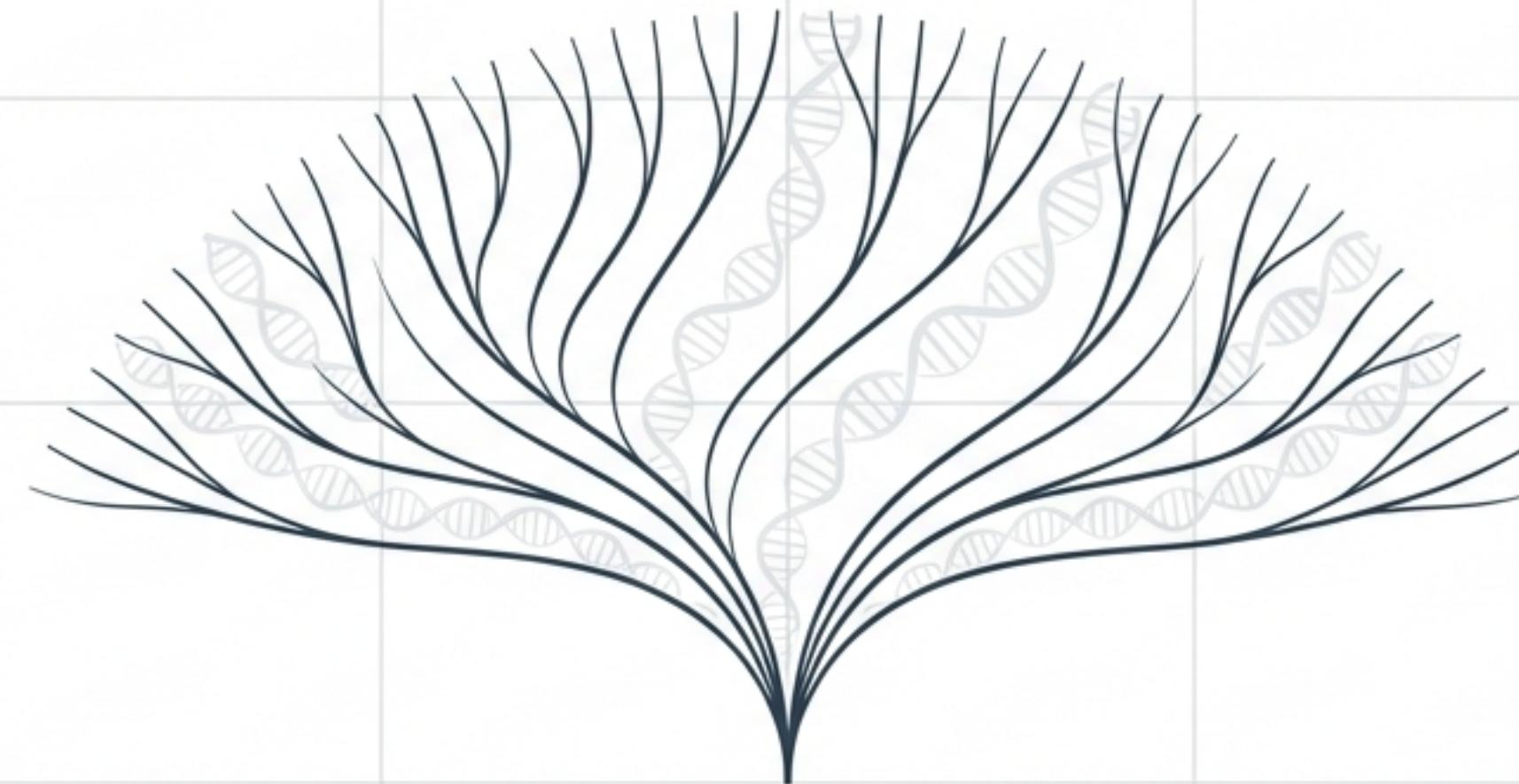


Phylogenetic Analysis and Genetic Variation in Arteriviridae

A Comparative Study of Replicase, Glycoprotein 5, and Nucleocapsid Protein Genes



ALI AHMAD, MSC CLINICAL AND MOLECULAR MICROBIOLOGY
SCHOOL OF LIFE SCIENCES, QUEEN'S MEDICAL CENTRE

Executive Summary

The Subject: The family Arteriviridae (order Nidovirales) comprises viruses affecting swine, equines, and simians. This study conducts a phylogenetic and distance analysis of three distinct genomic regions: Replicase (ORF1a) (Biological Blue, #0055AA), Glycoprotein 5 (ORF5) (Alert Red, #D32F2F), and Nucleocapsid (ORF7) (Structural Green, #2E7D32).

The Core Conflict: The prevailing biological hypothesis suggests the surface protein (GP5) should exhibit the highest genetic variation due to intense immune pressure from the host.

The Finding: Contrary to the hypothesis, the structural Nucleocapsid (N) protein and Replicase genes demonstrated significantly higher levels of genetic variation than the surface Glycoprotein 5.

Expected:



GP5

(High Variation)



N Protein

(Low Variation)

Observed:



N Protein

(High Variation)



GP5

(Low Variation)

**HYPOTHESIS
REFUTED**

Profile of the Family Arteriviridae



Equine Arteritis Virus (EAV)

Respiratory & reproductive disease.



Porcine Reproductive and Respiratory Syndrome Virus

Global agricultural impact.



Simian Haemorrhagic Fever Virus (SHFV)

High mortality in macaques.



Lactate Dehydrogenase-Elevating Virus (LaDV)

Lifelong viremia.



Wobbly Possum Disease Virus (WPDV)

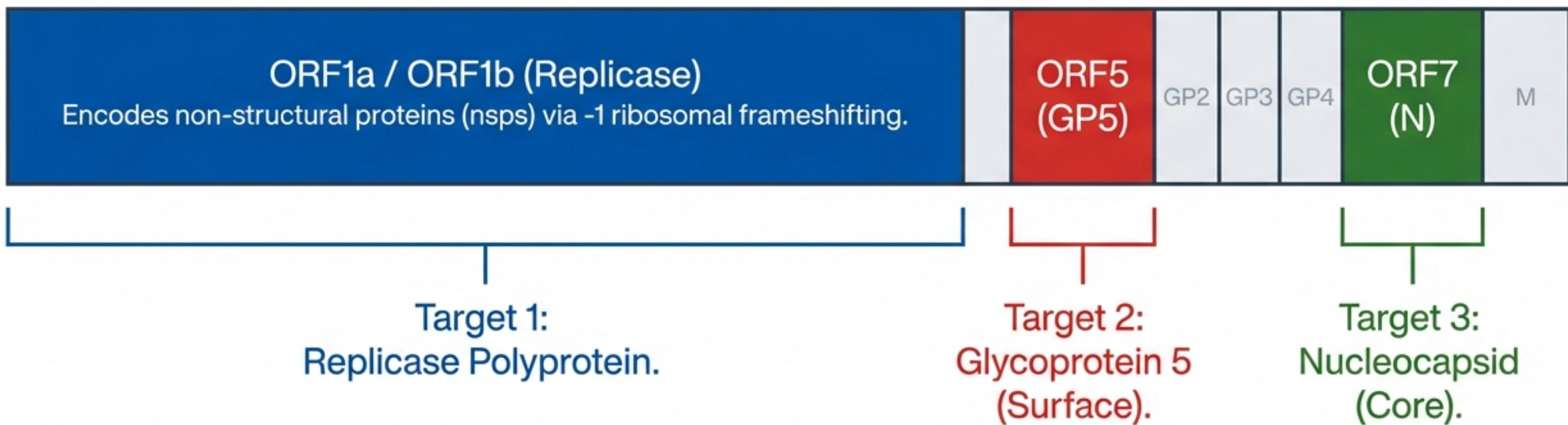
Neurological disease; highly divergent.

