

● BIOL4300

Phylogenetics of Human Herpesvirus 6B

McKay Jones
Kaden Crabb
Warren Averett
Lonnie Ernst

Department of Biology,
Utah Valley University
December 3rd, 2025

Introduction

- Human herpesvirus 6B (HHV-6B), an ubiquitous childhood virus, produces an incurable lifelong infection.
 - Adult seroprevalence: 95% (King et al., 2025)
 - Affects 90% of individuals within first 2 years of life (Zerr et al., 2005)
- A double-stranded DNA virus, HHV-6 is a member of the genus *Roseolovirus*, within the *Betaherpesvirinae* subfamily of the *Herpesviridae* family (Pellet et al., 2009).

Figure 1: Osmosis (2025)

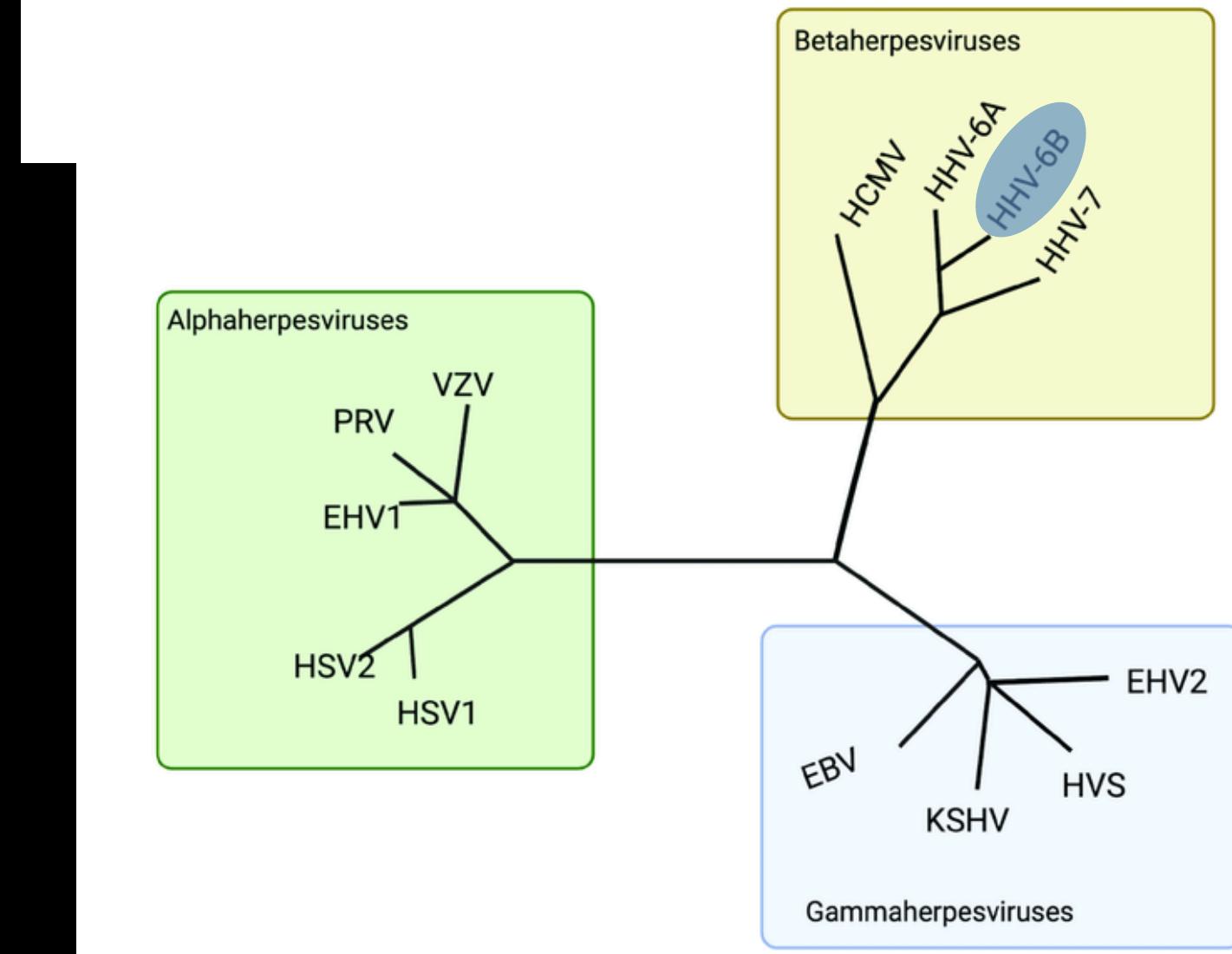
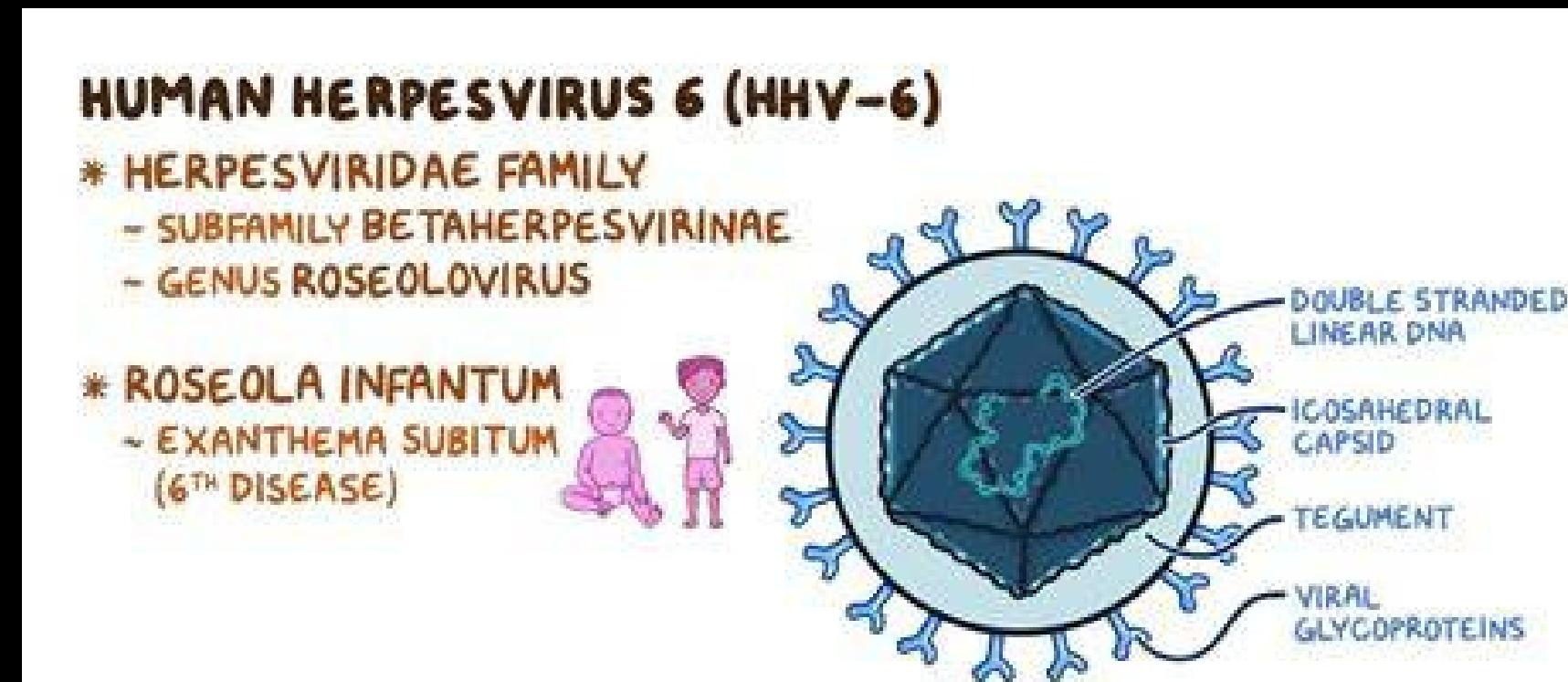


Figure 2: Alam (2021)

Why study HHV-6B?

- Like other human herpesviruses, HHV-6B establishes lifelong latency
(Dunn & Fogdell-Hahn, 2020; Leibovitch et al., 2018)
- Unlike other herpesviruses, HHV-6B can integrate into human chromosomes and can be passed on to offspring
(Greninger et al., 2018).
 - This is called iciHHV6 (inherited chromosomally integrated HHV6)
 - iciHHV6 occurs in 1-2% of the world population (158 million people!)
(Asward et al., 2020; Clark, 2016)

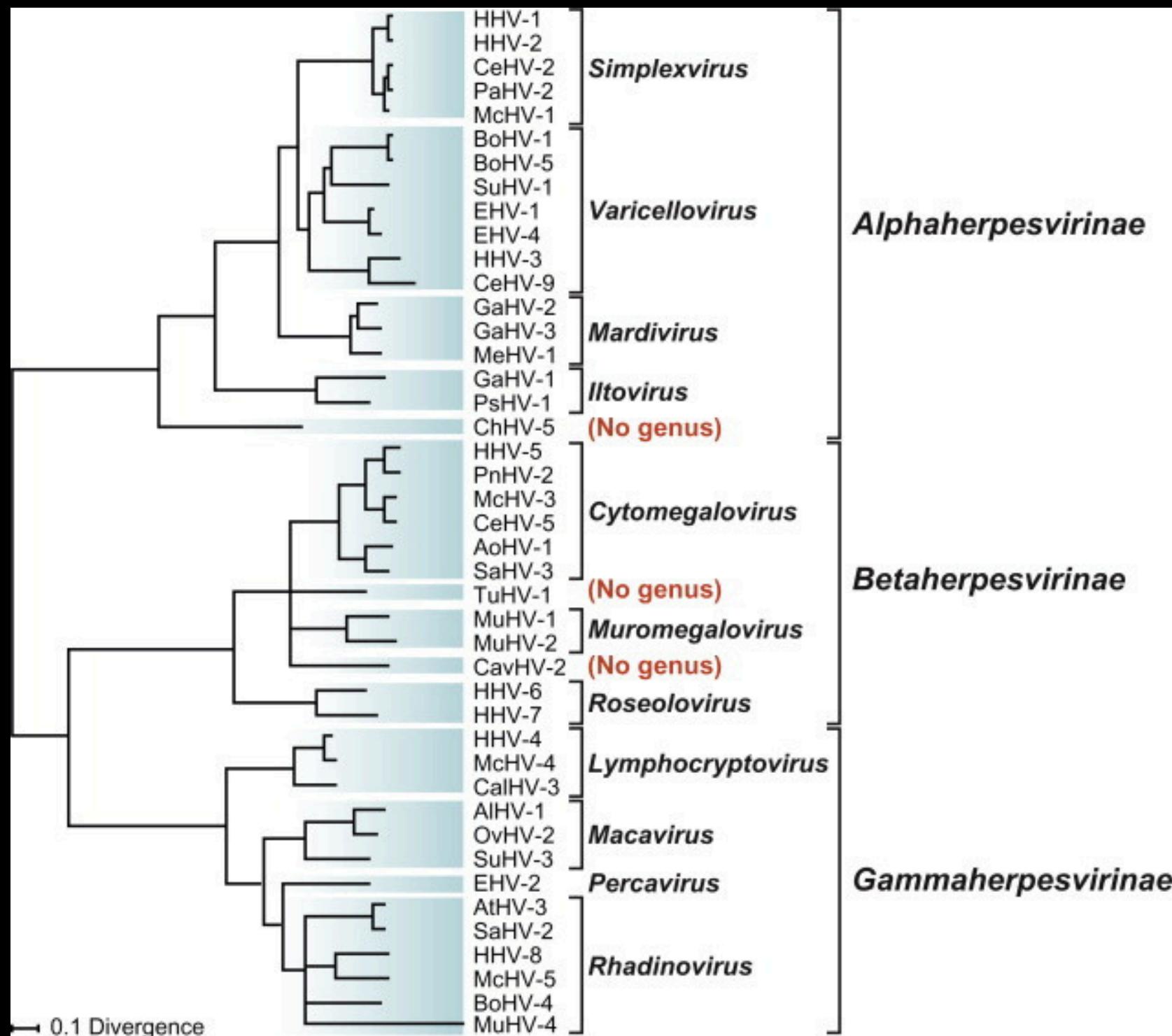


Figure 3: Phylogenetic relationships within the family *Herpesviridae*.
McGeoch, et al., (2008)

Clinical Manifestations

- Initially causes **roseola** (Mullins & Krishnamurthy, 2025)
- 5x increased risk of **Type 1 Diabetes** (Mistry et al., 2023)
- Linked with:
 - **Multiple sclerosis** (Cheng et al. 2012; Friedman et al. 1999)
 - **Celiac Disease** (Kemppainen et al. 2017)
 - **Hashimoto's thyroiditis** (Zatelli et al., 2012)
 - **Systemic lupus erythematosus** (Broccolo et al. 2009; Sokolovska et al. 2024)
 - **Epilepsy** (Bartolini et al., 2019)
 - Hematopoietic stem cell transplantations can result in **encephalitis** and **acute graft-versus-host disease** (Phan et al., 2018)
 - **Alzheimer's Disease** (Lazzarotto et al., 2014)

Phylogenetics+Research Question

- Previous gene annotations found limited sequence diversity of HHV-6B sequences within geographic clusters.

