

Primate phylogeny: molecular evidence for a pongid clade excluding humans and a prosimian clade containing tarsiers

1 Genetic non-equidistance to a more complex outgroup despite equidistance in time

Table S1 Human relationship with mollusks. The percentage identities in protein sequence between species (*Octopus vulgaris*, *Acanthocardia tuberculatum*, and *Homo sapiens*) are shown for 10 mitochondrial proteins

Proteins	Percentage identity		
	H.s.-O.v.	H.s.-A.t.	O.v.-A.t.
COX1	75	60	63
COX2	53	34	37
COX3	66	36	41
ND1	46	39	44
ND2	32	30	31
ND3	43	<20	32
ND4	38	<38	40
ND5	33	<33	43
COB	57	49	50
ATP6	40	19	24

Table S2 Human relationship with brachiopods. The percentage identities in protein sequence between species (*Terebratulina retusa*, *Lingula anatina*, and *Homo sapiens*) are shown for 10 mitochondrial proteins.

Proteins	Percentage identity		
	H.s.-T.r.	H.s.-L.a	T.r.-L.a.
COX1	74	55	51
COX2	56	<33	33
COX3	62	38.46	38.28
ND1	50	40	41
ND2	28	26	29
ND3	47	36	<36
ND4	38	36	39
ND5	37	35	38
COB	59	47	48
ATP6	26	23	25

Table S3 The reptile clade (including birds): human is closer to birds than to snakes. The percentage identities in protein sequence between species (birds, snakes, and humans) are shown for 10 mitochondrial proteins and 13 randomly selected proteins encoded by the nuclear genome. The number was from BLASTP analysis of bird or snake database from Genbank and represent the highest identity. The mitochondrial proteins show that snakes are more distant to humans than birds are ($P<0.05$). A random sampling of 13 nuclear genes also showed the same result ($P<0.05$).

	Percentage Identity		
	Hu.-Bird	Hu.-Sn.	Bird-Sn.
ND1	70	64	67
ND2	50	45	46
ND3	54	48	62
ND4	60	53	55
ND5	57	52	53
ND6	36	32	41
COB	73	64	67
COX1	86	76	77
COX2	68	56	58
COX3	76	72	70
Cytochrome C	91	87	81
Albumin	47	32	30
HBA	74	65	60
HBB	70	67	71
ACTB	100	99	93
MC1R	62	57	64
ENO1	93	90	94
FBP1	80	74	75
MOS	68	64	73
Rag1	76	74	75
Rag2	72	66	70
Jun	68	66	82
Adam1a	51	42	41

Table S4 The amphibian group. The percentage identities in protein sequence between species (*Xenopus laevis*, *Limnonectes fujianensis*, and *Homo sapiens*) are shown for 12 randomly selected proteins. The number was from BLASTP analysis of GenBank. The data show that *Xenopus laevis* is closer to humans than *Limnonectes fujianensis* is, but more proteins need to be sampled to confirm the significance of this trend ($P=0.06$). *Limnonectes fujianensis* is closer to *Xenopus laevis* than to humans ($P=0.01$), consistent with a closer phylogenetic relationship between the two frogs.

	Percentage identity		
	H.s.-X.l.	H.s.-L.f.	X.l.-L.f.
COX1	87	80	84
COX2	70	63	74
COX3	80	77	80
COB	73	69	79
ND1	64	64	76
ND2	50	44	59
ND3	58	50	68
ND4	59	<49	49
ND5	57	47	50
ATP6	52	53	66
Tyrosinase	71	67	72
Rhodopsin	84	79	86

Table S5 The teleost fish group: human is closer to the loach than to the three spined frogfish. The percentage identities in protein sequence between species (*Vaillantella maassi*, *Batrachomoeus trispinosus*, and *Homo sapiens*) are shown for 13 mitochondrial proteins. The mitochondrial proteins show that the loach *Vaillantella maassi* is significantly closer to humans than the three spined frogfish *Batrachomoeus trispinosus* is ($P=0.005$). The data suggest that some teleost fishes are closer to humans than others, presumably due to higher epigenetic complexity. Future work is needed to determine if the loach is indeed more complex than the frogfish. Also, the frogfish is closer to the loach than to humans ($P=0.03$), consistent with a closer phylogenetic relationship between the two fishes.

	Percentage identity		
	H.s.-V.m.	H.s.-B.t.	V.m.-B.t.
ND1	66	60	63
ND3	60	53	65
ND2	49	45	53
ND4	60	56	59
ND5	62	55	53
COB	70.79	67	70.13
COX1	85	82	84
COX2	68	59	62
COX3	80	71	72
ATP6	50	45	54
ND6	34	30	46
ND4L	52	39	59
ATP8	27	<27	55

Table S6 The echinoderm phylum. The percentage identities in protein sequence between species (*Strongylocentrotus purpuratus*, *Ophiura lutkeni*, and *Homo sapiens*) are shown for 11 mitochondrial proteins. Using COX1 and COB proteins of humans as query, the sea urchin (*Strongylocentrotus purpuratus*) was identified as among the closest to humans, while the starfish (*Ophiura lutkeni*) was found among the most distant. A sampling of 11 proteins shows that sea urchin is slightly closer to humans than the starfish is ($P=0.19$). Future work with more proteins will be needed to determine if this trend is significant. The starfish is slightly closer to sea urchins than to humans ($P=0.07$), consistent with a clade containing the starfish and sea urchins.

