

# Using Evolution to Predict Sites of Critical Function in a Viral Pathogenesis Protein

Jeremy M. Brown<sup>1,2</sup>, Paul J.F. Rider<sup>3</sup>, Lyndon Coghill<sup>1,2</sup>, Misagh Naderi<sup>1</sup>, Michal Brylinski<sup>1</sup>, and Konstantin G. Kousoulas<sup>3</sup>

<sup>1</sup> Department of Biological Sciences, <sup>2</sup>Museum of Natural Science, and

<sup>3</sup>School of Veterinary Medicine, Louisiana State University



## Introduction

Alphaherpesviruses are a major cause of morbidity and mortality in humans and other animals, and include herpes simplex viruses (HSV) and varicella zoster virus (VZV, the causative agent of chicken pox; Fig. 1). Glycoprotein K (gK), conserved in all alphaherpesviruses, is a multi-membrane spanning virion protein essential for many virus functions. However, little is known about which gK domains and residues are most important. We used phylogenetic and structural analyses to predict functionally important parts of gK, employing a novel model for evolutionary rate variation across residues. We verified these predictions through mutagenesis and functional assays.

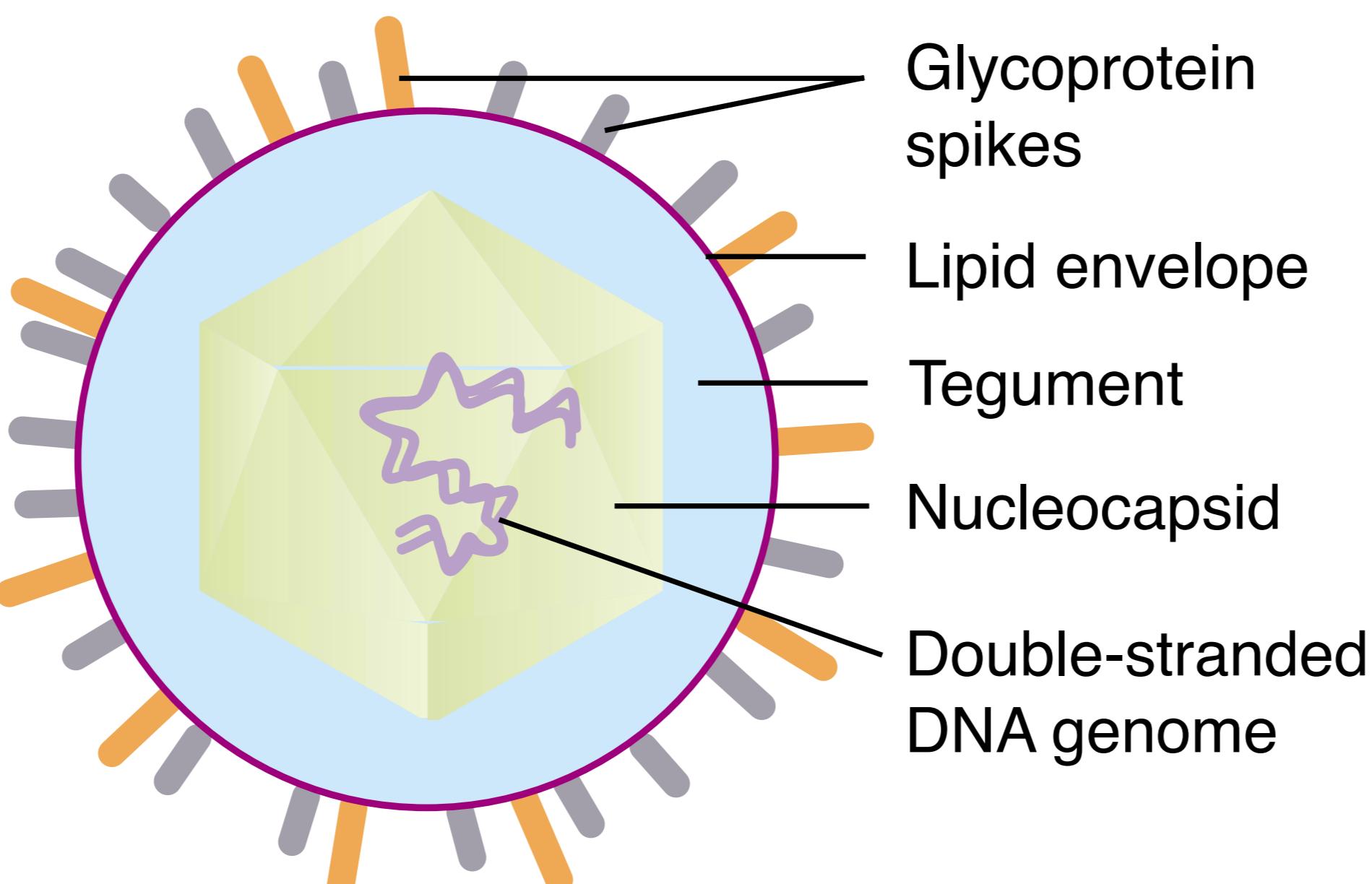


Fig. 1 - Structure of varicella zoster virus (VZV), the causative agent of chicken pox and shingles. Figure taken from Quinlivan and Breuer (2005).

## Novel Site-Specific Rate Model

We used a novel finite mixture model (FMM) where the number of rate categories ( $k$ ) is fixed *a priori*, although the assignment of sites to categories is random. Following Huelsenbeck and Suchard (2007), the assignment of sites to categories can be given as a vector of mappings  $\sigma = (\sigma_1, \dots, \sigma_n)$ , where  $n$  is the number of sites in the alignment and  $\sigma_i \in \{1, \dots, k\}$ . We explored two versions of this FMM: one where the prior probability of each site's assignment to each category is fixed and equal,  $P(\sigma_i=1, \dots, k) = 1/k$ , and another where a Dirichlet distribution is employed as a hyperprior on the probabilities of assignment,  $P(\sigma_i=1, \dots, k) \sim Dir(\alpha)$ . In the latter form,  $\alpha = (\alpha_1, \dots, \alpha_k)$ , where  $\alpha_i$  is the concentration parameter for category  $i$ . All of our analyses employed a flat Dirichlet with  $\alpha = (1, \dots, 1)$ .

