Protein structures unravel the signatures and patterns of deep time evolution

Ajith Harish
Independent Researcher, Uppsala, 756 57, Sweden.

Supplementary Material

Contents

Complete list of time reversible and nonreversible models of sequence substitution tested for the sequence alignment shown in Table 1 of the main text.

1. List of time reversible models tested

```
ModelFinder will test up to 702 protein models (sample size: 3983) ...
```

No.	Model	-LnL	df	AIC	AICc	BIC
1	LG	174208.021	71	348558.042	348560.656	349004.617
2	LG+I	172061.694	72	344267.389	344270.077	344720.254
3	LG+G4	166613.707	72	333371.413	333374.102	333824.278
4	LG+I+G4	166520.807	73	333187.614	333190.378	333646.769
5	LG+R2	168041.616	73	336229.232	336231.996	336688.387
6	LG+R3	166697.135	75	333544.271	333547.189	334016.005
7	LG+R4	166468.438	77	333090.877	333093.953	333575.191
8	LG+R5	166404.051	79	332966.101	332969.340	333462.995
9	LG+R6	166395.786	81	332953.571	332956.977	333463.044
21	LG+F+R5	166600.538	98	333397.077	333402.072	334013.476
22	LG+F+R6	166593.391	100	333386.782	333391.985	334015.761
34	WAG+R5	168014.780	79	336187.561	336190.799	336684.454
35	WAG+R6	168010.383	81	336182.767	336186.172	336692.240
47	WAG+F+R5	167858.361	98	335912.722	335917.718	336529.122
48	WAG+F+R6	167853.617	100	335907.233	335912.437	336536.212
60	JTT+R5	169515.329	79	339188.658	339191.896	339685.551
61	JTT+R6	169507.145	81	339176.290	339179.696	339685.763
73	JTT+F+R5	169929.928	98	340055.857	340060.853	340672.256
74	JTT+F+R6	169922.857	100	340045.713	340050.917	340674.692
86	Q.pfam+R5	166734.679	79	333627.357	333630.596	334124.251
87	Q.pfam+R6	166724.885	81	333611.769	333615.174	334121.242
99	Q.pfam+F+R5	166939.300	98	334074.599	334079.595	334690.999
100	Q.pfam+F+R6	166929.724	100	334059.448	334064.651	334688.427
112	Q.bird+R5	173921.387	79	348000.773	348004.012	348497.667
113	Q.bird+R6	173892.748	81	347947.497	347950.902	348456.970
125	Q.bird+F+R5	174714.386	98	349624.772	349629.768	350241.172
126	Q.bird+F+R6	174686.965	100	349573.929	349579.133	350202.908
138	Q.mammal+R5	172246.873	79	344651.746	344654.985	345148.640
139	Q.mammal+R6	172220.628	81	344603.256	344606.661	345112.729
151	Q.mammal+F+R5	172795.232	98	345786.463	345791.459	346402.863
152	Q.mammal+F+R6	172771.496	100	345742.993	345748.196	346371.972
164	Q.insect+R5	167287.315	79	334732.630	334735.868	335229.523
165	Q.insect+R6	167267.687	81	334697.374	334700.779	335206.847
177	Q.insect+F+R5	167955.772	98	336107.545	336112.541	336723.944
178	Q.insect+F+R6	167942.037	100	336084.074	336089.278	336713.053
190	Q.plant+R5	169529.537	79	339217.074	339220.312	339713.967
191	Q.plant+R6	169507.751	81	339177.502	339180.907	339686.975
203	Q.plant+F+R5	170115.587	98	340427.174	340432.170	341043.573
204	Q.plant+F+R6	170097.512	100	340395.025	340400.228	341024.004
216	Q.yeast+R5	167333.610	79	334825.221	334828.459	335322.114
217	Q.yeast+R6	167316.664	81	334795.327	334798.733	335304.800
229	Q.yeast+F+R5	167360.516	98	334917.033	334922.029	335533.432
230	Q.yeast+F+R6	167346.816	100	334893.631	334898.835	335522.610
242	JTTDCMut+R5	169459.669	79	339077.339	339080.577	339574.232
243	JTTDCMut+R6	169451.283	81	339064.567	339067.972	339574.040
255	JTTDCMut+F+R5	169832.627	98	339861.254	339866.250	340477.654
256	JTTDCMut+F+R6	169825.713	100	339851.426	339856.629	340480.405
268	DCMut+R5	170756.771	79	341671.542	341674.780	342168.435
269	DCMut+R6	170748.991	81	341659.981	341663.386	342169.454
281	DCMut+F+R5	170328.978	98	340853.955	340858.951	341470.355
282	DCMut+F+R6	170323.086	100	340846.171	340851.375	341475.150
294	VT+R5	168293.396	79	336744.792	336748.031	337241.685
295	VT+R6	168290.791	81	336743.581	336746.986	337253.054
307	VT+F+R5	168362.700	98	336921.401	336926.397	337537.800
308	VT+F+R6	168360.763	100	336921.526	336926.729	337550.505
320	PMB+R5	168569.915	79	337297.831	337301.069	337794.724
321	PMB+R6	168568.443	81	337298.887	337302.292	337808.360
333	PMB+F+R5	168777.017	98	337750.035	337755.030	338366.434
334	PMB+F+R6	168775.452	100	337750.905	337756.109	338379.884
346	Blosum62+R5	168142.228	79	336442.457	336445.695	336939.350
347	Blosum62+R6	168140.638	81	336443.275	336446.680	336952.748
359	Blosum62+F+R5	168488.206	98	337172.412	337177.407	337788.811
360	Blosum62+F+R6	168486.566	100	337173.132	337178.335	337802.111
372	Dayhoff+R5	170748.535	79	341655.070	341658.309	342151.964