	Percentage identity		
	H.s.-S.s.	H.s.-O.l.	S.s.-O.l.
COX1	76	71	73
COX2	62	50	57
COX3	63	58	57
COB	63	64	69
ND1	57	55	58
ND2	40	33	34
ND3	50	47	61
ND4	46	40	47
ND5	45	51	52
ND6	30	<30	32
ATP6	42	<40	40

Table S7 The arthropod phylum: human is closer to the dragonfly than to the louse. The percentage identities in protein sequence between species (*Orthetrum triangulare melania*, *Campanulotes bidentatus compar*, and *Homo sapiens*) are shown for 10 mitochondrial proteins. The wingless louse (*Campanulotes bidentatus compar*) was identified as among the most distant to humans as measured by a randomly chosen protein COX1. The dragonfly (*Orthetrum triangulare melania*) was identified as among the closest to humans among arthropods. The distance of these two species to humans was next determined using ten mitochondrial proteins. Humans are significantly closer to the dragonfly than to the louse ($P<0.05$). This suggests that the dragonfly is more complex than the wingless louse, which is consistent with fact that the former can fly. The louse is not significantly closer to dragonfly than to human ($P=0.35$), suggesting that the distance between the two insects is close to the maximum.

	Percentage identity		
	H.s.-O.t.m	H.s.-C.b.c.	O.t.m.-C.b.c.
COX1	77	69	68
COX2	54	47	49
COX3	62	52	51
COB	63	46	51
ATP6	46	37	45
ND1	49	41	44
ND2	37	18	23
ND3	46	<20	34
ND4	45	35	37
ND5	40	33	42

Table S8 The nematode phylum. The percentage identities in protein sequence between species (*Cooperia oncophora*, *Brugia malayi*, and *Homo sapiens*) are shown for 10 randomly selected proteins. Using COX1 and COB proteins of humans as query, *Cooperia oncophora* was identified as among the closest to humans, while *Brugia malayi* was found among the most distant. A sampling of 11 proteins showed that there is a trend ($P=0.06$) for a closer relationship between *Cooperia oncophora* and human. *Brugia malayi* is not significantly closer to *Cooperia oncophora* than to humans, suggesting that the distance between the two nematodes is close to the maximum.

	Percentage identity		
	H.s.-C.o.	H.s.-B.m.	C.o.-B.m.
COX1	73	49	56
COX2	48.48	49.09	47
COX3	42	29	32
COB	44	43	52
ND1	36	34	50
ND4	33	32	45
ND5	34	31	42
Actin b	97.59	97.33	96
Tubulin b	89.62	89.14	87
KCNMA1	59	58	78

Table S9 The porifera phylum: human is closer to the chicken liver sponge than to *H. lachne*. The percentage identities in protein sequence between species (*Chondrilla aff. nucula*, *Hippospongia lachne*, and *Homo sapiens*) are shown for 10 mitochondrial proteins. Using COX1 and COB proteins of humans as query, the chicken liver sponge (*Chondrilla aff. nucula*) was identified as among the closest to humans, while *Hippospongia lachne* was found among the most distant. A sampling of 10 proteins showed that humans are significantly closer to *Chondrilla aff. nucula* than to *Hippospongia lachne* ($P<0.05$). However, *Hippospongia lachne* is not the sister taxon of a human-*Chondrilla* clade since it is closer to *Chondrilla aff. nucula* than to humans ($P<0.05$).

	Percentage identity		
	H.s.-C.a.n.	H.s.-H.l.	C.a.n.-H.l.
COX1	71	66	74
COX2	57	47	54
COX3	58	48	60
COB	64	48	60
ATP6	40	38	46
ND1	50	47	63
ND2	31	27	41
ND3	39	38	54
ND4	34	<34	55
ND5	45	42	53

Table S10 The fungi kingdom: human is closer to the corn smut than to yeast. The percentage identities in protein sequence between species (*Ustilago maydis*, *Candida zemplinina* or *Candida*, and *Homo sapiens*) are shown for 20 random selected proteins. Using COX1 and COB of humans as query, the smut fungus *Ustilago maydis* was identified among the closest to humans, while the yeast *Candida zemplinina* was among the most distant to humans. A sampling of five proteins (few *C. zemplinina* protein sequences are known) showed that the smut fungus is closer to humans than the yeast. To confirm that the smut fungus is indeed closer to humans than the *Candida* genus, 15 more proteins were randomly sampled. Among different *Candida* species, the one showing the highest identity with human is shown in the Table. The smut is closer to humans than *Candida* is in 19 of 20 proteins ($P = 0.003$). The data suggest that the smut has higher epigenetic complexity than the yeast, consistent with the status of this fungus as 'Higher Fungi'. However, *Candida* is not an outgroup to a human-smut clade since it is closer to smut than to humans ($P = 0.04$).

	Percentage identity		
	H.s.-U.m.	H.s.-C.z.	U.m-C.z
COX1	65	51	56
COX2	48	46	60
COX3	51	40	42
COB	55	46	57
ATP6	35	30	44
		H.s.-Candida	
ND1	45	41	50
Actin b	91	88.8	88.59
Q71U36	76	73	72
Tubulin b	82	74.88	74.83
Calmodulin	89	70	71
MnSOD	54	45	50
Enolase 1	61	60	66
FBP1	53	47	56
AAH06168	51	45	50
PGK1	66.26	66.02	68
Pyruvate kinase	53	51	58
AAA02807	41	37	49
TPI1	58	51	59
ND5	39	42	43
ND4	31	30	46

Table S11 The protist alveolates superphylum. The percentage identities in protein sequence between species (*Plasmodium falciparum*, *Tetrahymena thermophila*, and *Homo sapiens*) are shown for 11 random selected proteins. Using COX1 of humans as query, the malaria parasite *Plasmodium* (phylum Apicomplexa) was identified among the closest to humans, while *Tetrahymena* (phylum Ciliophora) was among the most distant. However, a sampling of 11 proteins showed that, relative to *Tetrahymena*, *Plasmodium* is closer to humans in 5 proteins but more distant in 6 proteins. Thus, the two species are equidistant to humans. Coincidence and common selection may account for the large differences in identity to humans between the two species in some proteins such as COX1, COB, and GPDH. The two protists are also no closer than either is to humans, suggesting that the separation time for the two protists has been long enough for their genetic distance to reach the maximum cap.

