Gene-wide selection analysis using a branch-site method (BUSTED), HIV-1 *env*

```
hyphy busted --srv No --alignment
data/HIV-sets.nex --starting-
points 5
```

Produces *HIV-sets.nex.BUSTED.json* file
View in http://vision.hyphy.org/BUSTED or
https://observablehq.com/@spond/busted

Based on the likelihood ratio test, there **is** evidence of *episodic diversifying selection* in this dataset (p=0.000).

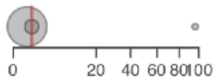
BUSTED analysis (v4.0) was performed on the alignment from /Users/sergei/Dropbox/Talks/VEME-current/data/HIV-sets.fas using HyPhy v2.5.40. This analysis **did not include** site-to-site synonymous rate variation.

Suggested citation: *Gene-wide identification of episodic selection*, Mol Biol Evol. 32(5):1365-71, *Synonymous Site-to-Site Substitution Rate Variation Dramatically Inflates False Positive Rates of Selection Analyses: Ignore at Your Own Peril*, Mol Biol Evol. 37(8):2430-2439

Evidence ratio threshold 10 Update

16 sequences in the alignment	**288** codon sites in the alignment	**1** partitions
26 median branches/partition used for testing	**3 classes** non-synonymous rate variation	**None** synonymous rate variation
0.0 p-value for episodic diversityfing selection	**13** Sites with ER≥10 for positive selection	**N/A:N/A** Multiple hit rates (2H:3H)

Alignment-wide results

Model	Log (L)	AIC-c	Params.	Rate distribution	Rate plot
Unconstrained model	−2039.96	4170.83	45	**Tested ω** 0.5596 (86.941%) 0.9885 (10.960%) 96.09 (2.0981%) Mean = **2.611**, CoV = **5.242**	
Constrained model	−2078.31	4245.48	44	**Tested ω** 1.000 (14.819%) 1.000 (20.229%) 1.000 (64.952%) Mean = **1.000**, CoV = **NaN**	

Murrell et al | Mol. Biol. Evol | 32(5) | 1365–1371

BUSTED inference

- Because BUSTED is a random-effects method, it **pools** information across multiple sites and branches to gain power
- The cost to this pooling is lack of site-level **resolution**, i.e., it is not immediately obvious which sites and/or branches drive the signal
- Standard ways to extract individual site contributions to the overall signal is to perform a post-hoc analysis, such as empirical Bayes, or “category loading”
- For BUSTED, “category loading” is faster and experimentally better
- Can also compute exploratory evidence for selection support along individual branches at specific sites

Murrell et al | Mol. Biol. Evol | 32(5) | 1365–1371