## Supplementary Information for "A comparison of one-rate and two-rate inference frameworks for site-specific dN/dS estimation"

Stephanie J. Spielman<sup>1\*</sup>, Suyang Wan<sup>2</sup>, and Claus O. Wilke<sup>1</sup>

Email: stephanie.spielman@gmail.com

## Supplementary Figures

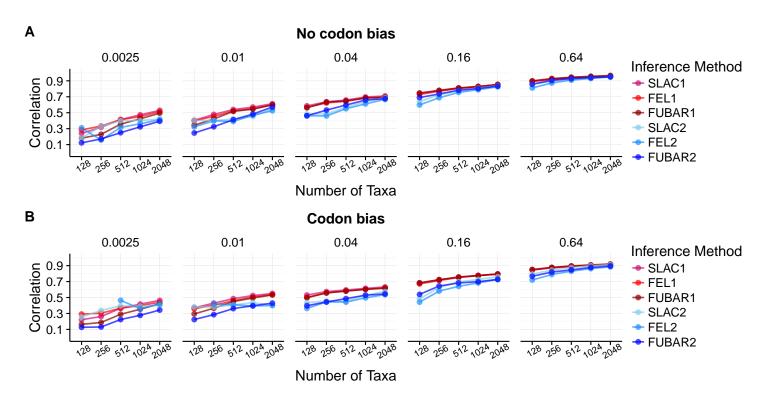


Figure S1 Pearson correlation between true and inferred dN/dS across methods and simulation conditions. Data shown in this figure corresponds to  $\Pi_{\text{equal}}$  simulations. A) Correlation between true and inferred dN/dS for simulations with no codon bias. B) Correlation between true and inferred dN/dS for simulations with codon bias. Note that certain FEL2 points ( $B = 0.0025/N \in \{128, 256\}$  and B = 0.01/N = 128, for codon bias simulations) are not present because FEL2 generally failed to converge under these conditions.

<sup>&</sup>lt;sup>1</sup>Department of Integrative Biology, Center for Computational Biology and Bioinformatics, and Institute of Cellular and Molecular Biology. The University of Texas at Austin, Austin, TX 78712, USA.

<sup>&</sup>lt;sup>2</sup>School of Physics and Astronomy, The University of Minnesota, Minneapolis, MN 55455, USA.

