localization of exon/intron boundaries in predicted

canine mRNA sequences for the candidate genes (ESPN,

XM\_546751; MYO3A, XM\_544234; SLC26A5, XM\_540393;

USH1C, XM\_860072) was performed using the mRNA-to-

genomic alignment program

SPIDEY

(http://www.ncbi.nlm.nih.

gov/IEB/Research/Ostell/Spidey/index.html).

localization of exon/intron boundaries in predicted

canine mRNA sequences for the candidate genes (ESPN,

XM\_546751; MYO3A, XM\_544234; SLC26A5, XM\_540393;

USH1C, XM\_860072) was performed using the mRNA-to-

genomic alignment program

SPIDEY

(http://www.ncbi.nlm.nih.

gov/IEB/Research/Ostell/Spidey/index.html).

The localization of exon/intron boundaries in predicted

canine mRNA sequences for the candidate genes (ESPN,

XM\_546751; MYO3A, XM\_544234; SLC26A5, XM\_540393;

USH1C, XM\_860072) was performed using the mRNA-to-

genomic alignment program

SPIDEY

(http://www.ncbi.nlm.nih.

gov/IEB/Research/Ostell/Spidey/index.html).

The localization of exon/intron boundaries in predicted canine mRNA sequences for the candidate genes was performed using the mRNA-to-genomic alignment program SPIDEY (<http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html>).

**— To eliminate the character '\_' from VCF file sample IDs, use the --const-fid 0 command (error message: Error: Multiple instances of '\_' in sample ID)**

plink --vcf yourfile.vcf --horse --allow-extra-chr --const-fid 0 --recode --out yourfile2

**— To extract sample IDs from a VCF file:**

module load bcftools

bcftools query -l file.vcf

**— To keep a subset of individuals in your PLINK file:**

module load plink

plink --file data --keep mylist.txt --recode --allow-extra-chr --horse --out outfile

**OR:**

/home/mccuem/shared/bin/vcftools --vcf New2M.vcf --keep keep.txt --recode --out arabtb.vcf

**— To extract fields from an annotated VCF using SnpSift:**

java -jar SnpSift.jar extractFields example.vcf "CHROM" "POS" "REF" "ALT" "AF" "ANN[\*].ALLELE" "ANN[\*].EFFECT" "ANN[\*].IMPACT" "ANN[\*].GENE" "ANN[\*].BIOTYPE"

**-- To extract a genomic region from a FASTA file (in this case, I used this command to extract ECA6 from EquCab3):**

module load samtools

samtools faidx build\_3\_i.ordered.fa

samtools faidx build\_3\_i.ordered.fa chr6:0-85000000 > [ECA6.fasta](https://eca6.fasta/)

**-- To obtain chromosome sizes from a FASTA file:**

cat build\_3\_i.ordered.fa | awk '$0 ~ ">" {print c; c=0;printf substr($0,2,100) "\t"; } $0 !~ ">" {c+=length($0);} END { print c; }'

**-- Blasting nucleotide sequence to EquCab3 in MSI**  
  
ssh mesabi (login in using internet password)  
  
module load ncbi\_blast  
  
vim ahr\_human.fasta (vim generates a new file)  
i — insert (allows you to insert text)  
>exon 1 (fasta format, copy and paste exon 1 sequence from other species)  
:wp — end and save  
Esc — press escape button to escape  
  
blastn -db ../beeso18/EquCab3 -query ahr\_human.fasta -out AHR\_gene (output is similar to blast website and will give location and chromosome)  
  
head -n 50 AHR\_gene (shows first 50 lines of file AHR\_gene)  
  
less AHR\_gene

**-- Removing values lower than 5 from hapQTL text output:**

awk '{if ($4 >= 5) print $1,$2,$3,$4}' infile > outfile