(Greninger et al., 2018)

- However, certain geographic regions had more divergence, with genomes from New York forming 2 distinct clusters.

(Greninger et al., 2018)

- Previous studies characterized gene U83 as the most divergent open reading frame between HHV-6A and HHV-6B

(Dominguez et al., 1999; Isegawa et al., 1999; Tweedy et al., 2015)

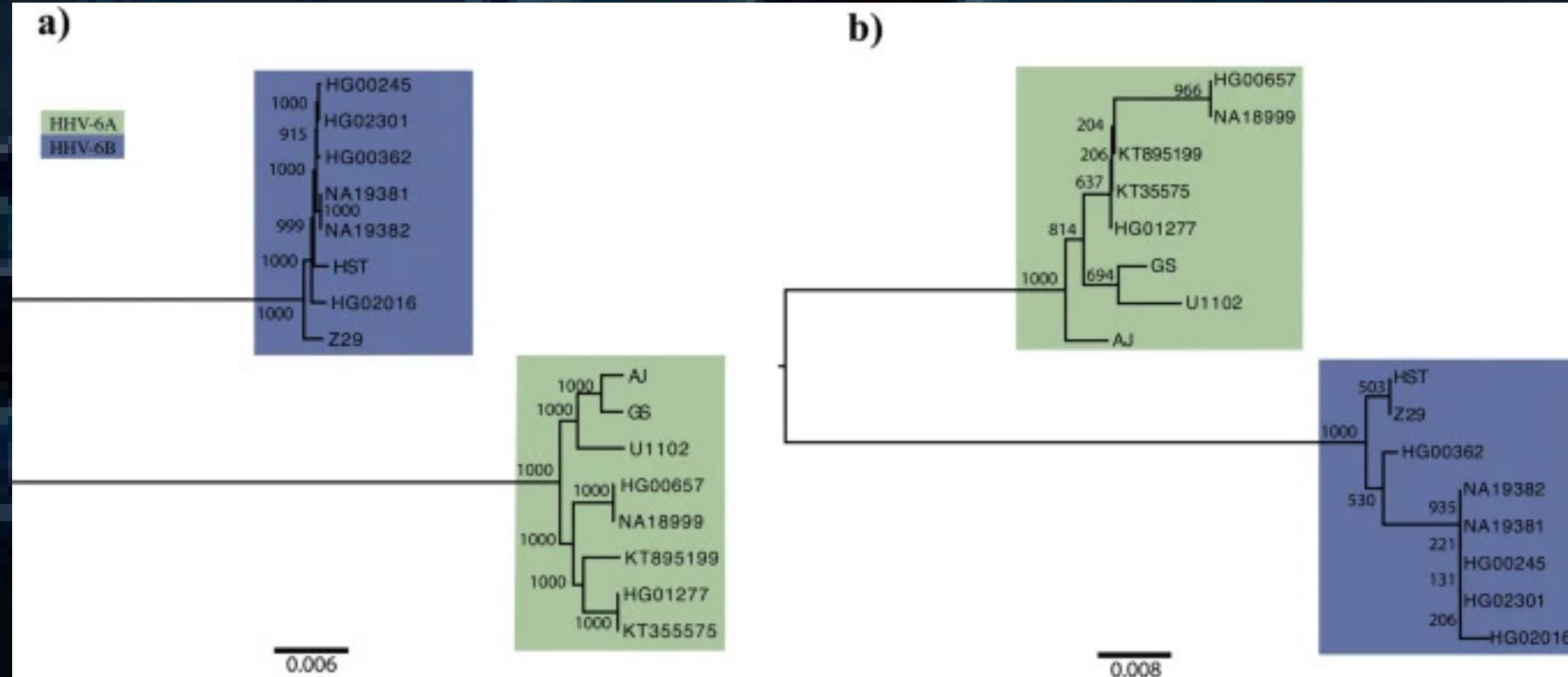


Figure 4: Clear separation between HHV-6A and –B in the U83 gene
(Telford et al., 2018)

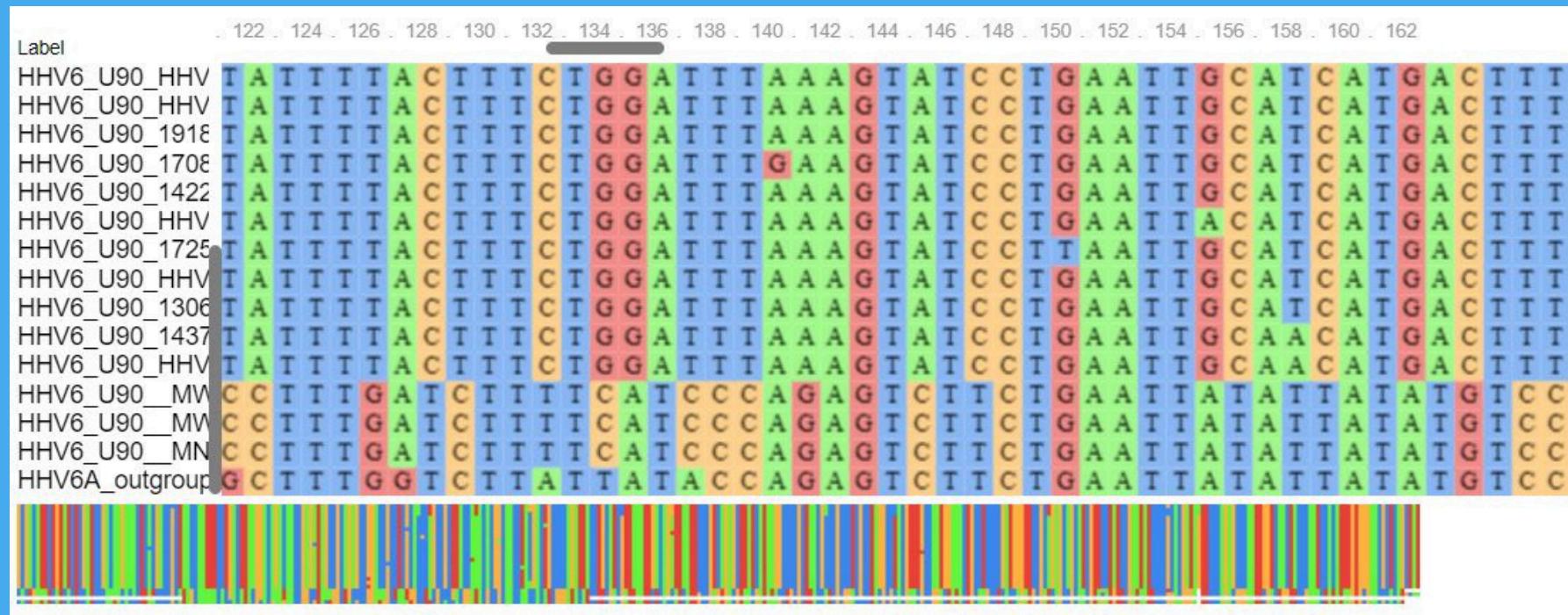
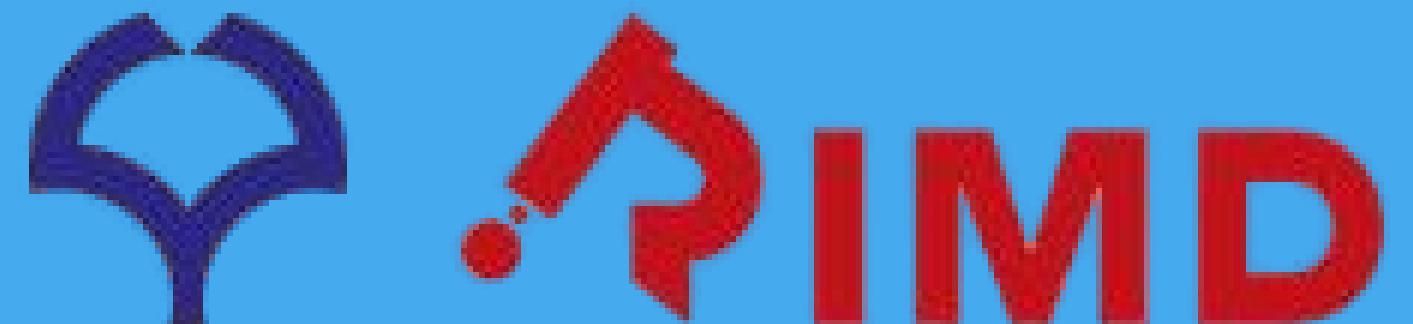
- Little is known about gene divergence within HHV-6B
- Q: Which genes in HHV-6B diverge the most, and why do they mutate?

Data Acquisition

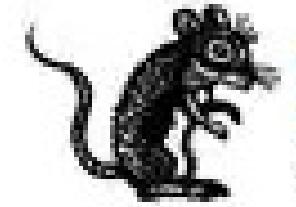
- We collected gene sequences in the NCBI GenBank database with the “rentrez” package in R.
- The search term “txid32604[ORGN]” AND {gene_name} [ALL] and cds[Title]” was used to fetch individual genes.
- Genes of interest:
 - U39, U47, U69, U83, U90
- The ‘genbankr’ package (of the Bioconductor ecosystem) was used to fetch complete genome sequences.
 - Python and the Bio.SeqIO interface was used to parse complete genome .gb files and to add individual gene sequences to individual gene FASTA files.



Alignment and Taxa Filtering



- We input individual gene FASTA files into the online version of MAFFT to generate alignments.
- The “ape,” “evobiR,” and “msaR” packages were utilized to read the aligned FASTA outputs from MAFFT, visualize a Multiple Sequence Alignment, and trim out unmatched regions (<80% gaps).



Beast2

Bayesian evolutionary analysis by sampling trees

Tree Inference

- Both IQ-TREE and BEAST v 2.7.8 were utilized to reconstruct trees.
- IQ-TREE found the HKY+F substitution model as the best substitution model for genes U47, U69, U83, and U90, and the TN+F+R3 model as the best for gene U39.
- The correct substitution models were ran in IQ-TREE with 1000 bootstraps, and in BEAST with a strict clock model and a coalescent constant population.
 - To achieve sufficient ESS (effective sample sizes) values, BEAST analyses were run for 10 million generations.



Tree Analyses

- FigTree and R (with the “ggtree”, “ape,” “viridis,” packages) were utilized to visualize and interpret output tree data from IQ-TREE and BEAST.
- The “rentrez,” “stringr” and “Biostrings” packages helped create a metadata .csv file for all of the filtered/output taxa.
 - the letterFrequency() function from “Biostrings” computed GC percentages for each taxa.
- All input/output data, scripts, and visuals can be found on the GitHub repository.



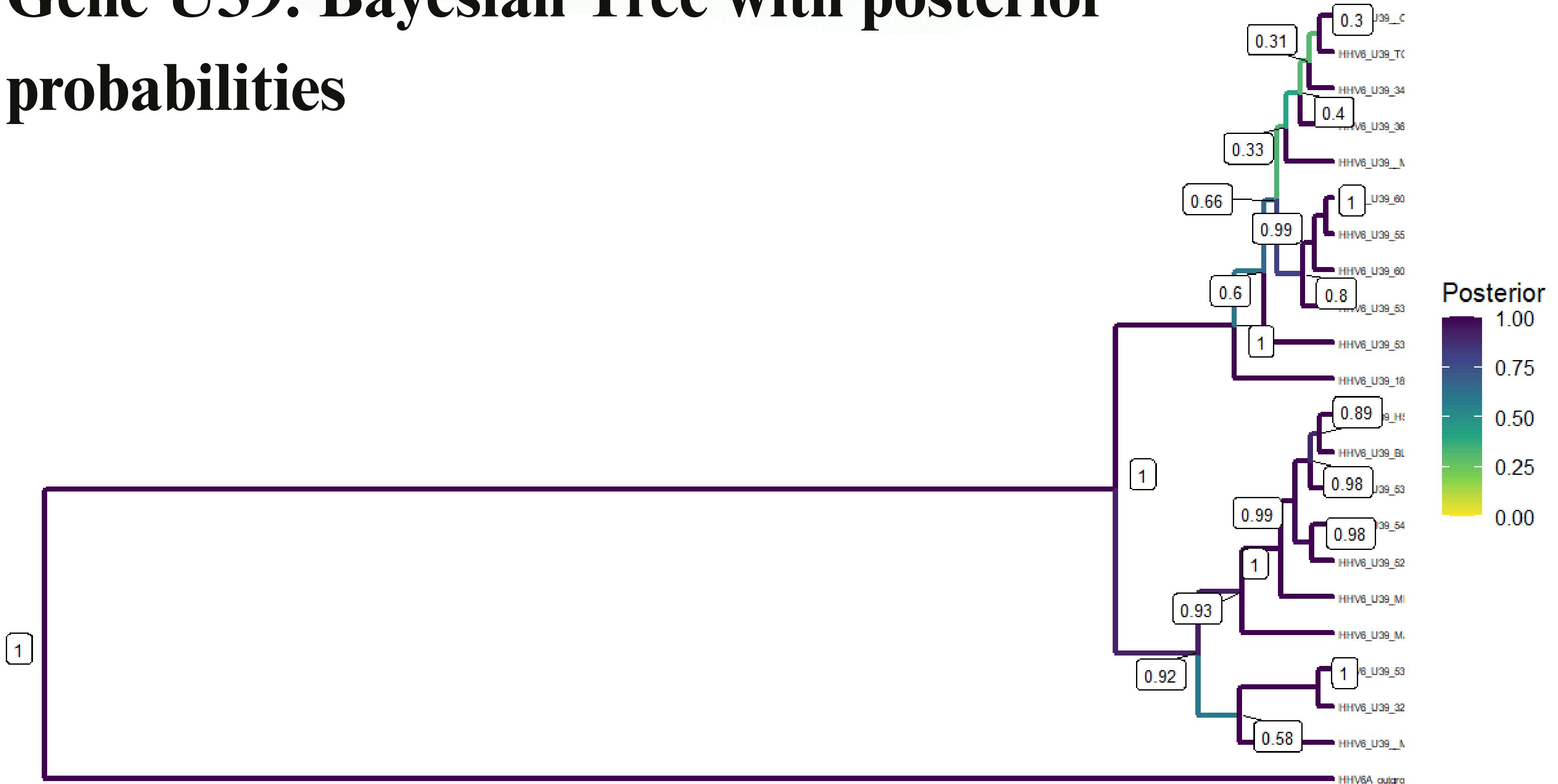
GitHub
(under construction)

