

maxAlike: Sequence Reconstruction by Maximum Likelihood Estimation – Supplementary materials

Peter Menzel, Peter F. Stadler, and Jan Gorodkin

February 1, 2010

Data sources

The multiple alignments and phylogenetic tree of the *MZ44* data sets were downloaded from <http://hgdownload.cse.ucsc.edu/goldenPath/hg18/multiz44way/>.

The multiple alignments and phylogenetic tree of the *ENC* data sets were downloaded from <http://hgdownload.cse.ucsc.edu/goldenPath/hg18/encode/MSA/DEC-2007/>.

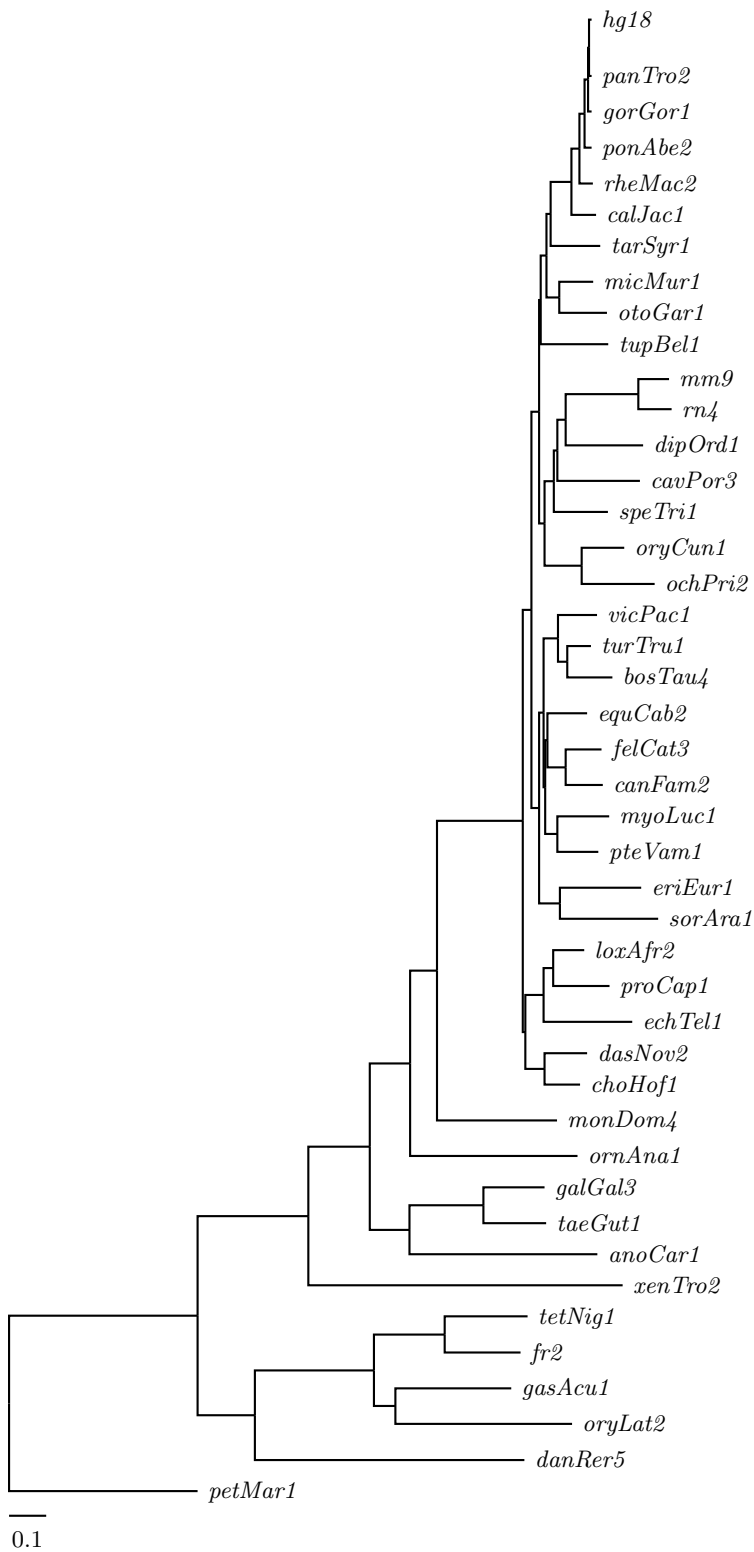


Figure 1: MZ44 phylogenetic tree

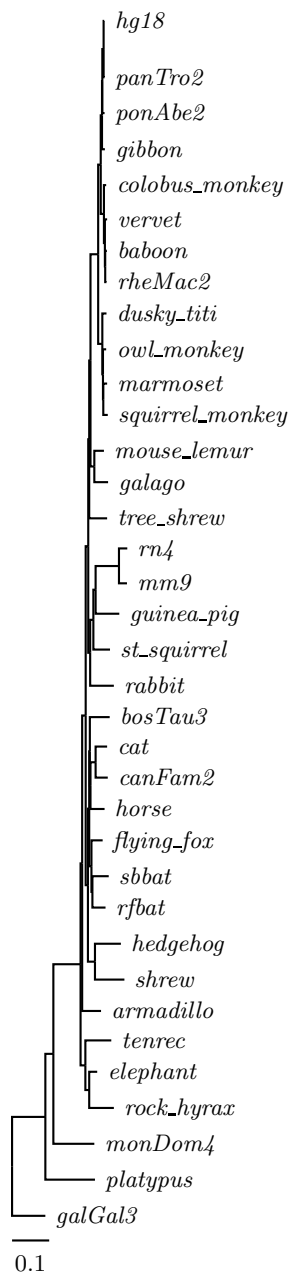


Figure 2: ENCODE phylogenetic tree

Species	Dist.	Threshold 0.5				No threshold		
		<i>Freq</i>	% ≥ 0.5	<i>ML</i>	% ≥ 0.5	<i>Freq</i>	<i>ML</i>	Nt.pred.
hg18	0.013	90.1	97.4	98.9	100.0	89.0	98.9	49076
panTro2	0.013	90.0	97.4	99.0	100.0	88.9	99.0	47964
gorGor1	0.018	90.2	97.7	99.0	100.0	89.2	99.0	32720
ponAbe2	0.037	89.9	97.5	97.2	100.0	88.8	97.2	47470
