## Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection

Martin D. Smith, <sup>1</sup> Joel O. Wertheim, <sup>2</sup> Steven Weaver, <sup>2</sup> Ben Murrell, <sup>2</sup> Konrad Scheffler, <sup>2,3</sup> and Sergei L. Kosakovsky Pond\*, <sup>2</sup> Mol. Biol. Evol. 32(5):1342–1353

- Fix the tree; estimate and fix some of the nuisance model parameters that are shared by all branches (GTR biases, frequency counts)
- Fit a simple baseline model (one ω per branch); use this model to get initial guesses for all other parameters
- Perform a greedy step-up procedure (like forward variable selection in regression models, but not as statistically bad)

- For each branch (longest first) try two  $\omega$  rate classes, then three  $\omega$  rate classes etc, until no more goodness-of-fit improvement (AIC-c)
- Fix the number of rates and move on to the next longest branch
- Perform selection testing on the overall model (different number of  $\omega$  classes on branches), using the likelihood ratio test
- Each branch specified a priori (could be all branches)
- Appropriate multiple testing correction

## HIV-1 env

## adaptive Branch Site REL results summary

INPUT DATA | HIV-sets.fas | 16 sequences | 288 sites



aBSREL found evidence of episodic diversifying selection on 3 out of 26 branches in your phylogeny.

A total of **26** branches were formally tested for diversifying selection. Significance was assessed using the Likelihood Ratio Test at a threshold of  $p \le 0.05$ , after correcting for multiple testing. Significance and number of rate categories inferred at each branch are provided in the **detailed results** table.

See here for more information about this method.

Please cite PMID 25697341 if you use this result in a publication, presentation, or other scientific work.

## Tree summary

ω rate classes	# of branches	% of branches	% of tree length	# under selection
1	21	81%	0.49%	0
2	5	19%	100%	3

This table contains a summary of the inferred aBSREL model complexity. Each row provides information about the branches that were best described by the given number of  $\omega$  rate categories.

hyphy absrel --alignment data/HIV-sets.nex