Taxonomy: Order Nidovirales (with Coronaviridae). The family contains a single genus but diverse sub-lineages based on host range.

Pathogenicity: No known human pathogenicity. Significant veterinary impact including interstitial pneumonia and abortion in mares.

Genomic Architecture

Polycistronic, positive-sense RNA genome structure



The Variables: Three Genes of Interest



Replicase (ORF1a)

Encodes enzymes for polyprotein cleavage. Historically considered the most conserved gene in the family.



Glycoprotein 5 (ORF5)

A glycosylated transmembrane protein (~25 kDa). Contains the major neutralization determinants. This is the primary target for host antibodies.

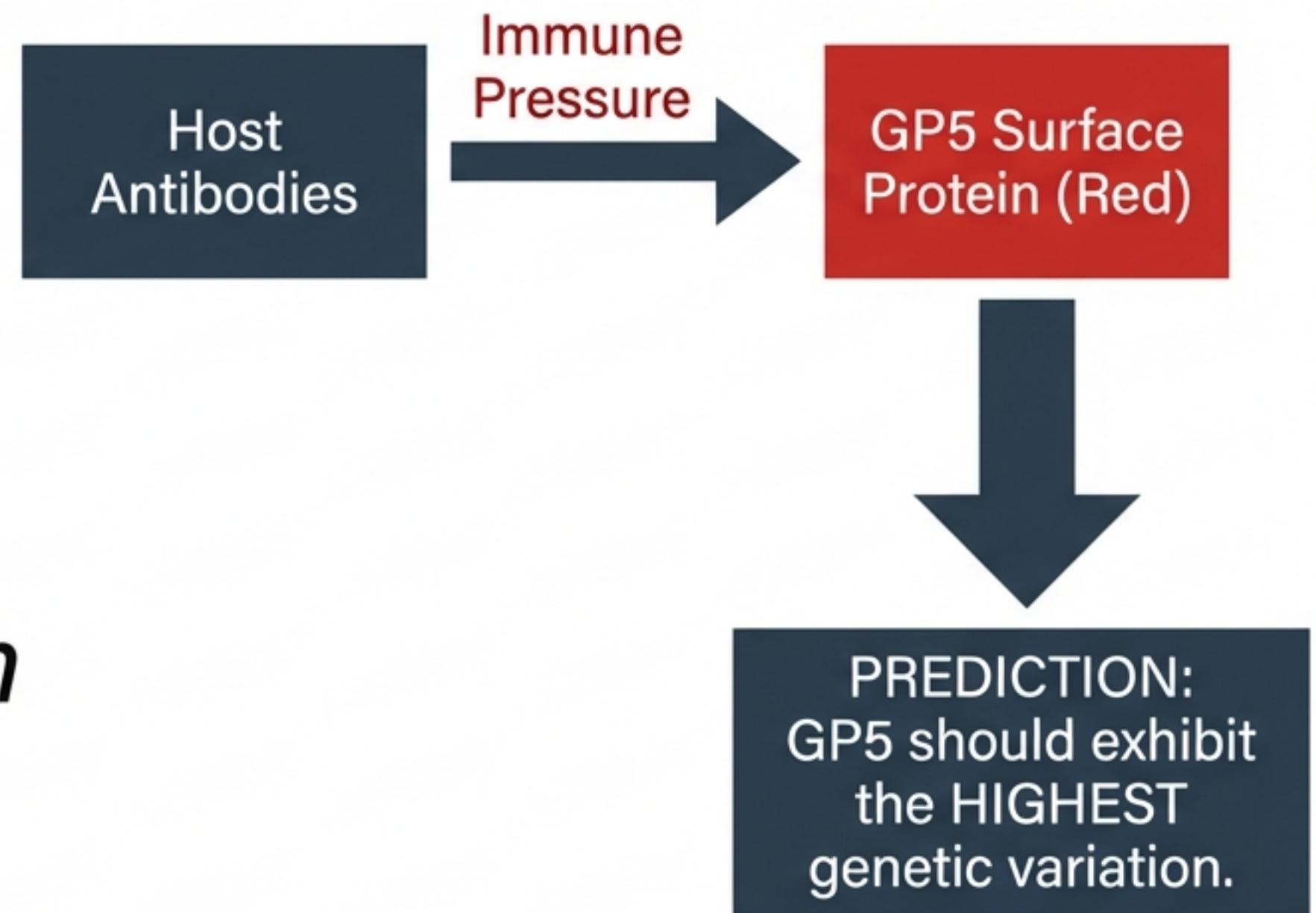


Nucleocapsid (ORF7)

Forms disulfide-bridged homodimers. Constitutes the viral nucleocapsid, the structural core of the virus.