rheMac2	0.069	88.8	97.5	94.9	99.9	87.7	94.9	48083
calJac1	0.120	87.3	97.6	91.6	99.7	86.3	91.4	44416
mm9	0.174	72.1	98.0	85.1	98.0	71.5	84.6	32796
rn4	0.174	72.0	97.9	86.1	98.8	71.4	85.9	30024
bosTau4	0.186	82.8	97.5	87.3	96.5	81.9	86.0	42322
turTru1	0.186	86.0	98.0	90.3	99.0	85.3	89.9	36696
vicPac1	0.195	84.3	97.5	87.9	98.0	83.5	87.2	31040
canFam2	0.199	83.8	97.6	86.7	98.1	82.9	86.1	43914
felCat3	0.199	83.7	97.7	86.9	98.0	82.9	86.1	26960
choHof1	0.211	82.6	97.6	84.9	96.9	81.9	83.9	26211
dasNov2	0.211	80.2	97.7	83.5	95.9	79.5	82.1	22849
micMur1	0.220	87.1	97.5	88.0	98.7	86.3	87.6	32992
otoGar1	0.220	83.5	97.6	84.8	98.0	82.8	84.1	31734
loxAfr2	0.236	82.3	97.5	85.6	96.9	81.5	84.5	27732
proCap1	0.236	78.0	97.6	82.8	93.7	77.2	80.6	26348
tarSyr1	0.246	85.1	97.7	85.9	98.5	84.3	85.4	32536
equCab2	0.247	86.8	97.6	89.2	99.0	86.0	88.8	45471
myoLuc1	0.253	82.4	97.9	85.2	98.2	81.6	84.5	21856
pteVam1	0.253	83.1	97.6	85.5	98.6	82.2	85.1	33725
