## 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 65 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.

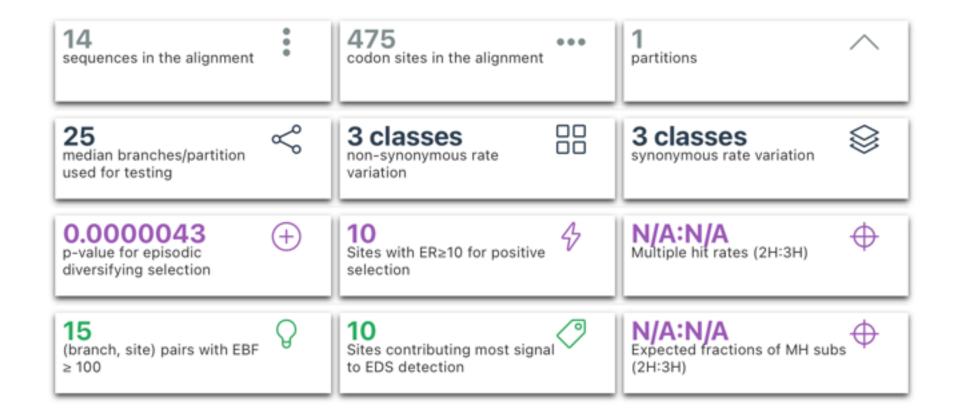
Article Open access Published: 18 December 2023

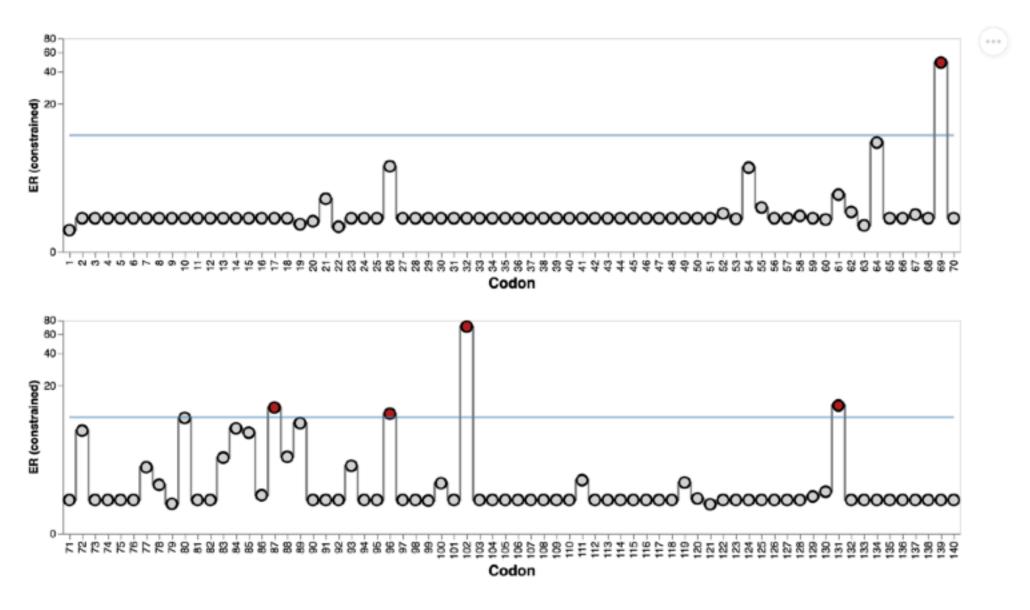
## Functional and evolutionary significance of unknown genes from uncultivated taxa

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 □

## IQCF1

 $\omega_1=0.4063$  (1.5589%)  $\omega_2=0.4076$  (93.816%)  $\omega_3=11.09$  (4.6249%)+ 0.0% error





## KRT8

 $\omega_1 = 0.00001848 (11.039\%)$   $\omega_2 = 0.04232 (75.696\%)$   $\omega_3 = 1.000 (12.675\%) +$   $0.58\% error (\omega_e > 106)$ 

11 sequences in the alignment	•	1516 codon sites in the alignment	1 partitions
18 median branches/partition used for testing	≪°	3 classes non-synonymous rate variation	3 classes synonymous rate variation
0.50 p-value for episodic diversifying selection	+	O Sites with ER≥10 for positive selection	N/A:N/A Multiple hit rates (2H:3H)
N/A (branch, site) pairs with EBF ≥ 100	Q	N/A Sites contributing most signal to EDS detection	N/A:N/A Expected fractions of MH subs (2H:3H)