A screenshot of a GitHub repository page. The repository is named "BIOL4300-project" and is described as "Public". The main branch is "main", there is 1 branch, and 0 tags. A search bar at the top right contains the placeholder "Go to file". Below the search bar, a list of commits is shown, starting with a commit from "mckayjones4" that updated "README.md". The commit message is "Update README.md". The date of the commit is "yesterday". Another commit from "mckayjones4" added raw tree files, with the commit message "add raw tree files" and the date "now". The repository has 149 commits in total, with the most recent being "08e645b · now". At the bottom of the page, the README file is displayed, titled "BIOL4300 Project: Phylogenetics of Human Herpesvirus 6B". The README text states: "Scripts and data necessary to complete gene-specific analysis of Human Herpesvirus 6B (HHV-6B)." A section titled "Folder organization" is also present at the bottom.

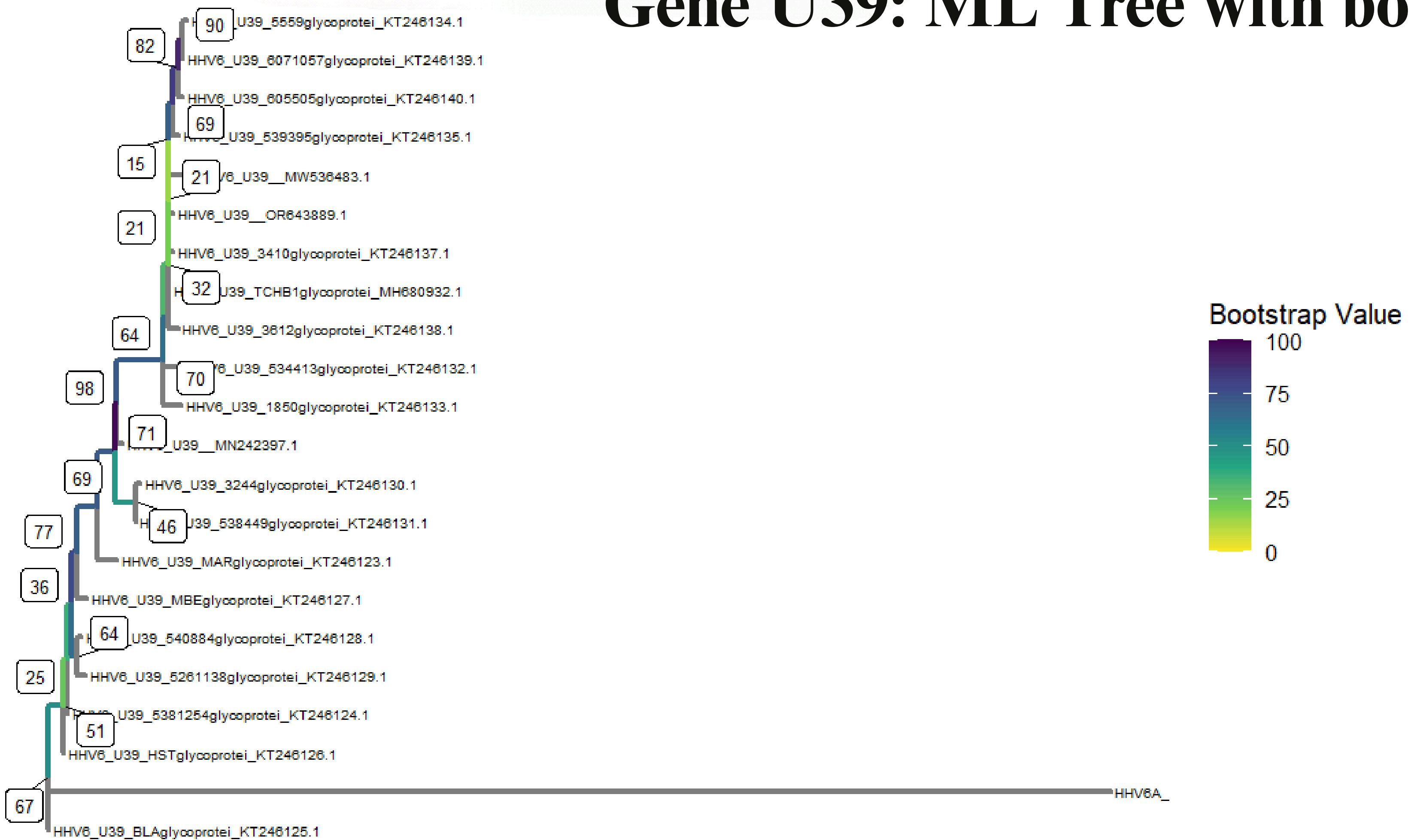
Results



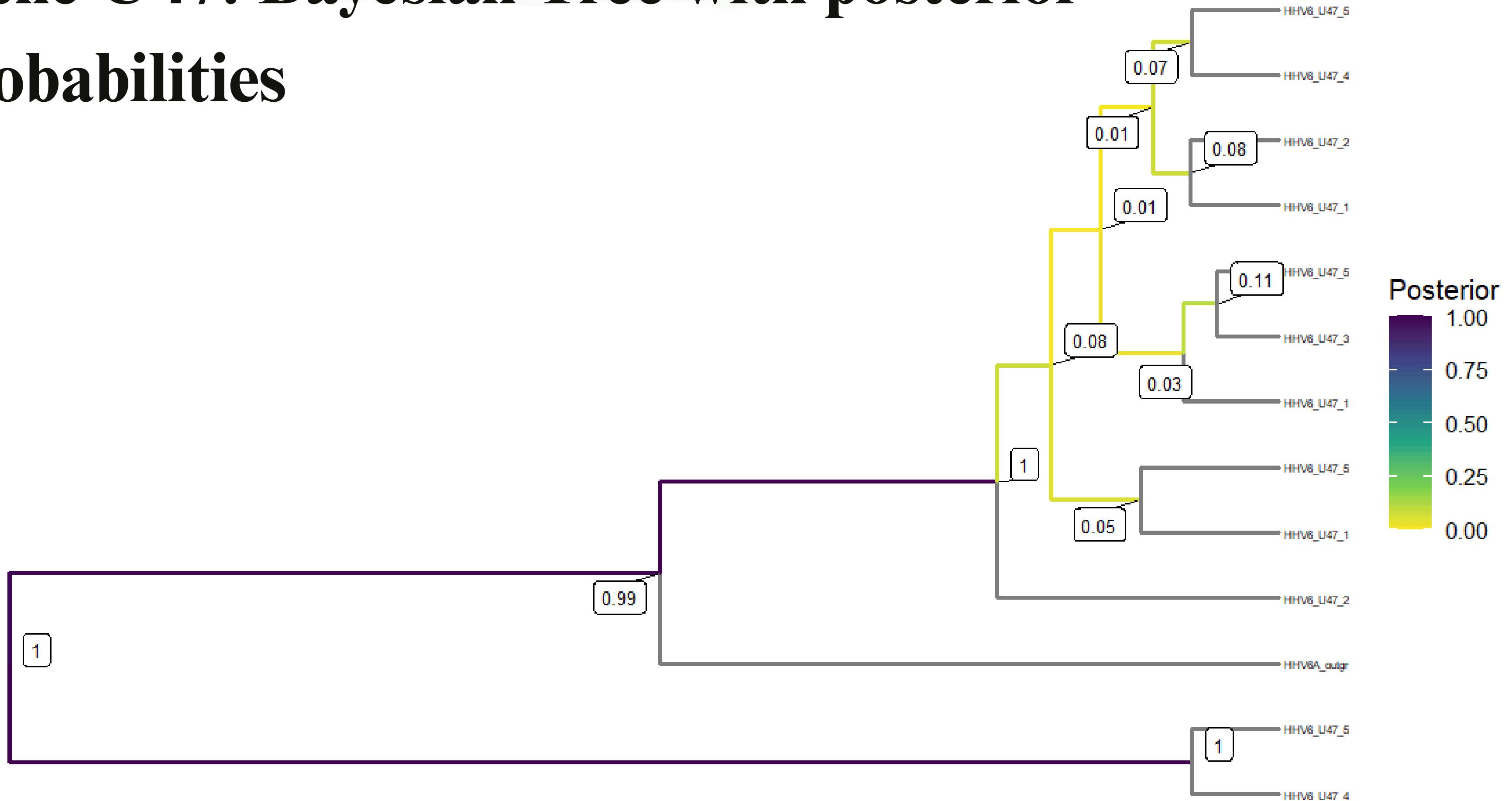
Gene U39: Bayesian Tree with posterior probabilities