ochPri2	0.314	74.0	97.8	78.4	87.6	73.3	74.7	23566
oryCun1	0.314	79.0	97.6	80.5	95.1	78.3	79.1	25726
tupBel1	0.321	82.3	97.4	82.5	97.9	81.4	81.8	31346
speTri1	0.330	80.6	97.4	81.5	96.9	79.8	80.4	26613
galGal3	0.337	58.1	93.8	73.6	68.7	56.4	61.7	2805
echTel1	0.350	73.9	97.7	77.5	93.3	73.2	75.5	18886
cavPor3	0.380	76.0	97.9	77.3	94.9	75.3	75.9	36299
dipOrd1	0.389	75.1	97.4	77.1	94.2	74.3	75.4	21518
eriEur1	0.407	76.4	97.3	78.5	94.9	75.6	77.3	15726
fr2	0.430	78.2	96.7	84.3	93.8	77.0	81.9	32322
tetNig1	0.430	50.7	91.9	72.2	75.8	49.2	65.6	4580
sorAra1	0.453	75.5	97.3	77.8	92.8	74.6	76.2	14707
monDom4	0.714	63.5	96.7	66.7	78.6	62.6	62.8	11263
gasAcu1	0.770	51.6	91.9	73.7	68.9	49.7	63.7	4581
oryLat2	0.792	51.0	91.8	73.4	65.2	49.4	62.7	4397
ornAna1	0.852	60.1	97.1	64.3	69.7	59.3	59.9	4239
anoCar1	0.876	54.3	93.6	66.9	60.1	52.6	55.6	3199
petMar1	1.023	53.6	90.9	67.4	40.9	51.3	47.3	2869
danRer5	1.426	48.5	92.1	66.8	46.1	46.9	52.5	4785
xenTro2	1.493	53.5	92.4	68.5	45.3	51.7	52.1	4947

