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

```
        sample;

do
  in_vcf=`ls results/snippy_pairwise/*/$sample/${sample}.raw.vcf`;
  homo=${in_vcf%%.*}.homo.txt;
  het=${in_vcf%%.*}.het.txt;
  echo $sample;


        > $homo;



        > $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/
```