

- The simple fix here is to include an **explicit error component** in the model
- We simply allow a small fraction of the alignment (e.g.  $\leq 1\%$ ) to evolve with abiologically high rates (e.g.  $dN/dS \geq 100$ )
- This is in addition to the standard model which allows negative,

neutral, and positive selection regimes.

- The primary goal of this analysis is to classify the selective regime on a gene (in the presence of specific types of errors)
- So error detection (and filtering) is a byproduct of an already useful analysis

---

## Calculation of dN/dS

Multiple-sequence alignments from each gene family were back-translated into codon alignments to reconstruct phylogenetic trees using FastTree2 with default parameters. The entire workflow was executed using ETE3 (ref. [64](#)) with options `ete3 build --nt-switch-threshold 0.0 --noimg --clearall --nochecks -w clustalo_default-none-none-none --no-seq-rename`. For calculation of selective pressure per family we ran HyPhy using the BUSTED model<sup>[65](#)</sup> with default parameters, codon-based nucleotide alignment and the phylogenetic tree generated previously, retrieving the dN/dS ratio under the full codon model. We discarded gene families with dN/dS values higher than 0.5.

Article | [Open access](#) | [Published: 18 December 2023](#)

## Functional and evolutionary significance of unknown genes from uncultivated taxa

[Álvaro Rodríguez del Río](#), [Joaquín Giner-Lamia](#), [Carlos P. Cantalapiedra](#), [Jorge Botas](#), [Ziqi Deng](#), [Ana Hernández-Plaza](#), [Martí Munar-Palmer](#), [Saray Santamaría-Hernando](#), [José J. Rodríguez-Herva](#), [Hans-Joachim Ruscheweyh](#), [Lucas Paoli](#), [Thomas S. B. Schmidt](#), [Shinichi Sunagawa](#), [Peer Bork](#), [Emilia López-Solanilla](#), [Luis Pedro Coelho](#) & [Jaime Huerta-Cepas](#) 