

**Supplemental Information for****DNA Sequences Are as Useful as Protein Sequences for Inferring Deep Phylogenies**

Paschalia Kapli, Ioanna Kotari, Maximilian J. Telford, Nick Goldman, and Ziheng Yang

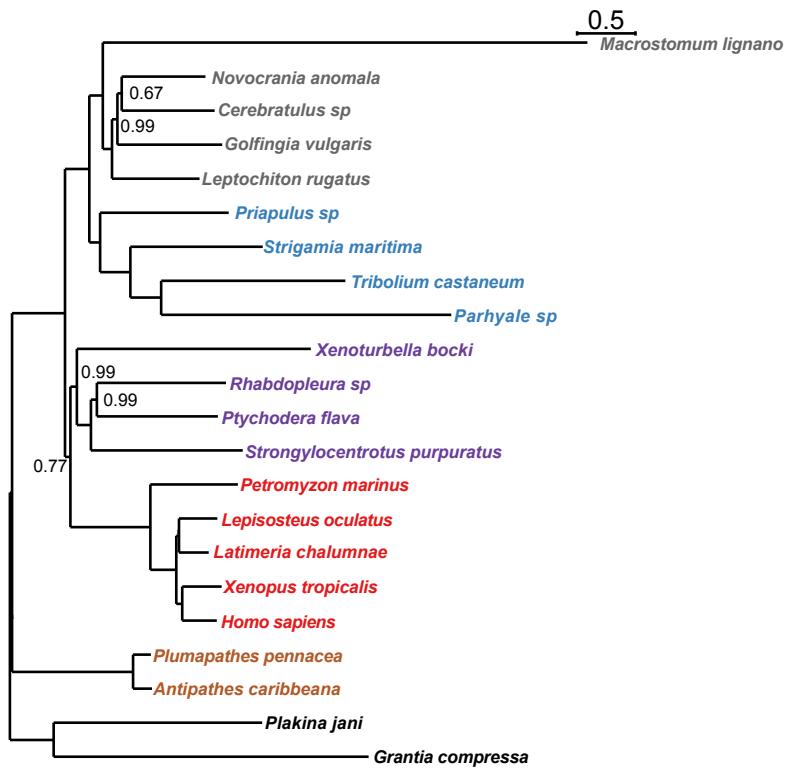


Fig. S1: Topology inferred with PhyloBayes under the CAT+GTR model using the second subset of  $4 \times 10^4$  randomly selected amino acid sites.

