Rates and Patterns of Indels in HIV-1 gp120 within hosts

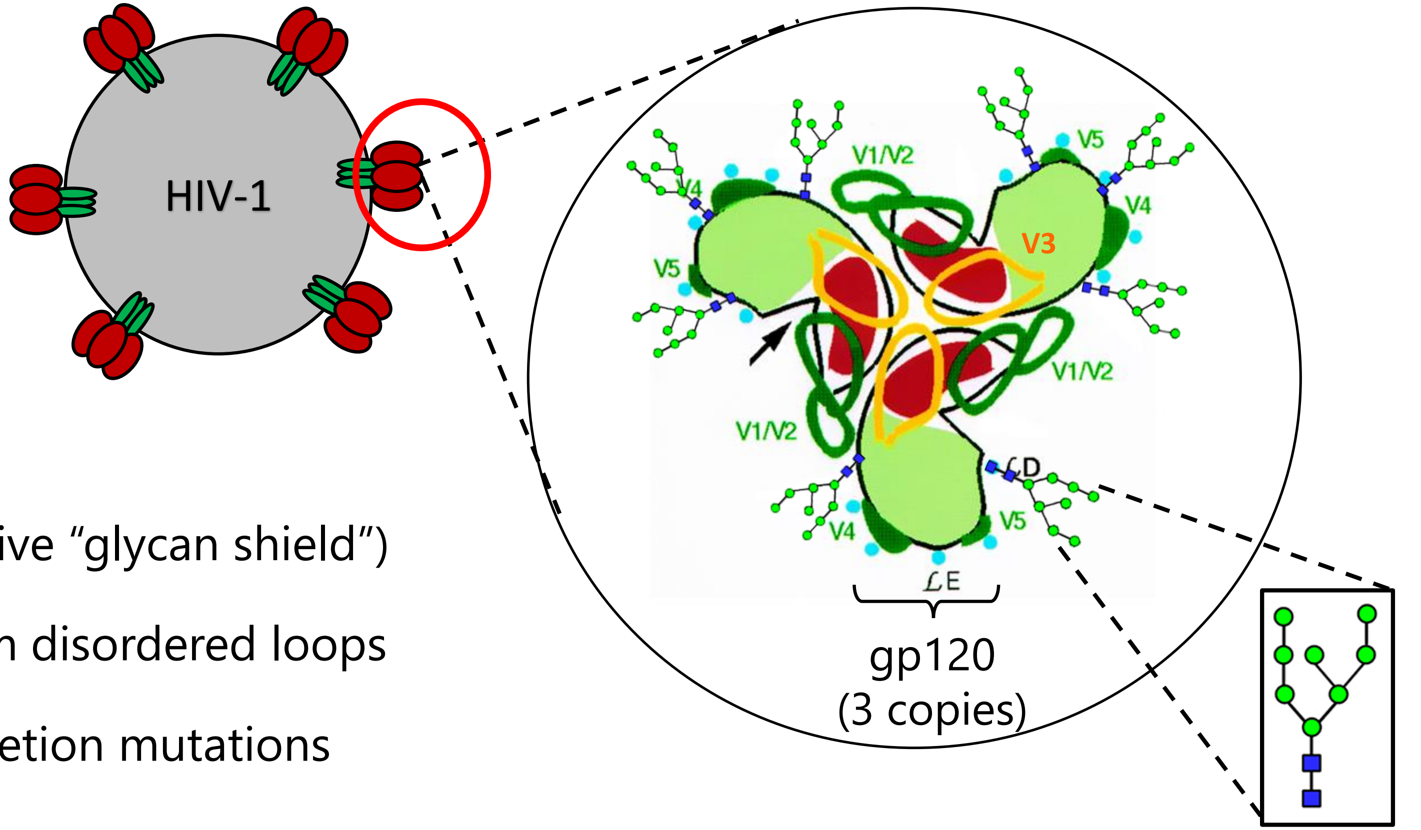
John Palmer[1], Art Poon[1]