	Percentage identity		
	H.s.-P.f.	H.s.-T.t.	P.f.-T.t.
COX1	46	36	34
COX2	54	<39	39
COB	42	77	23
Actin b	84	77	75
Tubulin b	89	90	95
Camodulin	87	90.13	90.6
MnSOD	42	45	38
Enolase	63	61	66
GPDH	43	27	28
Pyruvate kinase	44.17	44.33	47
TPI1	43	49	45

Other groups:

The annelida phylum

Using COX1 and COB of humans as query, seven species of annelida were matched with similar distance to humans. The distance between human and annelida is similar to the maximum distance within the phylum. The data suggests either that there is little difference in epigenetic complexity within this phylum or that not enough species have been sampled.

The platyhelminthes phylum

Using COX1 of humans as query, the *Pseudostylochus intermedius* (65% identity) was identified as among the closest to humans, while *Sparganum proliferum* (47%) was found among the most distant. However, few other proteins are known for these two species. It remains unclear therefore whether there exists a species in this phylum that is closer to humans than others are.

The cnidaria phylum

Using COX1 and COB proteins of humans as query, all the matched species of cnidaria (~ 30 species) are about equidistant to humans. However, the maximum distance among the species of cnidaria is smaller than the distance between cnidaria and human. This suggests either that not enough cnidaria species have been sampled or that cnidaria has evolved cnidaria-specific conserved domains since separating from the human line but before divergence of most species of cnidaria. About half of the six classes of cnidaria have not been sequenced at least for the COX1 and COB proteins.

The plant kingdom

In contrast to animal phyla where complex species show more identity with humans than simpler species, complex plants (flowering plants) that appeared later in evolution and simpler plants (mosses) that appeared earlier are about equidistant to mammals in several randomly analyzed genes (EF1a, Adh1a, EIF2b, Pin1, PP1, RPC1, and Cox1). However, there is a distinct difference between the plant kingdom and the animal phyla. The identity between flowering plants and mosses are much greater than between mammals and mosses (e.g., for EF1a, human is 77% identical to either mosses or apple tree but the identity between mosses and apple tree is 93%). This is in stark contrast to animal phyla where the maximum distance between human and a simple animal phylum is similar to the maximum distance of sister species of the simple animal phylum. Thus, plants have evolved plant-specific conserved domains since separating from humans but before divergence of mosses and flowering plants. Complex plants would show less genetic diversity but the conserved residues are distinctly plant specific. The biochemical pathways for building complex plants are different from those for building complex animals. In contrast, the pathways for building complex invertebrate animals are still shared with those for building complex vertebrate animals. Thus, complex invertebrates would have more sequences in common with complex vertebrates than simple invertebrates do. However, complex plants do not have more in common with complex vertebrates than simple plants do.

The bacteria kingdom

Using COX1 of humans as query, the bacterium *Magnetospirillum magnetotactillum* was identified among the closest to humans (59% identity), while *Gemmata obscuriglobus* was among the most distant to humans (38% identity). However, one can easily identify a randomly selected protein, such as GCAT, that shows more identity between human and *G. obscuriglobus* (56%) than between human and *M. magnetotactillum* (37%). Indeed, despite numerous efforts, no one has been able to identify a bacterium lineage that is significantly closer to humans in most genes than other sister lineages. According to the MGD hypothesis, the great genetic diversity of bacteria makes possible fortuitous resemblance between a bacterium protein and a human protein.

2 Difference between slow and fast evolving genes in phylogeny inference

Table S12 Difference between slow and fast evolving genes in phylogeny inference. The percentage identities between zebrafish (*D. rerio*) and pufferfish (*T. nigroviridis*), human (*H. sapiens*), or mouse (*M. musculus*) are shown for a number of lysine methyltransferases (KMTs) and ribosome proteins. Genes are considered as having reached maximum distance in fishes if the identity between the two fishes is equal to or slightly smaller than that between fish and mammal.

	Percentage Identity		
	<i>D. rerio</i> vs.		
	<i>T. nigroviridis</i>	<i>H. sapiens</i>	<i>M. musculus</i>
<i>Genes reached cap</i>			
KMT family			
Suv39H1/KMT1A	61	63	62
Smyd2/KMT3C	70	75	70
SET7/9/KMT7	71	73	73
PRDM11	61		64
PRDM4	57	59	59
PRDM15	60	63	63
Ribosome family			
L11	97	97	95
S2	97	97	96
<i>Genes not yet reached cap</i>			
KMT family			
KMT5B	59	53	54
EZH2/KMT6	82	77	76
PRDM2/KMT8	48	41	43
Ribosome family			
L13	92	87	86
S19	89	88	88
L12	93	91	90
L14	83	72	72
L9	91	89	89
S11	92	91	91
S3	96	95	95
S13	98	97	96
L3	92	89	89
L7	85	79	80

3 Pongo abelli is closer to Pan troglodytes than to Homo sapiens

Table S13 Divergence time between homo and *Pongo abelli* based on sequences from *Pongo abelli*, *Pan troglodytes*, and *Homo sapiens*. Of 733 randomly selected cDNA sequences from *P. abelli* (NCBI accession number, CAI29673 to CAI29581, CAH93520 to CAH93492, CAH92004 to 91825, CAH91005 to CAH90750, and CAH90602 to CAH90424), 218 sequences are informative and listed here. 68 have greater than 98% identity between *P. abelli* and *P. troglodytes* and are considered as slow evolving proteins, while the other 149 proteins have identities between *P. abelli* and *P. troglodytes* that are equal to or smaller than 98% and are considered fast evolving. Among fast evolving genes, 66 showed higher identity between orangutans and chimpanzees while 83 showed less ($P = 0.35 >> 0.05$). In contrast, among slow evolving genes, 53 showed higher identity between orangutans and chimpanzees while 15 showed less ($P < 0.001$).

