Starling-May18
Projects/Katarina Stuart/KStuart.Starling-Aug18/Sv10_NZstarlings/Data/2024-02-20.VarCalling

PDF Version generated by

Katarina Stuart (z5188231@ad.unsw.edu.au)

on

Sep 25, 2024 @01:00 PM AEST

Table of Contents

2024-02-20.VarCalling

Katarina Stuart (z5188231@ad.unsw.edu.au) - Feb 21, 2024, 3:10 PM GMT+11

Variant calling

index genome

```
module load BWA/0.7.17-GCC-11.3.0    #Check version. This is what A used

cd /nesi/nobackup/uoa02613/kstuart_projects/At4_MynaStarling/data/resources/genomes/
genome=Svulgaris_vAU_1.0.fasta
bwa index $genome
```

Read aligning

```
#!/bin/bash -e
#SBATCH --job-name=2024_02_21.align.sl
#SBATCH --output=%x_%j.errout
#SBATCH --mail-user=katarina.stuart@auckland.ac.nz
#SBATCH --mail-type=ALL
#SBATCH --time=1:00:00
#SBATCH --mem=4G
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=4
#SBATCH --account=uoa02613
#SBATCH --profile task
#SBATCH --array=1-202

module load SAMtools/1.16.1-GCC-11.3.0 BWA/0.7.17-GCC-11.3.0
ref=/nesi/nobackup/uoa02613/kstuart_projects/At4_MynaStarling/data/resources/genomes/Svulgaris_vAU_1.0.fasta
stats_dir=/nesi/nobackup/uoa02613/kstuart_projects/Sv10_NZstarlings/data/processing_rawdata/aligned/stats
out_dir=/nesi/nobackup/uoa02613/kstuart_projects/Sv10_NZstarlings/data/processing_rawdata/aligned

cd /nesi/nobackup/uoa02613/kstuart_projects/Sv10_NZstarlings/data/processing_rawdata/trimmed_reads_all
sample=$(find *fq.gz -printf "%f\n" | sed "${SLURM_ARRAY_TASK_ID}q;d")

echo Aligning ${sample}
sample_sai=${sample//.filtered.*fq.gz/.sai}
sample_sam=${sample_sai/.sai/.sam}

bwa aln -t 4 -B 5 $ref ${sample} > ${out_dir}/${sample_sai}
bwa samse $ref ${out_dir}/${sample_sai} ${sample} > $out_dir/${sample_sam}

samtools view -bS $out_dir/${sample_sam} | samtools sort -o $out_dir/${sample_sam%.sam}.sorted.bam
samtools index $out_dir/${sample_sam%.sam}.sorted.bam
samtools flagstat $out_dir/${sample_sam%.sam}.sorted.bam > $stats_dir/${sample_sam%.sam}.flagstat.txt
```

call snps

```
#!/bin/bash -e

#SBATCH --job-name=2024_02_21.mpileup.sl
#SBATCH --account=uoa02613
#SBATCH --time=00-72:00:00
#SBATCH --mem=6GB
#SBATCH --output=%x_%j.errout
#SBATCH --mail-user=katarina.stuart@auckland.ac.nz
#SBATCH --mail-type=ALL
#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=2
#SBATCH --profile task

module purge
module load BCFtools/1.16-GCC-11.3.0

ref=/nesi/nobackup/uoa02613/kstuart_projects/At4_MynaStarling/data/resources/genomes/Svulgaris_vAU_1.0.fasta
aligndir=/nesi/nobackup/uoa02613/kstuart_projects/Sv10_NZstarlings/data/processing_rawdata/aligned
outdir1=/nesi/nobackup/uoa02613/kstuart_projects/Sv10_NZstarlings/data/processing_rawdata/BCFtools

mpileup=mpileup_annotate.bcf
bcfcall=variant_calls_annotate.vcf.gz

cd ${aligndir}
echo "Create txt file with all sorted.bam files in folder"

ls *.sorted.bam > ALL_file_name.txt

echo "BCFtools mpileup"
bcftools mpileup -b ALL_file_name.txt -a "DP,AD,SP" --ignore-RG -Ou -f $ref -o ${outdir1}/${mpileup}

echo "BCFtools call"
bcftools call ${outdir1}/${mpileup} -mv -Oz -o ${outdir1}/${bcfcall} -f GQ
```