373	Dayhoff+R6	170740.487	81	341642.973	341646.379	342152.446
385	Dayhoff+F+R5	170325.239	98	340846.479	340851.475	341462.878
386	Dayhoff+F+R6	170319.179	100	340838.358	340843.562	341467.337
398	mtREV+R5	177282.048	79	354722.096	354725.335	355218.989
399	mtREV+R6	177263.475	81	354688.949	354692.354	355198.422
411	mtREV+F+R5	171176.959	98	342549.918	342554.914	343166.317
412	mtREV+F+R6	171154.013	100	342508.026	342513.229	343137.005
424	mtART+R5	176851.089	79	353860.177	353863.416	354357.071
425	mtART+R6	176798.171	81	353758.341	353761.746	354267.814
437	mtART+F+R5	171190.644	98	342577.288	342582.284	343193.687
438	mtART+F+R6	171150.538	100	342501.076	342506.279	343130.055
450	mtZOA+R5	172996.512	79	346151.023	346154.262	346647.917
451	mtZOA+R6	172955.698	81	346073.396	346076.802	346582.869
463	mtZOA+F+R5	169681.261	98	339558.523	339563.518	340174.922
464	mtZOA+F+R6	169653.004	100	339506.008	339511.212	340134.987
476	mtMet+R5	176860.496	79	353878.992	353882.230	354375.885
477	mtMet+R6	176824.307	81	353810.614	353814.019	354320.087
489	mtMet+F+R5	171853.874	98	343903.747	343908.743	344520.147
490	mtMet+F+R6	171826.803	100	343853.605	343858.809	344482.584
502	mtVer+R5	180392.323	79	360942.647	360945.885	361439.540
503	mtVer+R6	180347.770	81	360857.540	360860.945	361367.013
515	mtVer+F+R5	175583.834	98	351363.668	351368.664	351980.067
516	mtVer+F+R6	175534.285	100	351268.570	351273.773	351897.549
528	mtInv+R5	176342.257	79	352842.515	352845.753	353339.408
529	mtInv+R6	176316.203	81	352794.407	352797.812	353303.880
541	mtInv+F+R5	169709.382	98	339614.764	339619.759	340231.163
542	mtInv+F+R6	169696.566	100	339593.132	339598.335	340222.111
554	mtMAM+R5	182018.111	79	364194.222	364197.460	364691.115
555	mtMAM+R6	181949.848	81	364061.696	364065.102	364571.170
567	mtMAM+F+R5	176381.874	98	352959.748	352964.743	353576.147
568	mtMAM+F+R6	176335.472	100	352870.945	352876.148	353499.924
580	HIVb+R5	174514.243	79	349186.487	349189.725	349683.380
581	HIVb+R6	174489.019	81	349140.039	349143.444	349649.512
593	HIVb+F+R5	174764.747	98	349725.495	349730.490	350341.894
594	HIVb+F+R6	174738.608	100	349677.216	349682.420	350306.195
606	HIVw+R5	182974.016	79	366106.032	366109.271	366602.926
607	HIVw+R6	182919.116	81	366000.232	366003.637	366509.705
619	HIVw+F+R5	181520.941	98	363237.882	363242.877	363854.281
620	HIVw+F+R6	181478.971	100	363157.941	363163.145	363786.920
632	FLU+R5	173264.070	79	346686.139	346689.378	347183.033
633	FLU+R6	173232.329	81	346626.657	346630.062	347136.130
645	FLU+F+R5	173341.755	98	346879.510	346884.506	347495.909
646	FLU+F+R6	173304.927	100	346809.854	346815.057	347438.833
658	rtREV+R5	167875.540	79	335909.079	335912.318	336405.973
659	rtREV+R6	167866.371	81	335894.743	335898.148	336404.216
671	rtREV+F+R5	166829.356	98	333854.712	333859.708	334471.111
672	rtREV+F+R6	166822.638	100	333845.276	333850.480	334474.255
684	cpREV+R5	168671.553	79	337501.106	337504.344	337997.999
685	cpREV+R6	168657.311	81	337476.623	337480.028	337986.096
697	cpREV+F+R5	168672.126	98	337540.253	337545.249	338156.652
698	cpREV+F+R6	168658.257	100	337516.513	337521.717	338145.493