Gene U39: ML Tree with bootstrap values



Gene U47: Bayesian Tree with posterior probabilities

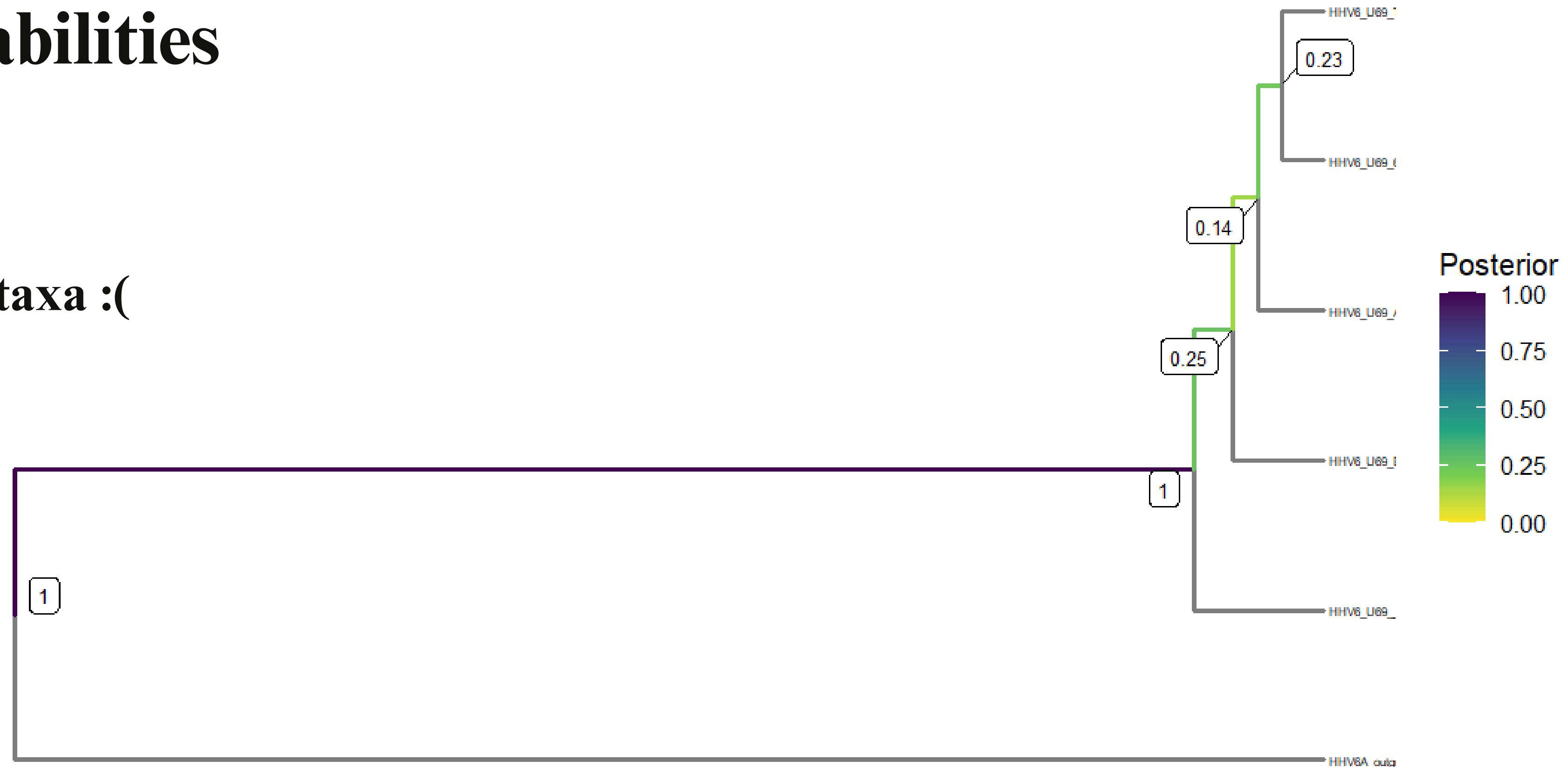


Gene U47: ML Tree with bootstrap values

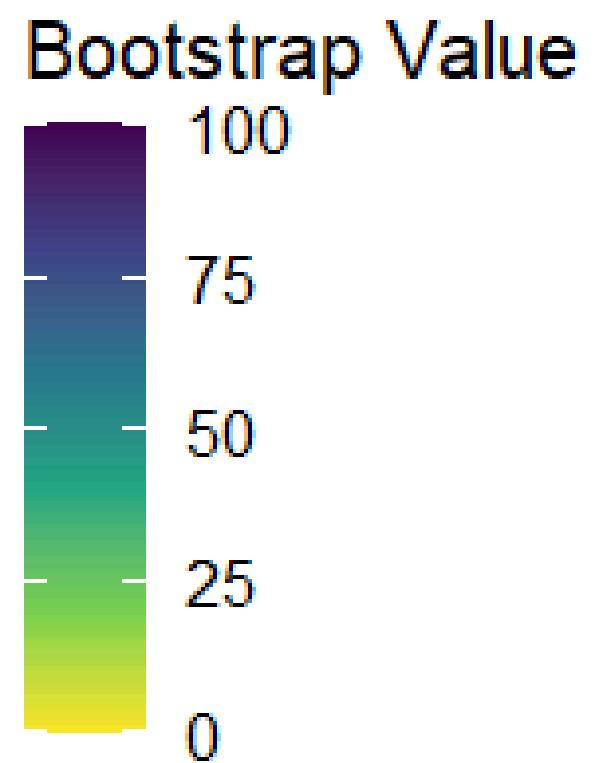
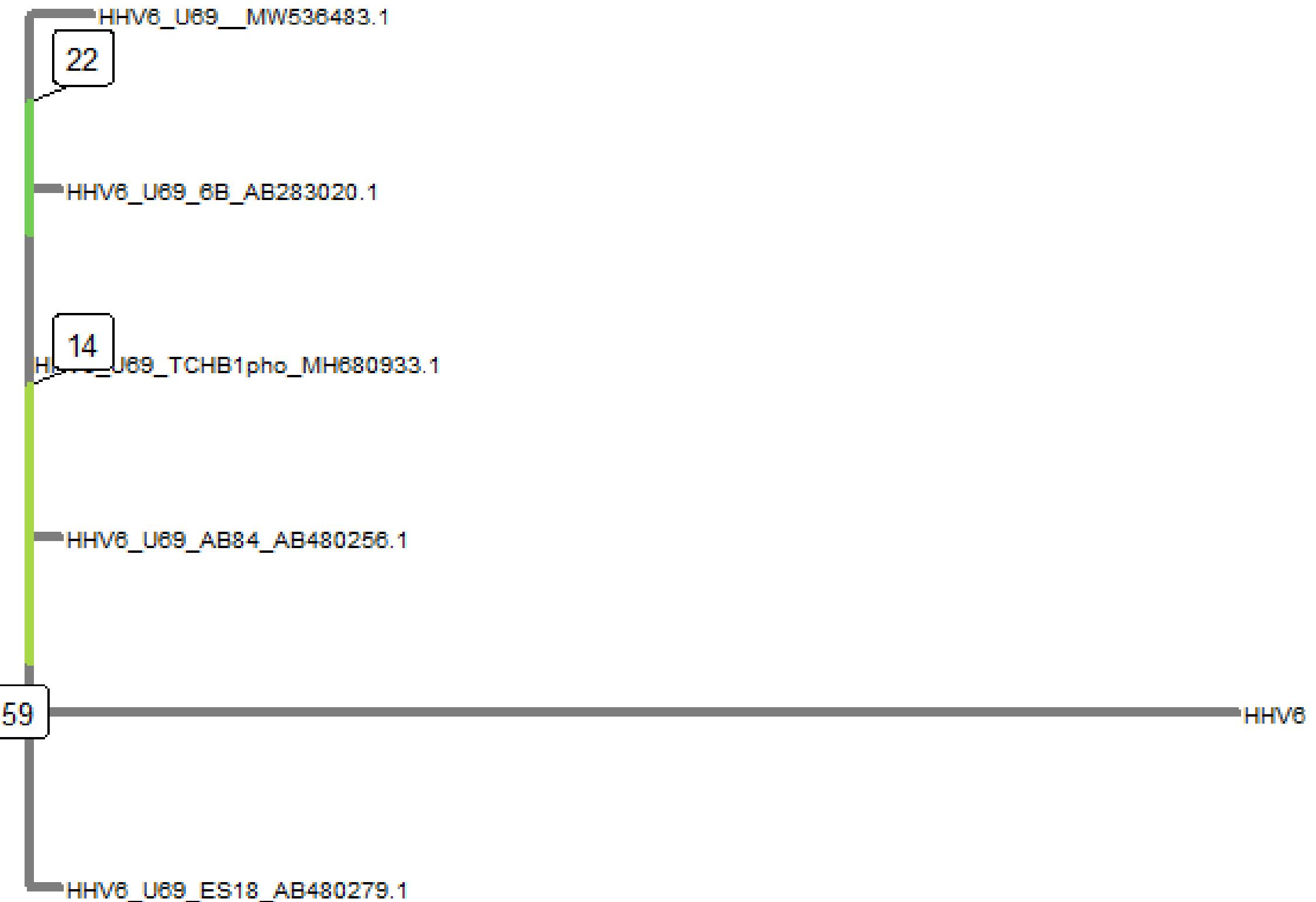


