

Supplementary data for: **Protein evolution is structure dependent and non-homogeneous across the tree of life** by Akanksha Pandey and Edward L. Braun

**Table S1.** Percentage of times clade-specific models were the best-fitting model for individual proteins in each training dataset

Best-fit model	Birds (1)	Birds (2)	Mammals	Plants	Oomycetes	Yeast	All Euk
<b>Bird (1)</b>	<b>49.2</b>	4.4	5.2	0.3	—	—	—
<b>Bird (2)</b>	16.8	<b>72</b>	23.4	0.3	—	—	—
<b>Mammal</b>	14	14	<b>59.1</b>	0.3	—	—	—
<b>Plant</b>	9.2	3.6	5.7	<b>88.1</b>	1.1	1	—
<b>Oomycete</b>	0.8	0.8	0.8	5.5	<b>86.5</b>	—	0.4
<b>Yeast</b>	—	0.4	—	2.6	1.1	<b>96.5</b>	—
<b>All Euk</b>	—	—	—	0.6	4	1	<b>93.5</b>
LG	—	0.4	—	0.6	3.6	1	5.2
JTT	3.2	1.6	4	—	0.4	—	—
JTT-DCMut	0.4	—	1.2	1.3	0.4	—	—
WAG	0.4	—	—	0.6	0.7	—	0.4
Dayhoff	—	0.4	—	—	—	—	—
BLOSUM	3.2	—	—	—	0.4	—	—
VT	1.6	—	—	—	0.4	—	—
HIVb	0.4	0.4	—	—	—	—	—
FLU	0.4	2	—	—	—	—	—
rtREV	—	—	—	—	0.4	—	—
mt models	0.4	—	0.4	—	1.1	0.5	0.4

Our clade-specific models are presented in bold. Values are rounded to the nearest 0.1%; ‘—’ indicates models that never had the best fit to any protein the specified validation dataset. Cases where mitochondrial models had the best fit were grouped (‘mt models’). This table only reports the best-fitting rate matrix; each protein was allowed to have any among-sites rate heterogeneity model.

**Table S2.** Percentage of times clade-specific XB mixture models were the best-fitting model for individual proteins in each training dataset

Best-fit model	Birds (1)	Birds (2)	Mammals	Plants	Oomycetes	Yeast	All Euk
<b>Bird (1)</b>	<b>45.2</b>	3.6	7.7	—	—	—	—
<b>Bird (2)</b>	19.2	<b>56</b>	25.1	—	—	—	—
<b>Mammal</b>	20	33.2	<b>57.4</b>	0.6	—	—	—
<b>Plant</b>	13.6	4	8.9	<b>90</b>	1.1	0.5	—
<b>Oomycete</b>	1.6	1.6	0.4	4.5	<b>83.4</b>	2	3.2
<b>Yeast</b>	0.4	0.4	—	2.6	2.9	<b>91.5</b>	0.8
<b>All Euk</b>	—	—	—	0.6	3.6	—	<b>69</b>
EX2	—	0.8	0.4	1.3	9	6	27

Our clade-specific mixture models are presented in bold. Values are rounded to the nearest 0.1%; ‘—’ indicates models that never had the best fit to any protein the specified validation dataset.

**Additional files:** Model files and other information are available from [https://github.com/ebraun68/clade\\_specific\\_prot\\_models](https://github.com/ebraun68/clade_specific_prot_models)