Table 1: **Data set MZ44-1:** Recovery rate in percent for consensus sequences derived from maxAlike probabilities (*ML* columns) and nucleotide frequencies (*Freq* columns), with a probability/relative frequency above a 0.5 threshold and without a threshold (i.e. the consensus sequence is made up from the nucleotides with highest probability/frequency at each position). A value of e.g. 70 means that 70% of the nucleotides in the consensus sequence were predicted correctly. Only species with at least 10 reconstructed sequences were considered. The last column (Nt.pred.) shows the total number of reconstructed nucleotides for each species, excluding fully conserved alignment columns. The fraction (in percent) of predicted nucleotides above the 0.5 threshold is shown in the columns "% ≥ 0.5 " for both methods. The Dist. column shows the distance to the phylogenetically nearest neighbor in the tree.

Species	Dist.	Threshold 0.5				No threshold		
		<i>Freq</i>	% ≥ 0.5	<i>ML</i>	% ≥ 0.5	<i>Freq</i>	<i>ML</i>	Nt.pred.
hg18	0.013	88.8	96.3	98.9	100.0	87.3	98.9	92089
panTro2	0.013	88.8	96.3	98.9	100.0	87.3	98.9	90693
gorGor1	0.018	89.1	96.4	99.0	100.0	87.5	99.0	61291
ponAbe2	0.037	88.5	96.3	97.1	100.0	87.0	97.1	88214
rheMac2	0.069	87.7	96.2	94.7	99.9	86.2	94.7	87992
calJac1	0.120	85.5	96.4	90.5	99.7	84.2	90.4	83033
mm9	0.174	65.9	96.7	82.7	96.2	65.0	81.7	65472
rn4	0.174	65.7	96.7	84.0	97.5	64.8	83.5	59323
bosTau4	0.186	79.3	96.3	85.8	95.1	78.1	84.0	76322
turTru1	0.186	83.6	96.5	89.2	98.2	82.4	88.4	75751
vicPac1	0.195	80.6	96.5	85.4	97.3	79.4	84.4	55289
canFam2	0.199	79.5	96.4	83.7	96.9	78.3	82.7	80552
felCat3	0.199	79.9	96.5	84.3	96.5	78.8	83.1	53158
choHof1	0.211	79.3	96.3	82.7	95.3	78.2	81.1	49130
dasNov2	0.211	77.2	96.4	81.0	94.0	76.2	79.0	48334
micMur1	0.220	84.4	96.3	85.3	97.8	83.2	84.5	63285
otoGar1	0.220	80.2	96.3	81.6	96.7	79.1	80.6	57445
loxAfr2	0.236	79.2	96.2	83.7	94.8	78.1	81.9	53365
proCap1	0.236	74.3	96.6	80.9	91.4	73.3	78.1	47590
tarSyr1	0.246	82.4	96.3	83.0	97.7	81.4	82.3	60541
equCab2	0.247	83.8	96.4	87.3	98.3	82.6	86.7	82307
myoLuc1	0.253	79.8	96.1	83.5	97.1	78.4	82.7	31309
pteVam1	0.253	80.1	96.4	83.8	98.0	78.9	83.1	59906
ochPri2	0.314	68.4	96.5	74.4	83.9	67.6	69.6	40361
oryCun1	0.314	74.0	96.5	76.4	92.5	73.1	74.0	49759
tupBel1	0.321	78.0	96.0	78.4	96.4	76.9	77.4	53481
speTri1	0.330	75.0	96.2	76.3	94.9	74.0	74.6	50565
galGal3	0.337	59.0	90.0	71.8	58.9	56.4	58.9	2000
echTel1	0.350	70.3	96.4	75.4	89.6	69.3	72.7	31358
cavPor3	0.380	69.4	96.5	71.9	91.0	68.5	69.7	70481
dipOrd1	0.389	68.6	96.4	71.6	90.7	67.7	69.2	41512
eriEur1	0.407	70.3	96.2	73.4	92.0	69.4	71.4	20764
fr2	0.430	77.6	95.8	83.0	93.2	76.3	80.6	56502
tetNig1	0.430	49.4	87.4	63.3	63.9	47.1	55.6	3203
sorAra1	0.453	69.1	96.1	72.4	89.1	68.1	70.3	22024
monDom4	0.714	58.1	95.3	61.5	70.3	56.8	56.6	11424
gasAcu1	0.770	48.0	89.2	63.1	60.4	46.6	52.7	3218
oryLat2	0.792	48.4	89.4	72.1	54.2	47.0	57.9	3234
ornAna1	0.852	56.7	96.8	59.5	59.2	55.9	54.3	3610
anoCar1	0.876	51.6	90.7	64.9	50.0	49.4	53.0	3021
petMar1	1.023	48.8	89.9	61.1	33.9	46.7	42.4	2833
danRer5	1.426	47.8	91.4	64.4	38.1	46.0	50.8	4378
xenTro2	1.493	54.3	89.5	70.1	38.1	52.0	51.4	3660

Table 2: **Data set MZ44-2:** Recovery rate in percent for consensus sequences derived from maxAlike probabilities (*ML* columns) and nucleotide frequencies (*Freq* columns), with a probability/relative frequency above a 0.5 threshold and without a threshold (i.e. the consensus sequence is made up from the nucleotides with highest probability/frequency at each position). A value of e.g. 70 means that 70% of the nucleotides in the consensus sequence were predicted correctly. Only species with at least 10 reconstructed sequences were considered. The last column (Nt.pred.) shows the total number of reconstructed nucleotides for each species, excluding fully conserved alignment columns. The fraction (in percent) of predicted nucleotides above the 0.5 threshold is shown in the columns "% ≥ 0.5 " for both methods. The Dist. column shows the distance to the phylogenetically nearest neighbor in the tree. (This table is the same as table 2 in the article.)

