# Synapomorphic Patterns in HIV Gag Gene Sequences

Samantha Elsey

## Background

- The Human Immunodeficiency Virus (HIV) is a sexually transmitted disease that emerged in the 1920's
- **The gag gene codes for multiple structural proteins** of the HIV virus and accounts for over 50% of the virus's mass
- **Synapomorphies are shared characteristics from a common ancestor**

## Objective

- This study aims to identify **group defining synapomorphies** in the HIV gag gene to show relationship in strains over time using multiple sequence alignments and a variety of phylogenetic analysis to compose a **phylogenetic tree of the strains relationships**

## Methods

1) NCBI BLAST used to identify 15 comparable HIV gag gene DNA sequences (outlier group from human SIV gag gene analog)
2) Uploaded entire 16 taxa DNA sequences to an online Multiple Sequence Alignment tool (WEBPrank)
3) Trimmed beginning and end of taxa's DNA sequences to align
4) Uploaded aligned sequences to NG Phylogeny using traditional TNT (parsimony) tree building parameters (with over 3 million rearrangements run)
6) Ran 1000 bootstraps to support arrangement of tree
7) Analyzed tree arrangements and find synapomorphic sequences between distinguished groups of HIV gag gene

## Results

HIV Gag gene tree. Most parsimonious tree generated for 16 taxa using TNT programming.
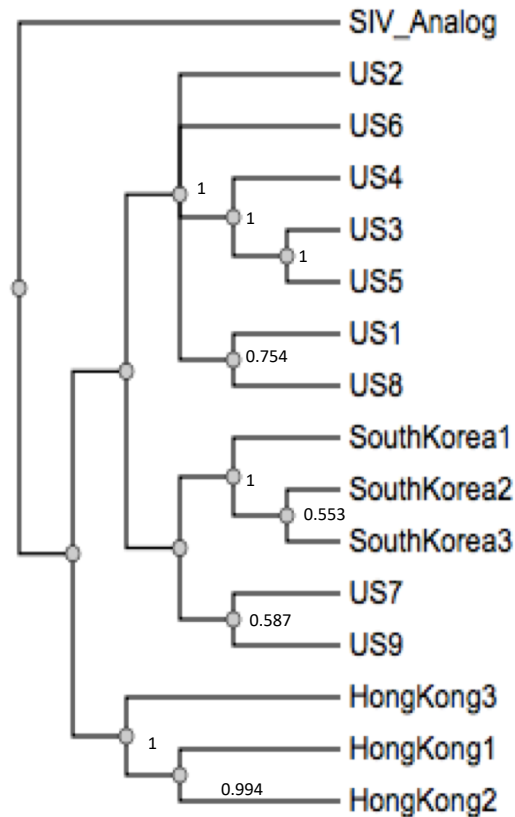


Unique insertions in genome found in each clade
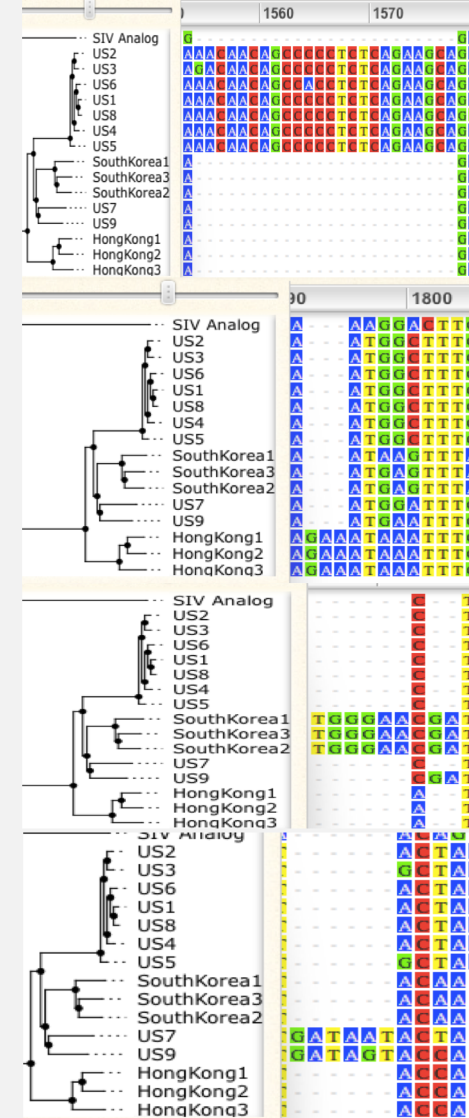


## Conclusion

- The grouping of Hong Kong 1, 2, & 3 have strong support values and come from a similar cluster of HIV-1 B groupings
- The grouping of South Korea 1, 2, & 3 have strong support values and come from similar HIV-1 type B clusters
- US 1, 2, 3, 4, 5, & 6 are also strongly grouped
- US 7 and 9 are less strongly grouped with other US variants, as they have a deactivated portion of their genome
- SIV Analog is the clear outgroup, as it is not an HIV-1 virus and has significantly less homology than the other taxa

## Future Study

- Do the insertions in each clade make virus more or less transmissible?
- Which insertions/ deletions/ base change cause changes in amino acids