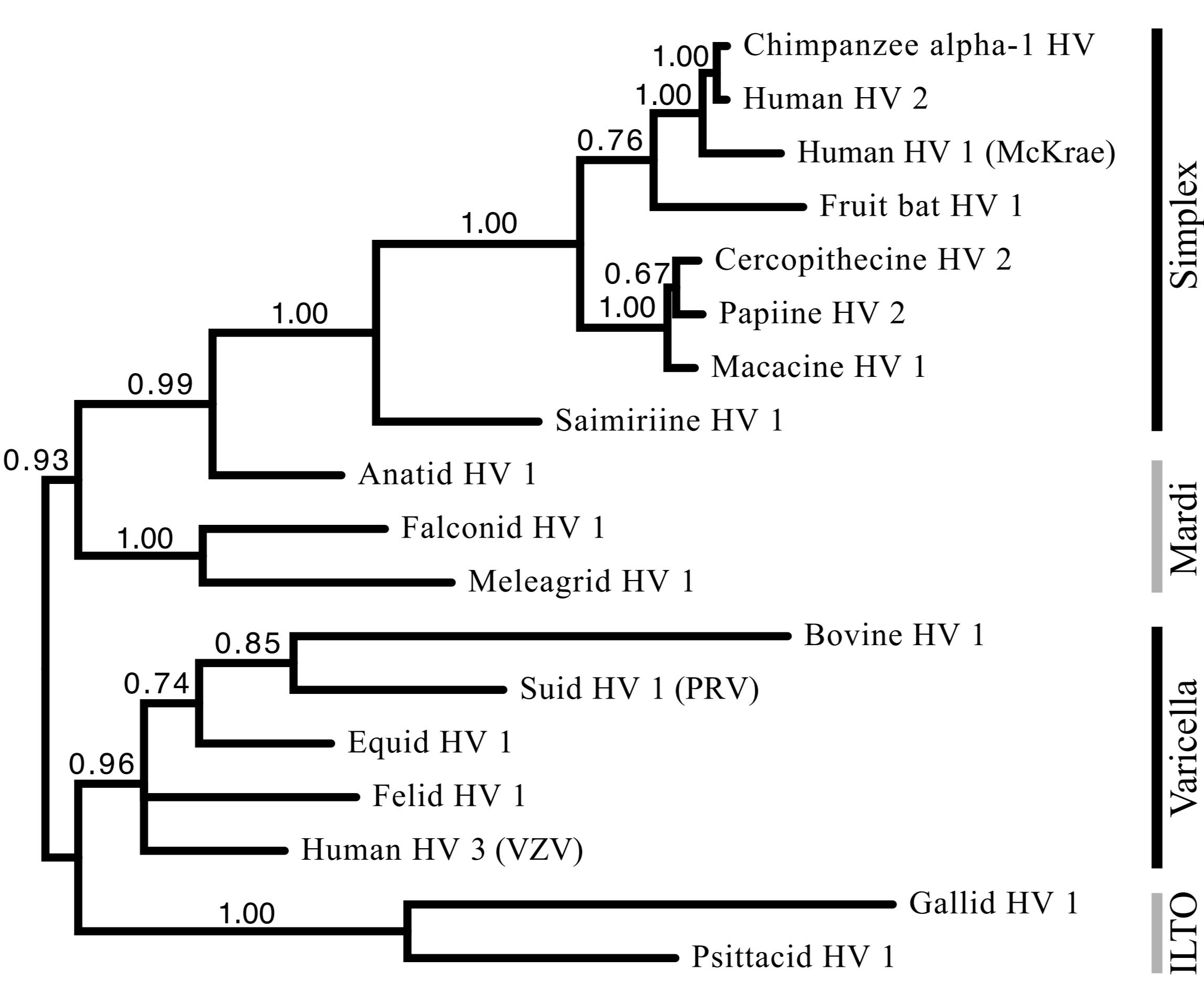


Fig. 2 - Phylogeny inferred from Bayesian phylogenetic analysis of gK orthologs from 18 alphaherpesviruses, arbitrarily rooted. Branches are labeled with posterior probabilities.

## Acknowledgments

This work was supported by NSF grants DBI-1262571 and DEB-1355071 to JMB. We thank Dr. Sebastian Höhna for help formulating the FMM in RevBayes.

## References

- Höhna S, Landis MJ, Heath TA, Boussau B, Lariolot N, Moore BR, Huelsenbeck JP, Ronquist F. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst Biol*. 65:726-736.
- Huelsenbeck JP, Suchard MA. 2007. A nonparametric method for accommodating and testing across-site rate variation. *Syst Biol*. 56:975-987.
- Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61:539-542.
- Quinlivan M, Breuer J. 2005. Molecular and therapeutic aspects of varicella-zoster infection. *Expert Reviews in Molecular Medicine*. 7:1-24.