Divergence time between orangutan and human was calculated based on the fossil split time of gorilla of 12 Myr ago. Since gorilla and chimpanzee are equidistant genetically to orangutans (see Table 4 of main text), they are also equidistant in time to orangutans. So the split time between chimpanzees and orangutans is also 12 Myr, which was used to calculate the divergence time between humans and orangutans. The average divergence time was calculated using slow evolving genes. Eight genes from the list showing greater similarity between orangutans and chimpanzees are excluded in the calculation because they are non-informative (ni) due to 100% identity between orangutans and chimpanzees. To compensate for this loss of genes showing the greatest

time of split between orangutans and humans, eight genes from the list showing less similarity between orangutans and chimpanzees are also excluded, which show the smallest distance between orangutans and humans.

	# identical a.a		% Identity	Div. time (Myr)
	P.a.-H.s	P.a.-P.t.	P.a.-P.t.	P.a.-H.s.
<i>P.a.-P.t. > P.a.-H.s., Slow evolving, 53 genes:</i>				
CAI29661	356	357/359	99	18
CAI29655	286	287/289	99	17.9
CAI29649	298	299/300	99	24.3
CAI29646	639	640/646	99	14.1
CAI29644	518	519/522	99	15.9
CAI29642	205	206/207	99	24.3
CAI29638	567	568/573	99	14.4
CAI29630	391	392/393	99	24.5
CAI29627	535	536/541	99	14.4
CAI29608	390	405/405	100	ni
CAI29586	710	712/714	99	24.1
CAH93510	450	452/452	100	ni
CAH93506	738	739/739	100	ni
CAH91980	346	347/348	99	24.7
CAH91971	468	469/471	99	17.8
CAH91961	380	381/383	99	18.1
CAH91957	495	497/500	99	20.1
CAH91948	234	235/237	99	18
CAH91922	299	300/302	99	18.2
CAH91891	569	571/574	99	20.2
CAH91877	906	907/910	99	16
CAH91875	241	242/243	99	24.2
CAH91872	402	403/407	99	15
CAH91839	398	399/402	99	16
CAH91837	463	465/469	99	18
CAH91836	481	482/483	99	23.7
CAH91832	302	303/303	100	ni
CAH91825	522	523/528	99	14.4
CAH90995	702	704/706	99	24.4
CAH90951	367	368/369	99	24.2
CAH90937	742	743/746	99	16.1
CAH90930	498	499/506	99	13.7
CAH90907	385	386/386	100	ni
CAH90905	364	365/366	99	24
CAH90891	531	532/535	99	16.1
CAH90860	516	517/522	99	14.4
CAH90849	730	732/739	99	15.5
CAH90848	297	298/298	100	ni
CAH90846	424	425/427	99	18
CAH90484	1383	1386/1393	99	17.3
CAH90800	879	881/888	99	15.5
CAH90788	342	345/347	99	30
CAH90758	407	409/412	99	20.1
CAH90750	263	264/266	99	18
CAH90597	300	301/304	99	16.1
CAH90590	216	217/217	100	ni
CAH90574	953	954/963	99	13.3
CAH90501	519	520/523	99	16
CAH90500	907	908/910	99	18
CAH90495	1524	1526/1539	99	20
CAH90473	703	704/707	99	16

CAH90462	332	333/333	100	ni
CAH90449	434	436/440	99	18.1
<i>P.a.-P.t. < P.a.-H.s., Slow evolving, 15 genes:</i>				
CAI29665	738	737/739	99	6.0ni
CAI29590	364	363/365	99	6.0ni
CAH93514	724	723/730	99	10.3
CAH93507	413	412/413	99	ni
CAH91998	475	471/475	99	ni
CAH91994	619	617/622	99	7.2
CAH91940	325	323/326	99	4.0ni
CAH91889	465	464/465	99	ni
CAH90999	876	873/876	99	ni
CAH90886	315	314/316	99	6
CAH90823	914	912/918	99	8
CAH90785	289	288/289	99	ni
CAH90589	1021	1020/1023	99	8
CAH90585	933	932/941	99	10.7
CAH90494	1060	1059/1066	99	10.3
Average				17.3±5.1
<i>P.a.-P.t. > P.a.-H.s., Fast evolving, 66 genes:</i>				
CAI29663	489	491/509	96	
CAI29658	560	561/567	98	
CAI29640	252	253/264	95	
CAI29629	480	482/494	97	
CAI29613	457	458/509	89	
CAI29605	162	163/165	98	
CAI29603	347	348/354	98	
CAI29599	337	339/346	97	
CAI29591	140	141/147	95	
CAI29589	144	145/147	98	
CAI29588	459	461/478	96	
CAI29581	674	675/687	98	
CAH91978	473	512/524	97	
CAH91970	236	239/243	98	
CAH91967	526	529/536	98	
CAH91955	521	523/530	98	
CAH91954	457	458/468	97	
CAH91938	153	155/161	96	
CAH91934	215	216/219	98	
CAH91890	230	233/238	97	
CAH91880	384	388/395	98	
CAH91852	298	300/305	98	
CAH91850	463	466/472	98	
CAH91846	526	529/538	98	
CAH91834	993	994/1013	98	
CAH91828	493	494/499	98	
CAH91001	418	419/425	98	
CAH90938	180	181/183	98	
CAH90926	683	686/701	97	
CAH90904	302	304/316	96	
CAH90894	1221	1222/1241	98	
CAH90889	569	570/581	98	
CAH90887	619	625/685	91	
CAH90867	235	237/245	96	
CAH90854	317	319/327	97	
CAH90837	456	457/463	98	