<sup>\*</sup>Corresponding author

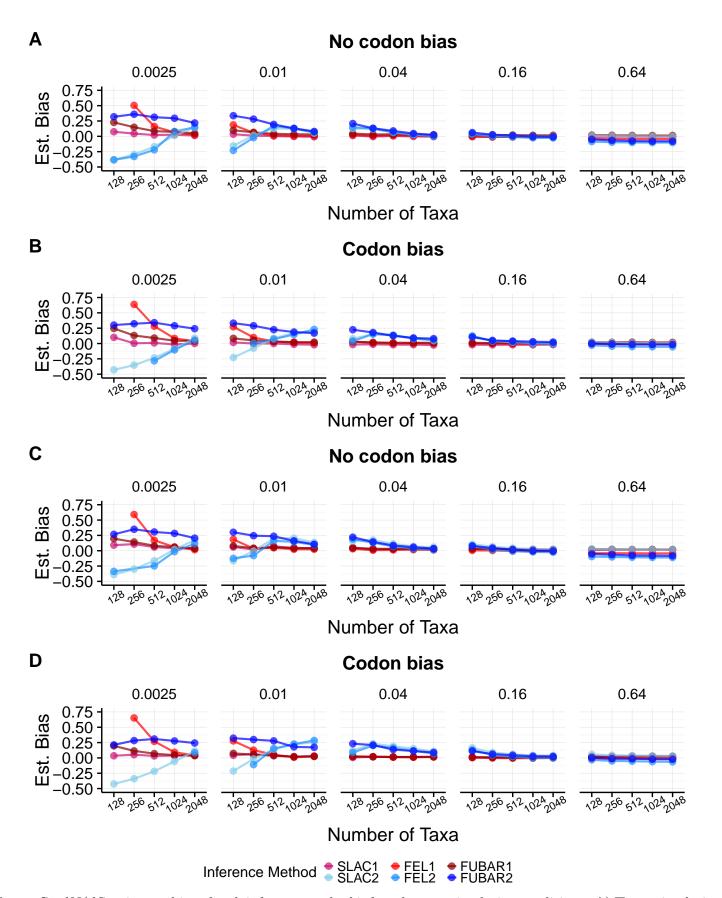


Figure S2 dN/dS estimator bias of each inference method inferred across simulation conditions. A)  $\Pi_{\text{equal}}$  simulations with no codon bias. B)  $\Pi_{\text{equal}}$  simulations with codon bias and equal nucleotide frequencies. C)  $\Pi_{\text{unequal}}$  simulations with no codon bias. D)  $\Pi_{\text{unequal}}$  simulations with codon bias. Note that certain FEL1 and FEL2 points are not present in the figure for conditions where the methods generally failed to converge.

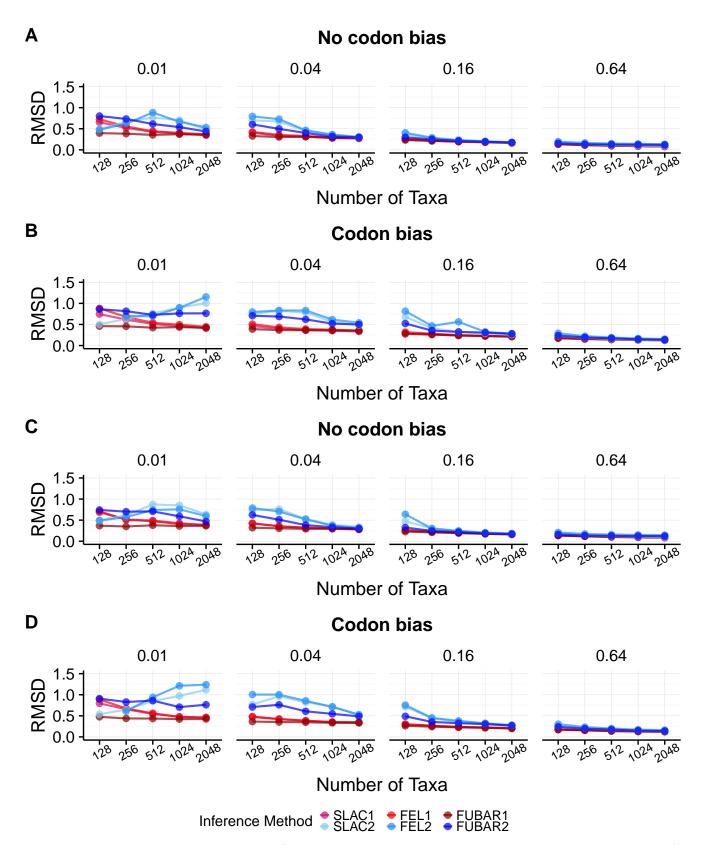


Figure S3 RMSD of inferred from true dN/dS across inference methods and simulation conditions. A)  $\Pi_{\text{equal}}$  simulations with no codon bias. B)  $\Pi_{\text{equal}}$  simulations with codon bias and equal nucleotide frequencies. C)  $\Pi_{\text{unequal}}$  simulations with no codon bias. D)  $\Pi_{\text{unequal}}$  simulations with codon bias. Note that certain FEL2 points are not present in the figure for conditions where the methods generally failed to converge.