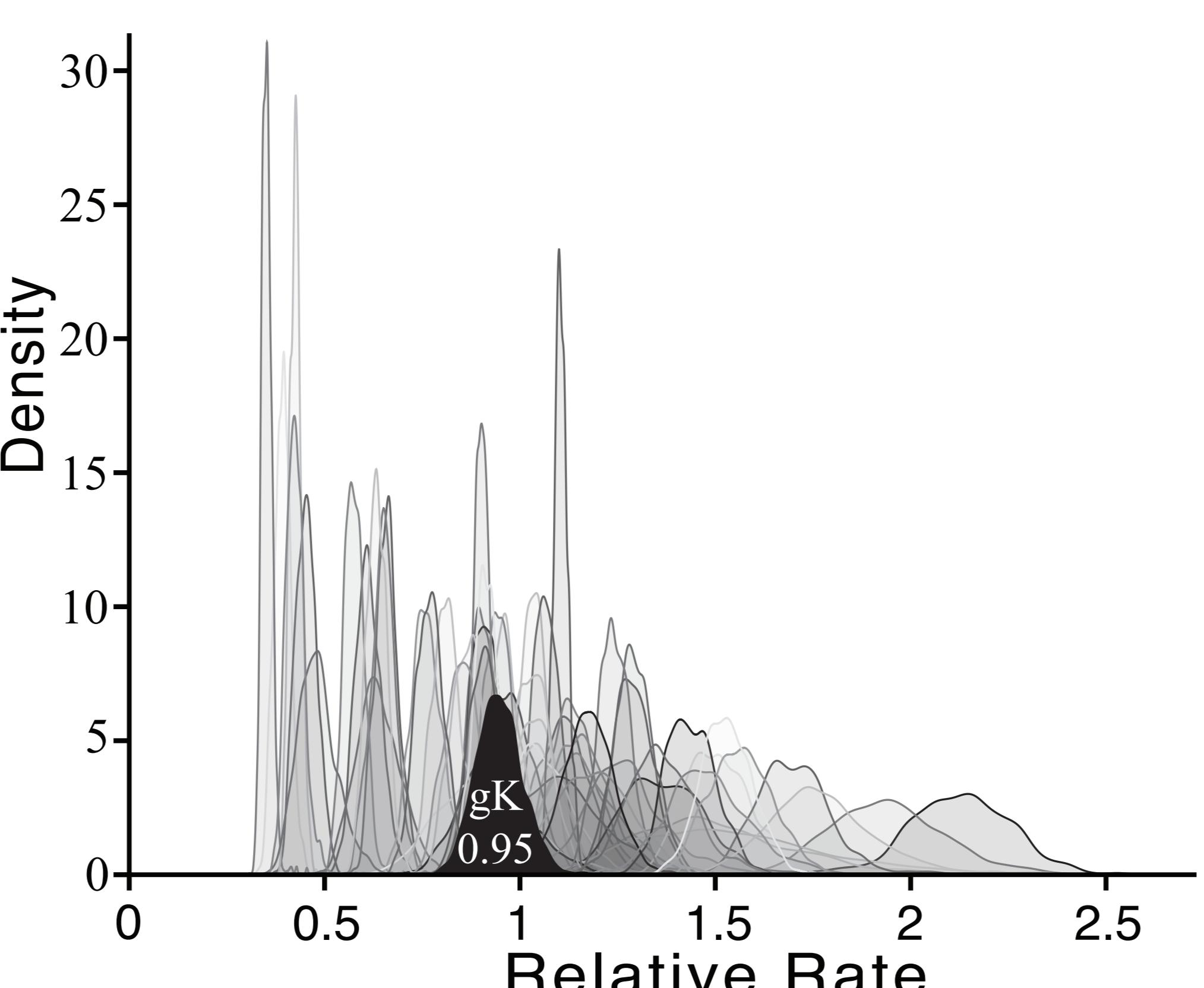


Fig. 3 - Relative rates of evolution for 59 genes from across the alphaherpesvirus genome, with gK highlighted for comparison.