CAH90833	132	133/135	98
CAH90808	323	325/330	98
CAH90798	816	824/833	98
CAH90783	120	123/126	97
CAH90776	378	379/385	98
CAH90774	522	525/536	97
CAH90773	653	657/663	98
CAH90765	383	386/395	97
CAH90595	658	660/687	96
CAH90562	131	132/137	96
CAH90562	131	132/137	96
CAH90551	169	168/172	98
CAH90543	110	111/117	94
CAH90535	456	457/468	97
CAH90520	778	779/789	98
CAH90518	176	177/178	99
CAH90515	571	572/585	97
CAH90514	1673	1676/1730	96
CAH90510	427	428/433	98
CAH90503	722	730/763	95
CAH90498	193	195/197	98
CAH90490	521	524/544	96
CAH90489	511	512/519	98
CAH90480	768	770/782	98
CAH90445	419	420/426	98
CAH90438	308	310/314	98
CAH90436	164	165/168	98
CAH90426	202	203/206	98
CAH90424	433	435/446	97
CAH90558	959	961/973	98

P.a.-P.t. < P.a.-H.s., Fast evolving, 83 genes:

CAI29671	644	636/647	98
CAI29664	485	446/493	90
CAI29659	484	482/489	98
CAI29657	374	365/403	90
CAI29643	283	274/286	95
CAI29628	466	463/476	97
CAI29620	143	142/144	98
CAI29600	349	346/352	98
CAI29597	454	453/462	98
CAI29587	236	233/239	97
CAH93511	1077	1070/1087	98
CAH93509	854	833/858	97
CAH93503	599	598/621	96
CAH93496	731	730/742	98
CAH91993	502	500/511	97
CAH91991	740	737/745	98
CAH91983	533	527/546	96
CAH91981	312	308/315	97
CAH91977	115	112/116	96
CAH91974	254	221/270	80
CAH91953	585	583/637	91
CAH91941	455	451/458	98
CAH91926	244	240/244	98
CAH91920	550	548/562	97
CAH91911	214	211/227	94

CAH91910	189	188/193	97
CAH91905	730	729/738	98
CAH91897	344	342/349	97
CAH91892	377	374/379	98
CAH91886	281	280/292	95
CAH91883	464	452/465	97
CAH91882	665	661/689	95
CAH91876	953	938/963	97
CAH91874	526	509/529	96
CAH91866	743	739/752	98
CAH91859	212	210/213	98
CAH91843	160	158/166	95
CAH91829	344	343/348	98
CAH91826	430	428/434	98
CAH90991	352	350/360	97
CAH90990	377	369/411	89
CAH90986	607	605/614	98
CAH90978	528	525/547	95
CAH90972	314	309/315	98
CAH90971	618	615/628	97
CAH90967	253	252/260	96
CAH90966	772	773/397	96
CAH90957	534	531/538	98
CAH90936	425	424/434	97
CAH90925	418	417/423	98
CAH90924	930	929/949	97
CAH90923	335	334/341	97
CAH90919	590	586/592	98
CAH90917	815	813/860	94
CAH90916	208	205/212	98
CAH90915	313	311/321	96
CAH90883	430	428/433	98
CAH90879	616	611/624	97
CAH90878	330	329/343	95
CAH90850	302	301/306	98
CAH90835	574	573/579	98
CAH90827	794	777/814	95
CAH90818	522	521/529	98
CAH90789	336	335/345	97
CAH90751	251	249/259	96
CAH90598	793	790/798	98
CAH90573	1159	1148/1166	98
CAH90566	443	438/445	98
CAH90561	475	472/477	98
CAH90560	475	472/477	98
CAH90549	275	271/278	98
CAH90544	573	571/578	98
CAH90531	542	541/550	98
CAH90524	407	406/417	97
CAH90513	663	662/685	96
CAH90511	503	489/535	91
CAH90493	1896	1892/1914	98
CAH90475	1066	1064/1088	97
CAH90472	782	772/790	97
CAH90471	511	510/524	97
CAH90470	505	495/514	96

CAH90448	551	550/558	98
CAH90434	257	256/263	97

Table S14 *Pongo abelli* is closer to *Pan troglodytes* than to *Homo sapiens*. Of 4330 random cDNA sequences of *P. abelli* available from Genbank, every 433 sequences based on their numerical order of appearance on the NCBI webpage were selected to form an experimental group. Orthologous genes with greater than 98% identity between *P. abelli* and *P. troglodytes* were considered as slow evolving proteins, while genes with identities between *P. abelli* and *P. troglodytes* that are equal to or smaller than 98% are considered fast evolving. The meaning of C-O > H-O: the percentage identity between chimpanzees (C) and orangutans (O) is greater than between humans (H) and orangutans. Numbers in parenthesis indicate *P* values from Fisher's exact test (2 tailed).