Gene U69: Bayesian Tree with posterior probabilities

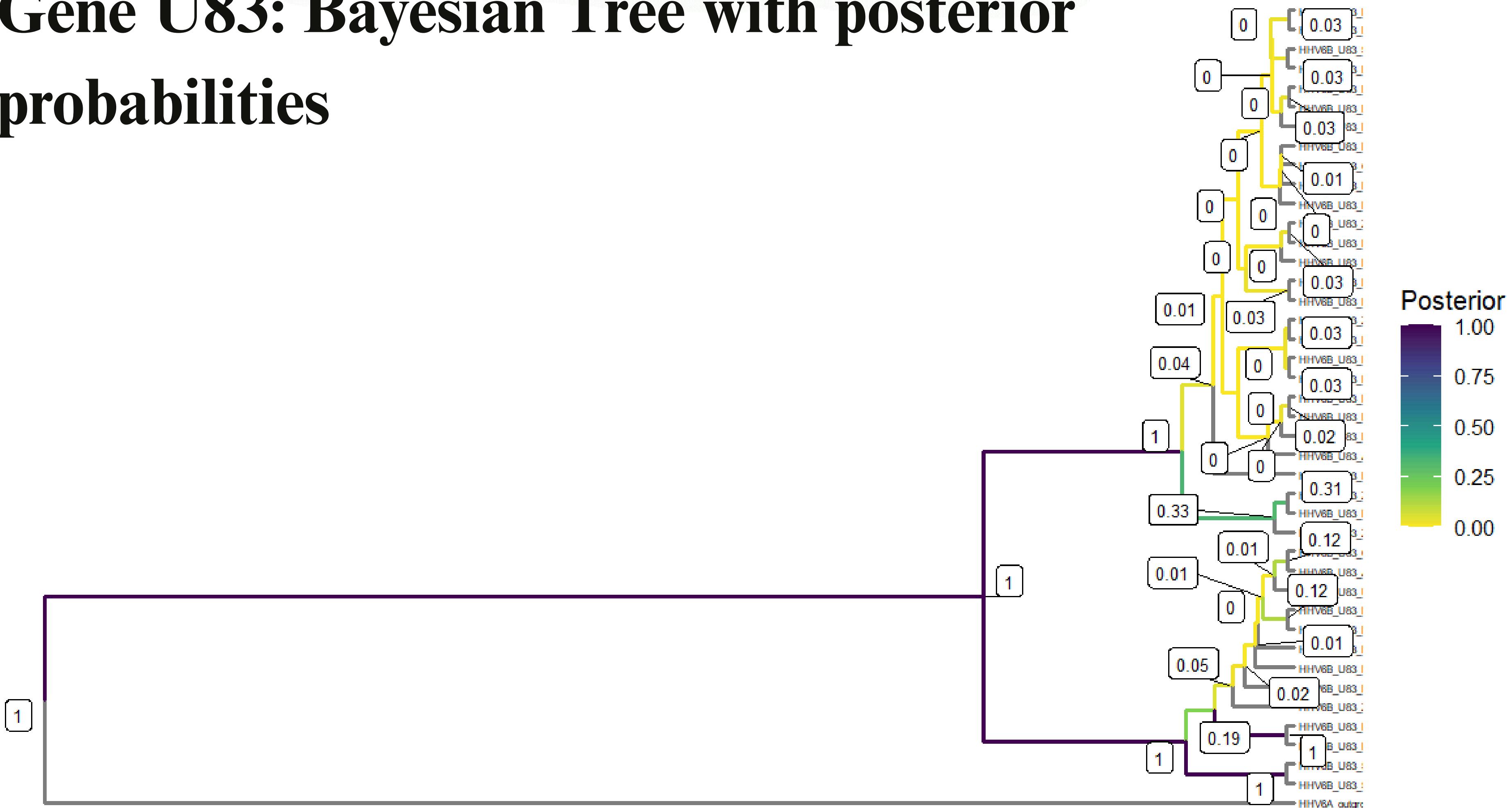
Only 6 taxa :(



Gene U69: ML Tree with bootstrap values

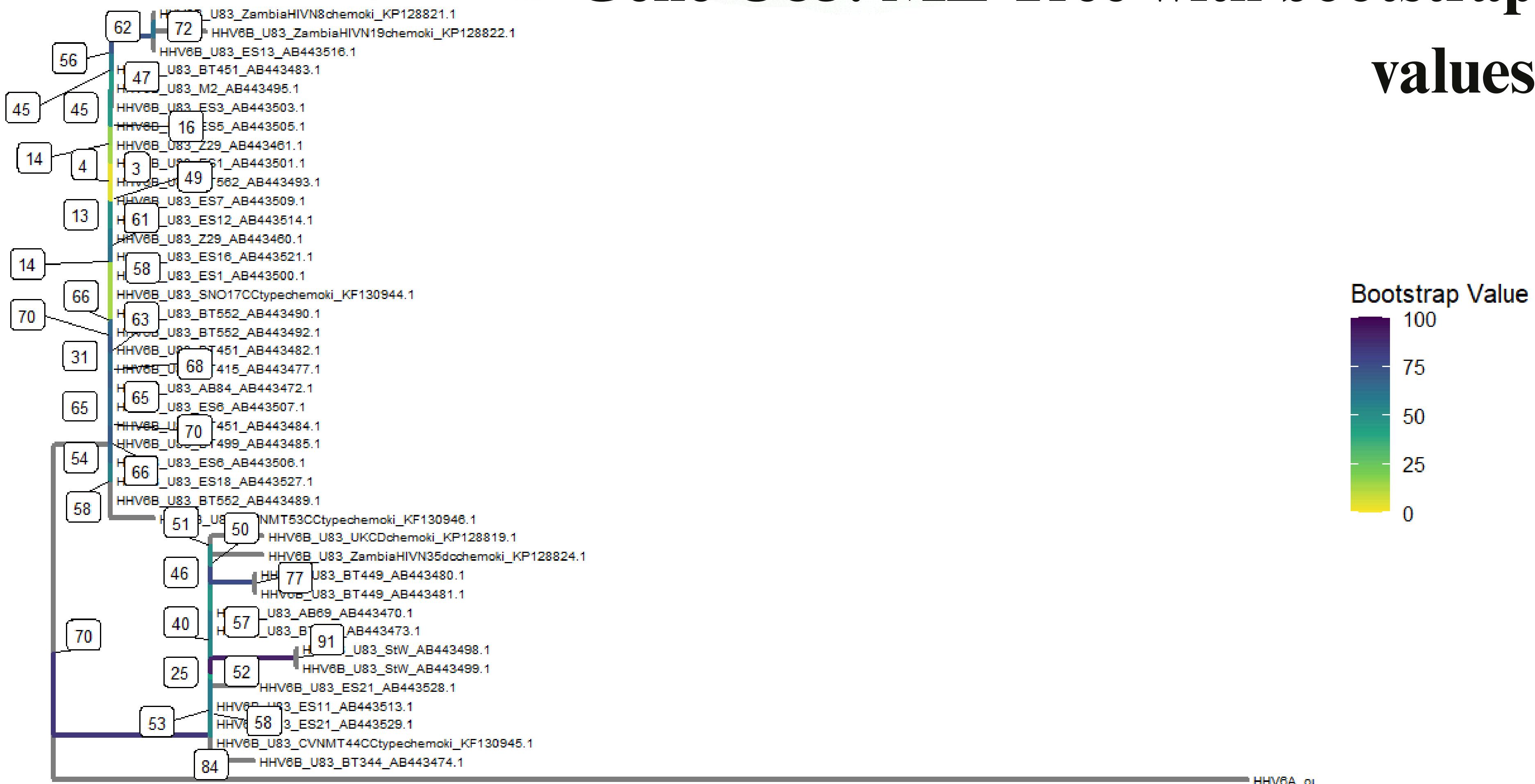


Gene U83: Bayesian Tree with posterior probabilities