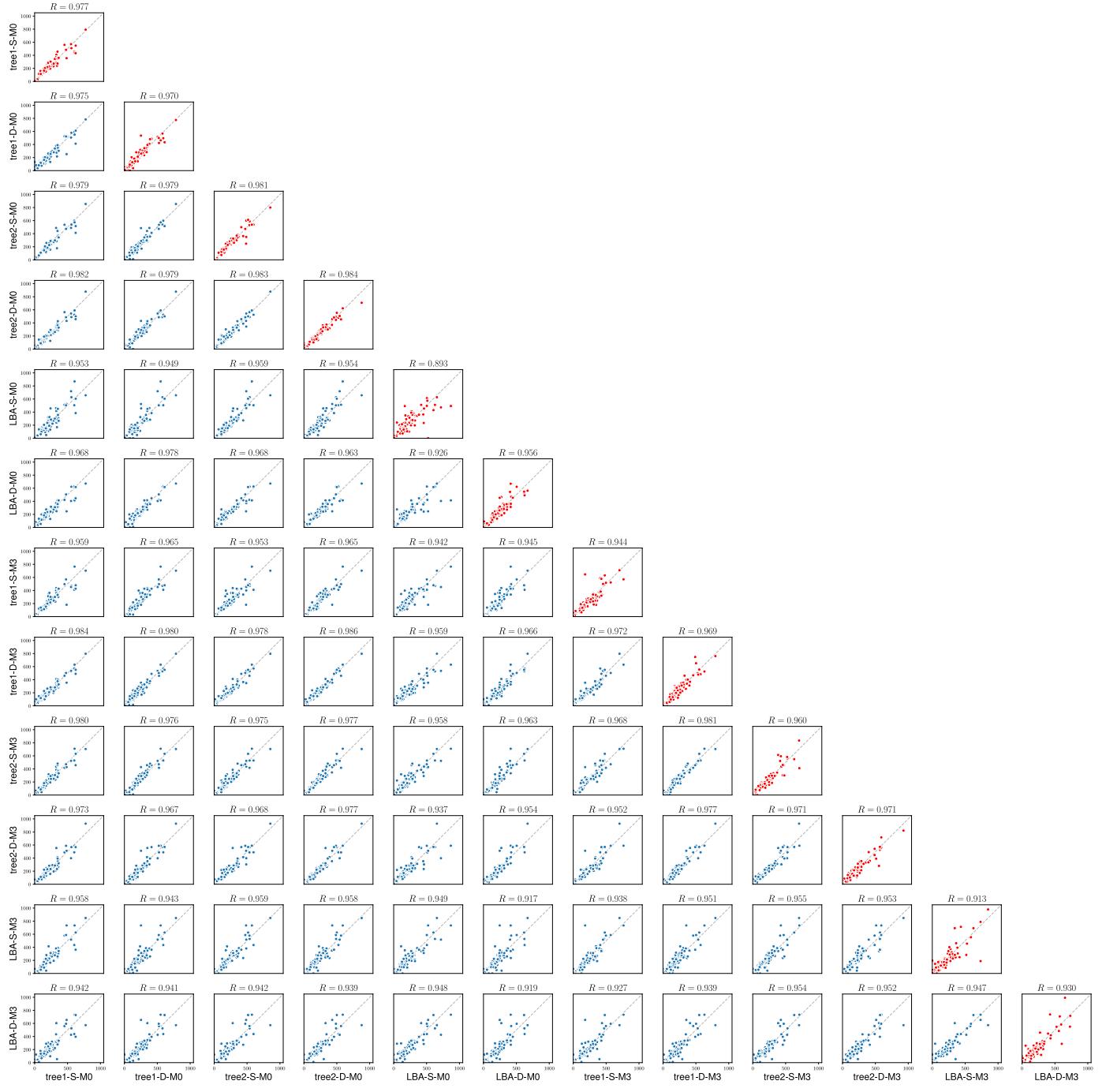


Fig. S2: Scatter plots of ML estimates of the 189 ( $= 20 \times 19/2 - 1$ ) exchangeability rates obtained using CODEML under the GTR+G<sub>5</sub> model from data simulated under different trees (tree1, tree2, LBA, and 'S' for shallow and 'D' for deep) and substitution models (M0 and M3). For each simulation condition (tree-model combination), one replicate dataset of 5000 codons was analyzed by ML to estimate the GTR rate matrix and branch lengths. On the diagonals (red dots) are for two replicate datasets simulated under the same condition to indicate sampling errors due to the finite number of sites in the alignment.