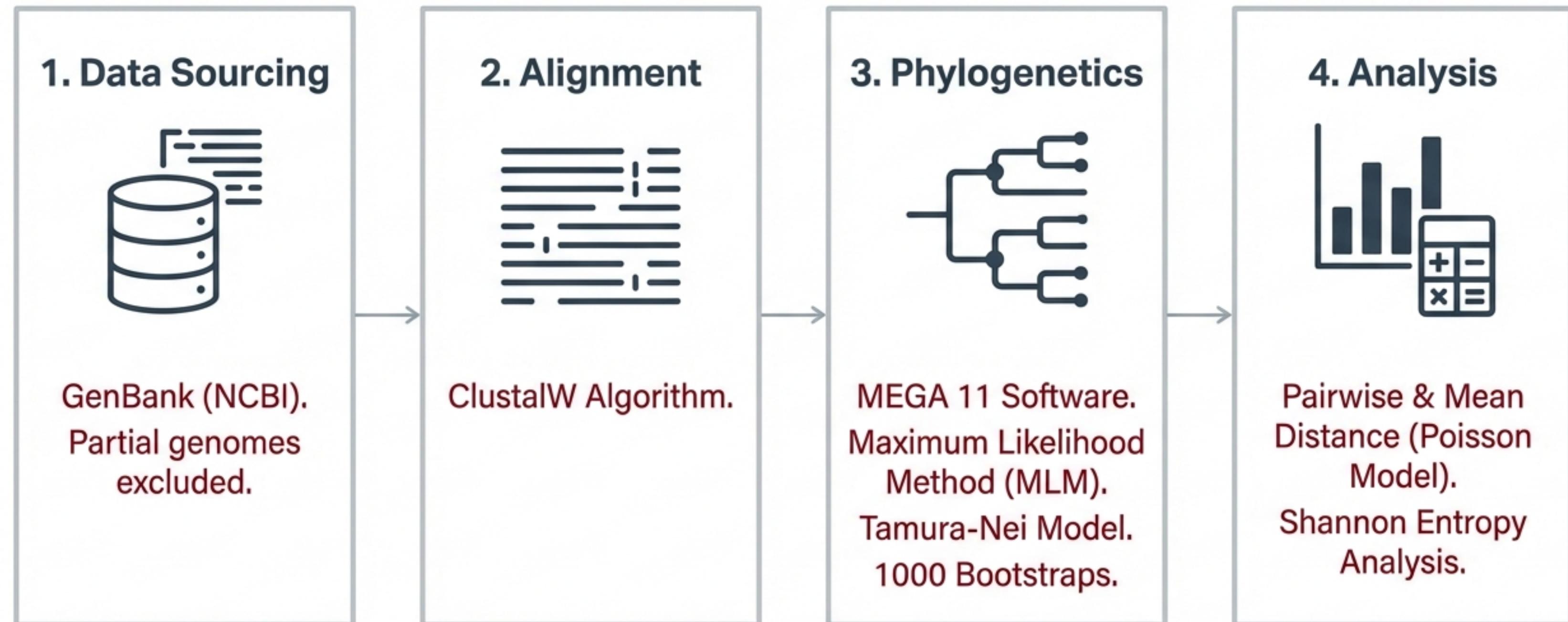
The Hypothesis

Because Glycoprotein 5 (GP5) contains major immune epitopes, it is subject to direct evolutionary pressure from the host immune system.



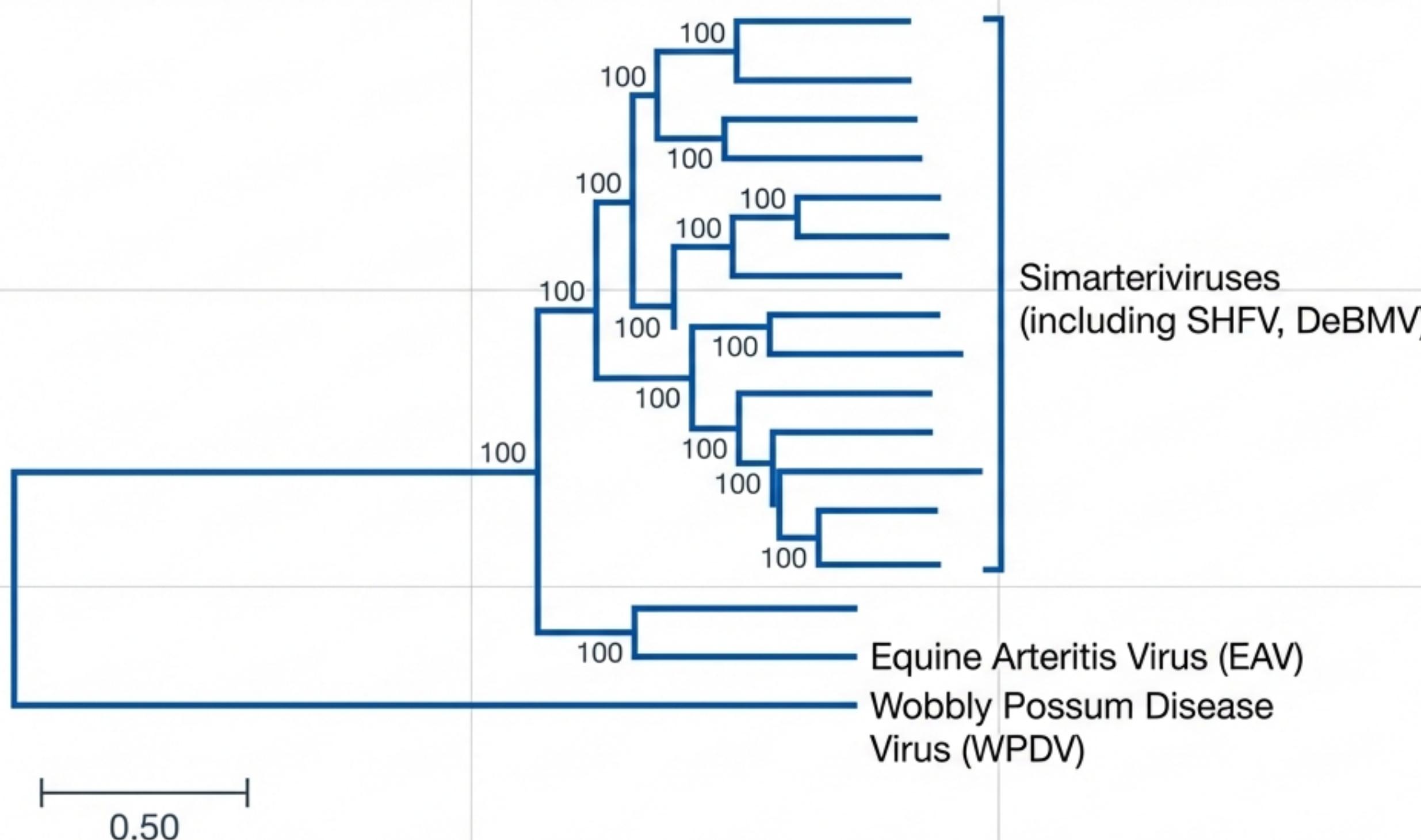
Null Hypothesis: There is no significant difference in variability, or structural genes (N) are more variable.

Methodology & Bioinformatics Workflow



Phylogenetics: Replicase (ORF1a)

Baseline relationships showing distinct clustering.



