Heterozygosity Experiment

This manuscript was automatically generated on June 15, 2021.

## Authors

# Heterozygosity Experiment

|  |  |
| --- | --- |
| Field | Value |
| Project | Plague Denmark |
| Date | 2021-06-14 |

## Objectives

**1. Why do Denmark samples have high counts of heterozygosity?**

Conclusions: - Sites flagged as heterozygous by snippy core primarily have a low genotype quality. - Low quality can occur because the Alternate Allele has low counts. - One explanation is DNA damage (ex. deamination of cytosines).

**2. How does this compare to other Second Pandemic samples?**

Conclusions: - All Danish samples have more homozygous sites than heterozygous sites - The number of heterozygous sites in Danish samples is equal to or less than other Second Pandemic samples.

## Methods

### Variant Calling (Pairwise)

snippy \  
 --prefix SAMPLE \  
 --reference GCA\_000009065.1\_ASM906v1\_genomic.fna \  
 --outdir SAMPLE \  
 --bam SAMPLE.bam \  
 --mapqual 30 \  
 --mincov 3 \  
 --minfrac 0.9 \  
 --basequal 20 \  
 --force \  
 --cpus 10 \  
 --report 2> SAMPLE.log; \

* Multiqc was run on the output directories of Snippy for all samples.

### Plot Site Distributions

head -n `awk 'END{print NR - 1}' results/snippy\_multi/all/snippy-multi.txt` results/snippy\_multi/all/snippy-multi.txt | tail -n+2 | cut -f 1 | while read sample;  
do  
 in\_vcf=`ls results/snippy\_pairwise/\*/$sample/${sample}.raw.vcf`;  
 homo=${in\_vcf%%.\*}.homo.txt;  
 het=${in\_vcf%%.\*}.het.txt;  
 echo $sample;  
 bcftools query -i 'TYPE="snp" & GT="1/1" & QUAL>=100' -f '%DP\n' $in\_vcf | sort -h > $homo;  
 bcftools query -i 'TYPE="snp" & GT="0/1" & QUAL>=100' -f '%DP\n' $in\_vcf | sort -h > $het;  
 /home/poinarlab/Projects/Plague/Denmark/scripts/plot\_homo\_het.py \  
 --homo $homo \  
 --het $het;  
done  
  
mkdir results/heterozygosity  
mv results/snippy\_pairwise/{local,sra}/\*/\*.png results/heterozygosity/