Table S1: Sources of the 22 sequences used in the empirical analyses

Species	Phylum	Source
<i>Homo sapiens</i>	Chordata	<a href="ftp://ftp.ensembl.org/pub/release-89/fasta/homo_sapiens/">ftp://ftp.ensembl.org/pub/release-89/fasta/homo_sapiens/</a>
<i>Latimeria chalumnae</i>	Chordata	<a href="ftp://ftp.ensembl.org/pub/release-89/fasta/latimeria_chalumnae/">ftp://ftp.ensembl.org/pub/release-89/fasta/latimeria_chalumnae/</a>
<i>Lepisosteus oculatus</i>	Chordata	<a href="ftp://ftp.ensembl.org/pub/release-89/fasta/lepisosteus_oculatus/">ftp://ftp.ensembl.org/pub/release-89/fasta/lepisosteus_oculatus/</a>
<i>Petromyzon marinus</i>	Chordata	<a href="ftp://ftp.ensembl.org/pub/release-89/fasta/petromyzon_marinus/">ftp://ftp.ensembl.org/pub/release-89/fasta/petromyzon_marinus/</a>
<i>Xenopus tropicalis</i>	Chordata	<a href="ftp://ftp.ensembl.org/pub/release-89/fasta/xenopus_tropicalis/">ftp://ftp.ensembl.org/pub/release-89/fasta/xenopus_tropicalis/</a>
<i>Parhyale hawaiensis</i>	Arthropoda	<a href="http://research.janelia.org/pavlopoulos/">http://research.janelia.org/pavlopoulos/</a>
<i>Strigamia maritima</i>	Arthropoda	<a href="ftp://ftp.ensemblgenomes.org/pub/release-36/metazoa/fasta/strigamia_maritima/">ftp://ftp.ensemblgenomes.org/pub/release-36/metazoa/fasta/strigamia_maritima/</a>
<i>Tribolium castaneum</i>	Arthropoda	<a href="ftp://ftp.ensemblgenomes.org/pub/release-36/metazoa/fasta/tribolium_castaneum/">ftp://ftp.ensemblgenomes.org/pub/release-36/metazoa/fasta/tribolium_castaneum/</a>
<i>Priapulus sp</i>	Priapulida	Max Telford
<i>Ptychoderia flava</i>	Hemichordata	<a href="http://marinegenomics.oist.jp/acornworm/viewer/download?project_id=33">http://marinegenomics.oist.jp/acornworm/viewer/download?project_id=33</a>
<i>Rhabdopleura</i>	Hemichordata	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX879690">https://www.ncbi.nlm.nih.gov/sra/SRX879690</a>
<i>Strongylocentrotus purpuratus</i>	Echinodermata	<a href="http://www.echinobase.org/Echinobase/SpDownloads">http://www.echinobase.org/Echinobase/SpDownloads</a>
<i>Xenoturbella bocki</i>	Xenacoelomorpha	Herve Philipp
<i>Cerebratulus sp</i>	Nemertea	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX872410">https://www.ncbi.nlm.nih.gov/sra/SRX872410</a>
<i>Golfingia vulgaris</i>	Annelida	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX872414">https://www.ncbi.nlm.nih.gov/sra/SRX872414</a>
<i>Leptochiton rugatus</i>	Mollusca	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX731464">https://www.ncbi.nlm.nih.gov/sra/SRX731464</a>
<i>Macrostomum lignano</i>	Platyhelminthes	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX872327">https://www.ncbi.nlm.nih.gov/sra/SRX872327</a>
<i>Novocrania anomala</i>	Brachiopoda	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX731477">https://www.ncbi.nlm.nih.gov/sra/SRX731477</a>
<i>Plumapathes pennacea</i>	Cnidaria	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX1716822">https://www.ncbi.nlm.nih.gov/sra/SRX1716822</a>
<i>Antipathes caribbeana</i>	Cnidaria	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX1716821">https://www.ncbi.nlm.nih.gov/sra/SRX1716821</a>
<i>Plakina jani</i>	Porifera	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX1719637">https://www.ncbi.nlm.nih.gov/sra/SRX1719637</a>
<i>Grantia compressa</i>	Porifera	<a href="https://www.ncbi.nlm.nih.gov/sra/SRX1719634">https://www.ncbi.nlm.nih.gov/sra/SRX1719634</a>

Table S2: Average bootstrap support for the ML tree in 1000 replicate datasets (5000 codons) simulated assuming the shallow trees of Figure 1

True tree	Tree 1				Tree 2				LBA tree			
	Model		M0 (1 $\omega$ )		M3 (3 $\omega$ s)		M0 (1 $\omega$ )		M3 (3 $\omega$ s)		M0 (1 $\omega$ )	
Inf. tree	correct	wrong	correct	wrong	correct	wrong	correct	wrong	correct	wrong	correct	wrong
<b>Homogeneous model (homo)</b>												
AA	85.8	74.4	81.2	72.3	84.3	78.1	81.3	76.0	77.1	55.5	71.9	54.2
DNA-123	87.2	77.7	84.0	75.6	91.8	85.9	87.2	81.7	90.2	50.0	81.1	53.1
DNA-123-P	87.3	78.6	83.5	75.7	93.2	87.2	89.2	83.9	96.4	63	92.2	58.6
DNA-12	83.8	75.2	79.1	72.4	90.6	85.1	85.3	80.0	89.1	57.2	80.3	49.7
DNA-12-P	83.9	75.3	79.1	72.2	91.0	85.3	86.0	80.8	94.1	59.0	88.7	61.8
Codon	88.7	80.4	88.2	77.6	94.0	88.3	88.4	82.7	94.0	51.4	82.9	54.6
<b>Site-heterogeneous model 1 (SH1)</b>												
AA	84.6	74.4	80.5	72.2	82.2	76.9	80.2	75.1	80.5	64.2	77.9	64.5
DNA-123	86.6	77.4	82.8	74.7	90.4	85.5	86.8	81.2	90.4	59.8	85.5	64.8
DNA-123-P	86.9	78.5	82.6	75.1	91.5	86.2	87.7	82.8	92.2	59.0	86.9	65.1
DNA-12	82.8	75.3	78.3	71.5	89.6	84.2	85.5	79.5	91.1	56.2	86.6	60.1
DNA-12-P	82.4	75.3	78.5	71.4	89.9	84.3	85.3	79.7	91.3	59.3	87.2	58.4
Codon	89.6	79.2	87.9	76.2	91.7	86.1	85.2	80.4	90.8	62.9	79.2	70.8
<b>Site-heterogeneous model 2 (SH2)</b>												
AA	80.1	72.4	77.5	69.2	78.3	73.7	76.2	70.2	73.5	70.0	71.8	71.2
DNA-123	81.7	73.6	79.8	71.8	84.2	79.3	81.4	75.3	81.3	64.7	78.5	65.8
DNA-123-P	84.4	75.5	81.7	72.0	85.5	80.3	82.5	76.0	83.0	64.3	79.7	67.3
DNA-12	79.9	72.6	76.4	69.5	81.8	76.9	78.3	71.6	78.8	63.0	77.1	64.3
DNA-12-P	80.6	72.4	76.7	69.6	82.5	77.3	78.4	72.0	79.3	63.2	76.9	65.8
Codon	86.9	76.3	84.5	74.0	84.5	79.3	80.0	74.4	79.7	67.5	72.4	73.2
<b>Branch-site heterogeneous model (BSH)</b>												
AA	74.8	71.0	70.8	70.7	79.0	76.6	71.6	73.1	79.7	68.6	75.4	71.3
DNA-123	-	98.4	-	99.0	-	90.5	-	91.7	-	100.0	-	100.0
DNA-123-P	64.0	86.7	-	92.4	74.8	82.6	-	83.0	55.3	96.7	-	99.5
DNA-12	74.2	70.0	70.9	69.6	81.6	80.5	74.4	75.2	81.8	66.7	77.9	70.0
DNA-12-P	74.4	70.4	70.9	69.7	82.2	81.1	75.3	75.8	82.1	66.7	78.1	69.6
Codon	-	97.8	-	99.0	-	88.7	-	90.0	-	99.7	-	99.9

