H3ABioNet
Pan African Bioinformatics Network for H3Africa

wellcome connecting science

# Next Generation Sequencing Bioinformatics Course 2021

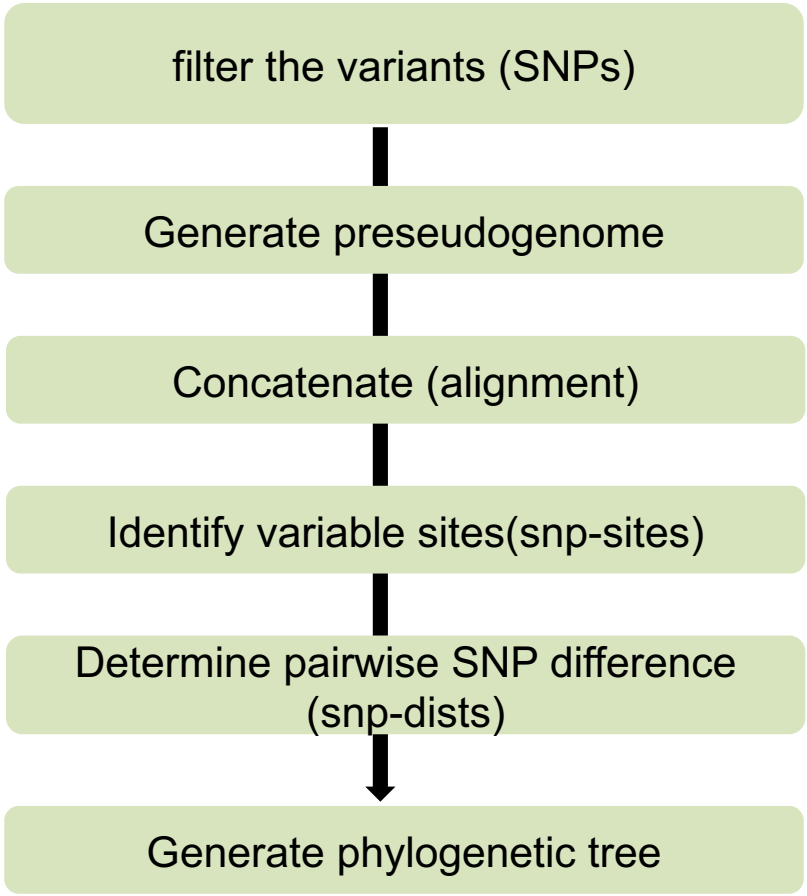
# Module 6 – Pathogen variant calling

# Session 3: Inferring genetic relatedness



CC BY NC SA

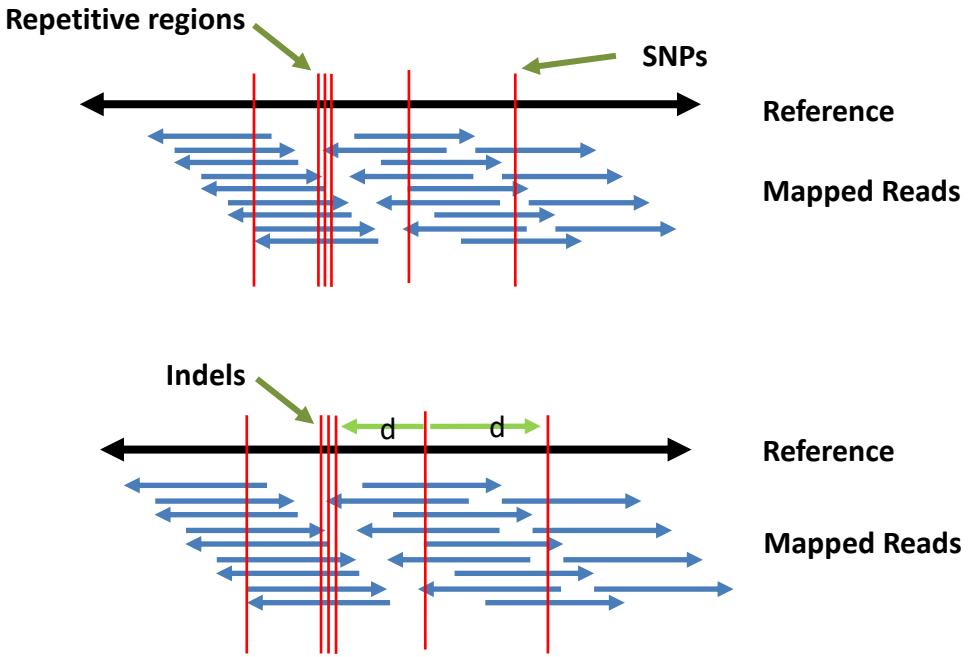
# Step3: Determining genetic relatedness