Groups	Genes	# inform. genes	# genes	
	analyzed		C-O > H-O vs. C-O < H-O	
	Start-end		>98%	< or = 98%
1	CAH89494-	105	33 vs. 8	29 vs. 35
	CAH93283		(0.004)	(0.60)
2	CAH93282-	95	16 vs. 8	36 vs. 35
	CAH92848		(0.38)	(0.93)
3	CAH92847-	97	30 vs. 8	24 vs. 35
	CAH92409		(0.02)	(0.36)
4	CAH92408-	119	29 vs. 7	35 vs. 48
	CAH91971		(0.01)	(0.35)
5	CAH91970-	98	28 vs. 8	25 vs. 37
	CAH91540		(0.03)	(0.38)
6	CAH91539-	105	28 vs. 9	33 vs. 35
	CAH91107		(0.03)	(1.00)
7	CAH91106-	106	22 vs. 6	33 vs. 45
	CAH90673		(0.05)	(0.42)
8	CAH90672-	117	22 vs. 10	41 vs. 44
	CAH90236		(0.20)	(0.88)
9	CAH90235-	102	20 vs. 11	30 vs. 31
	CAH89803		(0.31)	(1.00)
10	CAH89802-	112	19 vs. 5	49 vs. 39
	CAH89369		(0.07)	(0.55)
Total	4330	1056	247 vs. 80 (< 0.0001)	335 vs. 384 (0.21)

4 Gorillas are closer to chimpanzees than to humans

Table S15 Gorillas are closer to chimpanzees than to humans. Of the 69 informative gorilla proteins listed here, 35 have greater than 97% identity between gorillas (Go) and chimpanzees (Chimp) and are considered as slow evolving proteins, while the other 34 proteins have identities between gorillas and chimpanzees that are equal to or smaller than 97% and are considered fast evolving. Among fast evolving proteins, 18 showed higher identity between gorillas and chimpanzees than between gorillas and humans while 16 showed less (*P* >> 0.05). In contrast, among slow evolving genes, 27 showed higher identity between gorillas and chimpanzees while 8 showed less (*P* = 0.03).

	# identical a.a.		% identity
	Go.-Hu.	Go.-Chimp.	Go.-Chimp.
<i>Go.-Ch. > Go.-Hu., Slow evolving, 27 genes:</i>			
APOE	310	314/317	99
NDUFAF1	321	322/327	99
T2R38	307	309/310	99
ASIP	128	129/131	99
GSC	256	257/257	100
PCDH11X	1333	1336/1347	99
Myostatin	374	375/375	100
GPR56	674	677/687	98
BRCA1	1119	1129/1141	98

RNaseA	152	153/156	98
SCML2	175	176/176	100
ASPM	3421	3427/3447	98
CCR5	348	350/352	99
Trim5	478	484/493	98
MCPH1	816	820/835	98
Saitohin	126	127/128	99
T2R48	274	276/280	98
MAPT	768	769/776	99
Leptin	144	145/146	99
PTTG1	199	200/202	99
T2R49	302	303/309	98
KLF14	318	323/323	100
T2R50	255	256/260	98
IRBP	310	311/314	99
KCNS1	520	522/526	99
CMAH	495/501	599/600	99
ALDH5A1	533	534/535	99
<i>Go.-Ch. < Go.-Hu., Slow evolving, 8 genes:</i>			
HTR1F	362	361/365	98
CHRM3	588	586/590	99
CORTBP2	1653	1651/1663	99
Rh50	403	401/409	98
C5AR1	337	334/340	99
MATN4	578	575/581	98
CX3CR1	165	164/166	99
AFP	605	602/609	98
<i>Go.-Ch. > Go.-Hu., Fast evolving, 18 genes:</i>			
NACA2	199	205/210	97
MRGX2	316	322/329	97
T2R41	296	297/307	96
Twist	187	196/203	96
ND5	536	540/594	90
ND3	106	107/115	93
Syncytin 1	524	526/538	97
rcPSMB3	194	199/204	97
rcNIP30	236	246/254	96
PABP3	611	614/632	97
CDC14B2	442	448/458	97
POM121	136	138/142	97
Siglec9	214	215/224	95
Loc122650	194	196/205	95
AMACIL2	314	321/338	94
APOBEC3G	363	367/384	95
COX2	218	222/227	97
Cob	350	353/379	92
<i>Go.-Ch. < Go.-Hu., Fast evolving, 16 genes:</i>			
CHRM5	288	277/294	94
A4GALT	322	320/327	97
ND6	168	166/174	95
T2R10	274	272/279	97
T2R4	270	269/275	97
MC1R	304	303/317	95
OR1D2	305	304/312	97
ND4	434	428/459	94
ND1	300	299/316	94

ATP6	212	206/226	91
RNAse3	157	156/160	97
T2R14	287	285/292	97
rcCDC20	453	442/456	96
GMCL2	502	498/513	97
ZNF80	260	256/273	93
OR3A1	309	304/315	96

5 Gibbons are the outgroup to a pongid-hominid clade

Table S16 Gibbons are equidistant to orangutans and humans. Of the 53 informative gibbon (*Hylobates lar*) proteins shown here, 19 have greater than 95% identity between gibbons (Gi) and orangutans (Orang) and are considered slow evolving, while the other 34 proteins have identities between gibbons and orangutans that are equal to or smaller than 95% and are considered fast evolving. Among fast evolving proteins, 13 showed higher identity between gibbons and orangutans than between gibbons and humans while 21 showed less ($P \gg 0.05$). Similarly, among slow evolving genes, 12 showed higher identity between gibbons and orangutans than between gibbons and humans while 7 showed less ($P \gg 0.05$). The data show that gibbons are equidistant to orangutans and humans in both slow and fast evolving genes.

