

# Characterization of HERV9 Elements within the Human Genome

Gary A. Hovsepian<sup>1\*</sup>, Keylie M. Gibson<sup>2\*</sup>, Miguel de Mulder<sup>1</sup>, R. Brad Jones<sup>1</sup>, Keith A. Crandall<sup>2</sup>, Douglas F. Nixon<sup>1</sup>, Matthew L. Bendall<sup>1,2</sup>

<sup>1</sup>Department of Microbiology, Immunology & Tropical Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, D.C., USA;

<sup>2</sup>Computational Biology Institute, Milken Institute School of Public Health, The George Washington University, Washington, D.C., USA

## Introduction

- Human Endogenous Retroviruses (HERVs) are the genomic remains of ancient retroviruses that infected vertebrate genomes millions of years ago.
- Most are believed to be transcriptionally silent in normal human tissue due to an accumulation of mutations in the coding regions and long terminal repeats (LTRs).
- Recent evidence has shown several mechanisms by which HERV expression can **influence homeostatic processes** (alternative enhancers for protein coding genes, activation of non-coding genomic regions, and expression of retroviral transcripts or proteins).
- The **HERV9 family** is of particular interest because it represents one of the more recent endogenization events and is thus expected to **retain more of its functional capacity** than older HERV families.

## Methods

**Constructing HERV9 Organization**  
 1) Consensus sequence from dfam.  
 2) psiBLASTed ORFs and extended hits to identify homologous domains.  
 3) Determined slippage site by comparison to other retroviruses and frame change.

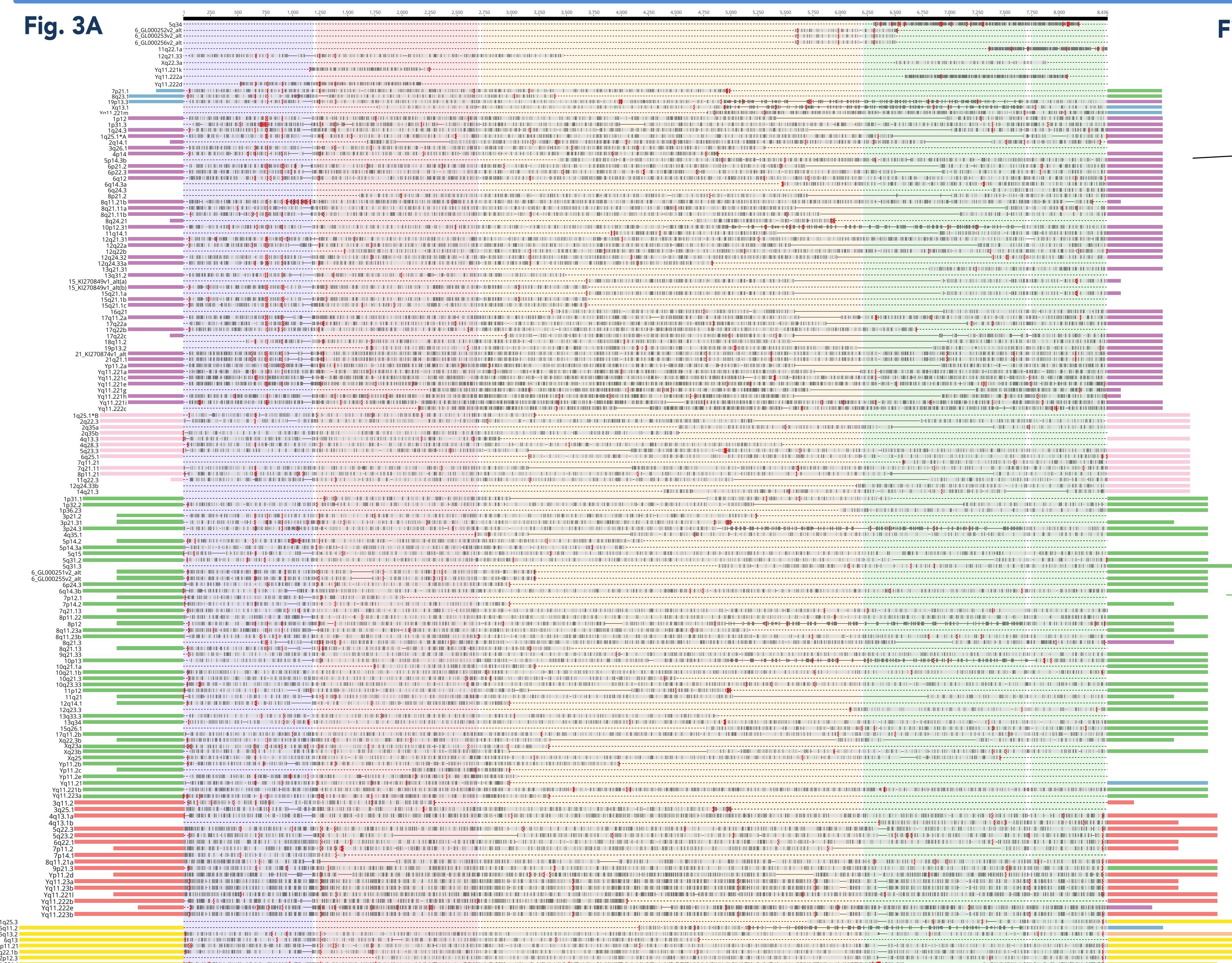
**Phylogenetic Trees**  
 1) Extracted homologous proteins.  
 2) Aligned using MAFFT with E-INS-i algorithm.  
 3) Reconstruction with RAxML with 100 bootstrap replicates.

