

# 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

- Best-in-class power
- Able to detect episodes of selection, not just selection on average at a branch
- Does not make unrealistic assumptions for tractability, improves statistical behavior
- Sample size is ~sites, branch level rate estimates could be imprecise
- Cannot reliably estimate which individual sites are subject to selection
- Exploratory testing of all branches leads to loss of power for large data sets (multiple test correction)

# 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  $\omega$  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