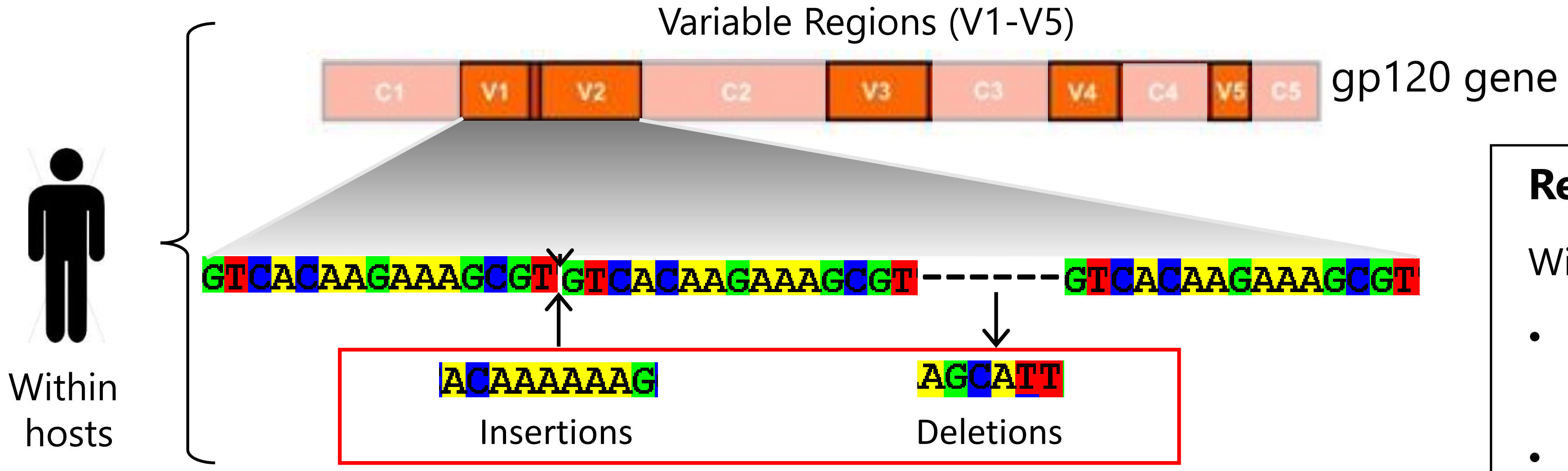
[1] Department of Pathology and Laboratory Medicine, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada

Introduction

gp120 glycoprotein

- Surface protein component of *env* gene
- Mediates HIV binding to CD4⁺ cells
- Abundant N-linked glycosylation (protective "glycan shield")
- Has five variable regions (V1-V5) that form disordered loops
  - Rapidly accumulate insertion and deletion mutations





Research Questions:

Within individual hosts…

- How quickly do indels accumulate in the gp120 variable regions?
- How frequently do they interact with N-linked glycans?

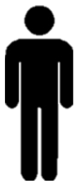
Methods

Dr. Vlad Novitsky (Harvard University)
25 patients sets
(n = 2,541)

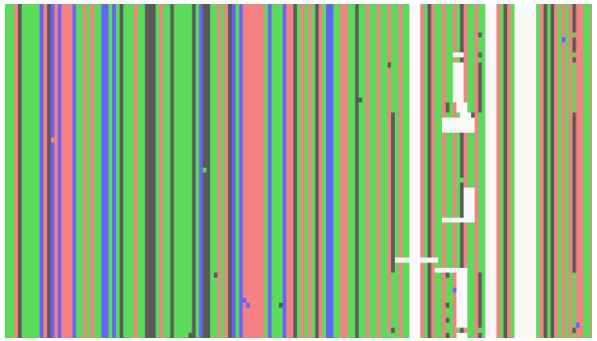
Los Alamos National Laboratory HIV Database
25 patient sets
(n = 11,265)

Filtering:
- Poor clock signal
- Insufficient time coverage

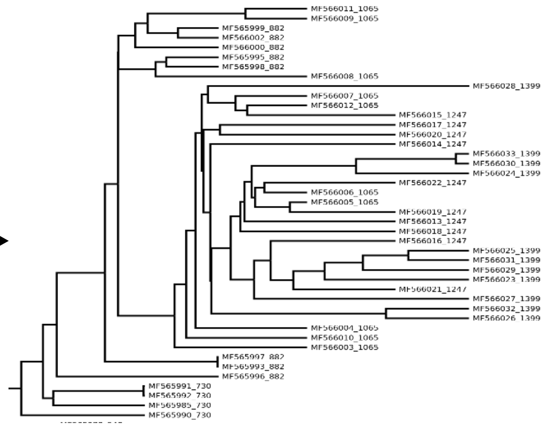
Final Data:
27 patients
(n = 5,577)