Note.— The average support was calculated separately depending on whether the ML tree matches the true tree (“correct”) or not (“wrong”). Dash (-) indicates the correct tree was not recovered in any replicates. See legend to Figure 4 for the simulation models and analysis strategies.

Table S3: Tree lengths (sum of branch lengths) averaged over replicate datasets

True Tree	Tree 1				Tree 2				LBA Tree			
	2000		5000		2000		5000		2000		5000	
Datasize	M0	M3	M0	M3	M0	M3	M0	M3	M0	M3	M0	M3
Model												
<b>Shallow trees</b>												
Homogeneous model (homo)												
DNA-123-p	1.38	1.38	1.38	1.37	1.38	1.37	1.38	1.38	0.40	0.40	0.40	0.40
DNA-12-p	1.07	0.99	1.07	0.99	1.07	0.99	1.07	0.99	0.31	0.29	0.31	0.29
Site-heterogeneous model 1 (SH1)												
DNA-123-p	1.34	1.32	1.34	1.32	1.34	1.32	1.34	1.32	0.40	0.39	0.40	0.39
DNA-12-p	1.15	1.06	1.15	1.07	1.15	1.06	1.15	1.07	0.34	0.31	0.34	0.31
Site-heterogeneous model 2 (SH2)												
DNA-123-p	1.34	1.34	1.34	1.34	1.35	1.34	1.35	1.35	0.39	0.39	0.39	0.39
DNA-12-p	0.68	0.61	0.68	0.61	0.68	0.61	0.68	0.61	0.20	0.18	0.20	0.18
Branch-site heterogeneous model 2 (BSH)												
DNA-123-p	1.78	1.84	1.77	1.83	1.77	1.83	1.77	1.83	0.45	0.44	0.44	0.44
DNA-12-p	0.71	0.65	0.71	0.65	0.71	0.65	0.71	0.65	0.17	0.16	0.18	0.16
<b>Deep trees</b>												
Homogeneous model (homo)												
DNA-123	5.35	6.80	5.36	6.87	5.35	6.84	5.36	6.86	1.59	1.58	1.59	1.58
DNA-12	4.15	3.83	4.16	3.83	4.15	3.83	4.15	3.83	1.25	1.14	1.24	1.14
Site-heterogeneous model 1 (SH1)												
DNA-123	4.53	4.23	4.53	4.23	4.53	4.23	4.53	4.24	1.42	1.32	1.41	1.31
DNA-12	4.25	3.78	4.26	3.79	4.25	3.79	4.26	3.79	1.30	1.18	1.30	1.18
Site-heterogeneous model 2 (SH2)												
DNA-123	5.41	6.28	5.40	6.28	5.43	6.30	5.43	6.32	1.34	1.30	1.34	1.29
DNA-12	2.31	1.95	2.31	1.95	2.31	1.95	2.32	1.96	0.69	0.60	0.69	0.60
Branch-site heterogeneous model 2 (BSH)												
DNA-123	4.34	4.20	4.33	4.18	4.34	4.16	4.32	4.16	1.31	1.26	1.30	1.25
DNA-12	2.52	2.15	2.52	2.15	2.52	2.15	2.53	2.16	0.70	0.62	0.70	0.62