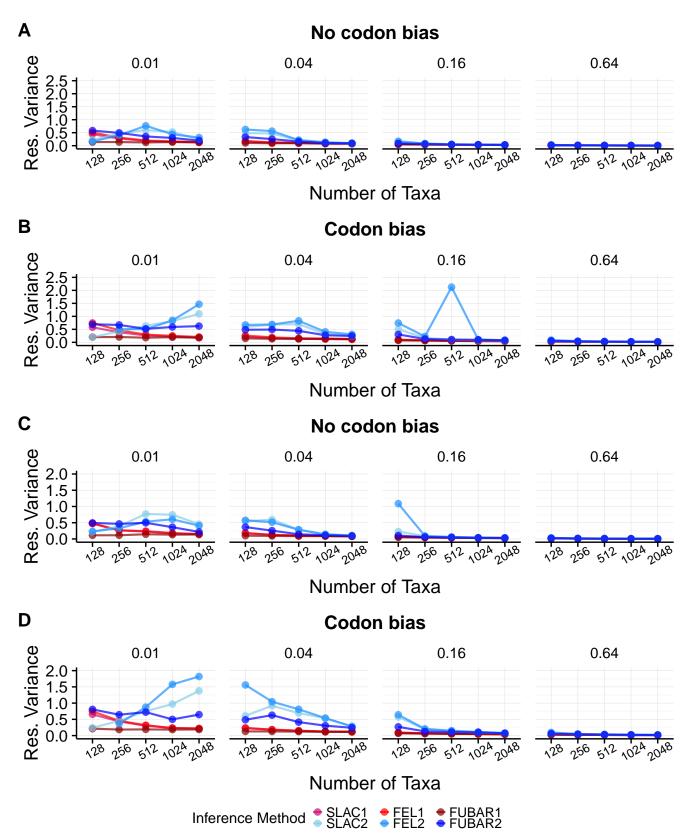


Figure S4 Variance in residuals for a linear model of inferred dN/dS regressed on true dN/dS across inference methods and simulation conditions. A)  $\Pi_{\text{equal}}$  simulations with no codon bias. B)  $\Pi_{\text{equal}}$  simulations with codon bias and equal nucleotide frequencies. C)  $\Pi_{\text{unequal}}$  simulations with no codon bias. D)  $\Pi_{\text{unequal}}$  simulations with codon bias. Note that certain FEL2 points are not present in the figure for conditions where the methods generally failed to converge.

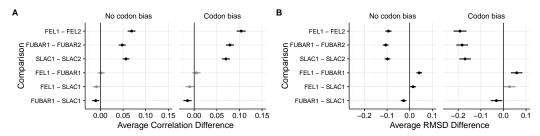


Figure S5 Pairwise comparisons of correlation strength and RMSD inference across methods, as determined through multiple comparisons tests. Results shown correspond to  $\Pi_{\text{equal}}$  simulations. A) Results for multiple comparison tests of correlation strength. B) Results for multiple comparison tests of RMSD.

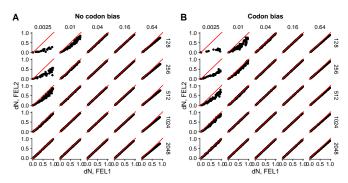


Figure S6 dN estimates are nearly equivalent between one-rate and two-rate frameworks. Each panel compares the point dN estimate for each simulation condition, with branch lengths from left to right and number of taxa from top to bottom. X-axes indicate dN estimates with FEL1, and Y-axes indicate dN estimates with FEL2. The line in each panel is x = y. Results correspond to  $\Pi_{\text{unequal}}$  simulations.

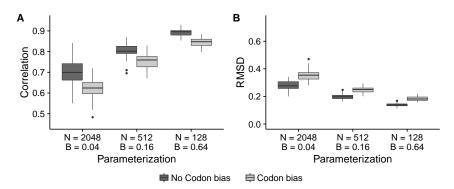
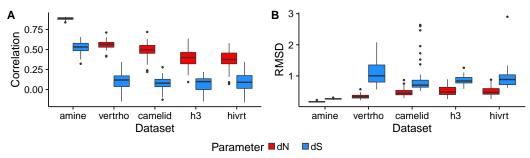


Figure S7 Divergence is more important than the number of sequences is for obtaining the equilibrium dN/dS value. Boxplots represent either A) correlation or B) RMSD, from FEL1 inference, across the 50 respective  $\Pi_{\text{equal}}$  simulation replicates.



**Figure S8** A) Correlation and B) RMSD for individual dN and dS parameter estimates with FUBAR2, for simulations with codon bias performed along empirical phylogenies.