Akaike Information Criterion: LG+R6
Corrected Akaike Information Criterion: LG+R6
Bayesian Information Criterion: LG+R5
Best-fit model: LG+R5 chosen according to BIC

2. List of time nonreversible models tested

ModelFinder will test up to 26 protein models (sample size: 3983) ...

No.	Model	-LnL	df	AIC	AICc	BIC
1	NONREV+FO	171827.048	451	344556.096	344671.560	347392.791
2	NONREV+FO+I	169999.443	452	340902.885	341018.894	343745.870
3	NONREV+FO+G4	165339.072	452	331582.143	331698.152	334425.129
4	NONREV+FO+I+G4	165254.455	453	331414.909	331531.465	334264.184
5	NONREV+FO+R2	166587.486	453	334080.971	334197.527	336930.247

```
  6  NONREV+FO+R3    165424.251   455 331758.501   331876.154   334620.356
  7  NONREV+FO+R4    165208.442   457 331330.883   331449.639   334205.318
  8  NONREV+FO+R5    165163.806   459 331245.612   331365.476   334132.626
  9  NONREV+FO+R6    165160.515   461 331243.029   331364.007   334142.623
 21  NONREV+F+R5     165163.323   478 331282.646   331413.332   334289.166
 22  NONREV+F+R6     165160.515   480 331281.029   331412.885   334300.129

Akaike Information Criterion:           NONREV+FO+R6
Corrected Akaike Information Criterion: NONREV+FO+R6
Bayesian Information Criterion:         NONREV+FO+R5
Best-fit model: NONREV+FO+R5 chosen according to BIC
```

From: **mcgill** mcgill@digizyme.com
Subject: Re: Request for permission to use figure in publication
Date: August 23, 2021 at 21:44
To: Ajith Harish ajith.harish@gmail.com

Hi Ajith.

