

BUSTED analysis - summary

- **West Nile Virus NS3 protein**

- No statistical support for selection; ML point estimate allocates a small proportion of sites ($\sim 1\%$) to the selected group ($dN/dS \sim 2$)
- The rest of the gene is very strongly conserved ($dN/dS = 0.004$)

- **HIV-1 transmission pair**

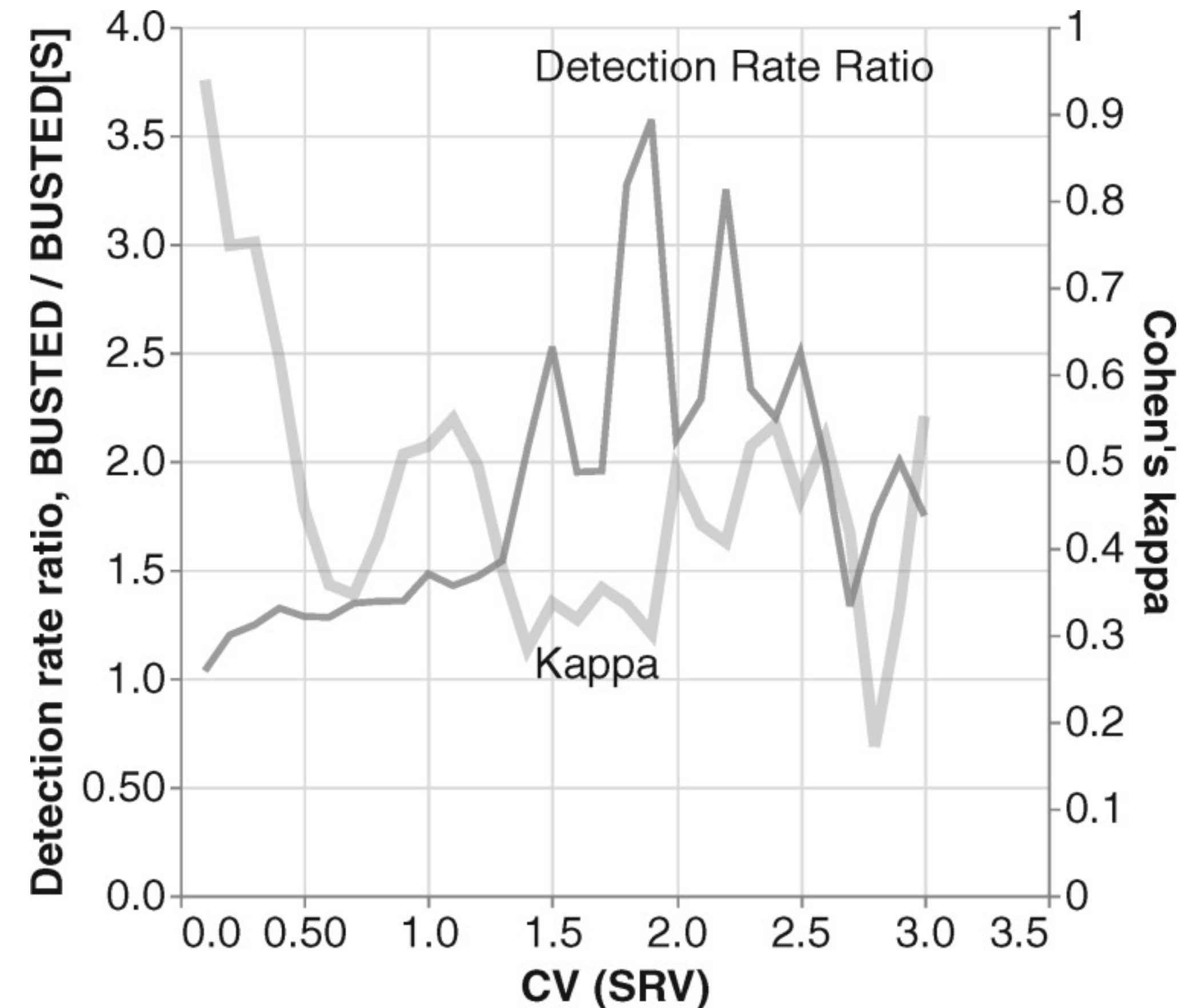
- Very strong evidence of strong episodic diversification ($dN/dS \sim 100$) on a small proportion of sites (2%)
- The rest of the gene evolves with weak purifying selection ($dN/dS = 0.6-0.7$)

- **SARS-CoV-2 spike**

- Evidence of episodic diversification ($dN/dS \sim 15$) on a small proportion of sites ($\sim 2\%$)
- Most of the rest of the gene evolves with purifying selection ($dN/dS = 0.2$)

Synonymous rate variation

- **dS** = constant for all sites (assumed by many models); this assumption appears to be nearly universally violated in biological data, due to e.g. secondary structure, localized codon usage bias, overlapping reading frames, etc.
- This can lead to, e.g. incorrect identification of relaxed constraint as selection and high false positive rates
- Most of HyPhy methods provide support for including dS



Synonymous Site-to-Site Substitution Rate Variation Dramatically Inflates False Positive Rates of Selection Analyses: Ignore at Your Own Peril

Sadie R Wisotsky^{1 2}, Sergei L Kosakovsky Pond², Stephen D Shank², Spencer V Muse^{1 3}

Affiliations + expand

PMID: 32068869 PMCID: [PMC7403620](#) DOI: [10.1093/molbev/msaa037](#)

[Free PMC article](#)