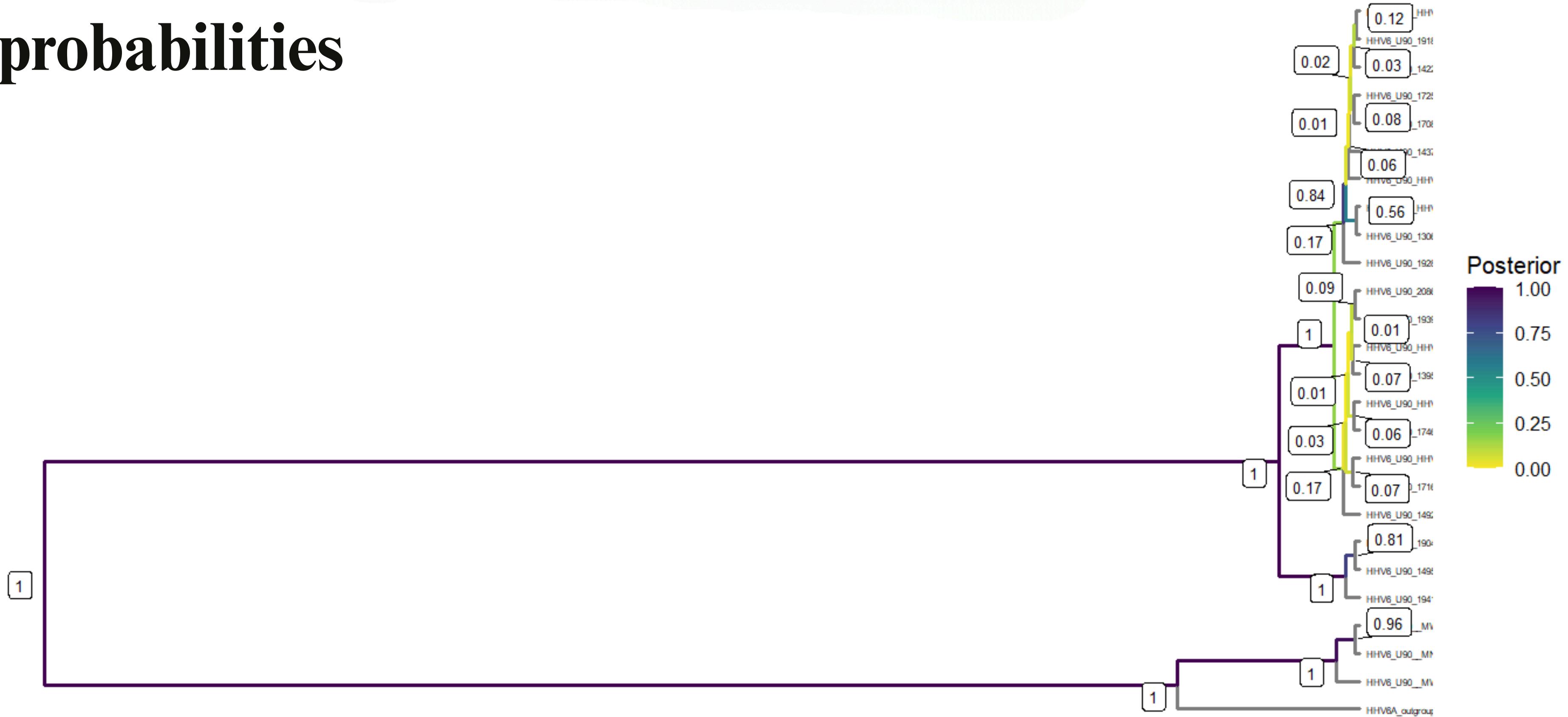


Gene U83: ML Tree with bootstrap

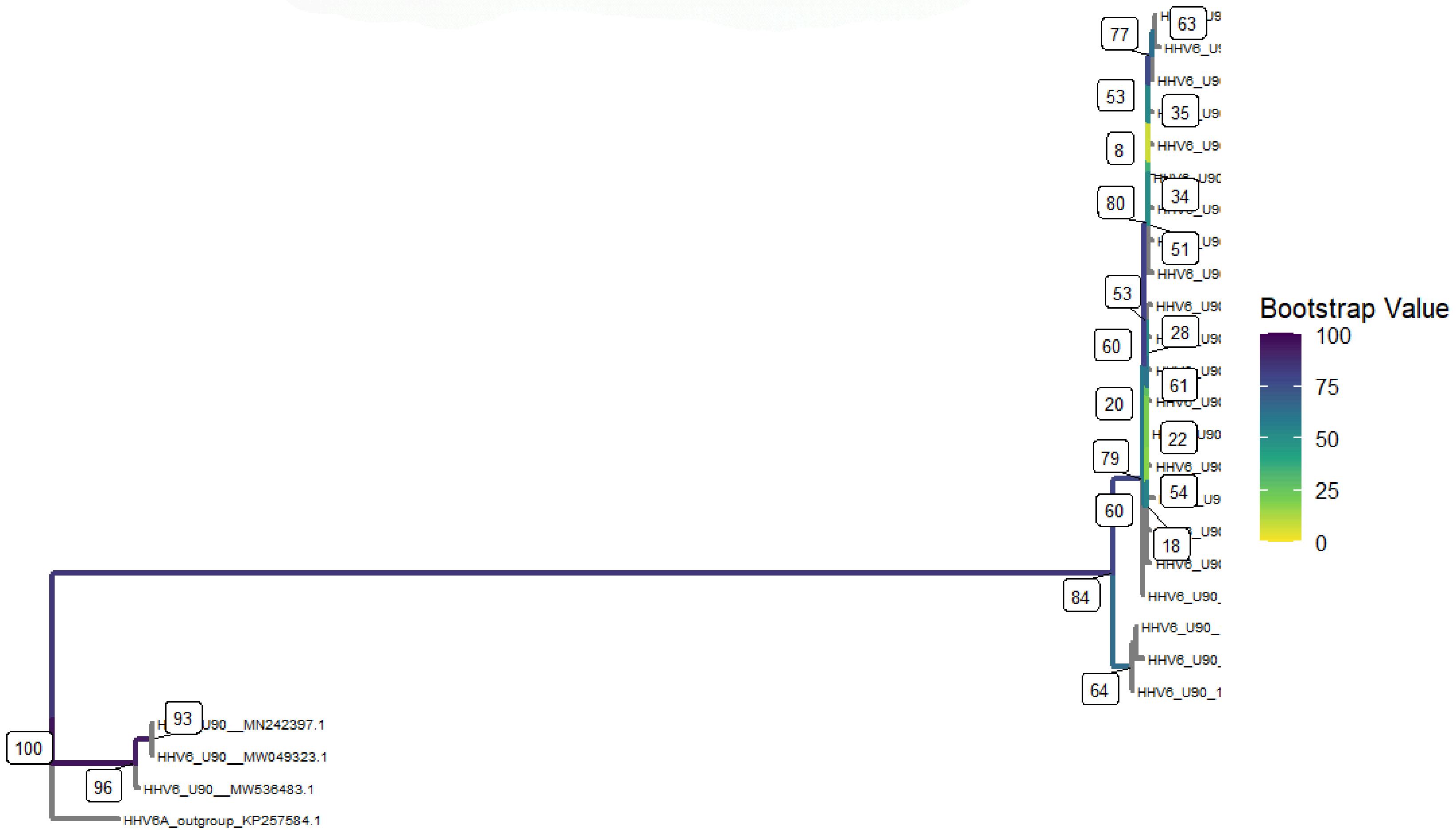
values



Gene U90: Bayesian Tree with posterior probabilities

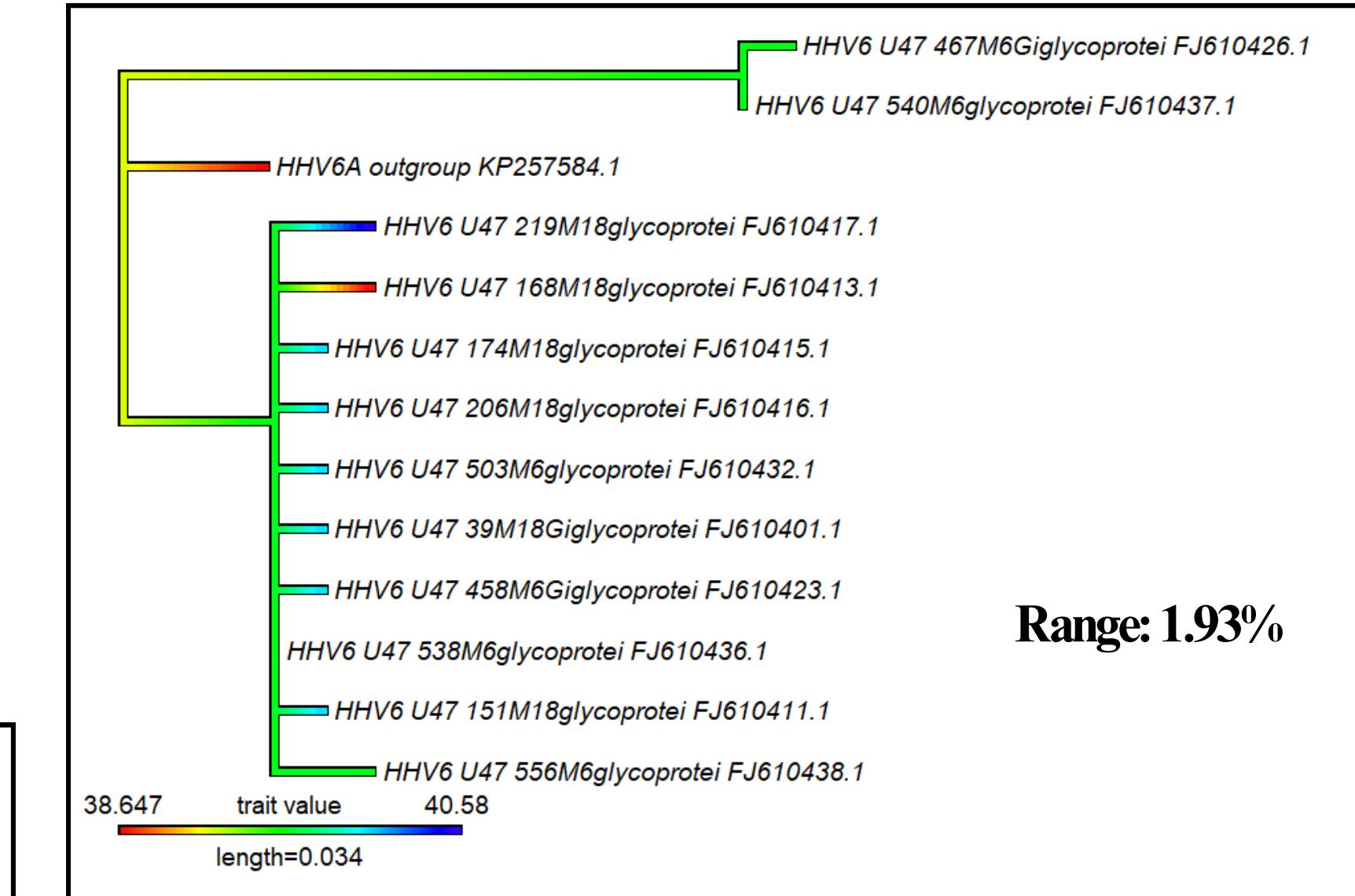
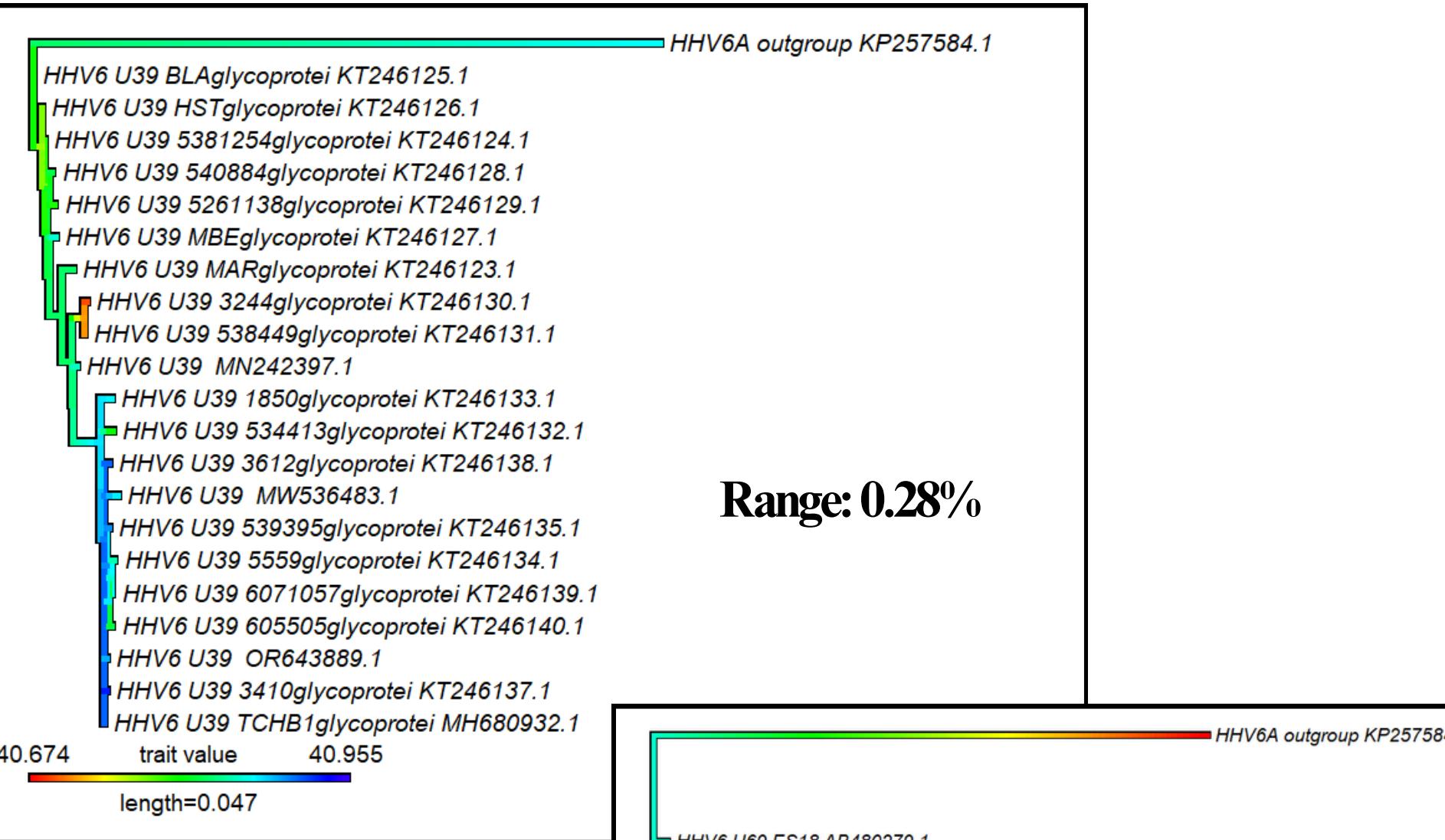