	# identical a.a.		% identity
	Gi.-Hu.	Gi.-Orang.	Gi.-Orang.
<i>Gi.-Or. > Gi.-Hu., Slow evolving, 12 genes:</i>			
Hepcidin	80	82/85	96
PML	862	866/883	98
ASPM	3382	3383/3477	97
RBM1	374	382/385	99
VAT1	262	264/264	100
USP9Y	235	240/243	98
Foxp2	710	712/713	99
HLA132	155	158/162	97
CXCR4	345	346/347	99
IfnG	141	142/143	99
LZM	141	143/148	96
GPX3	222	223/226	98
<i>Gi.-Or. < Gi.-Hu., Slow evolving, 7 genes:</i>			
COX1	499	498/513	97
PPIA	165	164/165	99
ALDH5A1	528	524/535	97
GPX1	198	197/201	98
CCR2	101	100/103	97
GJB2	226	224/226	99
CRYGB	170	170/175	97
<i>Gi.-Or. > Gi.-Hu., Fast evolving, 13 genes:</i>			
DEFB132	88	91/95	95
MC1R	294	303/317	95
HBG2	143	145/147	95
MCPH1	787	799/839	95
CD209L2	144	145/162	94
ND4	399	404/458	88
ESX1	164	175/222	79
DEFB125	131	133/158	84
SPRY	445	448/494	90
TRIM5	445	447/494	90
PABP3	581	591/635	93
Semg2	516/582	516/567	91
XL	151	152/166	91
<i>Gi.-Or. < Gi.-Hu., Fast evolving, 21 genes:</i>			

PGR	894	888/933	95
DEFB120	87	84/88	95
TAF1L	884	883/923	95
HBG2	142	141/147	95
MBL2	228	225/235	95
SMCY	458	452/475	95
Mapt	752	731/776	94
Fut5	359	353/374	94
Cd209	374	362/439	82
TGIF2LX	213	204/241	84
ND2	278	265/321	82
ATP6	199	187/226	82
COX3	249	242/261	92
ND3	99	95/115	82
ND5	502	488/590	82
ND6	151	148/174	86
COB	339	325/379	85
TSSK2	266	252/270	88
FCAR	85	81/96	84
MIC	228	226/274	82
SOD1	145	143/154	92

6 Old World monkeys are the outgroup to an ape-human clade

Table S17 Old World monkeys are equidistant to gibbons and humans. Of the 34 informative Old World monkeys (macaque) proteins shown here, 18 have greater than 92% identity between macaque (Ma) and gibbons (Gi) and are considered slow evolving, while the other 16 proteins have identities between macaque and gibbons that are equal to or smaller than 92% and are considered fast evolving. Among fast evolving proteins, 8 showed higher identity between macaques and gibbons than between macaques and humans while 8 showed less ($P \gg 0.05$). Similarly, among slow evolving genes, 7 showed higher identity between macaques and gibbons than between macaques and humans while 11 showed less ($P \gg 0.05$). The data show that macaques are equidistant to gibbons and humans in both slow and fast evolving genes.

	# identical a.a.		% identity
	Ma.-Hu.	Ma.-Gi.	Ma.-Gi.
<i>Ma.-Gi. > Ma.-Hu., Slow evolving, 7 genes:</i>			
EnvFRD	512	519/538	96
DEFB1	63	66/68	97
DEFB107	59	62/66	93
ALDH5A1	519	524/548	95
VAT1	257	259/263	98
GPX2	186	187/190	98
TSSK2	348	350/362	96
<i>Ma.-Gi. < Ma.-Hu., Slow evolving, 11 genes:</i>			
RhBG 1	451	431/458	94
Fyn	535	505/537	94
Lck	493	492/509	96
ADRA2B	350	349/365	95
PML	840	838/882	95
GPX4	196	188/197	95
GPX3	217	216/226	95
GPX1	198	196/201	97
USP9X	252	251/252	99
UTX	227	226/228	99
TAF1L	903	883/923	95
<i>Ma.-Gi. > Ma.-Hu., Fast evolving, 8 genes:</i>			
DARC	309	312/336	92

TRIM5	170	173/203	85
DEFB120	78	79/87	90
RHAG	366/409	387/428	90
DEFB128	82	83/93	89
TMPRSS2	158	165/189	87
XL	358/409	366/415	88
MBL2	210/234	213/234	91
<i>Ma.-Gi. < Ma.-Hu., Fast evolving, 8 genes:</i>			
DEFB119	78	77/84	91
CD209	363	350/393	89
TGIF2LX	208	201/249	80
FCAR	67	63/92	68
Lysozyme	131	130/148	87
SMCY	448	443/486	91
DEFB105	69	66/78	84
MC1R	296	290/317	91

7 New World monkeys are the outgroup to an Old World monkey-ape-human clade

Table S18 New World monkeys are equidistant to Old World monkeys and humans. Of the 39 informative New World monkeys (*Saguinus*) proteins shown here, 17 have greater than 90% identity between *Saguinus* (Sa) and macaque (Ma) and are considered slow evolving, while the other 22 proteins have identities between *Saguinus* and macaque that are equal to or smaller than 90% and are considered fast evolving. Among fast evolving proteins, 9 showed higher identity between *Saguinus* and macaques than between *Saguinus* and humans while 13 showed less ($P >> 0.05$). Similarly, among slow evolving genes, 8 showed higher identity between *Saguinus* and macaques than between *Saguinus* and humans while 9 showed less ($P >> 0.05$). The data show that New World monkeys are equidistant to macaques and humans in both slow and fast evolving genes.

	# identical a.a.		% identity
	Sa.-Hu.	Sa.-Ma.	Sa.-Ma.
<i>Sa.-Ma. > Sa.-Hu., Slow evolving, 8 genes:</i>			
DAZL	282	283/296	95
Twist	194	196/203	96
VDR	419	422/427	98
Prion	189	203/210	96
CCR5	331	338/339	99
HBB	139	140/146	95
AAC25658	351	352/364	96
PQBP1	165	167/167	100
<i>Sa.-Ma. < Sa.-Hu., Slow evolving, 9 genes:</i>			
PPIA	162	160/164	97
PML	826	808/881	91
Boule	272	271/283	95
CD81	231	230/236	97
CXCR4	329	328/334	98
GCR	748	743/777	95
KLK15	239/255	230/244	93
Cryopyrin	473	471/499	94
NOTCH2	455	451/462	97
<i>Sa.-Ma. > Sa.-Hu., Fast evolving, 9 genes:</i>			
ND4	171	175/234	74
MC1R	268	281/317	88
Interferon α	158	162/189	85
Epo	112	113/133	84
PKDREJ	1799	1800/2017	89
TAS1R2	418	420/530	79
CAMP	131	140/169	82