Species	Dist.	Threshold 0.5				No threshold		
		<i>Freq</i>	% \geq 0.5	<i>ML</i>	% \geq 0.5	<i>Freq</i>	<i>ML</i>	Nt.pred.
hg18	0.005	94.6	98.7	98.0	100.0	94.2	98.0	23755
panTro2	0.005	94.5	98.7	98.0	100.0	94.1	98.0	23374
baboon	0.008	93.4	98.8	98.1	99.7	92.9	97.9	23290
rheMac2	0.008	93.6	98.8	98.1	100.0	93.2	98.1	22342
vervet	0.011	93.8	99.2	97.7	100.0	93.4	97.7	14305
ponAbe2	0.012	93.9	98.7	96.4	100.0	93.5	96.4	23432
colobus_monkey	0.014	93.1	99.0	96.4	99.9	92.8	96.3	20930
gibbon	0.015	94.3	98.7	96.5	100.0	93.9	96.5	21867
marmoset	0.020	88.0	99.0	94.3	99.9	87.6	94.2	21468
owl_monkey	0.020	89.3	99.0	95.7	99.8	89.0	95.6	21569
squirrel_monkey	0.021	87.7	99.1	94.5	99.9	87.3	94.5	14941
dusky_titi	0.022	88.7	99.0	93.7	99.9	88.4	93.6	20990
mm9	0.046	57.9	98.5	83.3	99.6	57.5	83.2	5613
rn4	0.046	57.7	98.4	83.7	99.7	57.3	83.5	5610
galago	0.063	66.0	99.1	72.4	96.4	65.7	71.2	14184
mouse_lemur	0.063	72.9	99.1	79.4	96.9	72.6	78.3	12461
canFam2	0.064	63.5	98.6	77.3	95.4	63.0	75.7	9651
cat	0.064	63.4	98.7	78.5	95.6	63.0	77.1	8628
flying_fox	0.068	67.9	98.8	81.3	98.3	67.6	80.5	1928
rfbat	0.068	65.7	98.7	77.4	97.0	65.2	76.4	10475
sbbat	0.077	62.8	99.0	76.0	96.5	62.5	74.7	3294
horse	0.078	65.6	98.9	79.3	96.4	65.3	78.2	5446
elephant	0.088	64.0	98.6	70.5	95.6	63.6	69.0	4467
bosTau3	0.092	62.9	98.5	73.5	96.9	62.5	72.5	7652
armadillo	0.101	59.9	98.7	65.9	94.4	59.5	64.9	8513
st_squirrel	0.101	63.1	98.6	68.1	96.2	62.8	67.2	7612
rabbit	0.112	58.2	98.9	62.9	94.2	58.0	61.8	7426
guinea_pig	0.114	57.7	98.6	62.7	93.7	57.3	61.7	5690
shrew	0.139	51.6	99.0	59.1	90.7	51.3	57.8	2380

Table 3: **Data set *ENC-1***: Recovery rate in percent for consensus sequences derived from maxAlike probabilities (*ML* columns) and nucleotide frequencies (*Freq* columns), with a probability/relative frequency above a 0.5 threshold and without a threshold (i.e. the consensus sequence is made up from the nucleotides with highest probability/frequency at each position). A value of e.g. 70 means that 70% of the nucleotides in the consensus sequence were predicted correctly. Only species with at least 10 reconstructed sequences were considered. The last column (Nt.pred.) shows the total number of reconstructed nucleotides for each species, excluding fully conserved alignment columns. The fraction (in percent) of predicted nucleotides above the 0.5 threshold is shown in the columns "% \geq 0.5" for both methods. The Dist. column shows the distance to the phylogenetically nearest neighbor in the tree.

