HIV-phyloTSI: IAVI workshop

Andrea Brizzi

2023-11-12

Intrahost diversity and recency of infections
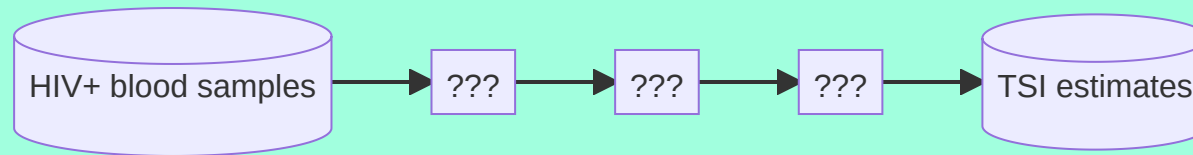
Recency of infection

- *What?*
  Determining whether a person testing HIV positive has been infected recently or not.

- *Why ?*
  - incidence estimation
  - study delays to diagnosis in subpopulations.

Code and slides

HIV-PhyloTSI bioinformatics pipeline

Demistifying the process

HIV+ blood samples → ??? → ??? → ??? → TSI estimates

- In this section, want to **describe** processes and **data manipulation** carried out by the pipeline.

- **Not** the actual **implementation** of the pipeline.

Main steps:

- Next Generation Sequencing (NGS)

- constructing phylogenies **MANY!**

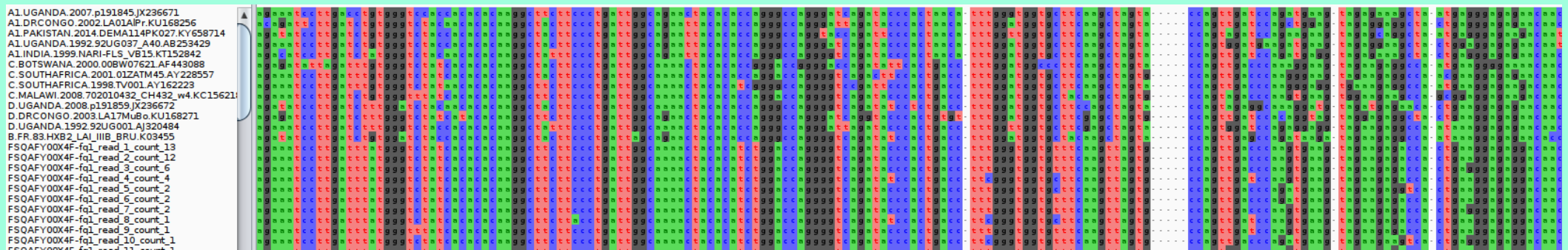- analyse phylogenies

Code and slides

Next Generation Sequencing

- simultaneous sequencing of multiple viral particles to capture within-host diversity

- PANGEA protocols: Gall (TOCITE) and Bonsall et al. (Bonsall et al. 2018)

- TSI pipeline starts with the outputs produced by `shiver` (Wymant et al. 2018): `*.bam`, `*.ref.fasta`
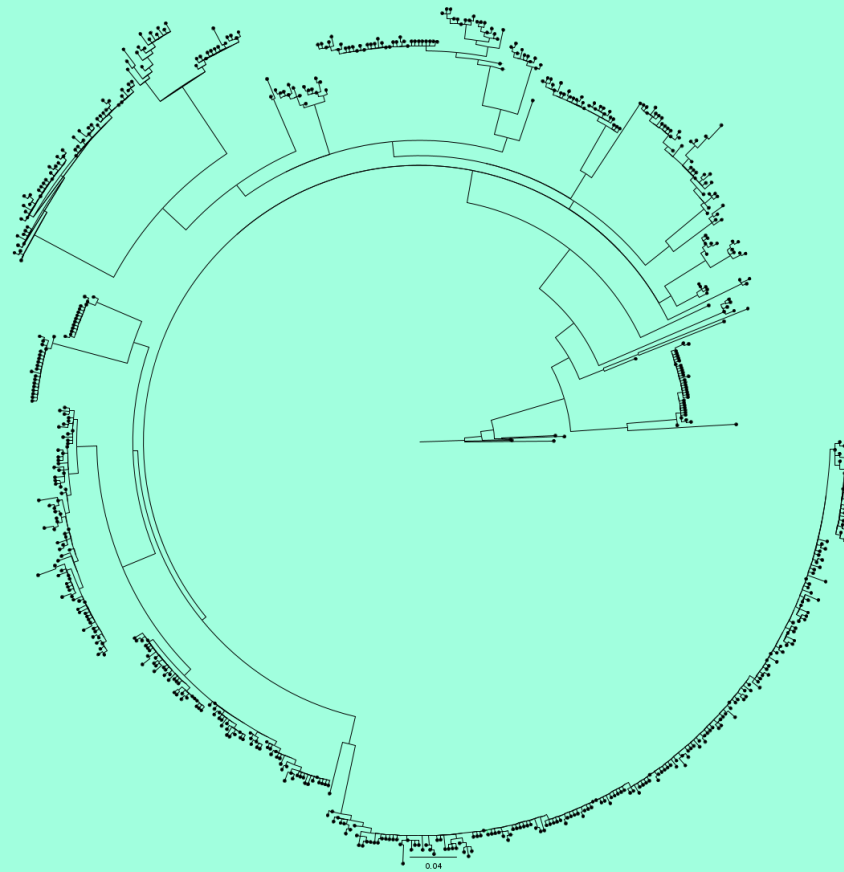