# Filtering SNPs in repetitive regions

```
bcftools filter -T MtbRepetitiveElemetsDrgenes.bed -i 'type="snp" && QUAL>=50 && FORMAT/DP>5 && MQ>=30
&& DP4[2]/(DP4[2]+DP4[0])>=0.80 && DP4[3]/(DP4[3]+DP4[1])>=0.80' -g10 -G10 variants.vcf -o repfiltered.vcf
```

- -repetitive regions
- -distance from indels (-G)
- -distance from other nearby SNPs (-g)






CC BY NC SA

# Generating pseudogenome

bcftools consensus -f reference.fa repfiltered.vcf.gz >consensus.fa

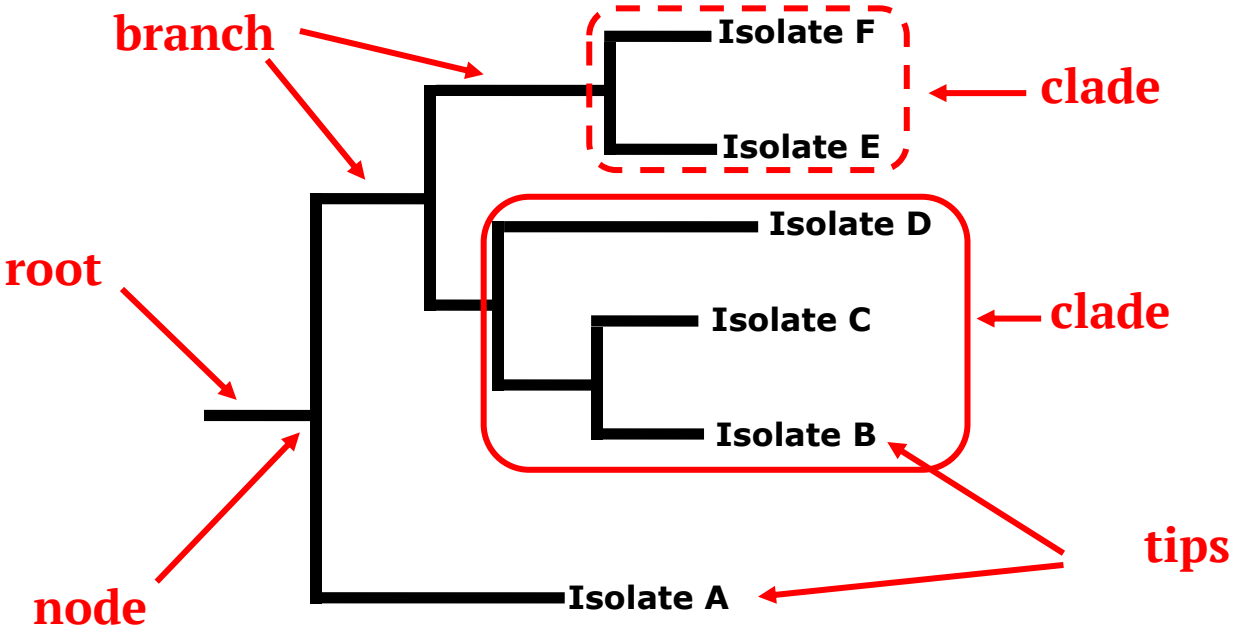
| Position | Ref | Alt |
|---|---|---|
| 4 | A | T |
| 11 | C | A |
| 18 | G | T |
| 26 | G | A |
| 36 | C | T |

# Pairwise SNP difference

snp-dists alignment.mfa >matrix.tsv

| Ids | Isolate A | Isolate B | Isolate C | Isolate D | Isolate E | Isolate F |
|---|---|---|---|---|---|---|
| Isolate A | 0 | 123 | 125 | 110 | 109 | 200 |
| Isolate B | 123 | 0 | 12 | 30 | 50 | 60 |
| Isolate C | 125 | 12 | 0 | 33 | 45 | 40 |
| Isolate D | 110 | 30 | 33 | 0 | 100 | 95 |
| Isolate E | 109 | 50 | 45 | 100 | 0 | 20 |
| Isolate F | 200 | 60 | 40 | 95 | 20 | 0 |

# Phylogenetic tree




Useful resource: https://evolution.berkeley.edu/evolibrary/article/0_0_0/evotrees_intro

# Summary: session3

- Identify high quality variants
- Generate pseudogenomes (with variants)
- Create alignment and extract variant sites
- Calculate pairwise SNP difference
- Generate phylogenetic tree
- Visualization and understanding


CC BY NC SA

# Thank you



CC BY NC SA