Species	Dist.	Threshold 0.5				No threshold		
		<i>Freq</i>	% \geq 0.5	<i>ML</i>	% \geq 0.5	<i>Freq</i>	<i>ML</i>	Nt.pred.
hg18	0.005	88.8	98.1	95.9	100.0	88.0	95.9	120226
panTro2	0.005	88.5	98.2	95.9	100.0	87.7	95.9	115311
baboon	0.008	85.1	98.2	94.5	97.7	84.3	93.4	106091
rheMac2	0.008	85.5	98.2	93.6	99.9	84.7	93.5	103849
vervet	0.011	84.5	98.4	91.8	100.0	83.8	91.8	26945
ponAbe2	0.012	87.5	98.1	91.6	99.9	86.7	91.5	110647
colobus_monkey	0.014	83.8	98.3	89.3	99.8	83.1	89.3	93992
gibbon	0.015	86.6	98.2	90.0	99.8	85.8	89.9	96875
marmoset	0.020	65.3	98.4	85.0	99.6	64.7	84.8	96971
owl_monkey	0.020	67.6	98.5	87.3	99.5	67.0	87.1	94971
squirrel_monkey	0.021	65.1	98.3	84.4	99.7	64.5	84.3	44555
dusky_titi	0.022	66.6	98.5	84.1	99.2	66.1	83.8	93652
mm9	0.046	52.4	93.9	77.4	94.2	50.5	74.4	5543
rn4	0.046	51.4	97.4	83.4	100.0	50.9	83.4	4090
galago	0.063	47.6	96.0	52.5	88.1	46.5	50.6	18686
mouse_lemur	0.063	56.5	96.5	61.0	87.5	55.2	58.5	17655
canFam2	0.064	55.8	97.1	68.9	92.6	54.9	67.2	11644
cat	0.064	54.2	97.6	69.3	92.7	53.8	67.8	8492
rfbat	0.068	60.0	98.0	72.3	93.3	59.5	70.3	11251
sbbat	0.077	56.9	98.2	74.7	94.4	56.4	73.2	2680
horse	0.078	56.3	98.2	66.4	90.9	55.9	64.7	3116
elephant	0.088	51.8	95.5	58.1	87.6	50.0	55.5	4261
bosTau3	0.092	54.7	97.7	64.8	94.4	54.3	63.5	7896
armadillo	0.101	52.8	97.7	56.9	90.9	52.4	55.8	7456
st_squirrel	0.101	52.1	97.2	56.6	91.6	51.5	55.6	7545
tenrec	0.102	50.0	94.2	58.0	76.2	47.7	53.9	2378
rabbit	0.112	53.9	96.4	58.6	91.4	52.7	56.4	5574
guinea_pig	0.114	50.9	97.8	56.1	88.9	50.7	54.9	7353
shrew	0.139	45.9	97.6	51.6	91.2	46.0	50.9	1625

Table 4: **Data set *ENC-2***: Recovery rate in percent for consensus sequences derived from maxAlike probabilities (*ML* columns) and nucleotide frequencies (*Freq* columns), with a probability/relative frequency above a 0.5 threshold and without a threshold (i.e. the consensus sequence is made up from the nucleotides with highest probability/frequency at each position). A value of e.g. 70 means that 70% of the nucleotides in the consensus sequence were predicted correctly. Only species with at least 10 reconstructed sequences were considered. The last column (Nt.pred.) shows the total number of reconstructed nucleotides for each species, excluding fully conserved alignment columns. The fraction (in percent) of predicted nucleotides above the 0.5 threshold is shown in the columns "% \geq 0.5" for both methods. The Dist. column shows the distance to the phylogenetically nearest neighbor in the tree.