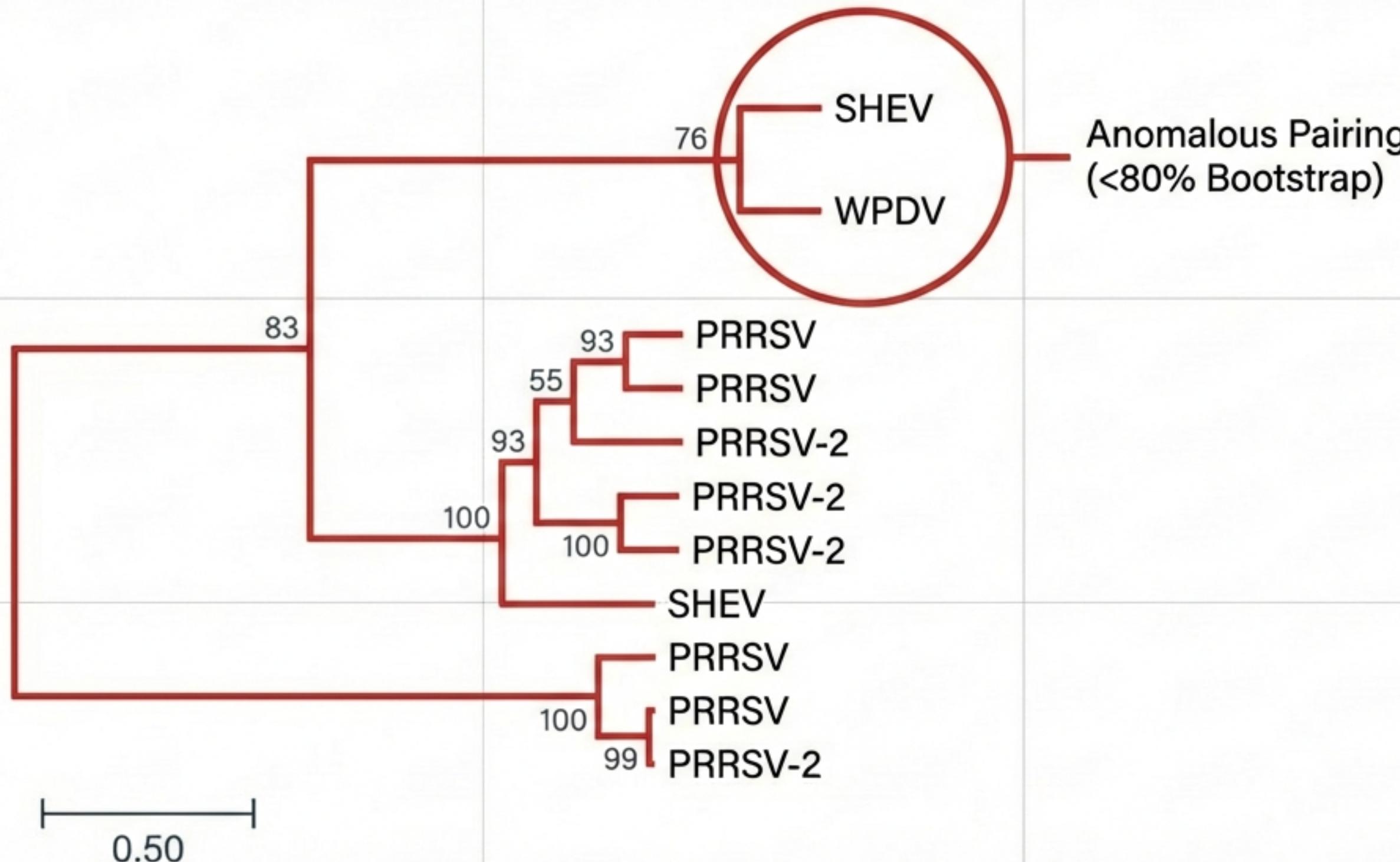
Observation: Six distinct genetic clusters.

Simarteriviruses: Form a strong monophyletic group.

The Outlier: WPDV appears as a sister clade to EAV, distantly related to all others.

Phylogenetics: Glycoprotein 5 (ORF5)

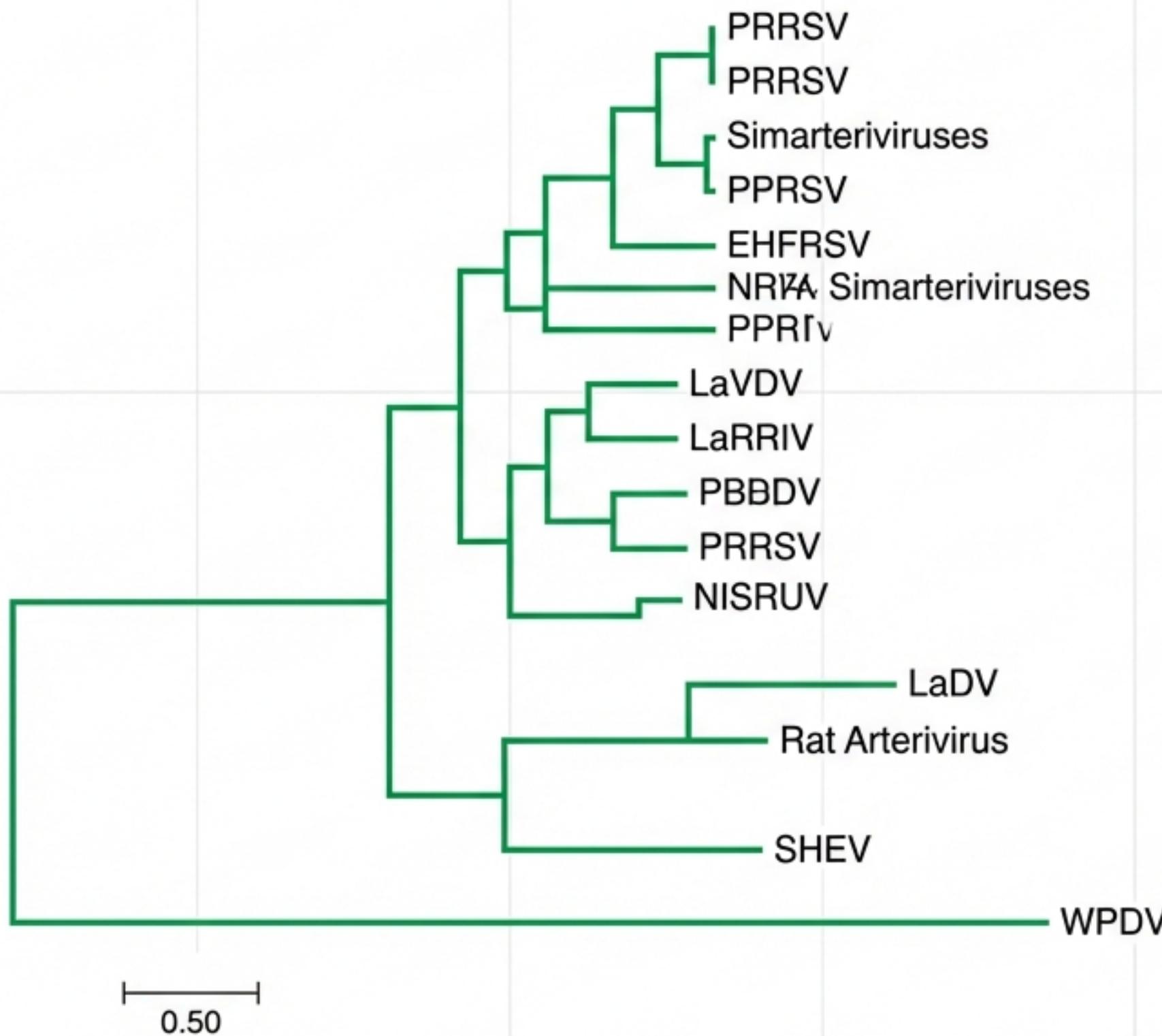
Topology shifts indicate divergent evolution.



