_D2\_sorted.bam

##############6) align to aael5

###Copy the AGWG\_AaegL5\_primary.fsa in the folder with the sorted.bam files

cp /data/black\_lab/fastqfiles/Tem-Chlor-Perm/Col-Tem/AGWG\_AaegL5\_primary.fsa /data/black\_lab/fastqfiles/Tem-Chlor-Perm/5Feb-Tem

###works 11/30/19

samtools mpileup -f AGWG\_AaegL5\_primary.fsa 5feb\_tem\_A1\_sorted.bam > 5feb\_tem\_A1.pileup

samtools mpileup -f AGWG\_AaegL5\_primary.fsa 5feb\_tem\_A2\_sorted.bam > 5feb\_tem\_A2.pileup

samtools mpileup -f AGWG\_AaegL5\_primary.fsa 5feb\_tem\_D1\_sorted.bam > 5feb\_tem\_D1.pileup

samtools mpileup -f AGWG\_AaegL5\_primary.fsa 5feb\_tem\_D2\_sorted.bam > 5feb\_tem\_D2.pileup

#####7)remove zeros from pileup 12/01/19

awk '$4 != 0 {print $0}' 5feb\_tem\_A1.pileup > 5feb\_tem\_A1.no\_zeros.pileup

awk '$4 != 0 {print $0}' 5feb\_tem\_A2.pileup > 5feb\_tem\_A2.no\_zeros.pileup

awk '$4 != 0 {print $0}' 5feb\_tem\_D1.pileup > 5feb\_tem\_D1.no\_zeros.pileup

awk '$4 != 0 {print $0}' 5feb\_tem\_D2.pileup > 5feb\_tem\_D2.no\_zeros.pileup

##########8) convert mpileup file to readcounts file using Varscan 12/01/19

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts 5feb\_tem\_A1.no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file 5feb\_tem\_A1.readcounts

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts 5feb\_tem\_A2.no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file 5feb\_tem\_A2.readcounts

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts 5feb\_tem\_D1.no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file 5feb\_tem\_D1.readcounts

java -jar /apps/VarScan-2.3.5/VarScan.v2.3.5.jar readcounts 5feb\_tem\_D2.no\_zeros.pileup --min-coverage 25 --min-base-qual 30 --output-file 5feb\_tem\_D2.readcounts

#########Remove data from readcounts files 12/5/2019 Start col\_perm #change name from readcounts to rcounts

awk '{ a[$1]++ } END { for (b in a) { print b } }' col\_chlorpyr\_d1.readcounts | sort | uniq > unique\_ids.txt

#replace in place exact string with another string

awk -i inplace '{ gsub("AGWG\_AaegL5\_hic\_chr\_1\_PBJ\_arrow", "1"); print }' col\_chlorpyr\_d1.readcounts

awk -i inplace '{ gsub("AGWG\_AaegL5\_hic\_chr\_2\_PBJ\_arrow", "2"); print }' col\_chlorpyr\_d1.readcounts

awk -i inplace '{ gsub("AGWG\_AaegL5\_hic\_chr\_3\_PBJ\_arrow", "3"); print }' col\_chlorpyr\_d1.readcounts

awk -i inplace '{ gsub("AGWG\_AaegL5\_hic\_scaff\_", ""); print }' col\_chlorpyr\_d1.readcounts

awk -i inplace '{ gsub("\_PBJ\_arrow", ""); print }' col\_chlorpyr\_d1.readcounts

# replace tabs with a single space

awk -i inplace '{ gsub(/\t/, " "); print }' col\_chlorpyr\_d1.readcounts