**Identifying HERV9 Elements**  
 1) Used TELESCOPE to extract elements from hg38.  
 2) HMMalign with HERV9 model.  
 3) Reconstruction with RAxML with 100 bootstrap replicates.

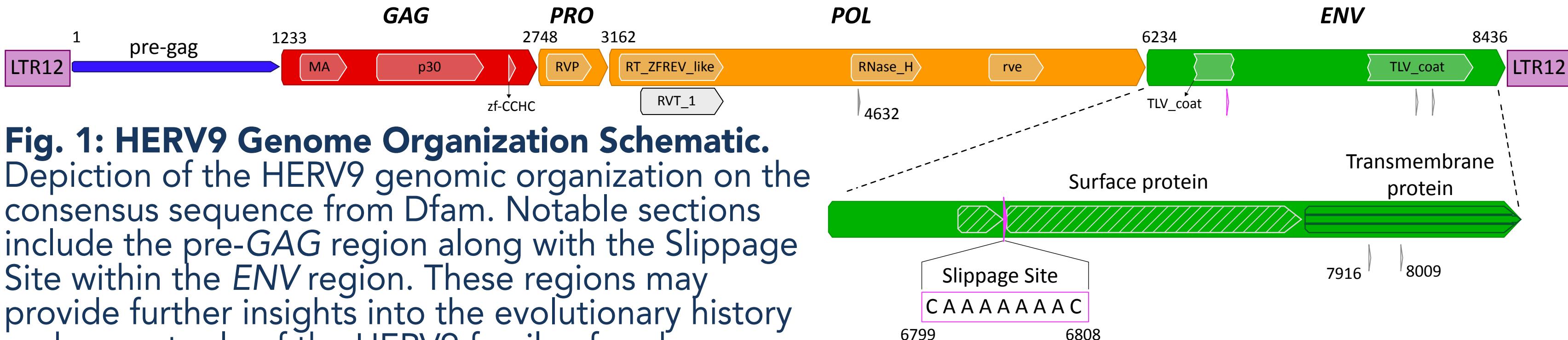
## Results

- In the present work, we identified and detailed the location and genomic context of 190 HERV9 elements in humans.
- This bioinformatic analysis has led to a **characterization of all near-complete HERV9 elements in the human reference genome** (hg38), with a report on the genomic and epigenomic context of their insertions throughout the genome and a phylogenetic classification of HERV9 subfamilies.
- Our exploratory analyses show dynamic connectivities within the HERV9 families.