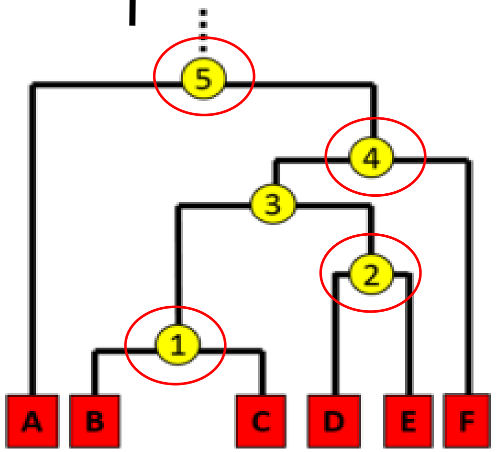
One per patient (x27)



Multiple sequence alignment (MAFFT)

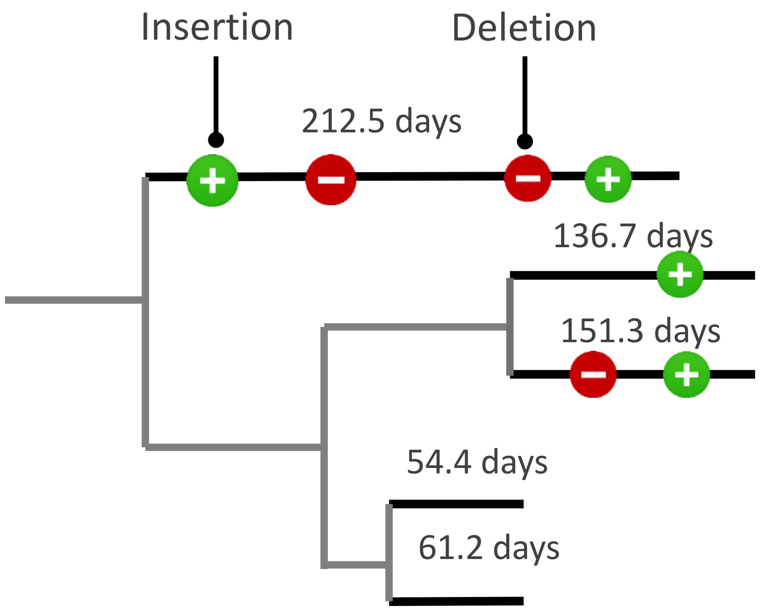


Within-host phylogenetic tree (BEAST)

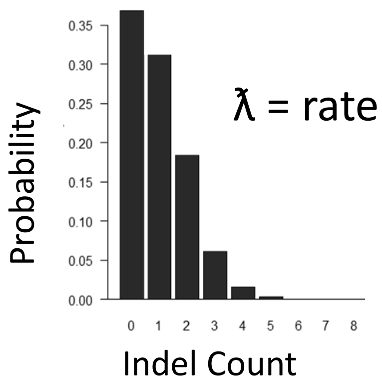


Reconstruct Ancestral Sequences (Historian)

Analysis



- Insertion/deletion rates (Poisson regression)