Table S4: Average bootstrap support of ML trees in 1000 replicate datasets (5000 codons) simulated under the deep trees of Figure 1

True tree	Tree 1				Tree 2				LBA tree			
	Model		M0 (1 $\omega$ )	M3 (3 $\omega$ s)	Model		M0 (1 $\omega$ )	M3 (3 $\omega$ s)	Model		M0 (1 $\omega$ )	M3 (3 $\omega$ s)
Inf. tree	correct	wrong	correct	wrong	correct	wrong	correct	wrong	correct	wrong	correct	wrong
<b>Homogeneous model (homo)</b>												
AA	64.7	57.6	77.2	68.9	64.9	61.1	77.5	72.1	80.9	57.7	81.5	61.6
DNA-123	75.0	66.9	73.2	66.5	75.9	70.9	74.4	69.7	74.4	70.4	68.8	80.0
DNA-123-P	75.6	67.4	74.4	68.0	79.1	74.6	79.8	74.9	91.1	59.2	83.7	61.0
DNA-12	73.6	65.8	72.9	66.5	75.7	71.7	77.6	72.5	81.2	61.9	77.9	65.4
DNA-12-P	73.9	65.9	73.8	66.3	77.1	72.8	78.9	73.3	88.8	59.1	83.6	60.3
Codon	76.9	69.6	77.4	68.5	82.4	77.5	75.3	71.0	90.9	55.7	68.0	90.9
<b>Site-heterogeneous model 1 (SH1)</b>												
AA	62.1	54.7	75.3	67.3	63.5	57.4	75.0	69.9	76.6	62.7	79.0	60.9
DNA-123	72.6	65.2	72.2	63.6	73.4	69.3	70.8	67.5	69.7	78.5	65.5	87.3
DNA-123-P	73.0	66.3	73.7	66.0	74.9	70.9	75.9	71.0	74.1	72.6	68.0	82.1
DNA-12	71.6	65.7	71.8	65.9	73.8	69.7	75.6	71.0	81.2	63.2	77.1	63.6
DNA-12-P	71.3	65.7	72.2	65.8	74.5	70.3	76.4	71.4	84.5	60.3	79.2	60.5
Codon	75.0	68.0	76.3	66.3	76.3	72.3	71.8	67.9	75.4	70.8	59.0	97.9
<b>Site-heterogeneous model 2 (SH2)</b>												
AA	68.6	58.3	71.3	62.3	65.1	58.7	68.3	62.3	50.0	97.8	65.5	95.8
DNA-123	70.8	63.2	68.6	60.5	69.5	64.7	66.9	62.2	66.9	83.0	66.2	88.2
DNA-123-P	74.9	65.7	72.0	63.3	71.4	66.8	70.2	65.0	66.2	87.2	64.6	89.7
DNA-12	71.7	63.7	69.4	60.7	69.4	64.4	68.6	62.5	66.5	85.9	63.5	85.7
DNA-12-P	72.9	64.2	69.6	61.4	69.5	64.8	69.5	63.2	67.3	85.7	63.6	86.3
Codon	78.0	67.9	76.0	65.7	70.5	67.3	68.1	62.8	62.5	91.5	0*	98.3
<b>Branch-site-heterogeneous model (BSH)</b>												
AA	61.8	64.7	63.6	64.1	-	64.6	60.9	62.6	64.0	93.8	63.2	90.1
DNA-123	-	74.0	-	68.7	-	77.4	-	72.8	-	100.0	-	100.0
DNA-123-P	-	81.6	-	77.4	-	82.3	-	79.6	-	99.8	-	99.9
DNA-12	57.3	64.2	56.3	60.5	65.3	63.3	63.0	62.7	66.0	86.3	68.4	84.0
DNA-12-P	58.6	65.4	57.7	61.3	63.8	63.9	63.3	63.2	65.6	85.7	67.1	85.0
Codon	-	84.3	-	82.2	-	83.5	-	82.2	-	100.0	-	100.0

Note.— See legend to table S2.