Gene U90: ML Tree with bootstrap values



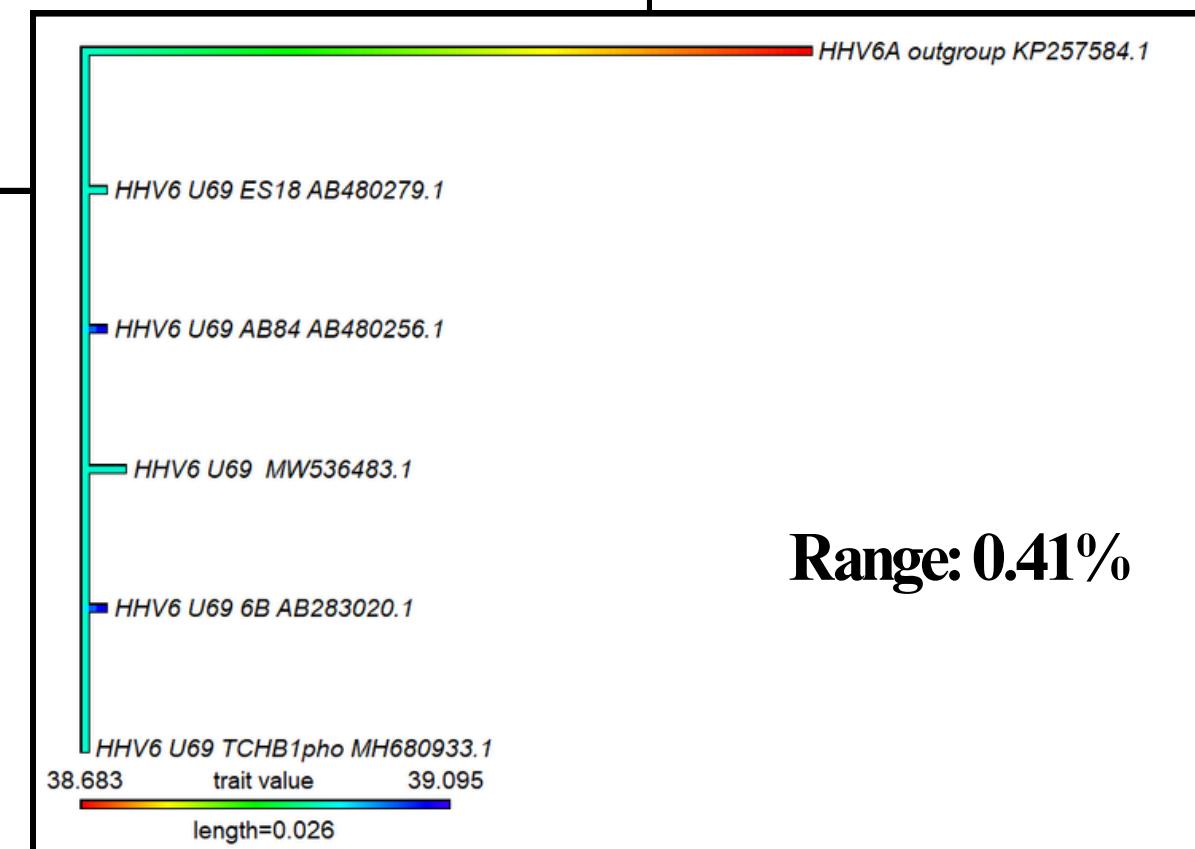
GC Percentage ContMaps

Gene U39



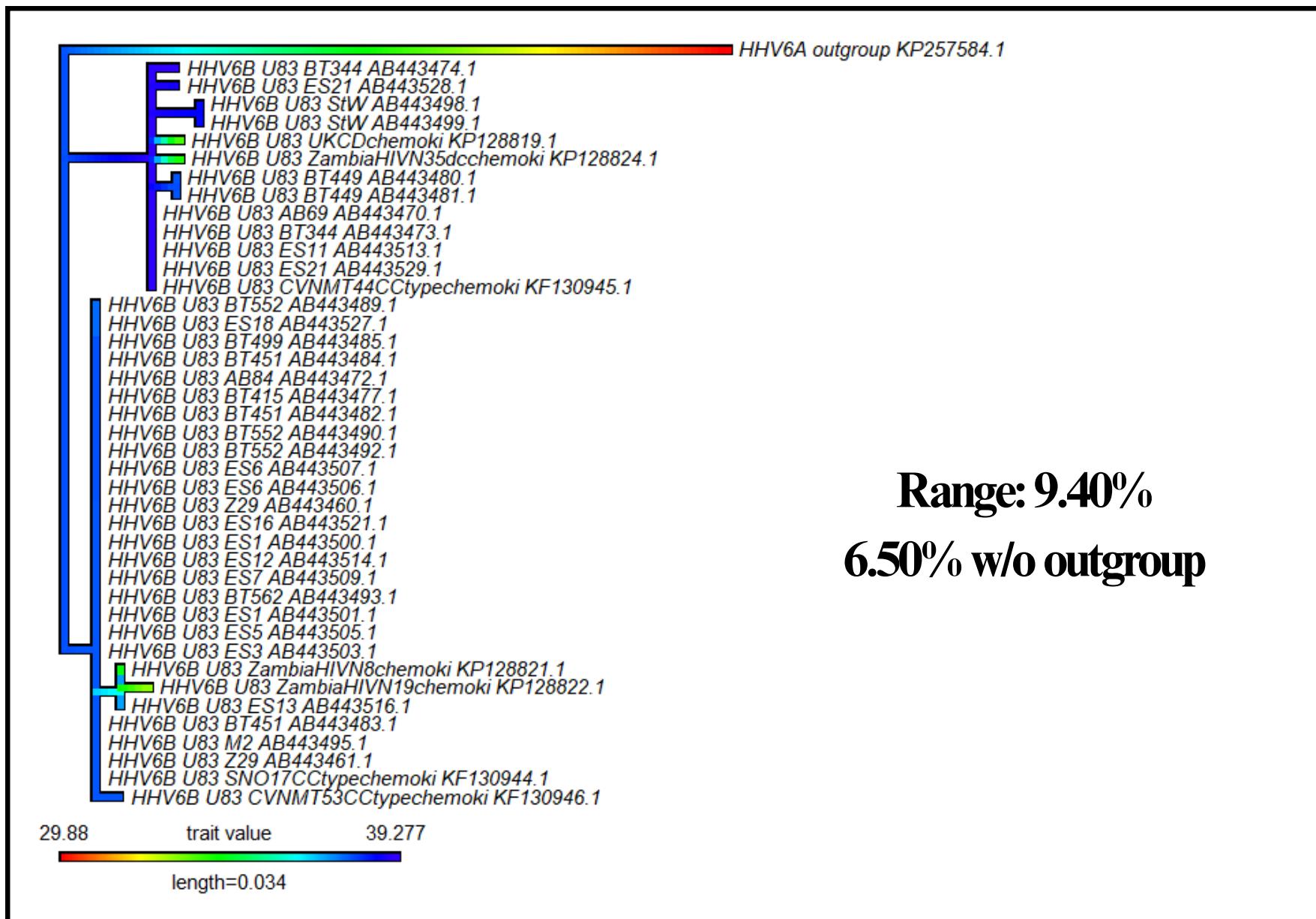
• Small amount of variation

Gene U69

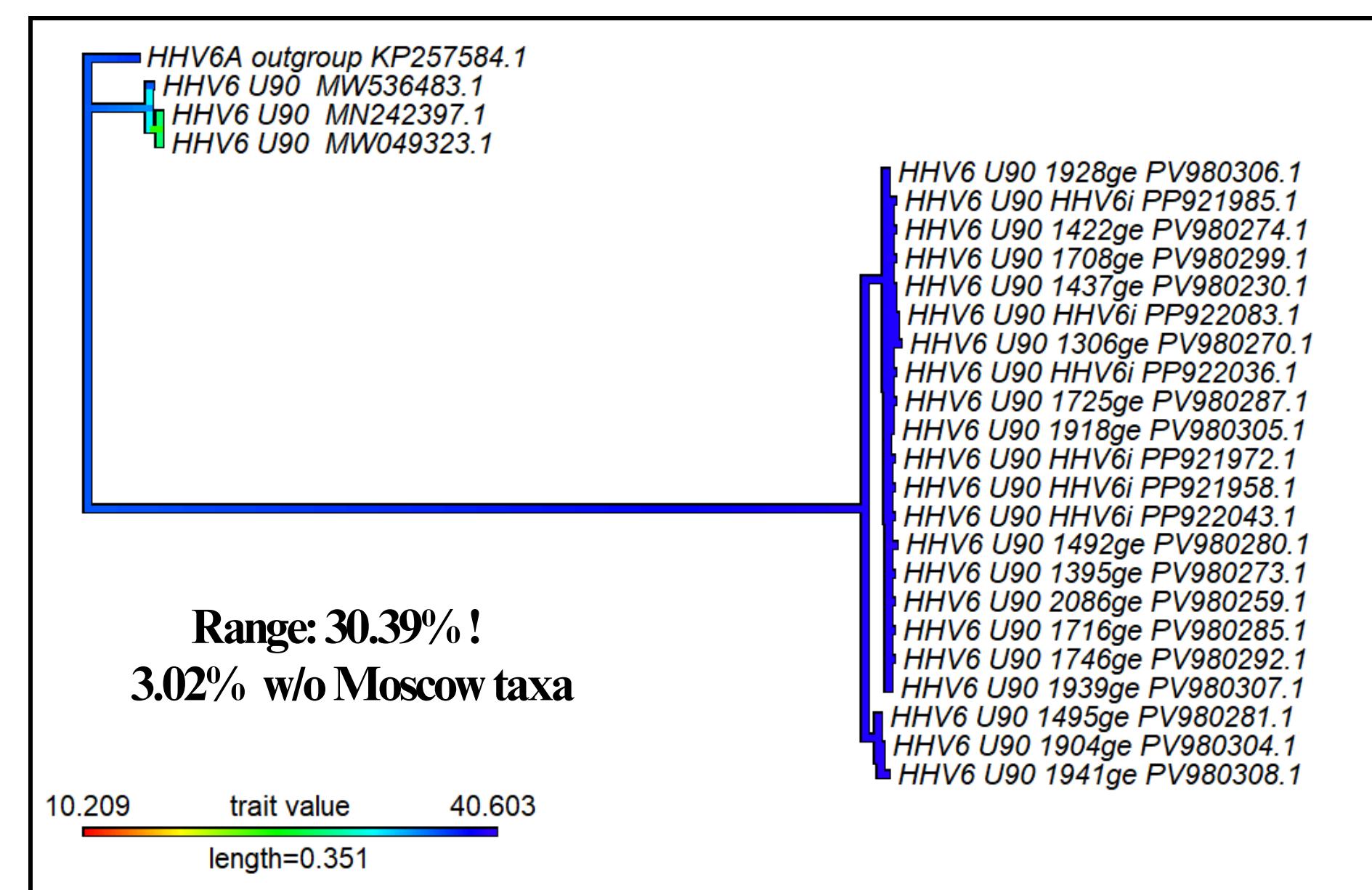


GC Percentage ContMaps

Gene U83

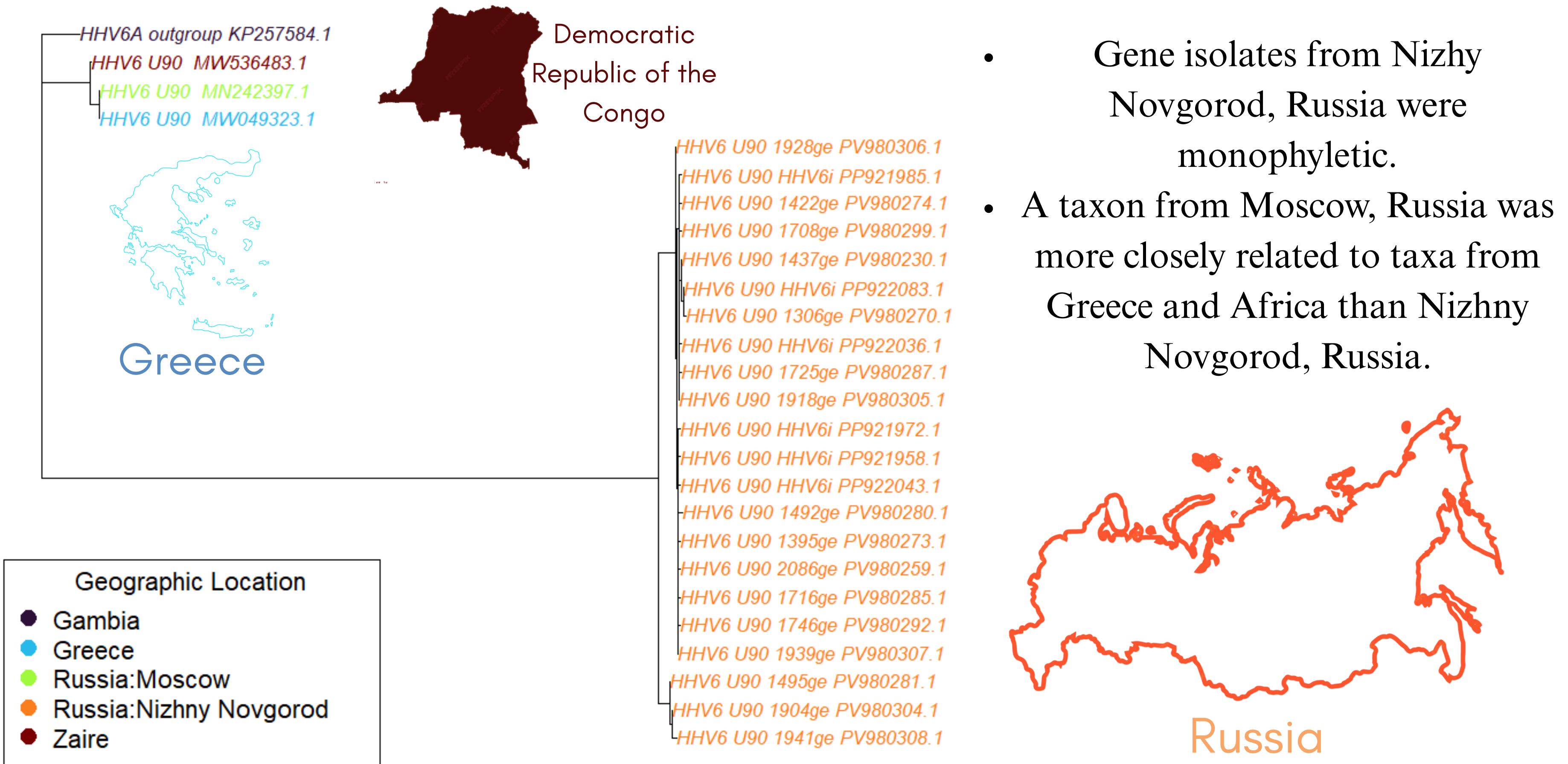


Gene U90

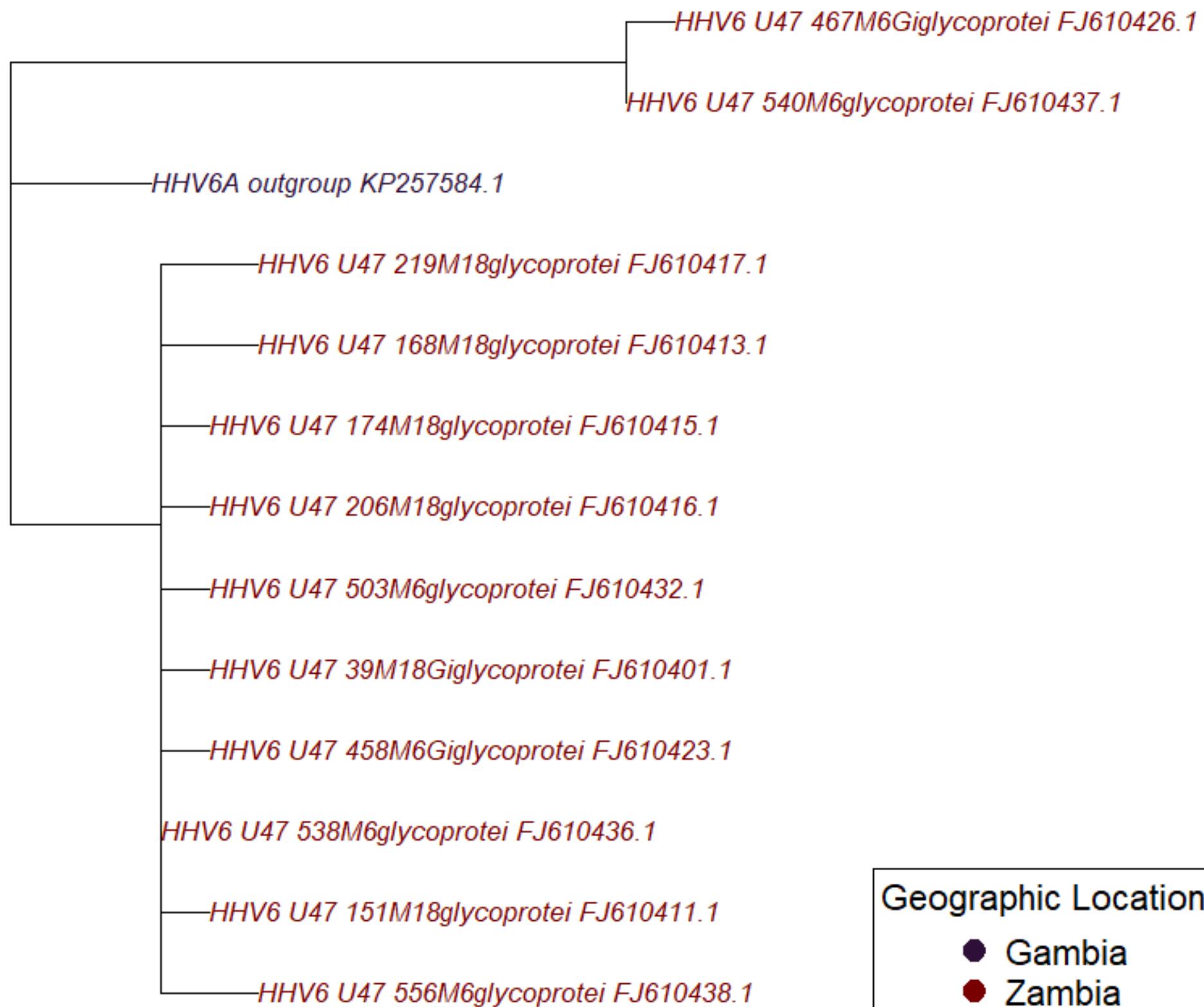


- Genes U83 and U90 have much higher differences in GC Percentage.