Species	Dist.	<i>ML</i>	<i>Freq</i>	Δ
hg18	0.013	1.000	0.964	0.036
panTro2	0.013	1.000	0.965	0.035
gorGor1	0.018	1.000	0.965	0.035
ponAbe2	0.037	1.000	0.963	0.037
rheMac2	0.069	0.969	0.957	0.012
calJac1	0.120	0.951	0.943	0.008
mm9	0.174	0.926	0.826	0.100
rn4	0.174	0.924	0.830	0.094
bosTau4	0.186	0.925	0.906	0.020
turTru1	0.186	0.941	0.925	0.017
vicPac1	0.195	0.929	0.914	0.015
canFam2	0.199	0.922	0.913	0.010
felCat3	0.199	0.930	0.908	0.022
choHof1	0.211	0.912	0.896	0.017
dasNov2	0.211	0.895	0.876	0.019
micMur1	0.220	0.930	0.928	0.003
otoGar1	0.220	0.904	0.903	9e-04
loxAfr2	0.236	0.924	0.901	0.022
proCap1	0.236	0.895	0.869	0.027
tarSyr1	0.246	0.905	0.909	-0.004
equCab2	0.247	0.930	0.927	0.003
myoLuc1	0.253	0.903	0.892	0.011
pteVam1	0.253	0.909	0.905	0.004
ochPri2	0.314	0.868	0.831	0.038
oryCun1	0.314	0.892	0.867	0.024
tupBell	0.321	0.885	0.892	-0.006
speTri1	0.330	0.882	0.884	-0.002
galGal3	0.337	0.840	0.765	0.075
echTel1	0.350	0.852	0.830	0.021
cavPor3	0.380	0.853	0.847	0.006
dipOrd1	0.389	0.856	0.846	0.010
eriEur1	0.407	0.863	0.852	0.011
fr2	0.430	0.854	0.736	0.118
tetNig1	0.430	0.838	0.715	0.123
sorAra1	0.453	0.859	0.843	0.016
monDom4	0.714	0.793	0.771	0.022
gasAcu1	0.770	0.861	0.726	0.135
oryLat2	0.792	0.839	0.723	0.115
ornAna1	0.852	0.783	0.737	0.046
anoCar1	0.876	0.794	0.718	0.076
petMar1	1.023	0.805	0.711	0.095
danRer5	1.426	0.816	0.690	0.126
xenTro2	1.493	0.806	0.710	0.097

Table 5: **Data set MZ44-1:** Median MATCH scores of the maxAlike PSSMs (*ML*) and the frequency PSSMs (*Freq*) for 10 randomly selected 30nt windows from each alignment. The last column shows the difference between both medians. Only species with more than 100 predicted PSSMs were included.

Species	Dist.	<i>ML</i>	<i>Freq</i>	Δ
hg18	0.013	1.000	0.957	0.043
panTro2	0.013	1.000	0.957	0.043
gorGor1	0.018	1.000	0.958	0.042
ponAbe2	0.037	0.997	0.954	0.043
rheMac2	0.069	0.967	0.946	0.021
calJac1	0.120	0.934	0.925	0.010
mm9	0.174	0.895	0.760	0.135
rn4	0.174	0.895	0.762	0.134
bosTau4	0.186	0.901	0.870	0.030
turTru1	0.186	0.935	0.908	0.026
vicPac1	0.195	0.898	0.880	0.018
canFam2	0.199	0.890	0.869	0.021
felCat3	0.199	0.903	0.876	0.027
choHof1	0.211	0.890	0.867	0.023
dasNov2	0.211	0.872	0.847	0.025
micMur1	0.220	0.901	0.899	0.002
otoGar1	0.220	0.871	0.864	0.006
loxAfr2	0.236	0.900	0.867	0.034
proCap1	0.236	0.867	0.822	0.045
tarSyr1	0.246	0.875	0.884	-0.008
equCab2	0.247	0.908	0.905	0.003
myoLuc1	0.253	0.881	0.866	0.015
pteVam1	0.253	0.885	0.877	0.008
ochPri2	0.314	0.811	0.769	0.042
oryCun1	0.314	0.850	0.822	0.028
tupBell	0.321	0.840	0.847	-0.008
speTri1	0.330	0.833	0.830	0.002
galGal3	0.337	0.840	0.758	0.083
echTel1	0.350	0.829	0.794	0.035
cavPor3	0.380	0.795	0.781	0.014
dipOrd1	0.389	0.794	0.777	0.017
eriEur1	0.407	0.799	0.787	0.012
fr2	0.430	0.780	0.688	0.091
tetNig1	0.430	0.745	0.696	0.048
sorAra1	0.453	0.795	0.774	0.021
monDom4	0.714	0.710	0.680	0.030
gasAcu1	0.770	0.761	0.674	0.087
oryLat2	0.792	0.835	0.699	0.136
ornAna1	0.852	0.727	0.686	0.041
anoCar1	0.876	0.791	0.686	0.105
petMar1	1.023	0.732	0.649	0.082
danRer5	1.426	0.767	0.648	0.119
xenTro2	1.493	0.817	0.732	0.085

Table 6: **Data set MZ44-2:** Median MATCH scores of the maxAlike PSSMs (*ML*) and the frequency PSSMs (*Freq*) for 10 randomly selected 30nt windows from each alignment. The last column shows the difference between both medians. Only species with more than 100 predicted PSSMs were included. (This table is the same as table 1 in the article.)

