

Supplementary Information: Relative evolutionary rates in proteins are largely insensitive to the substitution model

Stephanie J. Spielman^{1*} and Sergei L. Kosakovsky Pond¹

Address:

¹Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA 19122.

*Corresponding author

Email: stephanie.spielman@temple.edu

Supplementary Figures

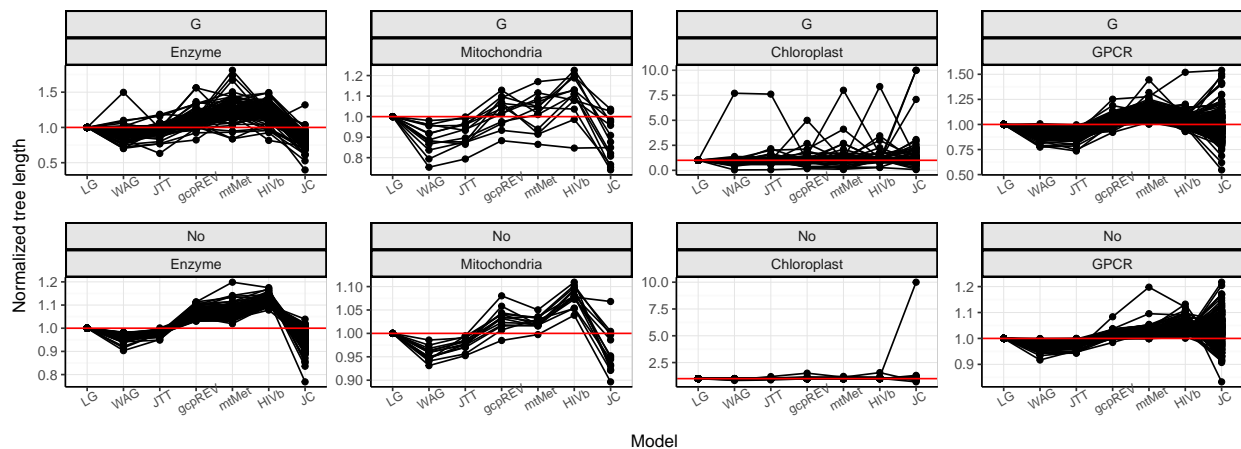


Figure S1 Optimized tree lengths by LEISR, relative to LG, across models and datasets. All values have been normalized to LG-inferred branch lengths, respectively for each rate variation setting, such that the all tree lengths under LG are equal to 1. Each line represents the optimized relative tree length for a given alignment. For visual clarity, relative tree lengths have been right-censored at 10. The top row shows relative tree lengths for inferences without rate variation, and the bottom row shows relative tree lengths for +*G* inferences.

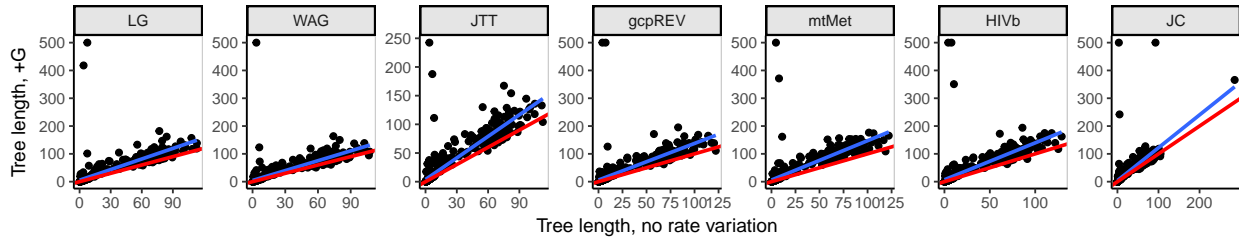


Figure S2 Relationship between LEISR-optimized tree lengths with and without + G , across models. Each point represents a given alignment across all datasets. The red line in each panel represents the line $x = y$, and the blue line in each panel represents the linear regression line. In general, optimized + G tree lengths are larger compared to optimized tree lengths without rate variation. This trend becomes increasingly pronounced as overall tree length increases. For visual clarity, tree lengths have been right-censored at 500.

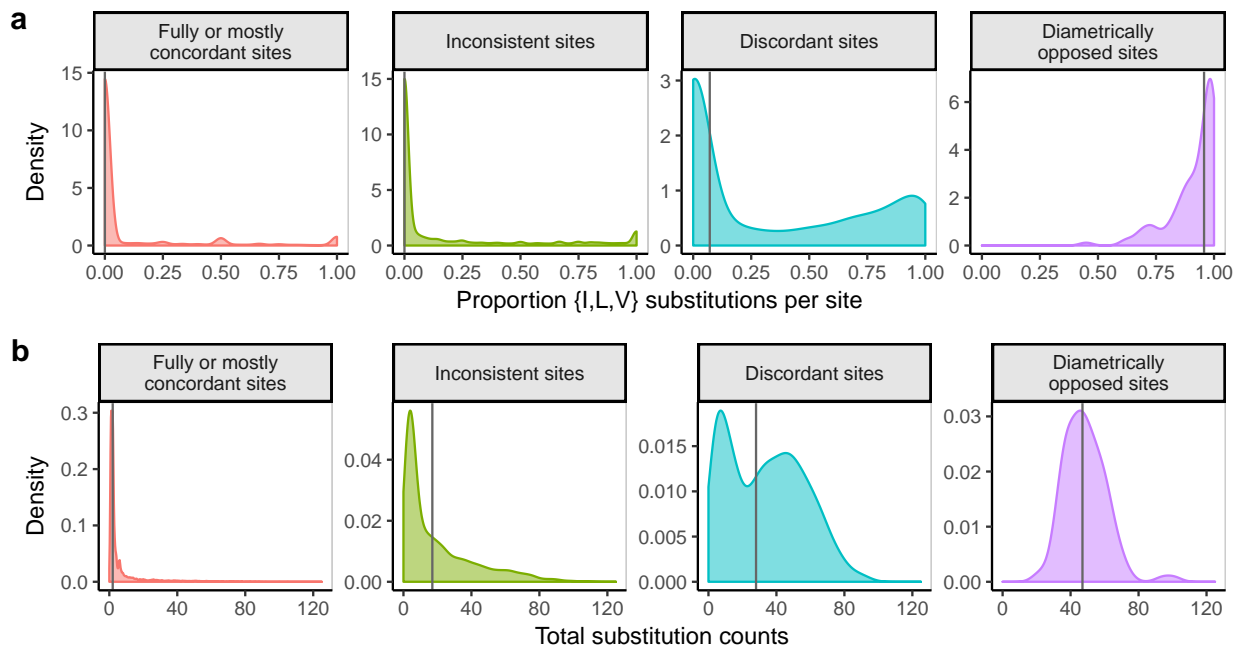


Figure S3 Distribution of substitution counts across sites. a) Proportion of ILV among all sites, across site classifications. Only sites which had experienced at least one substitution are shown. The vertical line in each panel represents the median proportion of substitutions among I, L, and V. b) Total number of substitutions across site classifications, again only considering which had experienced at least one substitution. The vertical line in each panel represents the median total number of substitutions. Results in this figure were obtained using substitution counts under the JC model.