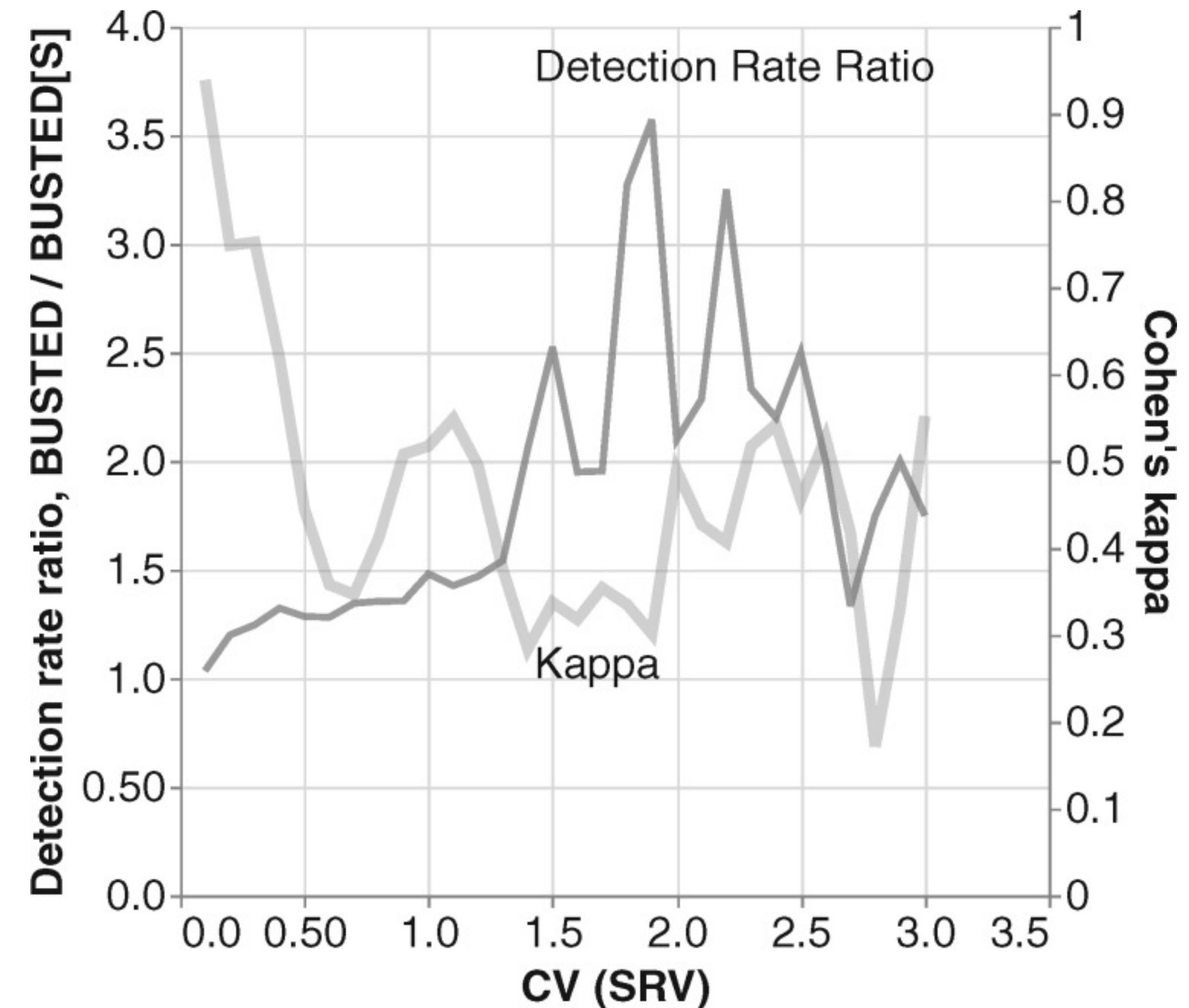


Synonymous rate variation

- **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](https://pubmed.ncbi.nlm.nih.gov/PMC7403620/) DOI: [10.1093/molbev/msaa037](https://doi.org/10.1093/molbev/msaa037)

[Free PMC article](#)

Allowing multi-nucleotide substitutions

- Some of the methods (e.g. BUSTED, aBSREL, RELAX) can extend substitution models to allow instantaneous double- and triple-“hits” (e.g. ACC to AGG)
- Sometimes multi-nucleotide changes along short branches at a single site can drive selection signal (possible false positives?)
- HyPhy includes a simple standard analysis for estimating alignment-wide multiple-hit rates.

Extra base hits: Widespread empirical support for instantaneous multiple-nucleotide changes

Alexander G. Lucaci , Sadie R. Wisotsky , Stephen D. Shank, Steven Weaver, Sergei L. Kosakovsky Pond 

Published: March 12, 2021 • <https://doi.org/10.1371/journal.pone.0248337>

[See the preprint](#)