## HERV9 Elements in the Human Genome



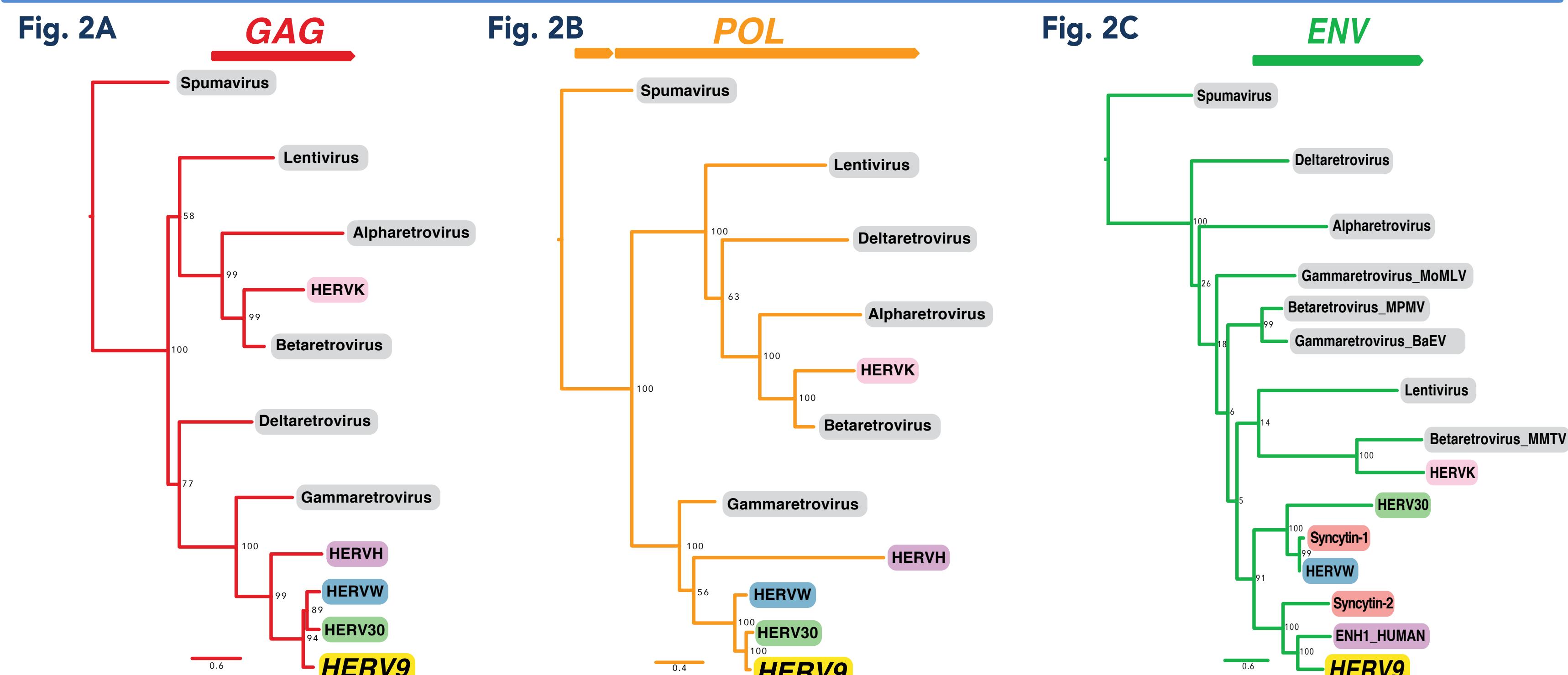
## HERV9 Genome Organization



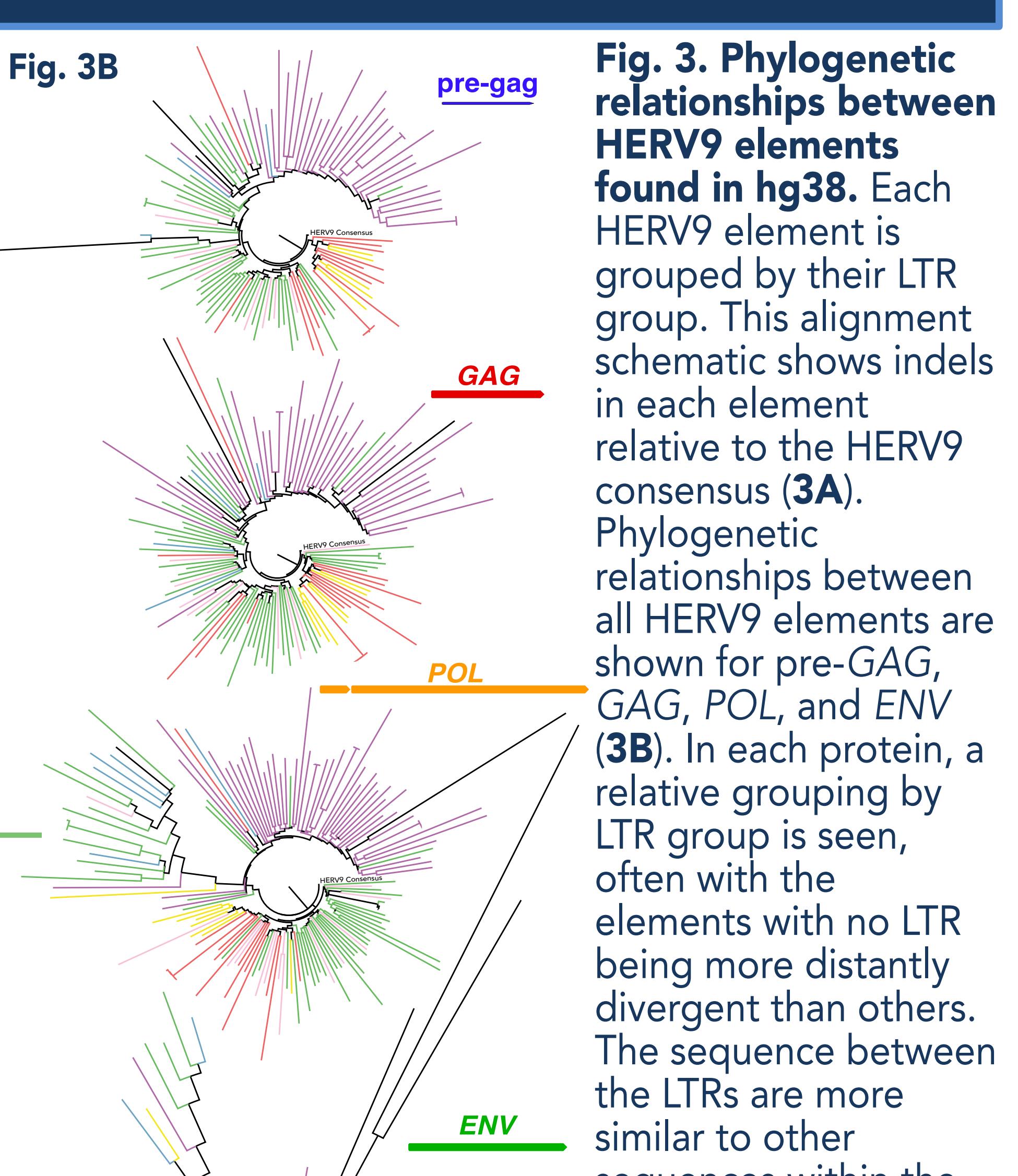
**Fig. 1: HERV9 Genome Organization Schematic.**

Depiction of the HERV9 genomic organization on the consensus sequence from Dfam. Notable sections include the pre-GAG region along with the Slippage Site within the ENV region. These regions may provide further insights into the evolutionary history and current role of the HERV9 family of endogenous retroviruses in humans today.

## Evolutionary Relationships of Retroviral Families with HERV9



**Fig. 2. Phylogenetic relationship of HERV9 genes.** The evolutionary relationships of retroviral proteins GAG (2A), POL (2B), and ENV (2C) were investigated between HERV9, other HERVs, and retrovirus families. Spumavirus was the outgroup. Between 2A and 2B, there is near uniformity, except regarding the placement of the Deltaretrovirus family. In 2C, there is a greater divergence between HERV9 and HERV30, with HERV9 being more closely related to HERVH in the ENV protein compared to 2A and 2B.



## Conclusions

The goal of our study is to provide an **exhaustive reference library for HERV9** to be used in understanding its role in both pathology and cooption throughout human evolution. This exhaustive reference library characterizing complete HERV9 sequences within the genome could be used in understanding its role in both pathology and cooption throughout human evolution.

## Acknowledgments

This work was supported by National Institute of Health grants (CA206488 and AI76059) and the Luther Rice Undergraduate Research Fellowship from GWU. Thank you to my mentors for their support, specifically Matthew L. Bendall, Keylie M. Gibson, and Douglas F. Nixon for their direct oversight on this project.

The Department of  
MICROBIOLOGY,  
IMMUNOLOGY, &  
TROPICAL MEDICINE

Computational  
Biology Institute  
THE GEORGE WASHINGTON UNIVERSITY

Milken Institute School  
of Public Health  
THE GEORGE WASHINGTON UNIVERSITY