APOBEC3H	127	131/181	72
SRY	151	152/208	73
<i>Sa.-Ma. < Sa.-Hu., Fast evolving, 13 genes:</i>			
ND1	260	249/318	78
COB	303	282/375	75
Trim5	350	348/503	69
CD46	282	278/369	75
TGIFLX	164	156/242	65
DMP1	264	258/293	89
TNF	139	138/154	89
RNase1	145	131/156	83
Angiogenin	103	102/145	70
SLAM	290	286/336	85
FUT1	252	251/281	89
Enamelin	657	652/776	84
TRIM22	423	412/479	86

8 Simian primates are the sister taxon to a loris-tarsier clade

Table S19 Lorises are closer to tarsiers than to humans but are equidistant to New World monkeys and humans. Most of the protein sequences of lorises available at the Genbank were selected for comparison with humans, tarsiers, and New World monkeys (NWM). Of the 40 informative proteins as shown here, 22 have greater than 85% identity between lorises and tarsiers and are considered slow evolving, while the other 18 proteins have identities between lorises and tarsiers that are equal to or smaller than 84% and are considered fast evolving.

	# identical amino acid		% identity	# identical. a.a.	% identity
	Lo-Hu	Lo-Ta	Lo-Ta	Lo-NWM	Lo-NWM
<i>Lo.-Ta. > Lo.-Hu., Slow evolving, 19 genes:</i>					
PAX9	337	338/341	99	333/341	98
COX1	468	487/512	95	469/512	91
Cyt c	91	99/105	94	90/105	85
Cnr1	282	283/299	94	285/299	95
ISP	186	190/202	94	181/202	89
HBA	131	132/141	93	130/141	92
COX5A	122	127/136	93	123/136	90
Epsilon-globin	60	62/67	92	61/67	91
Amelogenin	122	123/134	92	ni	
COX3	224	234/261	89	197/261	75
LHB	90	108/119	88	85/119	69
IRBP	256	267/301	88	254/301	84
COX4I1	112	120/137	87	112/137	81
Tyr	121	122/140	87	118/140	84
Growth horm.	119	150/174	86	116/174	67
COB	298	322/379	85	301/379	79
COX2	159	185/220	85	163/220	74
COX6c	54	64/75	85	60/75	78
COX8A	49	58/68	85	50/68	73
<i>Lo.-Ta. < Lo.-Hu., Slow evolving, 3 genes:</i>					
EDG1	151	144/158	91	ni	
HBB	135	130/146	89	133/146	91
AAR2B	336	321/368	87	324/368	88
<i>Lo.-Ta. > Lo.-Hu., Fast evolving, 10 genes:</i>					
Pyrin	124	132/160	82	126/160	78
ND4L	73	81/98	82	70/98	71
ATP6	169	184/226	81	159/226	70
ND1	249	259/317	81	253/318	71

HBG	120	119/148	80	123/147	83
ND5	406/591	445/572	77	411/593	69
ND4	329	349/457	76	321/458	70
ND3	79	85/115	76	82/115	71
ND2	194/324	209/318	65	194/343	56
ATP8	32	42/67	62	37/64	57
<i>Lo.-Ta. < Lo.-Hu., Fast evolving, 8 genes:</i>					
ADORA3	107/127	90/107	84	90/107	84
Atp7a	173	168/205	82	170/205	83
COX7AH	15	13/16	81	14/16	87
AR	386	384/476	80	378/491	76
MSX1	131	117/149	78	132/150	88
VWF	355	319/407	78	343/407	84
D4DR	11	10/16	62	ni	
ND6	106	100/175	58	100/177	56

9 Calculation of the divergence time between chimpanzees and gorillas

Table S20 The divergence time between chimpanzees and gorillas. The 27 slow evolving genes as listed in Table 4 were used to calculate the divergence time between chimpanzees and gorillas. This calculation assumes that the mutation rates in these genes are similar in gorillas and orangutans, which is highly likely given the close relationship between the two apes. Calculation based on the gorilla fossil split time of 12 Myr ago was performed using the formula: Divergence time of chimpanzees and gorillas = 12 x the Poisson correction distance between gorillas and chimpanzees divided by the Poisson correction distance between gorilla and orangutan. ni: most of the non-informative genes show 100% identity either between chimpanzees and gorillas or between gorillas and orangutans. Two genes (KLK3 and CCR5) shows more identity between gorilla and orangutans than between chimpanzees and gorillas and has likely reached cap of diversity and is therefore non-informative.

	# identical a.a.		Chimp.-Go.
	Chimp.-Go.	Go.-Orang.	Div. time (Myr)
APOE	314	311/317	6
MBP1	234	229/235	2
KLK3	ni		
T2R38	330	325/333	4.5
ASIP	ni		
WNT7A	ni		
FSHB	ni		
GSC	ni		
Myostatin	ni		
GPR56	677	670/687	overlap
BRCA1	1125	1108/1141	overlap
RNAseA1	153	151/156	7.2
MAOA	ni		
HNMT	116	113/117	3.2
SCML2	ni		
CXCR4	ni		
UTY	223	217/226	4
MBL2	234	227/235	1.5
Oxytocin re.	287	284/289	4.8
CXCR2	349	342/355	5.5
ASPM	3426	3403/3477	8.3
CCR5	ni		
FUT2	341	331/343	2
Prion	252	249/253	3
TPMT	244	236/245	1.3
Globin a2	ni		
COX1	504	497/512	overlap
Average of 16 informative genes:			4.1 +/- 1.3