Gene U90 Geographic Locations



Gene U47 Geographic Locations



- All taxa for gene U47 (other than the outgroup) originated from Zambia, Africa



Gene U39 Geographic Locations

HHV6A outgroup KP257584.1

HHV6 U39 BLAglycoprotei KT246125.1

HHV6 U39 HSTglycoprotei KT246126.1

HHV6 U39 5381254glycoprotei KT246124.1

HHV6 U39 540884glycoprotei KT246128.1

HHV6 U39 5261138glycoprotei KT246129.1

HHV6 U39 MBEglycoprotei KT246127.1

HHV6 U39 MARglycoprotei KT246123.1

HHV6 U39 3244glycoprotei KT246130.1

HHV6 U39 538449glycoprotei KT246131.1

HHV6 U39 MN242397.1

HHV6 U39 1850glycoprotei KT246133.1

HHV6 U39 534413glycoprotei KT246132.1

HHV6 U39 3612glycoprotei KT246138.1

HHV6 U39 MW536483.1

HHV6 U39 539395glycoprotei KT246135.1

HHV6 U39 5559glycoprotei KT246134.1

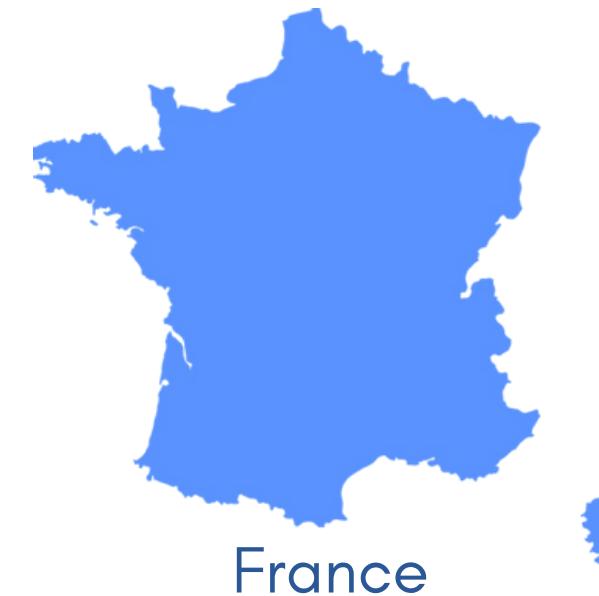
HHV6 U39 6071057glycoprotei KT246139.1

HHV6 U39 605505glycoprotei KT246140.1

HHV6 U39 OR643889.1

HHV6 U39 3410glycoprotei KT246137.1

HHV6 U39 TCHB1glycoprotei MH680932.1



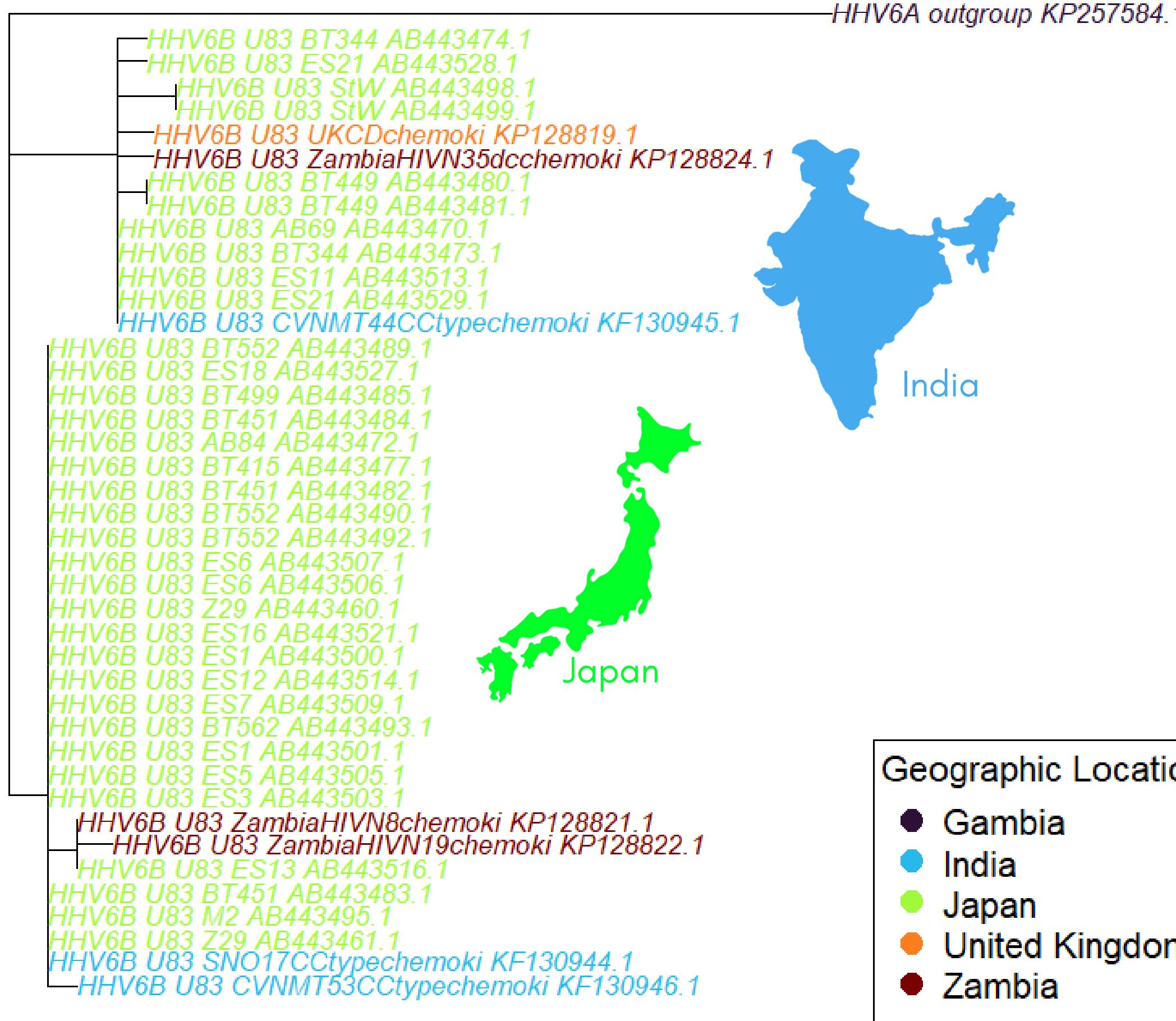
HHV6A outgroup KP257584.1

- Most taxa originated from France (with a few exceptions).

Geographic Location

- Finland
- France
- Gambia
- Japan
- Russia
- USA
- Zaire

Gene U83 Geographic Locations



- Many taxa from the same country were paraphyletic, but there were a few polyphyly groups.
- For example, taxa from Zambia and India did not share a recent common ancestor.

Why?

- Isolates that “jump” across geographic regions reflect global migration and HHV-6B’s ability to recombine.

(Finkel et al., 2020; Asward et al., 2020)

- *ici*HHV-6B sequences may confound our results.

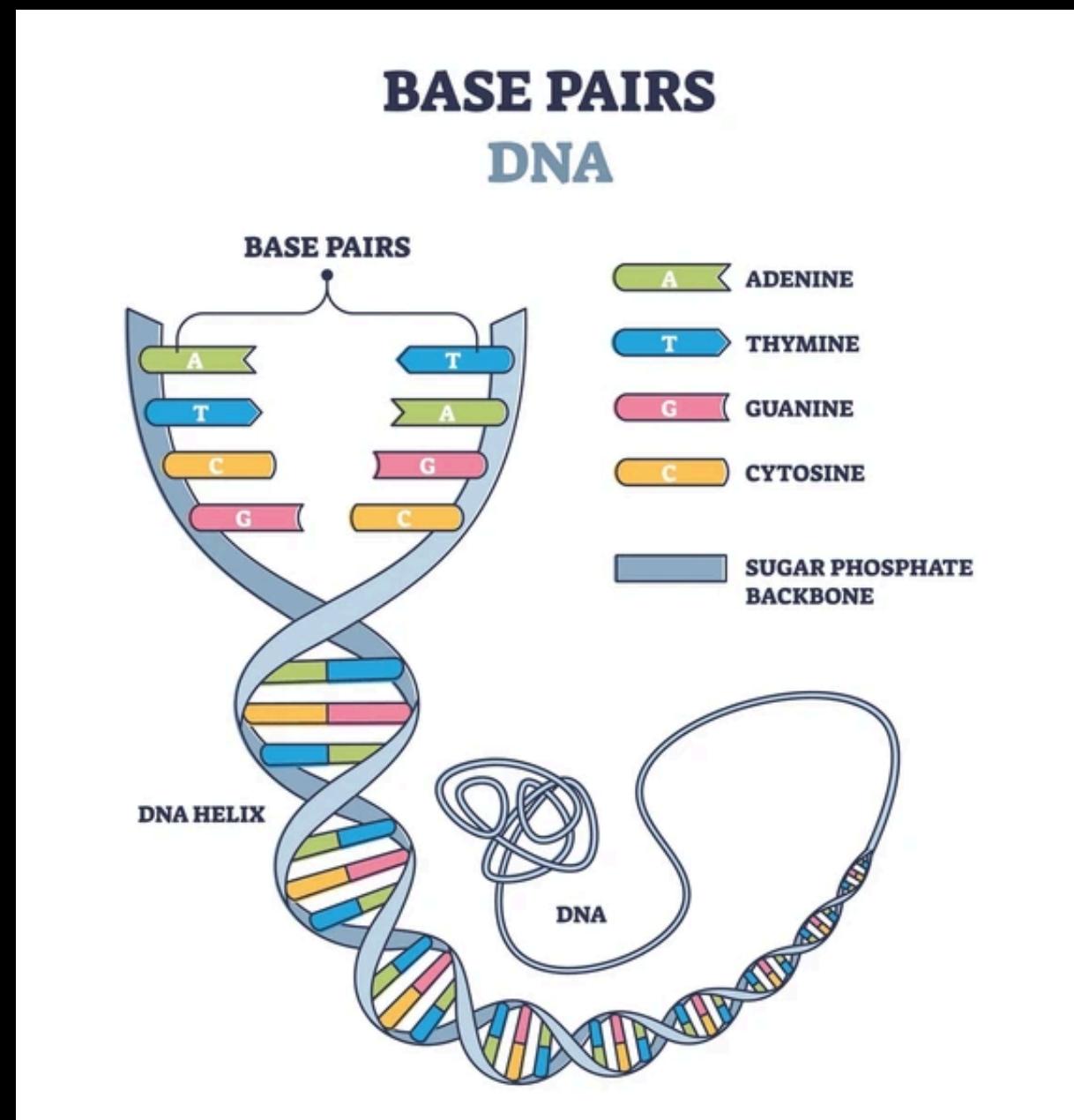
(Wood et al., 2025)

Interpretations/Biological Meaning

Our biggest take away is the GC% differences for each gene, which show that genes U90 and U83 may evolve more.

This has a couple of biological meanings

- It tells us which nucleotide sequences may be changing more rapidly
- Could help find what pressures are causing these changes
- May indicate parts of the chromosome are more likely to experience change than others



Why would Genes U90 and U83 have larger GC% differences?



Gene U90 is an immediate early (IE) gene, which are the first viral genes transcribed after infection.

(Stinski & Meier, 2007)

- Differences within IE genes may exist to meet regulatory needs of different cells.
- Host-facing genes/functions face higher evolutionary pressure and variability across strains.

(Finkel et al., 2020)

Gene U83 encodes for a chemokine-like protein that stimulates the immune system, which has documented sequence diversity from immune-driven selection and recombination.

(Zou et al., 1999; Sjahril et al., 2009)

Project Limitations

- Limited Data
 - We selected the genes that had the most available data/taxa.
 - Some genes showed overrepresentation of certain geographic clusters, labs, and dates.
 - Our potentially distorted geographic signal is not representative of the world population.
- Gene trees != Genome trees
 - Recombinations within distinct genes may misrepresent whole genome relationships.
 - There are only 6 complete HHV-6B genomes on GenBank, limiting future analyses.
- GC% is not a direct evolutionary metric.
 - It cannot distinguish synonymous vs nonsynonymous mutations.

Future Directions

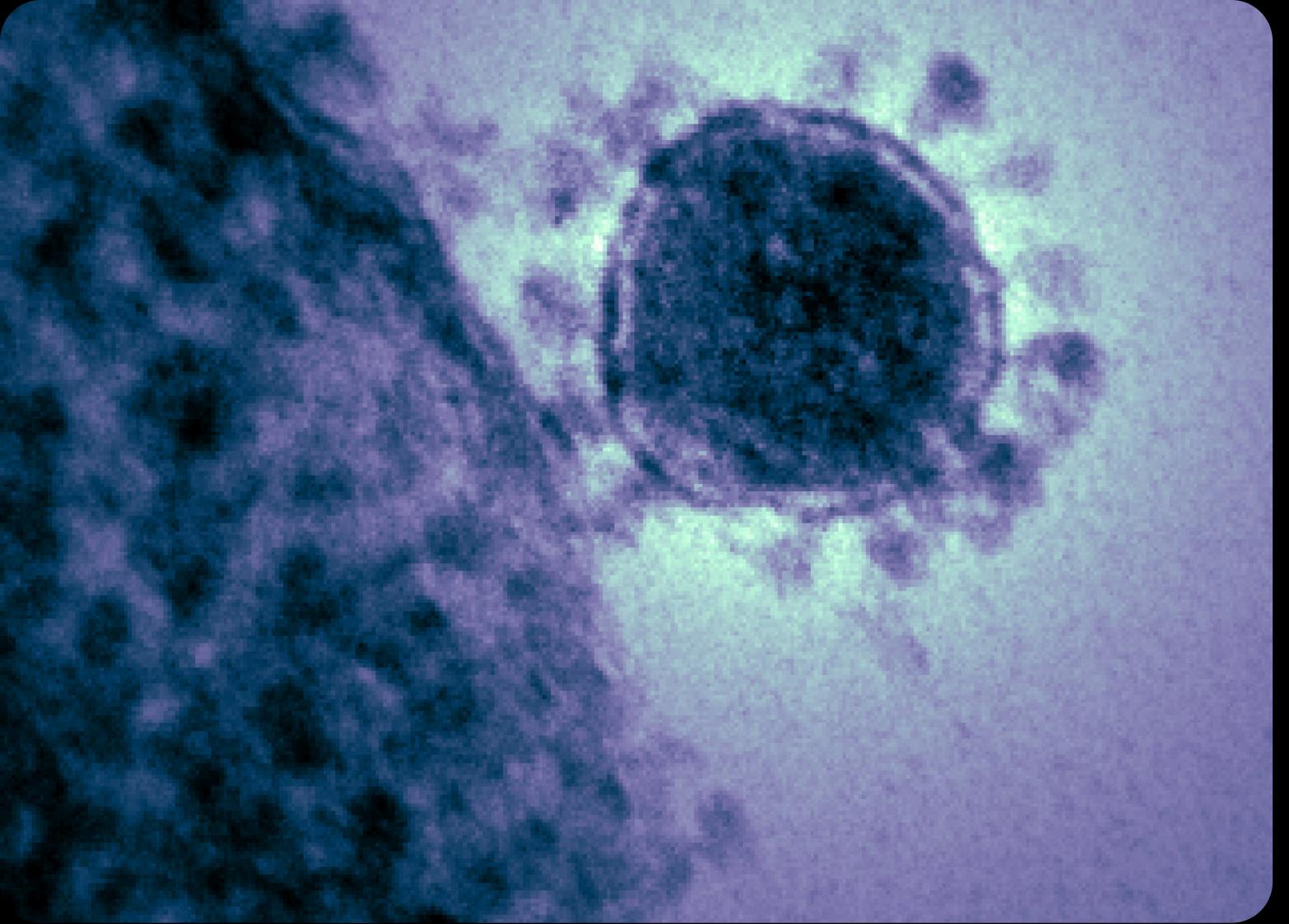
- Explore other parameters beyond GC% changes to research gene mutations.
 - For example, investigate dN/dS ratios within each gene.
- Comparative genomics with other subfamilies of herpesviruses.
- Analyze the differences between circulating HHV-6B and inherited chromosomally integrated HHV-6B sequences.

How existing iciHHV-6B studies compare with our HHV-6B genes

iciHHV-6 Comparisons

- Gene U90 had the highest dN/dS ratio in iciHHV-6B
- Signs of accelerated evolution were found in genes U24, U47, and U90 of iciHHV-6A
- U39 (in iciHHV-6A) had low dN/dS values (consistent with our low GC% range)

(Telfrod et al., 2018)



e BIOL4300

Thank you!



References