- **A Shift in Relationship:** Topology changes compared to Replicase.
- **The Anomaly:** SHEV moves out of the Simarterivirus cluster to group with WPDV.
- **Implication:** Suggests potential divergent lineage or recombination events in the surface protein gene.

Phylogenetics: Nucleocapsid (ORF7)

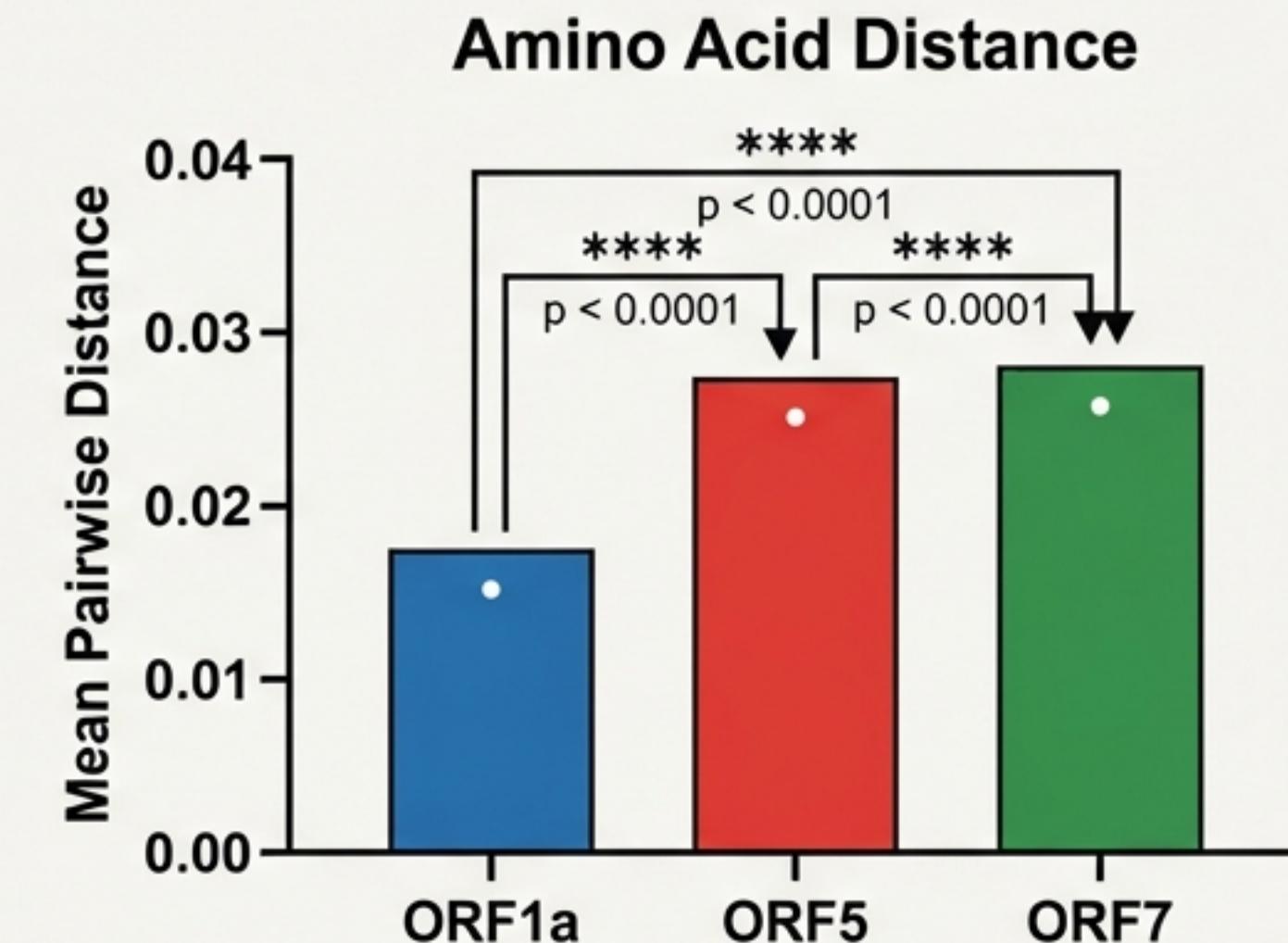
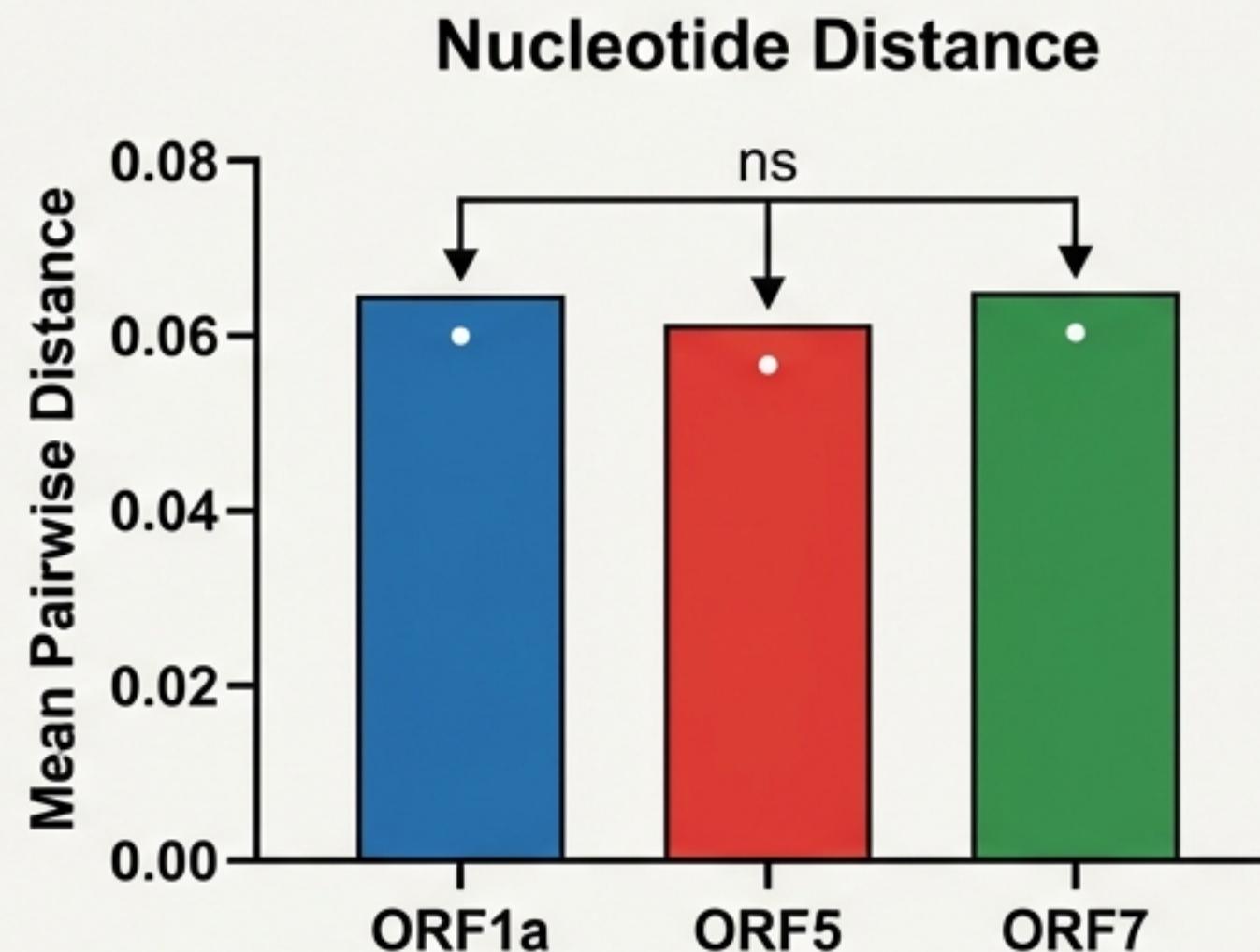
Confirming structural divergence.