Code and slides
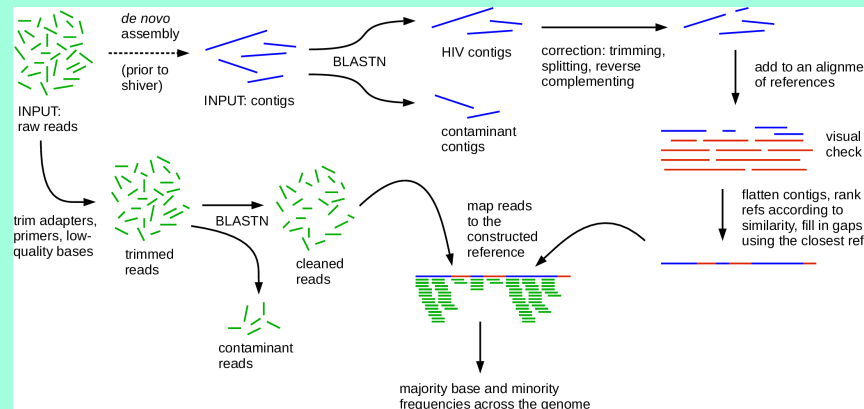
Grouping & multiple sequence alignment

- Many sequences, many hosts, many genome windows -> **group by host!**

- aligning to compare "apples to apples": MAFFT TOCITE

Inferring phylogenies

- Phylogenies are made through IQTREE by group and window through IQTREE (TOCITE)
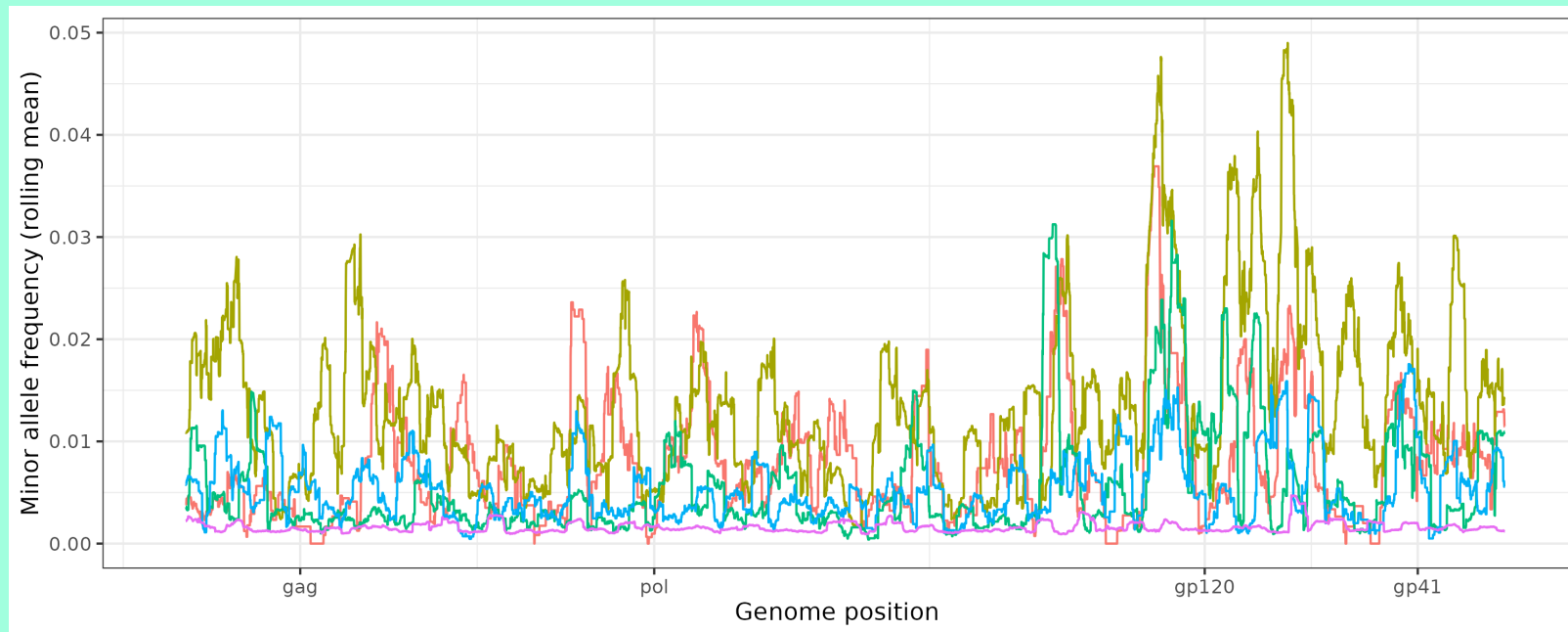