Species	Dist.	<i>ML</i>	<i>Freq</i>	Δ
hg18	0.005	1.000	0.988	0.012
panTro2	0.005	1.000	0.990	0.010
baboon	0.008	1.000	0.981	0.019
rheMac2	0.008	1.000	0.982	0.018
vervet	0.011	1.000	0.985	0.015
ponAbe2	0.012	1.000	0.983	0.017
colobus_monkey	0.014	1.000	0.980	0.020
gibbon	0.015	1.000	0.985	0.015
marmoset	0.020	0.968	0.958	0.011
owl_monkey	0.020	0.970	0.962	0.008
squirrel_monkey	0.021	0.968	0.958	0.010
dusky_titi	0.022	0.968	0.962	0.007
mm9	0.046	0.923	0.777	0.146
rn4	0.046	0.916	0.774	0.142
galago	0.063	0.879	0.847	0.032
mouse_lemur	0.063	0.900	0.868	0.032
canFam2	0.064	0.895	0.838	0.057
cat	0.064	0.904	0.836	0.068
flying_fox	0.068	0.908	0.866	0.042
rfbat	0.068	0.897	0.846	0.051
sbbat	0.077	0.883	0.847	0.037
horse	0.078	0.906	0.859	0.047
elephant	0.088	0.865	0.836	0.029
bosTau3	0.092	0.861	0.817	0.044
armadillo	0.101	0.828	0.802	0.027
st_squirrel	0.101	0.842	0.814	0.028
rabbit	0.112	0.809	0.783	0.026
guinea_pig	0.114	0.810	0.780	0.029

Table 7: **Data set *ENC-1***: Median MATCH scores of the maxAlike PSSMs (*ML*) and the frequency PSSMs (*Freq*) for 10 randomly selected 30nt windows from each alignment. The last column shows the difference between both medians. Only species with more than 100 predicted PSSMs were included.

Species	Dist.	<i>ML</i>	<i>Freq</i>	Δ
hg18	0.005	1.000	0.987	0.013
panTro2	0.005	1.000	0.986	0.014
baboon	0.008	1.000	0.978	0.022
rheMac2	0.008	1.000	0.980	0.020
vervet	0.011	0.991	0.978	0.012
ponAbe2	0.012	0.986	0.983	0.002
colobus_monkey	0.014	0.974	0.973	8e-04
gibbon	0.015	0.970	0.980	-0.010
marmoset	0.020	0.966	0.939	0.027
owl_monkey	0.020	0.969	0.945	0.024
squirrel_monkey	0.021	0.965	0.938	0.028
dusky_titi	0.022	0.965	0.940	0.025
mm9	0.046	0.859	0.742	0.118
rn4	0.046	0.906	0.744	0.162
galago	0.063	0.801	0.777	0.024
mouse_lemur	0.063	0.857	0.830	0.026
canFam2	0.064	0.863	0.808	0.055
cat	0.064	0.867	0.802	0.064
rfbat	0.068	0.880	0.836	0.043
sbbat	0.077	0.892	0.829	0.063
horse	0.078	0.869	0.828	0.041
elephant	0.088	0.817	0.781	0.036
bosTau3	0.092	0.817	0.786	0.032
armadillo	0.101	0.798	0.774	0.024
st_squirrel	0.101	0.786	0.769	0.018
tenrec	0.102	0.792	0.734	0.058
rabbit	0.112	0.770	0.752	0.018
guinea_pig	0.114	0.768	0.742	0.026
shrew	0.139	0.763	0.733	0.030
monDom4	0.230	0.591	0.564	0.027

Table 8: **Data set *ENC-2***: Median MATCH scores of the maxAlike PSSMs (*ML*) and the frequency PSSMs (*Freq*) for 10 randomly selected 30nt windows from each alignment. The last column shows the difference between both medians. Only species with more than 100 predicted PSSMs were included.