- **Structural Divergence:** WPDV remains the most distantly related genus.
- **Consistent Pattern:** SHEV again tracks closer to WPDV than other Simarteriviruses.
- **Inference:** Trees show relationship, but to solve the hypothesis, we must quantify the exact genetic distance.

Quantifying Variability: Pairwise Distance

Amino acid analysis reveals hidden variation.



Nucleotide Analysis: No significant difference ($p > 0.05$).

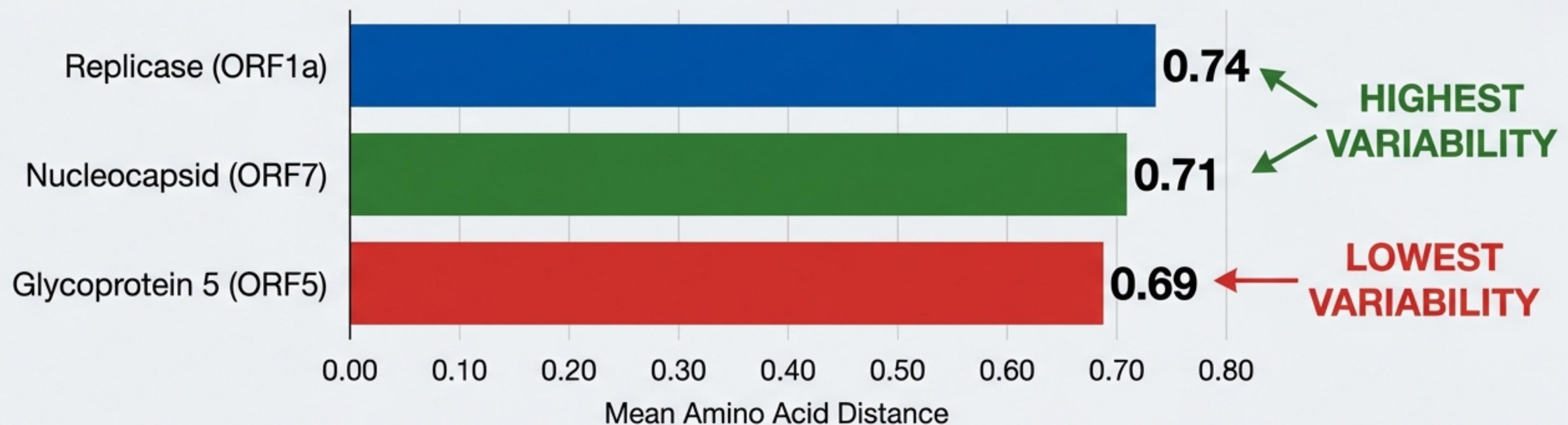
Amino Acid Analysis: Highly significant differences ($p < 0.0001$) between ORF1a/ORF7 and ORF5/ORF7. Variation is driven by protein structure changes.

The Twist: Overall Mean Distance Results

The Expectation

~~Hypothesis: GP5 > ORF7 > ORF1a~~

The Reality (Amino Acid Distance)

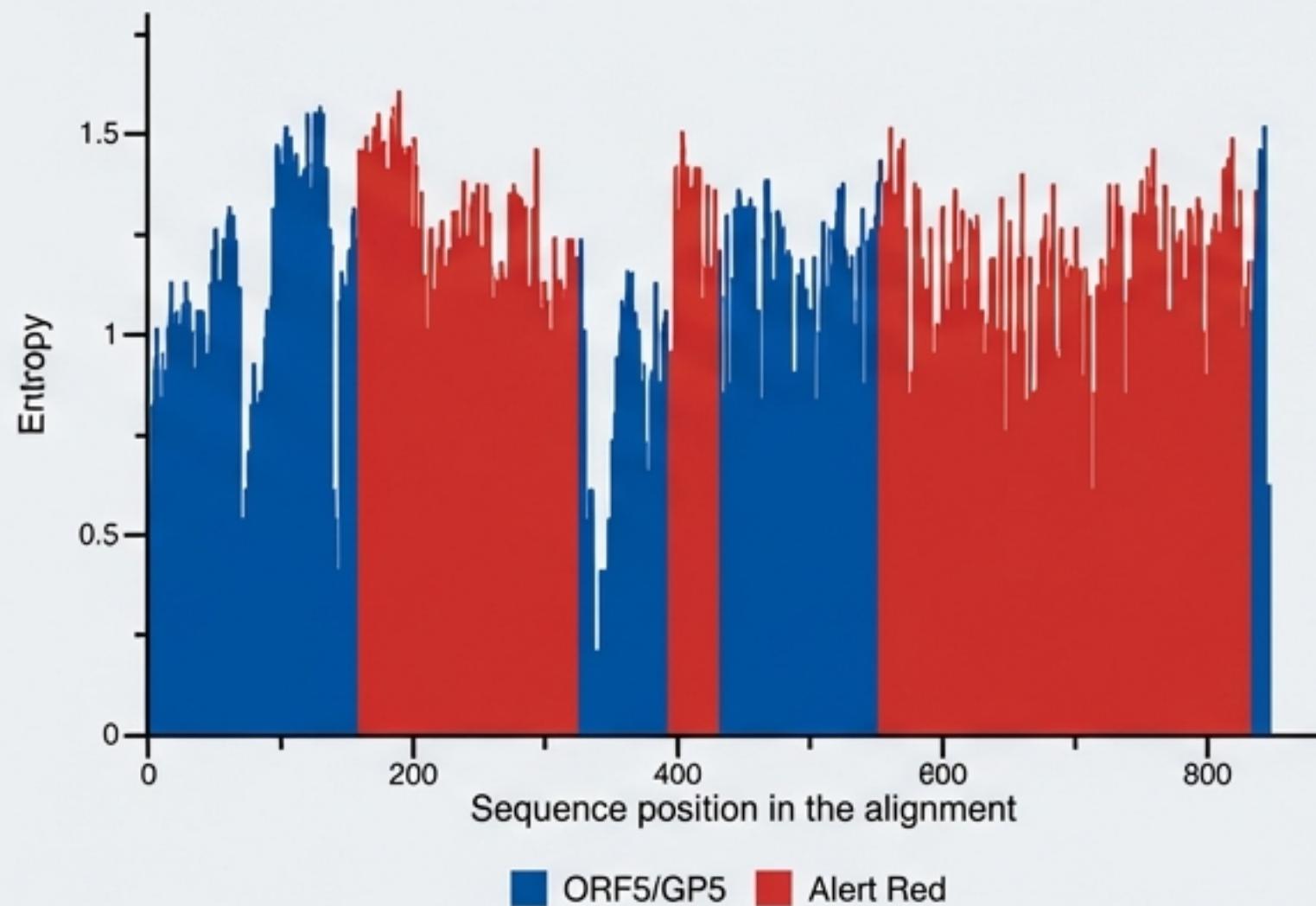


Verdict: The null hypothesis is invalid. The surface protein (GP5) is significantly less variable than the structural Nucleocapsid (N) protein.

Visualising Chaos: Shannon Entropy Plots

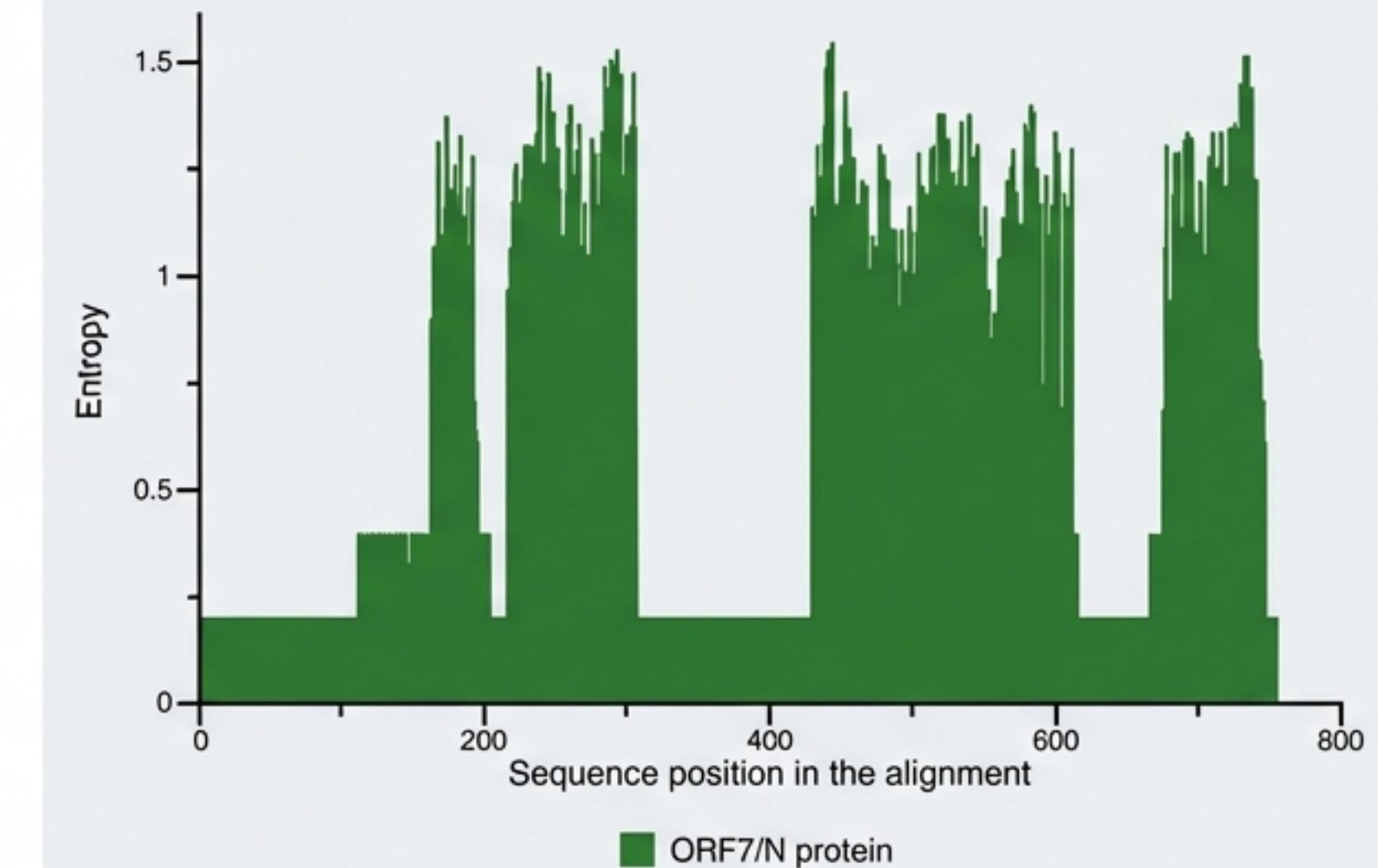
Comparing disorder across the alignment.

ORF5 (Glycoprotein 5)



Comparative conservation despite immune pressure.

ORF7 (Nucleocapsid)

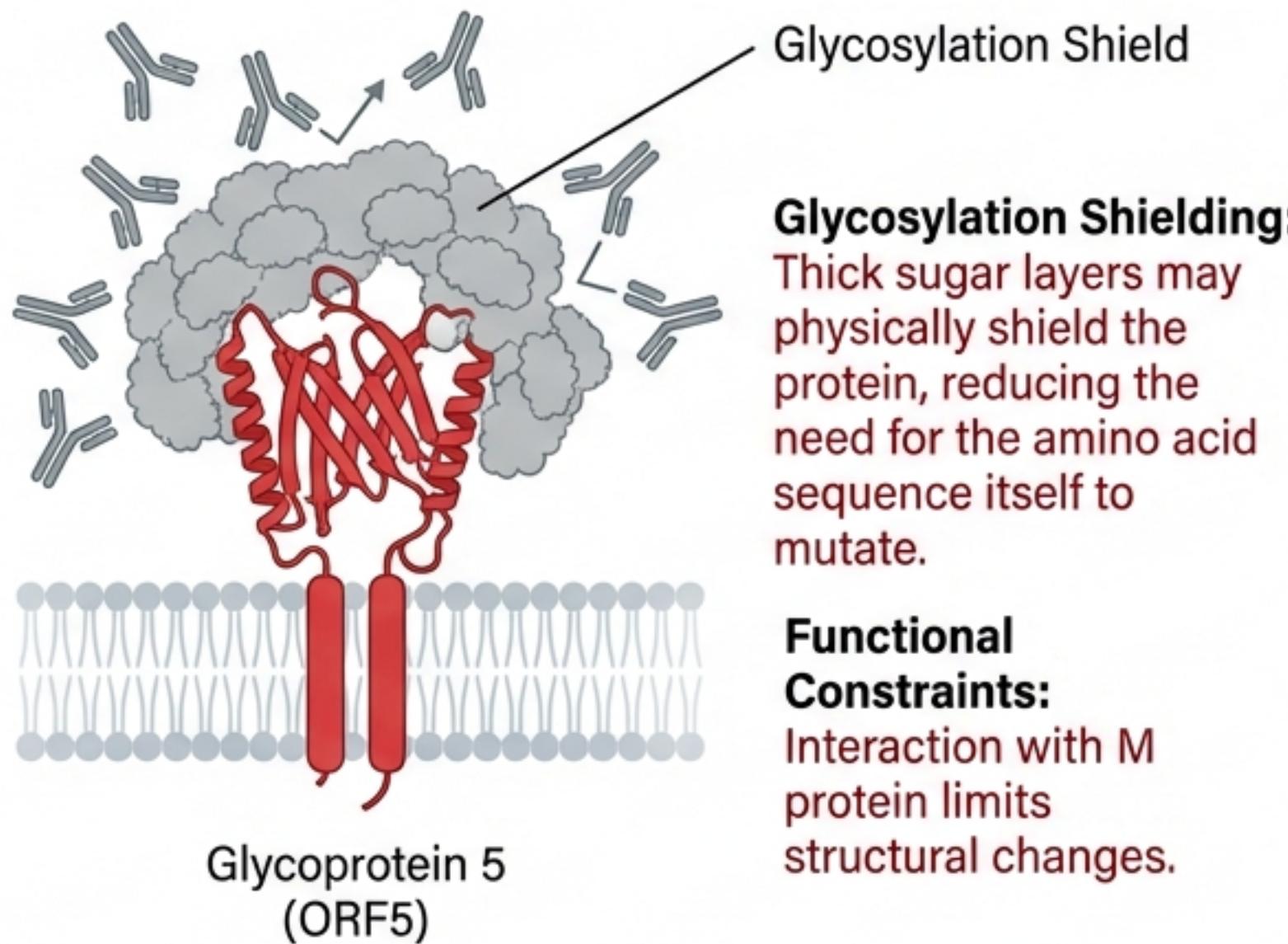


Sustained high-entropy regions indicating frequent substitution.

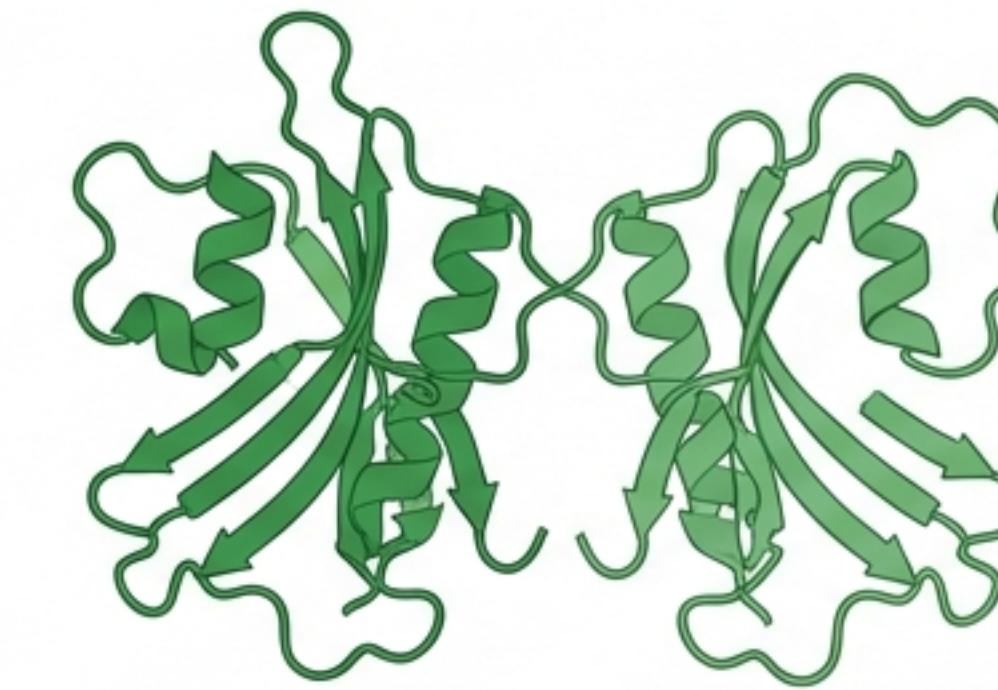
Interpreting the Inversion

Why is the 'shield' conserved while the 'core' mutates?

GP5 Conservation



N Protein Variability



Nucleocapsid (ORF7)

Tolerance for Mutation: The Nucleocapsid gene (Highest nucleotide distance 0.58) tolerates high mutation rates, suggesting structural flexibility in homodimer formation.

Conclusions & Implications

Key takeaways and future research directions based on the analysis of WPDV genetic variation.



Final Ruling: The Nucleocapsid (N) protein gene demonstrates greater genetic variation than the Glycoprotein 5 (GP5) gene, overturning the initial hypothesis.



Taxonomic Insight: Wobbly Possum Disease Virus (WPDV) represents a highly divergent lineage, challenging current classification boundaries.



Future Research: Understanding the stability of GP5 is critical for developing broad-spectrum vaccines targeting conserved regions.