## Results

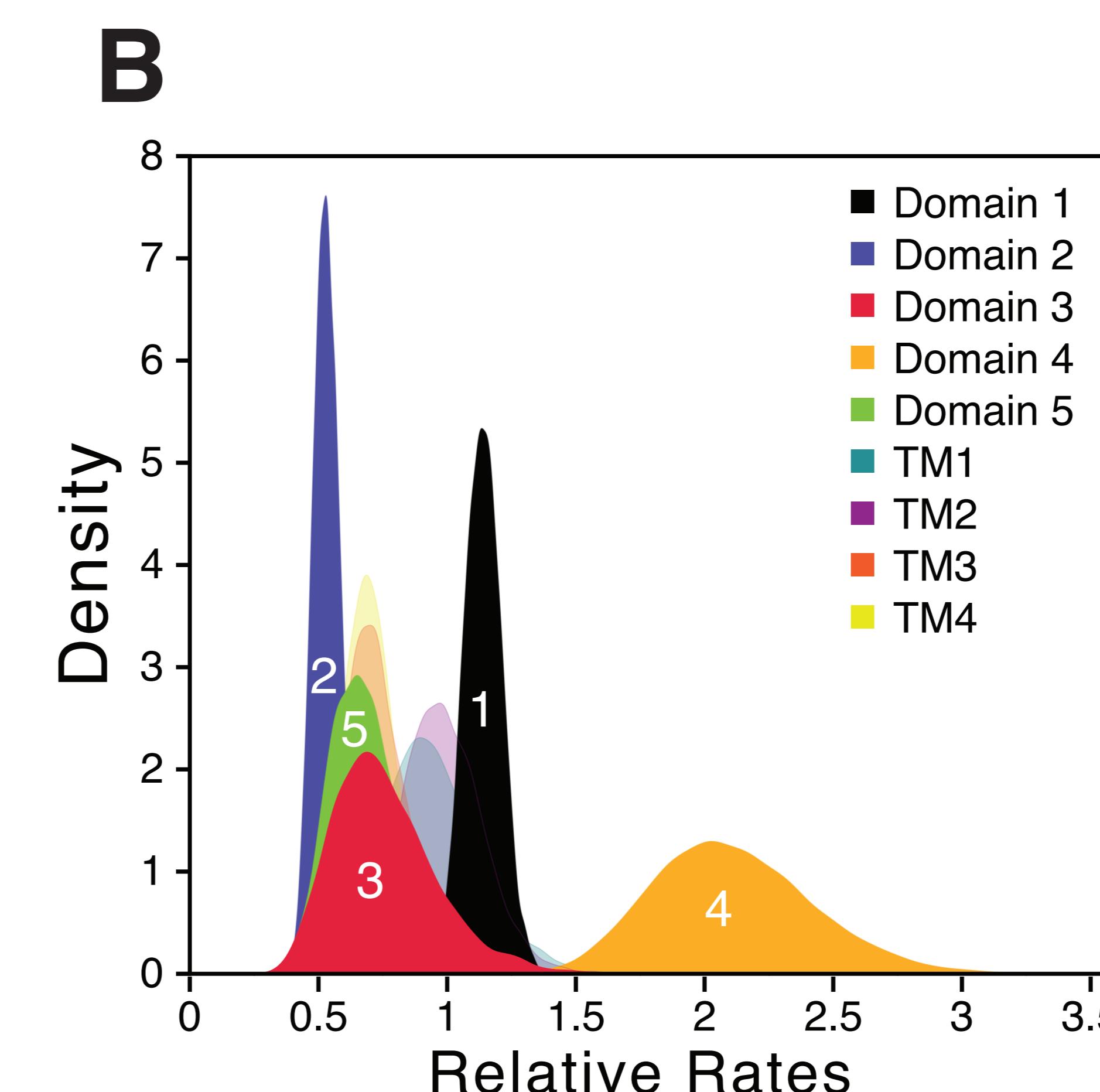