- Other analyses…

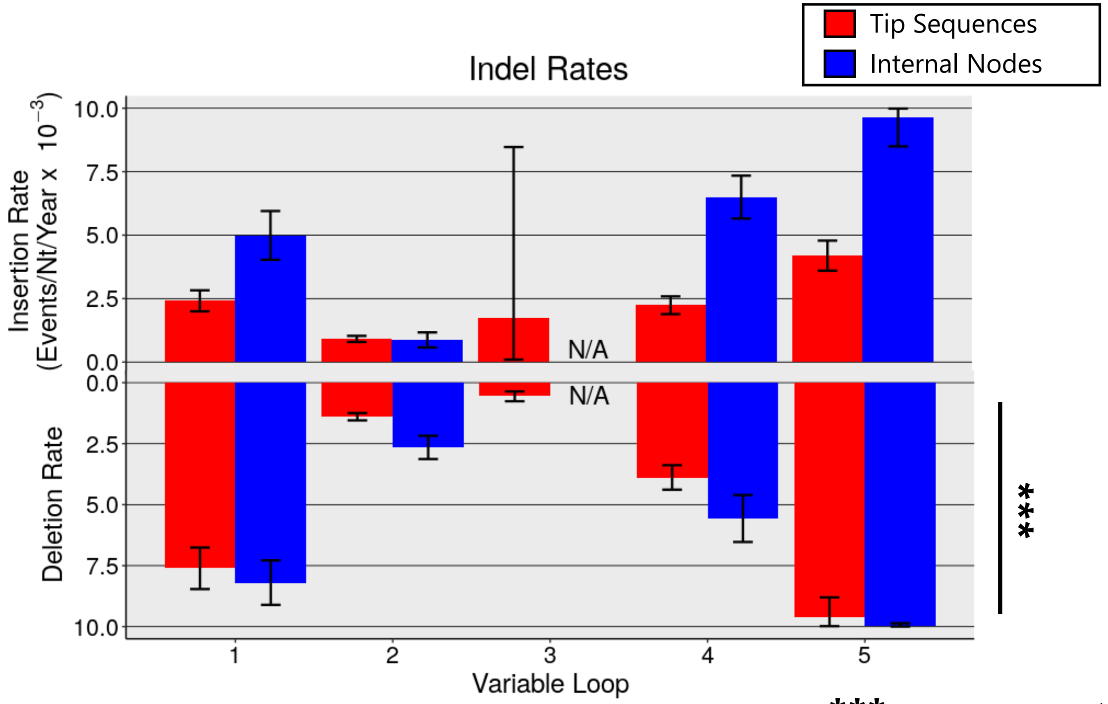
Results

Indel Rates

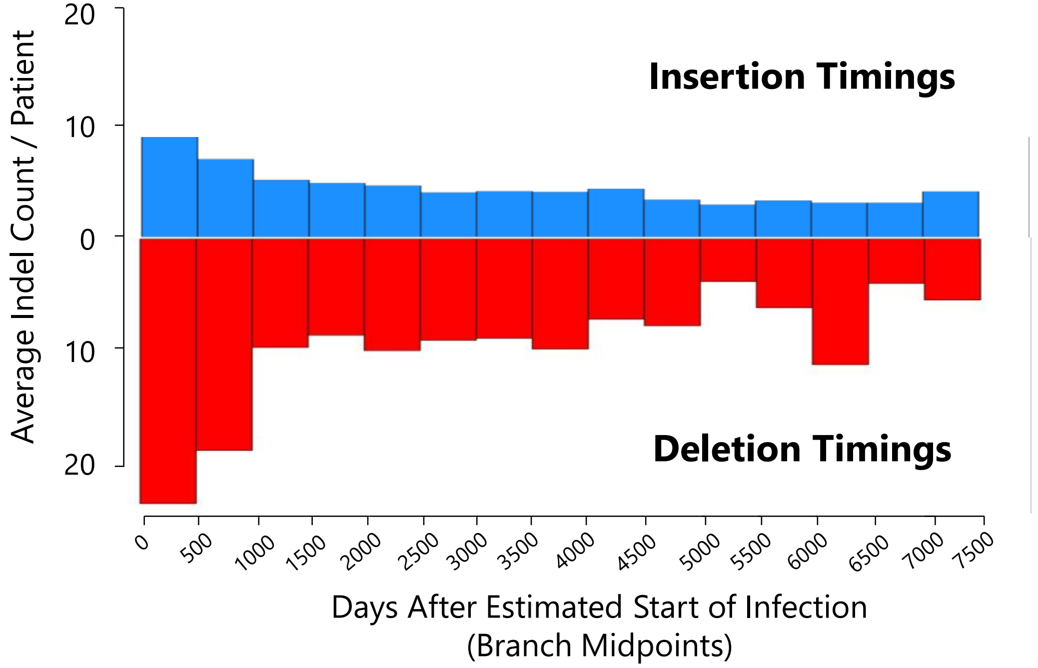
- Deletion rates were significantly higher than insertion rates across all v-loops ($p < 2.2 \times 10^{-16}$)
- N/A values are due to insufficient data to generate a Poisson regression estimate
- Error bars represent preliminary 95% confidence intervals
- Significantly higher than among-host indel rate estimates from previous study ($p = 8.5 \times 10^{-4}$)
[Palmer & Poon, Virus Evolution, 2019. doi://10.1093/ve/vez022]