Code and slides

Analysing phylogenies

- `phyloscanner` (TOCITE) summarizes each tree through summary statistics:

- `patStats.csv`: contains LRTT, number of tips, etc…
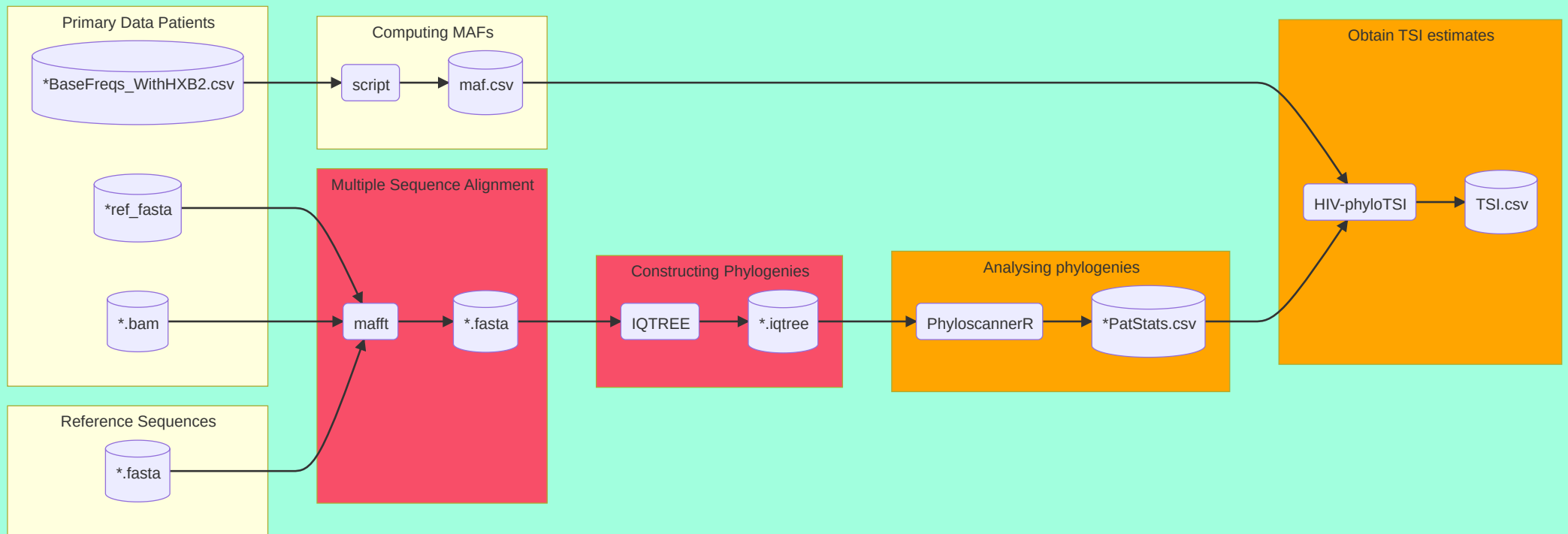
Minor Allele Frequencies

- $\text{MAF} = (1 - \text{proportion of majority bases})/\text{dept}$
- Evaluated at first 2 codos positions (MAF12c) and/or third codon position (MAF3c)

HIV-phyloTSI

- Focus of the practical session

- *ML algorithm* reads features of phylogenies and MAFs to estimate TSIs.

Code and slides

Summary

But what is HIV-phyloTSI doing?

HIV-PhyloTSI obtaining TSI estimates

Test

which color is this

Possible follow-up analyses

And this?

References

Bonsall, David, Tanya Golubchik, Mariateresa De Cesare, Mohammed Limbada, Barry Kosloff, George MacIntyre-Cockett, Matthew Hall, et al. 2018. "A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in a Global Health Setting." Preprint. Genomics. https://doi.org/10.1101/397083.

Wymant, Chris, François Blanquart, Tanya Golubchik, Astrid Gall, Margreet Bakker, Daniela Bezemer, Nicholas J Croucher, et al. 2018. "Easy and Accurate Reconstruction of Whole HIV Genomes from Short-Read Sequence Data with Shiver." *Virus Evolution* 4 (1): vey007. https://doi.org/10.1093/ve/vey007.

Code and slides

Code and slides