Thank you for this additional information.
You are welcome to use the image in your article - I have attached a high resolution version.

For the credits/legend, please use the following:

"Eukaryotic cellular landscape, by Evan Ingersoll & Gaël McGill, Ph.D. (Digizyme Inc.) using Molecular Maya software. Created for Cell Signaling Technology, Inc., and inspired by the stunning art of David Goodsell, this 3D rendering of a eukaryotic cell is modeled using X-ray, nuclear magnetic resonance (NMR), and cryo-electron microscopy datasets for all of its molecular actors."

I would very much appreciate if you could share the finished article when ready.

Thank you,

-Gael.

From: Ajith Harish <ajith.harish@gmail.com>
Sent: Friday, August 13, 2021 2:43 AM
To: mcgill <mcgill@digizyme.com>
Subject: Re: Request for permission to use figure in publication

Hi Gael,

Thank you for your reply. I am an evolutionary biologist and my article is about the evolution of cellular life, in the context of the universal tree of life (ToL). I want to show the two basic types of cells: eukaryotes and prokaryotes, and compare the different levels of complexity of two cell types as well as show the molecular. The more specific context is how the use of protein structure-based phylogenetic studies are beginning to show a new view of the TOL - A Two-Domain tree rather than the Three-Domain tree. If helps to give a better idea, here is one of my recent articles (https://f1000research.com/articles/9-112)

The figure I have in mind is to compare your image side-by-side with Goodsell's E. coli image (https://pdb101.rcsb.org/sci-art/goodsell-gallery/escherichia-coli-bacterium) as sectional views of the eukaryote and prokaryote cell organizations. Both these will be presented as a section of a more generic eukaryote and prokaryote type cell diagrams (similar to textbook images). Hope it gives an ideal of what I plan to do.

Needless to say, your image is the most detailed and most stunning depiction of the eukaryote cell that is also scientifically accurate.

Sincerely,
Ajith

> On Aug 13, 2021, at 03:03, mcgill <mcgill@digizyme.com> wrote:
>
> Hi Ajith,
>
> Thanks for your interest in our image.
>
> It really depends on the context of use - I would need to gain a better understanding of what you plan to use it for (i.e. what is your article about and what points do you intend to make in reference to the image).
> Forgive my caution, but this image has gone viral in recent months... after it was stolen from our site without permissions, credits and, often, connected to a bogus legend. We have been receiving hundreds of requests for its use in various contexts.
>
> Sincerely,
> -Gael.
>
> ---
> Gaël McGill, Ph.D.
>
> President & CEO
> Digizyme, Inc.
> http://www.digizyme.com
> http://www.clarafi.com
>
> Faculty & Director of Molecular Visualization
> Center for Molecular & Cellular Dynamics
> Harvard Medical School

Harvard Medical School
http://cmcd.hms.harvard.edu/faculty.php.html
http://www.visabli.org

Digital Director & Co-Author
E.O. Wilson's 'Life on Earth'
https://books.apple.com/us/book/e-o-wilsons-life-on-earth-unit-1/id888107968#see-all/author-other-books

---

From: Ajith Harish <ajith.harish@gmail.com>
Sent: Wednesday, August 11, 2021 7:37 AM
To: mcgill <mcgill@digizyme.com>
Subject: Request for permission to use figure in publication

Dear Dr. McGill,

I am in independent researcher, I am writing to request permission to use one of the gorgeous figures hosted at https://www.digizyme.com/cst_landscapes.html in a publication (specifically the image attached here). The publication will most likely be a open access publication (depending the which journal accepts my article). Or I might pre-print (also depending on the journal).

Could you let me know if there are any charges associated with the permissions. Please let me know if there is a formal request form/procedure.

Looking forward to hear.
Best wishes,
Ajith

<my7kljdlhux51.jpeg>

Digizyme_CellLandscap….jpg.zip