Indel Timings

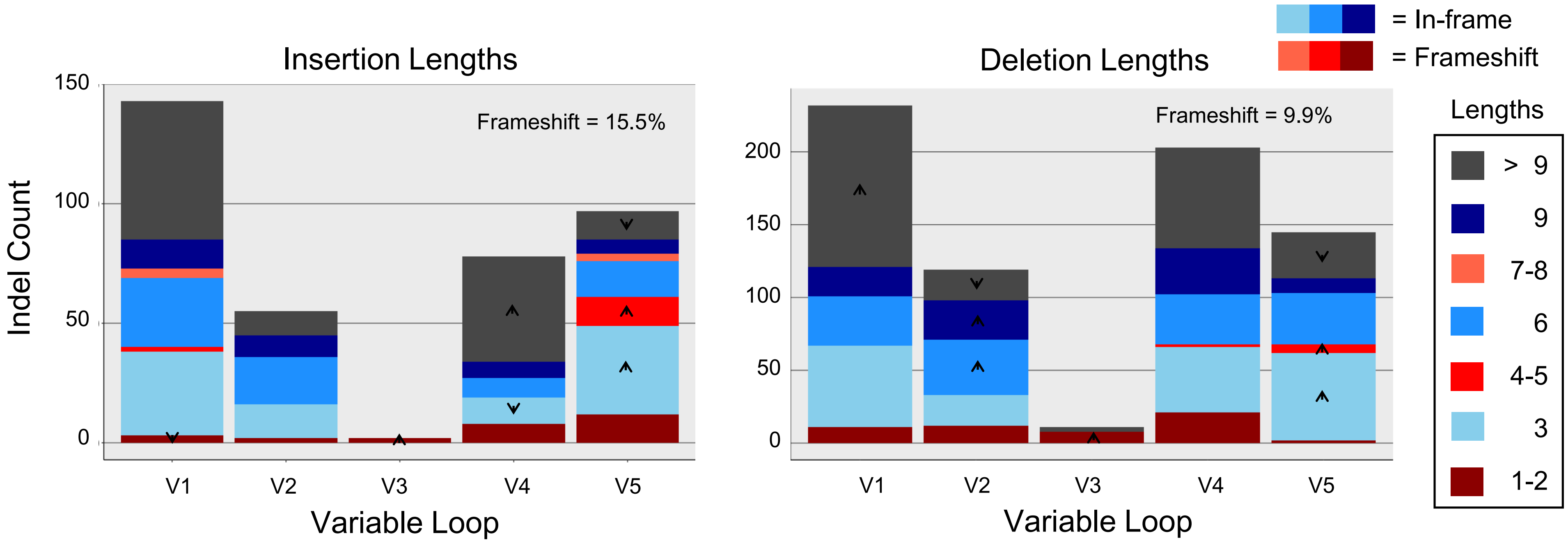
- Adjusted for the amount of available patient data at higher time intervals
- Suggest that insertions / deletions accumulate faster at the beginning of infection and gradually decline in rate
- As reflected in rates above, deletion counts appear higher than insertions



*** = $p < 2.2 \times 10^{-16}$



Results



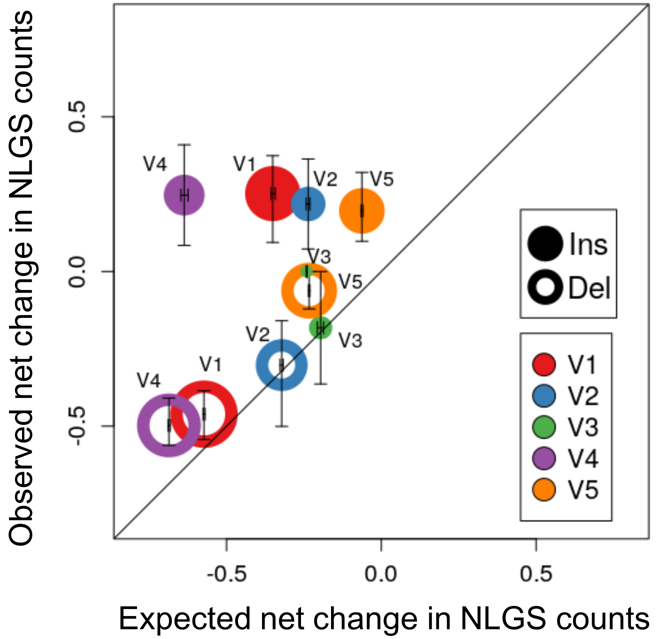
Indel Lengths

- Relatively few frameshift-inducing insertions (15.5%) and deletions (9.9%) demonstrates considerable purifying selection within hosts
- Arrows indicate significantly higher or lower counts based on Pearson chi-squared residuals

N-linked glycosylation site changes

- Expected changes were generated by placing every insertion/deletion back into 100 randomly-sampled locations within its v-loop sequence of origin and examining the average change to N-linked glycosylation site (NLGS) counts
- Error bars represent the 95% confidence intervals; point sizes indicate sample size
- Insertions in V1, V2, and V4 appear to induce higher net changes to NLGS counts than we would expect by chance (deviation from line)

N-linked glycosylation site (NLGS) changes



Thanks for reading!
Questions?

Email: jpalme56@uwo.ca

Conflict of Interest Disclosure

I have no conflicts of interest.