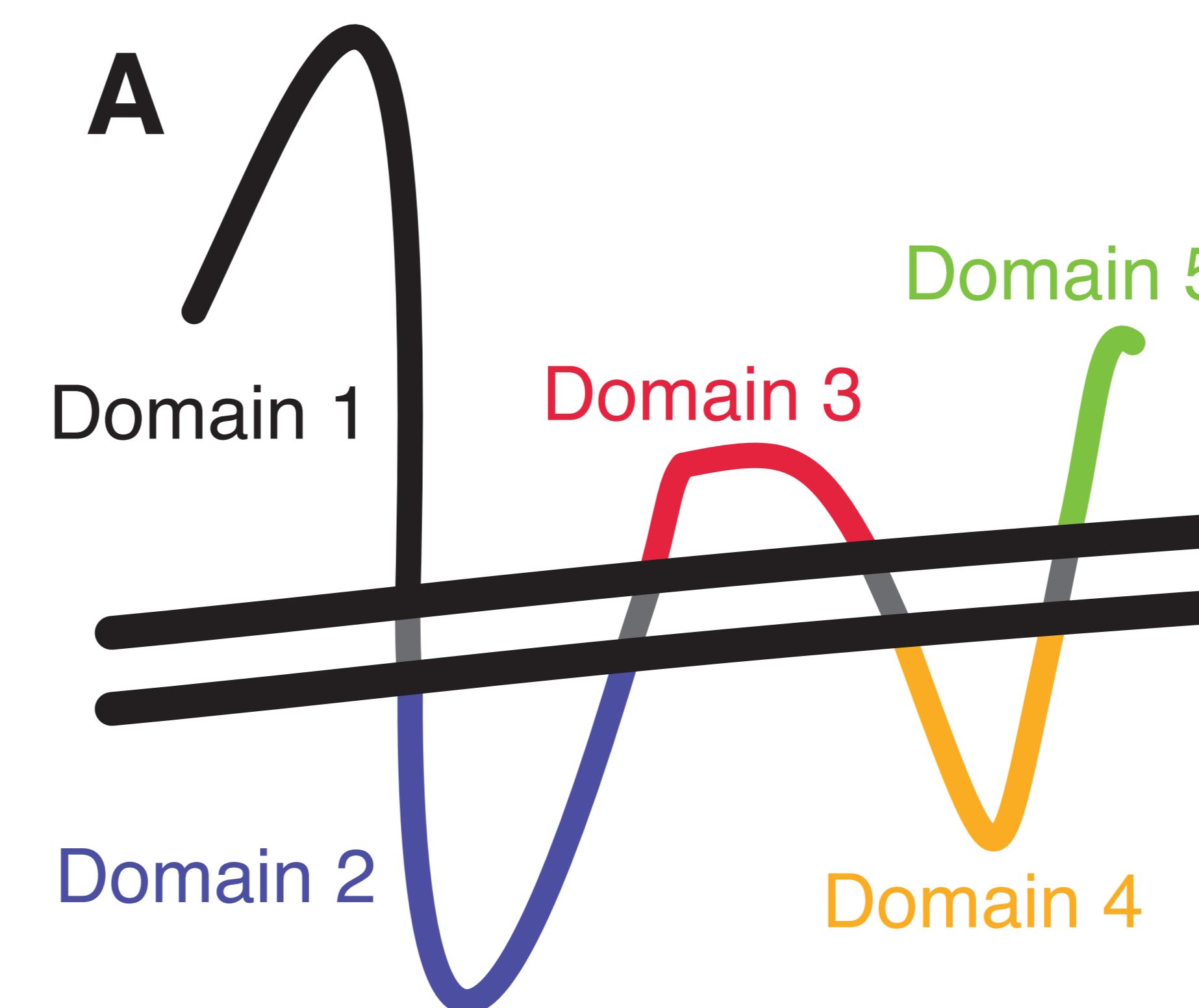


Fig. 4 - (A) A cartoon of gK's structure. (B) Relative rates of evolution for the 5 extramembrane and 4 transmembrane domains of gK.

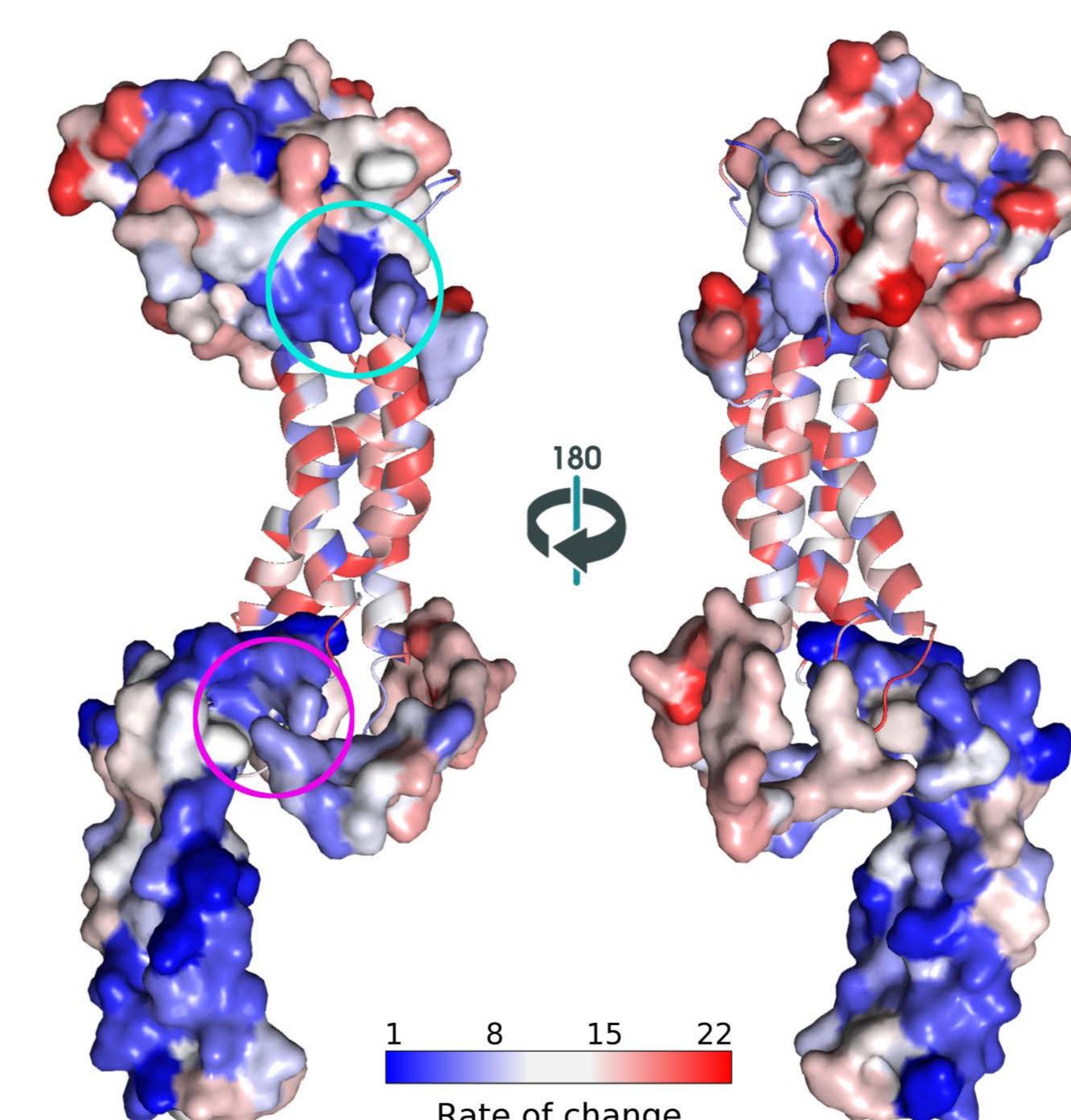


Fig. 5 - The predicted structure of HSV-1 gK with site-specific rates mapped onto residues. Cyan and magenta circles show residues with low rates of evolution at the interface between different domains.

## Questions and Methods

### Question 1: How quickly does gK evolve relative to the rest of the genome?

We inferred gene-specific rates in MrBayes (Ronquist et al., 2012) using a fixed-partition model for 59 genes from 8 viruses (Fig. 3).

### Question 2: How do rates vary across gK domains?

We inferred domain-specific rates in MrBayes using a fixed-partition model and sequences from 18 viruses (Figs. 2 and 4).

### Question 3: How do rates vary across gK residues?

We inferred site-specific rates across 18 viruses using a novel finite mixture model in RevBayes (Höhna et al., 2016) and mapped them on the predicted structure (Fig. 5).

### Question 4: Are slowly evolving residues functionally important?

We used mutagenesis and functional assays to determine whether changes to several slowly evolving residues in interesting structural positions produce viruses with impaired function (data not shown). Spoiler alert: they do!

## Conclusions

- Combined evolutionary and structural prediction can be a powerful tool for understanding molecular function.
- The finite mixture model for among-site rate variation is both quite general, yet also tractable.
- Constraint imposed by selection for function due to protein-protein interactions may be very site-specific. Broader scale rate variation may be determined by overall structural constraints.
- The average rate of evolution for a gene is not a good predictor of functional importance.

Find this poster and RevBayes code here:

