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

Shi Huang

State Key Laboratory of Medical Genetics
Xiangya Medical School
Central South University
110 Xiangya Road
Changsha, Hunan 410078, China

shuangtheman at yahoo.com

Key words: Genetic non-equidistance, genetic equidistance, evolution, molecular clock, maximum genetic diversity, great apes, pongid, orangutans, tarsiers, primate phylogeny, paleontology

Running title: Primate phylogeny

Abstract Interpretations of molecular data by the modern evolution theory are often sharply inconsistent with paleontological results. This is to be expected since the theory is only true for microevolution and yet fossil records are mostly about macroevolution. The maximum genetic diversity (MGD) hypothesis is a more coherent and complete account of evolution that has yet to meet a single contradiction. Here, molecular data were analyzed based on the MGD to resolve key questions of primate phylogeny. A new method was developed from a novel result predicted by the MGD: genetic non-equidistance to a simpler taxon only in slow but not in fast evolving sequences given non-equidistance in time. This 'slow clock' method showed that humans are genetically more distant to orangutans than African apes are and separated from the pongid clade (containing orangutan and African apes) 17.3 million years ago. Also, tarsiers are genetically closer to lorises than simian primates are, suggesting a tarsier-loris clade to the exclusion of simian primates. The validity and internal coherence of the primate phylogeny here were independently verified. The molecular split time of human and pongid calibrated from the fossil record of gorilla, or the fossil times for the radiation of anthropoids/mammals at the K/T boundary and for the Eutheria-Metatheria split in the Early Cretaceous, were independently confirmed from molecular dating calibrated using the fossil split times of tarsier-loris and two other pairs of mammals (mouse-rat and opossum-kangaroo). This remarkable and unprecedented concordance between molecules and fossils provides the latest confirmation of the inseparable unity of genotype and phenotype and the unmatched value of MGD in a coherent interpretation of life history.

Introduction

Two kinds of sequence alignment can be made using the same set of sequence data. The first aligns a recently evolved organism such as a mammal against those simpler or less complex species that evolved earlier such as amphibians and fishes. The second aligns a simpler outgroup organism such as fishes against those more complex sister species that appeared later such as reptiles and mammals. In the early days of molecular evolution studies, genetic distance was represented by percent identity in protein sequence alignments.

The first alignment indicates a near linear correlation between genetic distance and time of divergence, implying indirectly a very similar mutation rate among vastly different species. For example, human is closer to mouse, less to bird, still less to frog, and least to fish. The second alignment shows the genetic equidistance result where sister species are approximately equidistant to the simpler outgroup. For example, human, mouse, bird, and frog are all equidistant to fish in any given protein dissimilarity. Since all of the sister species are also equidistant in time to the outgroup fish, this directly triggered the idea of constant or similar mutation rate among different species, no matter how different they may be. Since both alignments use the same sequence data set, certain information may be revealed by either alone. But the data that most directly and obviously support the interpretation of a constant mutation rate is the genetic equidistance result.

The molecular clock hypothesis was first informally proposed in 1962 based largely on data from the first alignment [1]. Margoliash in 1963 performed both alignments and made a formal statement of the molecular clock after noticing the genetic equidistance result [2,3]. However, the constant mutation rate interpretation of the genetic equidistance result is in fact a tautology since it has not been verified by any independent observation and has on the contrary been contradicted by a large number of factual observations [4,5,6,7,8,9,10,11,12,13,14,15].

Nonetheless, people have treated the molecular clock as a genuine reality and have in turn proposed a number of theories to explain it [16,17,18,19,20,21].

The 'Neutral Theory' has become the favorite [19,20,21], even though this theory is widely acknowledged to be an incomplete explanation [8,22]. However, it has never occurred to anyone that the failure to explain the clock after 46 years of extensive effort is because there is in reality no such thing as similar mutation rate per year among vastly different species. Indeed, no one has even attempted to explain the real original empirical fact, the genetic equidistance result, without presupposing a constant mutation rate.

Besides the numerical feature in terms of percent identity, the other characteristic of the equidistance result is the overlap feature where most of the mutant positions relative to the outgroup are shared between the sister lineages. For example, yeast is approximately equidistant to drosophila (67/104 identity) and to human (66/102 identity) in cytochrome c. Among those 36 residue positions different between yeast and human, 31 are also different between yeast and drosophila. This nearly complete overlap in mutated residue positions in separate sister lineages has been completely overlooked in the past 46 years. The molecular clock and the neutral theory were invented based on a complete ignorance of the overlap feature. They would not have been invented in the first place if people had paid attention to this feature because they are clearly contradicted by it. They predict a much smaller number of overlapped positions [23].

The first kind of alignment performed by Zuckerkandl and Pauling using hemoglobin also shows the overlap feature, as would be expected since both alignments use the same sequence information and should tell similar stories. For example, human hemoglobin alpha is 17/142 identical to horse and 42/142 to chicken. Of the 17 variant positions between human and horse, 14 are also variants between human and chicken. Molecular clock can only account for 5 or 6 overlaps, far short of the observed 14 [24]. Thus, Zuckerkandl, Pauling, and Margoliash all could have noticed the overlap feature. If they had done that, the molecular clock would never have been invented for macroevolution. It may still be invented and useful for microevolution as long as it only means similar mutation rates for identical or very similar species. But its impact on the understanding of molecular macroevolution would be trivial.

The modern evolution theory consists of the Neo-Darwinian theory of natural selection and the neutral theory. The Neo-Darwinian theory is largely useless or irrelevant in understanding molecular evolution or the key phenomenon of molecular evolution, the genetic equidistance result, which in fact contradicts it. As a result, Neo-Darwinists are forced to accept an anti-selection theory, the neutral theory, in order to at least have an ad hoc understanding of molecular evolution. However, as discussed above, the molecular clock and the neutral theory are completely mistaken for macroevolution. Thus, the modern evolution theory is largely useless in understanding molecular macroevolution, and is in fact contradicted or falsified by the major facts of molecular macroevolution, chief among which is the overlap feature of the genetic equidistance result. The theory is largely correct for microevolution for the simple fact that it has not a single contradiction in this domain. But by the same standard, it is also largely incorrect for macroevolution for the simple fact that it is contradicted by numerous facts of macroevolution (though one contradiction is sufficient to doom any theory).

Unlike the modern evolution theory, the recently proposed maximum genetic diversity (MGD) hypothesis is self-evident and explains all major facts of evolution in a coherent fashion via a single universal theme or axiom [7,25]. Phenotypes are determined by both genetics/DNA and epigenetics with epigenetics playing a more important role in complex organisms. Since DNA is never free of proteins/RNAs in any cellular organisms at any stage of life cycle, it cannot be said that DNA is ultimately more important than proteins/RNAs or epigenetics. Genetic diversity is inversely related to epigenetic complexity or organism complexity [7,25]. It is self-evident that genetic diversity cannot increase indefinitely with time and has a maximum limit being restricted by function or epigenetic complexity. While the idea of functional constraint on mutation or genetic diversity/distance is widely accepted, I have now expanded the scope of functional constraint to include epigenetic functions.

A gene may function in many different cell types or epigenetic states (each cell type represents a distinct epigenetic state). The more cell types in which a gene functions, the more functions it performs and the more functional constraints on the genetic diversity/mutation of the gene. The same gene encounters more functional constraints in complex organisms than in simple organisms because complex organisms have more cell types. The maximum genetic diversity of simple organisms is greater than that of complex organisms. However, a given population of a species may not always show the maximum genetic diversity due to recent common ancestry and/or homogeneous environmental selection [26].

The MGD hypothesis but not the modern evolution theory predicts the existence of Complexity-Associated-Protein-Sectors (CAPS) or Sequence-Sectors (CASS) as a group of correlated residues that is more conserved in complex than in simple organisms. The first example of such CAPS has recently been discovered for the S1A family protease, which is more conserved in vertebrates than in invertebrates [27,28]. Epigenetic complexity puts maximum CAPS on sequence divergence.

Over long evolutionary time, the genetic distance between sister species and a simpler outgroup taxon is mainly determined by the maximum genetic diversity of the simpler outgroup, although over short time scales it is mainly determined by time, drift, environmental selection, and the neutral mutation rates of the simpler outgroup as well as to a smaller extent by the rates of the sister taxa.

The MGD hypothesis predicts either genetic equidistance or non-equidistance to an outgroup depending on the epigenetic complexity of the outgroup, whereas the molecular clock predicts only genetic equidistance to the outgroup regardless whether the epigenetic complexity of the outgroup is more or less complex than the sister clade (Figure 1, prediction 1 and 2). According to the MGD hypothesis, the genetic distance between a complex outgroup and a simple taxon is mainly determined by the genetic diversity of the simple taxon. If one of the sister taxa is more complex than the others, it would have lower genetic diversity and would show higher sequence similarity to a more complex outgroup species. Here, I present a large number of cases of genetic non-equidistance to a complex outgroup despite equidistance in time.

Genetic distance would no longer correlate with time of separation after reaching maximum cap. Since fast evolving genes reach cap faster, they are non-informative for inferring genealogy in many cases. The molecular clock however predicts no difference between fast and slow evolving genes in their utility in genealogy as long as the orthologous genes have not changed so much that they cannot be recognized as orthologous by sequence alignments. In contrast, the MGD hypothesis predicts the phenomenon of genetic non-equidistance to a simpler taxon only in slow but not in fast evolving sequences given non-equidistance in time (Figure 1, predictions 3 and 4; also Figure 2). Examples of this novel phenomenon are here shown.

Paleontologists have long suggested that human is the outgroup to a pongid (orangutan-gorilla-chimpanzees) clade and diverged from pongids ~ 18 million years (Myr) ago [29,30,31,32,33,34]. The 14 Myr old *Ramapithecus* was considered the earliest human fossil [29,30,31]. However, interpretations of molecular similarity suggest that humans and chimpanzees belong to the same clade to the exclusion of other great apes and shared a common ancestor merely 5 Myr ago [35,36,37]. As paleoanthropologist Schwartz commented: “To paleoanthropologists, this was sheer blasphemy.” [34].

As shown by Wilson and Sarich [37], the molecular interpretation relied on two observations plus two unproven premises. First, human is closer to chimpanzees than to monkeys as measured by percent identity in protein sequences. This was taken to mean that human is genealogically closest to chimpanzees based on the premise that higher sequence similarity necessarily means closer genealogical relationship. Second, humans, chimpanzees, and monkeys are equidistant to the outgroup horses in both gene sequence and time of separation. This was interpreted to mean that these different primates have the same mutation rate based on the premise that the equidistance result is the outcome of a constant mutation rate. This justified using the mutation rate of monkeys to yield the divergence time of chimpanzees and humans.

The same two premises underlie all molecular dating analysis and have created some major contradictions with paleontological results, including, just for the mammals, the position of great apes and tarsiers, the timing of mammal radiation, and the split between Eutheria and Metatheria mammals [38,39,40,41,42,43,44,45,46,47]. These unproven premises have now been falsified by the MGD hypothesis and numerous facts. Thus, molecular phylogeny needs to be reevaluated by new and correct molecular methods.

The key questions in primate phylogeny concern the origins of humans, anthropoids, and tarsiers. Given the complexity of morphological features and extensive convergent evolution, these questions cannot be easily resolved by paleontological analysis alone [40,41].

The genetic non-equidistance to a more complex outgroup despite equidistance in time or genealogy, as described here, shows that higher sequence similarity to a complex outgroup cannot be used to infer closer genealogical relationships. The correct way to infer genealogy from sequence similarity must make use of the novel phenomenon of genetic non-equidistance to a simpler taxon in slow but not in fast evolving sequences given non-equidistance in time or genealogy. I have here used this ‘slow clock’ method to perform a complete reevaluation of primate phylogeny. The new primate phylogeny here was further independently verified for its internal coherence as well as consistency within Theria mammals by the remarkable concordance between molecular and fossil dating on the key diversification events within primates and Theria mammals. The results support the original views of paleontologists on the pongid clade and resolve the controversial position of tarsiers, leading to novel insights into the origins of humans, anthropoids, tarsiers, and mammals.

Results

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

To test prediction 2 in Figure 1, the most complex animal, human, was used as the outgroup to compare with sister species from a simpler clade or group. For each group, where possible, two sister species were identified with one representing a simple organism and the other more complex. Complexity is inferred from time of appearance in the fossil record (complex organisms generally appeared later), advanced nervous system, and greater number of cell types [7,25]. Genetic equidistance of A and B to an outgroup C can be established if the number of genes showing greater similarity between A and C than between B and C is similar to the number of genes showing less similarity between A and C than between B and C ($P > 0.05$). Similarity was simply measured by percent identity since the result of the equidistance testing method here is independent of distance correction such as the Poisson correction distance or other distance correction methods. If A is closer to C than B is in percent identity, distance correction would only change quantitatively how close A to C is relative to B to C, but would not change qualitatively the fact that A is closer to C than B is. The equidistance testing method here only needs to know that A is closer to C than B is but does not need to know by how much.

The mollusk phylum. The bivalves have existed since the Cambrian period. The octopuses have complex nervous systems and are considered among the most intelligent invertebrates. As shown in Table 1, a sampling of 10 mitochondrial proteins showed that humans are significantly closer to octopus (*Octopus vulgaris*) than to cockle (*Acanthocardia tuberculatum*) (10 showed more similarity between human and octopus than between human and cockle while 0 showed less, $P < 0.05$). This example shows that higher similarity to humans does not necessarily mean closer genealogical relationship with humans. This genetic non-equidistance to a more complex outgroup is very different from the genetic equidistance to a simpler outgroup. By the same equidistance testing method, it can be easily shown that two distinct vertebrates (mouse and chicken) are equidistant to the simpler invertebrate outgroup cockle (6 mitochondrial proteins showed more similarity between cockle and chicken than between cockle and mouse while 3 showed less, $P > 0.05$).

The brachiopod phylum. The inarticulate brachiopod genus Lingula (*Lingula anatina*) is the oldest, relatively evolutionarily unchanged animal known. The oldest Lingula fossils are found in Lower Cambrian rocks dating to roughly 550 Myr ago. Terebratulids (*Terebratulina retusa*) are modern articulate brachiopods and appeared 430 Myr ago. As shown in Table 2, humans are significantly closer to Terebratulina than to Lingula ($P < 0.05$). Lingula is equidistant to Terebratulina and human ($P = 0.64$). This suggests that the time of separation between the two brachiopods has been long enough for their genetic distance to reach a maximum cap that is similar to the maximum distance between brachiopods and humans. In this case, if the results were interpreted using the molecular clock hypothesis, it would lead to the absurd conclusion that Lingula is the outgroup to a Terebratulina-human clade. This example shows that sequence similarities are not always informative for genealogy. Once the maximum distance is reached, sequence dissimilarity would no longer correlate with time of separation.

The reptile group (including birds). Snakes maybe simple reptiles without limbs whereas birds have complex flying capacities. A sampling of 10 mitochondrial proteins shows that snakes are significantly more distant to humans than birds are ($P < 0.05$). A random sampling of 13 nuclear genes also showed the same result ($P < 0.05$) (Supplementary Table S1). The combined result from both mitochondrial and nuclear genes is highly significant ($P < 0.0001$). Thus, mitochondrial proteins can reveal certain genetic relationships that are similar to those revealed by nuclear genes. For many of the analyses here, mitochondrial proteins were used because most species have available only sequences of mitochondrial proteins.

Other major groups of organisms. As shown in the Supplementary Information, significant non-equidistance to humans was found for sister species within the teleost fish clade, the arthropod phylum, the porifera phylum, and the fungi kingdom, but was not found for the amphibian group, the echinoderm phylum, the annelida phylum, the nematode phylum, the platyhelminthes phylum, the cnidaria phylum, the plant kingdom, the protist alveolates superphylum, and the bacteria kingdom. The failure to detect non-equidistance could be due to several reasons. In some cases, such as amphibians, echinoderms, and nematodes, a trend of non-equidistance was found for some sister species and future availability of more sequences could easily confirm the trend to be statistically significant. Some groups such as bacteria may have little difference in epigenetic complexity or genetic diversity among sister species. Some clades have few sister species that have been sampled such as the platyhelminthes phylum and the cnidaria phylum. Some group, such as plants, has evolved group-specific domains since separating from humans but before divergence of sister species within the group.

In all five cases (except plants) where difference in complexity of the sister species can be inferred (octopus vs. cockle, Terebratulina vs. Lingula, bird vs. snake, dragonfly vs. louse, and smut vs. yeast), the more complex species always show greater sequence similarity to humans, fully conforming to the predictions of the MGD hypothesis (Figure 1, prediction 2).

Genetic non-equidistance to a simpler taxon in slow evolving sequences given non-equidistance in time

Slow evolving genes are defined as having high identity between species. Table 3 shows that slow evolving genes are less likely to have reached the maximum cap on diversity than fast evolving genes. Most histone lysine methyltransferases (KMTs) (6 of 9) have identities between zebrafish and pufferfish that are equal to or slightly lower than that between zebrafish and human or mouse, showing that these proteins have reached the maximum cap on diversity for fishes. In contrast, only 2 of 12 ribosomal proteins have reached the cap. Thus, the KMT family is significantly different from the ribosome family in having more proteins reaching the cap ($P = 0.03$). This correlates well with the fact that the average identity between the two fishes for the KMT family is significantly smaller than that of the ribosome family (65.1 ± 8.5 vs. 92.1 ± 4.7, $P < 0.001$).

Since fast evolving genes show that pufferfish and human are equidistant to zebrafish, they are non-informative of the fact that human is the outgroup to the two fishes. Only slow evolving genes are informative: human is the outgroup because zebrafish is closer to pufferfish than to human in slow evolving genes such as ribosomal proteins (Table 3). This result shows an example of genetic non-equidistance to a simpler taxon in slow evolving sequences due to non-equidistance in time (Figure 1, predictions 3 and 4 by the MGD hypothesis). Human and pufferfish are non-equidistant to zebrafish in slow evolving genes (but not in fast evolving genes) because they are non-equidistant in time to zebrafish.

The phenomenon of genetic non-equidistance to a complex outgroup despite equidistance in time as described in Table 1 and 2 shows that, for any three species, A, B, and C, with A being most complex and C least complex, a smaller distance between A and B relative to A and C cannot be used to group A and B to the exclusion of C. To infer genealogy, one must rely on the genetic distance to C as measured by slow evolving genes (Figure 2A, time T1). Only when A and B are equidistant to C in slow evolving genes, they can be grouped in the same clade to the exclusion of C (Figure 2A, Model II). If, however, B is closer to C than A is, then B and C would belong to the same clade to the exclusion of A (Figure 2A, Model I).

Slow evolving genes are genes that show high identity between the simpler taxon C and a more complex taxon that is most similar to C in phenotypes. If B is more similar to C than A is, then B should be used for comparison with C to identify slow evolving genes. Large dissimilarity in phenotypes between A and C may indicate longer time of separation. Thus, relative to a list of high identity genes between B and C, a list of high identity genes between A and C would contain more genes that have reached cap and would not be informative for genealogy.

The genetic distance of A or B to C in slow evolving genes is mainly determined by the neutral mutation rate of C within the neutral diversity range of C (i.e., 20% for the example here in Figure 2). Since the neutral mutation rate of C should be roughly constant over evolutionary time, the genetic distance of A or B to C should reflect the time of separation with C. In Model I of Figure 2A, knowing the mutation rate of C based on the fossil split time of B (or A) can be used to calculate the split time for A (or B). Here fast evolving genes should not be used as they would have reached maximum cap on diversity and would show that C is equidistant to A and B even if B and C belong to the same clade (Figure 2B). After extremely long evolutionary time, even slow evolving genes would reach cap and become useless for inferring genealogy (Figure 2A, time T2). Other information such as paleontology would then become critical.

Primate phylogeny

Humans are the sister taxon to a pongid clade

To use slow evolving sequences in phylogeny analysis is here termed the "slow clock" method. I here used this method to reevaluate primate phylogeny. I randomly picked a set of orangutan proteins to determine whether gorillas or chimpanzees are closer to orangutans than humans are in slow evolving genes. These proteins were about equally divided into two groups of fast and slow evolving genes (Table 4). Among fast evolving genes, 14 showed higher identity between orangutans and gorillas than between orangutans and humans while 16 showed less (*P* >> 0.05). In contrast, among slow evolving genes, 27 showed higher identity between orangutans and gorillas than between orangutans and humans while 7 showed less (*P* = 0.02), suggesting that orangutans are significantly closer to gorillas than to humans. Thus, human is the sister taxon to an orangutan-gorilla clade.

The divergence time of humans and orangutans was next calculated using the fossil estimate of the gorilla split of 12 Myr ago as calibration point [39]. Assuming a constant mutation rate for the orangutan lineage during its entire history of existence, I calculated a human split of 17.3 ± 6.7 Myr ago (Table 4), congruent with the original paleontological estimate. This time was not significantly affected by the Poisson correction or other distance correction methods, because the time is relatively too short and the mutation rate too slow for multiple amino acid substitutions to occur at the same sites. The excellent match between paleontological and molecular results independently confirms the validity of the slow clock method.

Orangutans were also found to be closer to chimpanzees than to humans. As shown in Table 4, among fast evolving genes, 8 showed higher identity between orangutans and chimpanzees than between orangutans and humans while 10 showed less (*P* >> 0.05). In contrast, among slow evolving genes, 17 showed higher identity between orangutans and chimpanzees while 3 showed less (*P* < 0.05).

To independently verify the closer relationship between orangutans and chimpanzees, I randomly picked 733 cDNA sequences of *Pongo abelli* that were randomly generated by the German cDNA consortium. About 29.7% of these were informative (Supplementary Table S10). 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). Furthermore, calculations based on these slow evolving genes, assuming a 12 Myr split from orangutan for the African ape clade, gave a human split from orangutan of 17.3 ± 5.1 Myr ago. Thus, two independent and different data sets gave remarkably similar result on the spilt time of humans. Together, these observations show that humans are the sister taxon to a pongid clade containing orangutans and African apes.

To verify that results from a small set of genes is representative of a much larger set of genes or even the whole genome, I analyzed all available 4330 cDNAs of *Pongo abelli* available at the Genbank that were generated by random cDNA sequencing effort of the German cDNA consortium. I arbitrarily divided these cDNAs into 10 groups, with every 433 cDNAs forming a group based on their numerical order of listing in the Genbank. As shown in Table 5, for fast evolving genes, 2 groups (group 2 and 10) showed that orangutan is slightly closer to chimpanzees than to humans while 8 groups showed that orangutan is slightly closer to humans than to chimpanzees (*P* > 0.05). In contrast, for slow evolving genes, all 10 groups showed that orangutan is closer to chimpanzees than to humans (*P* < 0.05), suggesting that orangutan is significantly closer to chimpanzees than to humans in slow evolving genes.

None of the 10 groups individually showed that orangutan is non-equidistant to humans and chimpanzees in fast evolving genes based on the *P* value cutoff of 0.05. However, for slow evolving genes, 6 groups (groups 1, 3-7) each individually showed that orangutan is significantly closer to chimpanzees than to humans. The other 4 groups all showed that the number of genes with greater similarity between orangutans and chimpanzees is at least 2 fold greater than the number of genes with greater similarity between orangutans and humans. But unlike the slow evolving genes, none of the 10 groups of fast evolving genes showed that the number of genes with greater similarity between orangutans and chimpanzees is more than 1.5 fold greater or less than the number of genes with greater similarity between orangutans and humans. The combined result of the 10 groups of fast evolving genes is non-significant (335 vs. 384, *P* > 0.05). In contrast, the combined result of the 10 groups of slow evolving genes is extremely significant (247 vs. 80, *P* < 0.0001). The result of all 1046 informative genes, combining fast and slow evolving genes, also showed that orangutan is significantly closer to chimpanzees than to humans (582 vs. 464, *P* < 0.05).

This large scale analysis confirmed that a statistically significant result (*P* < 0.05) derived from a small set of genes by using the equidistance testing method here is equivalent to results from a much larger set of genes. The data also showed that there is little chance (*P* < 0.05) for variation in gene selection to produce an arti-factual non-equidistance result since none of the 10 groups of fast evolving genes showed statistically significant non-equidistance.

To further confirm that humans are the sister taxon to a pongid clade, I determined the genetic distance to gorillas of humans and chimpanzees using a set of randomly selected gorilla proteins (Supplementary Table S11). 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). The data thus show a sister grouping of gorillas and chimpanzees to the exclusion of humans.

Orangutans are the outgroup to a gorilla-chimpanzee clade

I next determined the relationship of the three great apes of the pongid clade using the data shown in Table 4. Among fast evolving genes, 11 showed higher identity between orangutans and gorillas than between orangutans and chimpanzees while 18 showed less (*P* >> 0.05). Similarly, among slow evolving genes, 12 showed higher identity between orangutans and gorillas while 14 showed less (*P* >> 0.05). So, orangutans are equidistant to gorillas and chimpanzees in both fast and slow evolving genes and are therefore the outgroup to a gorilla-chimpanzee clade given the well established closer sequence similarity between gorilla and chimpanzee than either is with orangutan.

Gibbons are the outgroup to a pongid-human clade

Similar analysis confirmed that the lesser ape gibbons (*Hylobates lar*) are the outgroup to a pongid-human clade (Supplementary Table S12). Among fast evolving proteins, 9 showed higher identity between gibbons and orangutans than between gibbons and humans while 15 showed less (*P* >> 0.05). Similarly, among slow evolving genes, 16 showed higher identity between gibbons and orangutans while 14 showed less (*P* >> 0.05). So, gibbons are equidistant to orangutans and humans in both fast and slow evolving genes. Gibbons are also equidistant to gorillas and humans as well as equidistant to chimpanzees and humans (data not shown).

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

Gibbons and humans are equidistant to the Old World monkey (OWM) *M. mulatta* in both fast and slow evolving genes (Supplementary Table S13). Together with the well-established closer sequence similarity between humans and gibbons, the data suggest that monkeys are an outgroup to a clade containing gibbons and humans.

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

Old World monkeys and humans are equidistant to New World monkeys (NWM) in both fast and slow evolving genes (Supplementary Table S14). Together with the well-established closer sequence similarity between humans and OWM, the data suggest that NWM are the outgroup to a clade containing OWM and humans.

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

The position of tarsiers is controversial among paleontologists while molecular biologists, based on the mistaken molecular clock hypothesis, group tarsiers with simian primates [40,41]. As shown in Table 6, among fast evolving genes, 10 showed higher identity between lorises and tarsiers than between lorises and humans while 8 showed less (*P* >> 0.05). In contrast, among slow evolving genes, 19 showed higher identity between lorises and tarsiers than between lorises and humans while only 3 showed less (*P* < 0.05), suggesting a loris-tarsier clade to the exclusion of higher primates. As an independent confirmation of this important conclusion, Table 6 also shows that loris is closer to tarsier than to the New World monkey marmoset *C. jacchus* in slow evolving genes (17 vs. 3, *P* < 0.05) but not in fast evolving genes (10 vs. 5, *P* > 0.05).

Lorises are the outgroup to a simian primate clade

Table 6 also shows that lorises are the outgroup to a simian primate clade. Among fast evolving genes, 6 showed higher identity between lorises and New World monkeys than between lorises and humans while 10 showed less (*P* >> 0.05). Similarly, among slow evolving genes, 9 showed higher identity between lorises and New World monkeys while 11 showed less (*P* >> 0.05). The data show that lorises are equidistant to New World monkeys and humans and are therefore the outgroup to a New World monkey-human clade given the well-established closer similarity between humans and New World monkeys than either is to lorises.

Verification of the validity and internal coherence of the primate phylogeny

A true phylogeny should give a coherent picture of different divergence times that are well established by independent methods. A molecular clock calibrated from one fossil split time should produce divergence times consistent with other fossil records and other independently calibrated molecular clocks. Here, I first calculated the split time between lorises and New World monkeys by using a molecular clock calibrated by the fossil split time of 40 Myr between tarsier and loris, based on the oldest fossils of tarsier and loris from the middle Eocene [48,49]. As shown in Table 7, the slow evolving genes of Table 6 were used for this calculation and produced a divergence time of 66.7 Myr, consistent with the fossil based estimation of anthropoid-prosimian split around the K-T boundary 65.5 Myr ago [50,51,52,53]. It is likely that anthropoid and prosimian simultaneously emerged as part of the same radiation that produced all the major mammals around the K/T boundary [43].

Because different species may contribute differently to genetic distance, it is important to have at least one species in common when calculating one split time from another. From tarsier-loris split to produce NWM-loris split time, loris is the common species. I next used the NWM-loris split time 66.7 Myr as calibration to calculate the divergence time between OWM and NWM with NWM here the common species. Using the same list of genes as shown in Table 7, the OWM-NWM split time was calculated to be 47.8 Myr, somewhat older than the age of the oldest OWM fossils of late Eocene such as Catopithecus [54]. This time was next used to calculate the orangutan-OWM split time of 29.7 Myr (Table 7), consistent with the age range of the first fossil ape *Proconsul* [47]. This time was then used to produce a human-orangutan split time of 17.3 Myr (Table 7), in remarkable agreement with the time independently calculated from calibration using the fossil split time of gorillas as described above. These results, therefore, suggest that the primate phylogeny here is extremely coherent and well supported by a number of independent observations.

Table 7 also shows a split time of 63.6 Myr between loris and cattle, consistent with mammal radiation at the K/T boundary. To further confirm this radiation and its coherence with the primate phylogeny here, I next used the newly derived human-pongid split time of 17.3 Myr, together with the well established fossil split time of 12.3 Myr between mouse and rat [47], to calculate the human-mouse divergence time. The mutation rate of the lineage leading to human was assumed to be similar to the average between human and orangutan and calculated using the human-orangutan spilt time of 17.3 Myr (R_{human} = D/2/17.3, where D is the distance between human and orangutan), while the mutation rate of the lineage leading to mouse was assumed to be similar to the average between mouse and rat and calculated using the mouse-rat split time of 12.3 Myr (R_{mouse} = D/2/12.3, where D is the distance between mouse and rat). Thus, the division time between human and mouse can be calculated as T = D/(R_{mouse} + R_{human}), where D is the distance between human and mouse. As shown in Table 8, a group of randomly selected slow evolving genes gave a human-mouse divergence time of 65.7 Myr, thus independently confirming the coherence of the human-pongid split with the mammal radiation at the K/T boundary.

To further examine the internal coherence of the primate phylogeny, I next determined whether the molecular split time of 65.7 Myr between human and mouse is consistent with the well established fossil split time between Eutheria and Metatheria mammals [46,47]. The mutation rate of the lineage leading to Eutheria mammals was assumed to be similar to the average between human and mouse lineages during their 65.7 Myr of separation and calculated as $R_{eutheria}$ = D/2/65.7, where D is the distance between human and mouse. The mutation rate of the lineage leading to Metatheria was assumed to be similar to the average between kangaroo and opossum during their 66.4 Myr of separation as determined from the fossil record [47] and calculated as $R_{metatheria}$ = D/2/66.4, where D is the distance between kangaroo and opossum.

For fossil time, I assume that the real time is close to the minimum constraint time plus 10% of the minimum time, e.g., the minimum age of gorilla is 10.5 Myr and its real age is estimated as 12 Myr [39]. If such time calculation happens to be close to the maximum constraint time such as in the case of mouse-rat fossil split (minimum 11.0 vs maximum 12.3 Myr), I use the maximum time. If it is close to the average of minimum and maximum, I use the average such as in the case of kangaroo and opossum (minimum 61.5 vs maximum 71.2 Myr, average 66.4).

As shown in Table 9, a group of randomly selected slow evolving genes gave a human-opossum split time of 131.7 Myr, in remarkable agreement with the fossil record of 124.6 to 138.4 with an average of 131.5 Myr [46,47].

Discussion

Sequence similarity is not necessarily genealogy

If genetic distance between a simple organism and a complex organism over long evolutionary time is determined by the maximum genetic diversity of the simple organism, then it is not necessarily related to the time of divergence. After reaching maximum distance, genetic distance would no longer correlate with time. We must then rely on fossil records and other biological features to infer genealogy. Based on sequence similarity to humans per se, we cannot infer, for example, that humans are genealogically closer to yeasts than to bacteria, because sequence similarity per se could also lead us to the absurd conclusion that humans are genealogically closer to one mollusk than to another mollusk (Table 1).

Inferring genealogy from molecular data has in the past relied on sequence similarity to the most complex taxon within the group of species analyzed. For example, within hominoids, the taxon that is closest in sequence similarity to human is considered to be genealogically also the closest to human. The sister grouping of chimpanzees and humans really has no other non-ambiguous support other than sequence similarity as measured by percent identity. The premise for this approach has now been nullified by the phenomenon of genetic non-equidistance to a more complex outgroup despite equidistance in time or genealogy. The same premise for grouping an ape (chimpanzee) with human to the exclusion of another ape (orangutan) would equally justify the obviously absurd grouping of human with a mollusk (octopus) to the exclusion of another mollusk (cockle), or with a brachiopod (Terebratulina) to the exclusion of another brachiopod (Lingula), or with a reptile (bird) to the exclusion of another reptile (snake).

Instead, the correct approach is to use sequence similarity, as measured by slow evolving genes, to the simplest taxon among a group of closely related taxa (Figure 2). Humans and African apes are equidistant to orangutans in fast evolving genes, but African apes are closer to orangutans in slow evolving genes. So, the net distance in all genes between orangutans and African apes remains smaller than that between orangutans and humans. This result could only be interpreted by the sister grouping of humans and pongids. Speculating a higher mutation rate for humans relative to the African apes would not work since it is not possible to imagine that the higher mutation rate should specifically apply only to slow but not fast evolving genes. In fact, it is commonly thought that humans have slower mutation rate than the African apes [55].

Past studies used average distance of all sampled genes to infer genealogy [56]. This cannot be informative because the average distance is more heavily determined/weighted by fast evolving genes that tend to show greater distances. For the data shown in Table 4, the average identity for all 64 proteins is 94.44 +/- 4.08 between orangutan and human and 94.64 +/- 4.41 between orangutan and gorilla (P >> 0.05). So, average distance of all genes masks the difference between slow and fast evolving genes. Since previous studies made no distinction between fast and slow evolving genes, it is not unexpected that the evidence here for a sister grouping of humans and pongids or of tarsiers and lorises was not found in previous multigene analyses.

Genetic non-equidistance is distinct from what is known as ‘variable molecular clock’

The variable molecular clock concept is mainly associated with two kinds of results. The first is the greater genetic distance between two sister taxa such as mouse and rat than between two other sister taxa such as human and gibbons even though the two rodents have diverged more recently based on the fossil records. The second result is related to the genetic equidistance to a simpler taxon, including both equidistance to a simpler outgroup and equidistance to a simpler taxon in fast evolving genes despite non-equidistance in time. Some of the slight differences in distance among taxa to a simpler taxon are interpreted to represent significant variations in ‘mutation rate’. Thus, the variable molecular clock associated with the second result represents a kind of ‘genetic non-equidistance (to a simpler outgroup) despite equidistance in time’, which is distinct and must be differentiated from the ‘genetic non-equidistance (to a complex outgroup) despite equidistance in time’. The former is not as real as the latter and may be merely insignificant variations of genetic equidistance (to a simpler outgroup). More importantly, it also must be differentiated from the ‘real’ non-equidistance to a simpler taxon associated with non-equidistance in time. Humans and chimpanzees are non-equidistant to orangutans because of different split time with orangutans.

The constant mutation rate or molecular clock hypothesis was originally proposed to explain the genetic equidistance to a simpler outgroup. Since the equidistance is approximate, it shows small deviations from an exact equidistance. The most striking fact about the equidistance result is that the deviations from exact equidistance are rarely large, hence giving rise to the idea of an ‘approximately constant clock’. The ‘variable molecular clock’ interpretation of the slight deviations may not be biologically meaningful since it was based on statistical tests (the relative rate test) that contain false premises. The tests incorrectly assume that two diverging lineages gradually accumulate genetic distance without maximum cap. The tests also do not consider sampling variations. No results associated with the variable molecular clock concept really represent true violations of the original ‘approximately constant clock’ idea if the idea is taken as a tautology or restatement of the genetic equidistance result.

Humans have been found to have slower ‘mutation rate’ relative to other great apes [55]. While humans and chimpanzees are approximately equidistant to orangutans as measured by fast evolving intron and intergenic regions, humans can be shown to be *slightly* closer to orangutans [55]. The MGD hypothesis can explain this phenomenon not in terms of mutation rate variations. Chimpanzees have higher genetic diversity range than humans. The genetic distance between orangutans and chimpanzees or humans is primarily contributed by the genetic diversity range of orangutans and to a less degree by the genetic diversity of chimpanzees or humans. Since the genetic diversity of chimpanzees is higher than that of humans, chimpanzees contribute slightly more than humans to the maximum distance with orangutans.

Because the difference is extremely small, it requires large amount of sequences to observe a slightly higher similarity between humans and orangutans than between chimpanzees and orangutans in fast evolving sequences. Even more than 1 million aligned bases of introns and intergenic regions in chromosome 21 were not enough to reveal a significant difference [55]. The analysis here in Table 4 using 30 fast evolving proteins (equivalent to ~ 45000 nucleotides assuming an average gene size of 500 amino acids) did not show significant violation of equidistance. Analysis of 719 fast evolving proteins (equivalent to ~1, 078, 500 nucleotides) in Table 5 also did not show violation of equidistance. In contrast, the *real* non-equidistance of humans and chimpanzees to orangutans can be easily shown using only ~20 slow evolving proteins (Table 4). Thus, the genetic non-equidistance to a simpler taxon due to non-equidistance in time as measured by slow evolving genes is categorically distinct from the tiny deviations from exact equidistance to a simpler taxon as measured by fast evolving genes. It is much more pronounced. Whether humans truly have slower mutation rates than the great apes due to longer generation times is irrelevant to the distance between humans and orangutans in slow evolving genes since that distance is largely determined by the mutation rate of orangutans.

The meaning of 'most recent common ancestor'

Based on the fossil record, there exist two kinds of diversification from an ancestor. One is slow and gradual and the other is fast and explosive. From fish to amphibian is a slow process. The oldest fish fossil is ~530 Myr old while the oldest amphibian fossil is ~340 Myr old. Here the most recent common ancestor (MRCA) of fish and amphibian is an individual fish from ~340 Myr ago. This MRCA would account in theory for all extant amphibians but only a tiny fraction of all extant fishes. In contrast, when diversification proceeds via radiation or explosion, such as during the Cambrian Explosion or during the placental mammal radiation at the K/T boundary, the MRCA of two extant species may account for all living individuals of these two species and may not look like either species.

It is important to keep in mind these two different kinds of MRCAs when one is looking for fossil MRCAs and the time of diversification. For speciation via radiation, one may not be able to identify the MRCA fossil since it may not look like any living species. And the estimation of divergence time may be inferred from the oldest fossil of any one of two extant species. However, for slow and gradual speciation, one extant species would have existed longer than another (fish is older than frog). Here, the oldest fossil for the older lineage will not be informative to divergence time but only the oldest fossil of the younger lineage will.

Most researchers today do not make a distinction between the two kinds of MRCAs and often in practice treat most speciation as the radiation kind due to the undue influence by the recent popularity of the cladistic method. The cladistic method is only good for identifying sister relationships but not for ancestor-descendant relationship. It was originally a method invented for living species only which can only have sister relationships. But a fossil species can be either sister or ancestor to a living species. While the positive identification of a fossil as a sister of a living species by cladistic analysis also implies a possibility for it to be an ancestor, a failure to do so cannot exclude it as an ancestor. So, even if some researchers may be right that the 10.5 Myr old gorilla-like fossil *Chororapithecus* is not a sister of living gorillas due to the lack of shared-derived features, it has no bearing on the fossil being a gorilla ancestor.

A living descendant that has highly derived features is more likely to have a fate of extinction rather than serving as ancestor to future descendants that will have different derived features. A derived feature is less likely to change or moldable than a less derived, generic, or stem feature. An ancestor capable of giving rise to multiple distinct descendants is like a stem cell while a living descendant with derived features is like a differentiated cell. Thus, a true stem ancestor fossil may not share any derived features with its living descendants and would not be identifiable by the cladistic methods. Any fossil that could be identified as sisters to a living species by the cladistic method is unlikely to be a stem ancestor capable of giving rise to distinct descendants. The MRCA responsible for the mammal radiation at the K/T boundary may not and should not share any derived features with any of the living mammals. Thus, a heavy reliance on the cladistic method in studying fossils can be extremely misleading.

In the view of the followers of the cladistic method, the individual who was the direct ancestor of A cannot be many years apart from the individual who was the direct ancestor of the sister taxon B. In fact, the MRCA of a clade is commonly viewed as a single individual [57]. A split of N Myr ago between A and B means that the first appearance of A like or B like species occurred about N Myr ago.

This notion is needed in order to make sense of the molecular data in terms of the molecular clock hypothesis. If the direct ancestor of A lived many years after the direct ancestor of B, then the maximum genetic distance within B would be greater than the minimum distance between A and B, according to the molecular clock hypothesis (Figure 3). For example, the maximum genetic distance within gorillas would be greater than the minimum distance between gorillas and chimpanzees. However, the fact is that the maximum genetic distance within a taxon is never greater than the genetic distance between the taxon and its sister taxon.

So, to accommodate this fact, the molecular clock hypothesis requires that the direct ancestor of A and the direct ancestor of B were either the same individual or had not lived many years apart. Thus, a gorilla fossil from 10.5 Myr ago was interpreted to mean that gorillas had diverged from its sister taxon chimpanzees at least 10.5 Myr ago [39]. A platypus fossil from 120 Myr ago was interpreted to mean that platypus and echidna had parted at least 120 Myr ago [58]. But such interpretations could be completely false if the diversifications in these cases were in fact not the radiation kind, and all indications show that they were not.

In contrast to the molecular clock hypothesis, the MGD hypothesis explains both kinds of MRCA without requiring that the ancestor of B cannot be in existence many years prior to the split of A (Figure 3). Given long enough time or for fast evolving sequences, the maximum genetic distance within B could never be greater than the minimum genetic distance between A and B. The distance between A and B is determined by the maximum genetic diversity of B and should be the same as the maximum genetic distance within B and cannot be smaller as long as time is long enough for most genes to reach maximum diversity.

According to the MGD hypothesis, there is no conflict between the fossil record and the molecular data. Based on the emergence of gorillas 12 Myr ago, my calculation showed that chimpanzees diverged from gorillas 4.5 Myr ago (Supplementary Table S15). Thus, gorilla-like apes living 12 Myr ago may be the ancestors of all extant gorillas while gorilla-like apes living 4.5 Myr ago may be the most recent ancestors of chimpanzees and the more recent ancestors of some extant gorillas. The extant gorillas that shared a MRCA with chimpanzees 4.5 Myr ago may not be distinguishable from other extant gorillas that are descendants of gorilla-like apes living 12 Myr ago.

Thus, the new molecular clock based on the MGD hypothesis differs from the old one in the concept of MRCA for gradual diversifications. Assuming B changed less than A in epigenotypes during gradual diversification, the MRCA of A and B should look like B and is an individual of the B-like lineage. The B-like lineage could have existed many years before the MRCA. While the direct ancestor of A is an individual (or one pair) from the B-like lineage, the ancestors of B could be many individuals from the B-like lineage living at different times. The MRCA of A and B marks the first appearance of A but not the first appearance of B or B-like lineage. It accounts for all extant individuals of A but only a fraction of all extant individuals of B if some of its descendants had remained as B. For such gradual diversifications, the concept of clade with a single MRCA individual accounting for all extant members of the clade is inaccurate.

The new MRCA concept for most gradual diversification processes reconciles the fossil records with molecular phylogeny. It explains the 10.5 Myr old gorilla-like fossil and the much later split of chimpanzees from gorillas at 4.5 Myr ago as estimated by the new molecular clock method. It is also consistent with the trend during gradual diversification that one of the sister taxa is always more similar than the other to the ancestor lineage. Gorillas are the sister taxon of chimpanzees and are more similar to orangutans [59].

Just like the clade concept is inaccurate for gradual diversification, the ‘sister taxon’ concept is also not accurate. The concept is accurate only if a single *individual or pair* is the MRCA to all individuals of each sister taxon of a clade, as may be the case during radiation. But during gradual diversification, one of the sister taxa is also the ancestor of the other taxon. Only a fraction of extant gorillas are sisters with chimpanzees or shared a common individual gorilla ancestor with chimpanzees. For gradual diversification, the concept of sister taxon should only mean that a fraction of the population of one taxon is the sister of the other sister taxon. Chimpanzees are the sisters of gorillas while gorillas are both ancestors and sisters to chimpanzees. Humans are the sisters of pongids while pongids are both ancestors and sisters to humans.

Premises of the slow clock method

The new molecular clock, the “slow clock”, approach here also has two premises. The first is that sequence similarity sometimes (not always) reflects genealogical relationship, which is self-evident and an easily proven fact. Only slow evolving genes, i.e., slow clocks, that have not yet reached maximum distance are informative. The second is the approximate constancy of neutral substitution rate in protein or DNA sequence within the neutral diversity range for any *single* lineage over its evolutionary life time, which is self-evident and much more likely to be true than the old premise that assumes *different* lineages to have the same substitution rate. The well-known ‘stasis’ phenomenon of the fossil record supports this new premise since it indicates morphological stability and by inference molecular stability for any given single lineage. The best evidence for the new premises is the *complete* congruence of the new molecular interpretation with the well-established paleontological results.

While some of the results of the old approach are similar to those of the new approach here, it merely indicates coincidence rather than mutual validation of the two approaches. Even a wrong hypothesis may by chance or by its ad hoc or tautological nature explain a small part of reality. Given the false premises, most of the interpretations of the old approach, even if correct, must be considered inconclusive. It is therefore imperative to perform a complete reevaluation of primate phylogeny by using the new approach.

Primate phylogeny

This reevaluation found no evidence of a gorilla-chimpanzee-human clade with orangutan as the outgroup, a chimpanzee-human clade with gorilla as the outgroup, or a tarsier-simian primate clade with lorises as the outgroup, all controversial clades claimed by the old approach but either contradicted or unresolved/unresolvable by the fossil records. The results indicate the non-existence of these clades rather than inappropriate method of analysis, since the same method positively identified an orangutan-gorilla-chimpanzee clade with human as the outgroup, a gorilla-chimpanzee clade with orangutan as the outgroup, and a loris-tarsier clade with simian primates as the outgroup, all consistent with paleontological findings or traditional views of paleontologists before the molecular clock era when such views were more independent or less biased and more reflective of the fossil record per se.

Within the pongid clade, the orangutans have the largest genetic diversity and genetic distance to humans while chimpanzees the smallest [60,61,62]. This is expected from the MGD hypothesis and the phenomenon of genetic non-equidistance to a more complex outgroup despite equidistance in time.

If the presently popular notion of a human-chimpanzee clade is real, it should be able to be shown by four independent tests using slow evolving genes. First, humans and chimpanzees should be equidistant to gorillas; chimpanzees cannot be closer to gorillas than humans are. Second, humans and chimpanzees should be equidistant to orangutans. Third, humans and gorillas should be equidistant to orangutans. Fourth, orangutans should not be equidistant to chimpanzees and gorillas to the exclusion of humans. However, none of these tests gave results that support a human-chimpanzee clade. In contrast, all support a sister grouping of human and pongids, as well as a sister grouping of gorillas and chimpanzees to the exclusion of humans and orangutans.

The higher similarity between orangutans and gorillas or chimpanzees than between orangutans and humans, as revealed by the genes listed in Table 4, is a result of random selection of genes since the same gene set also showed, as an internal and positive control for randomness, the expected results that orangutans are equidistant to gorillas and chimpanzees. Similarly, the higher similarity between lorises and tarsiers than between lorises and humans, as revealed by the genes listed in Table 6, is a result of random selection of genes since the same gene set also showed, as an internal and positive control, the expected results that lorises are equidistant to New World monkeys and humans. One cannot question the randomness in the selection of these genes without also invalidate a legitimate real result/fact.

The randomness of the selection is also evidenced by the fact that the selection of fast evolving genes is indeed random enough to be able to produce the expected equidistance result. Since the enrollment of a gene into the test was made before it was known or classified as fast or slow evolving genes, evidence for randomness of selection in fast evolving genes is also evidence for the same in slow evolving genes. Thus, for inferring genealogy by equidistance testing to a simpler taxon, an internal control for randomness in gene selection is that the set of fast evolving genes among the selected genes should give equidistance result. When the set of fast evolving genes are random enough to be able to produce an equidistance result, any result associated with the set of slow evolving genes can only be due to randomness in the selection of genes.

To further confirm that the result of Table 4 is independent of gene selections, I analyzed an independently selected list of genes in a previous study, which showed that humans and gorillas are about equidistant to orangutans in average protein identity [56]. I became aware of this list only after the result of Table 4. Only one third of the proteins in this list are also found in Table 4. Among fast evolving proteins, 2 showed more protein identity between orangutan and gorilla than between orangutan and human while 9 showed less. In contrast, among slow evolving proteins, 15 showed more identity between orangutan and gorilla while 6 showed less. The difference between fast and slow evolving proteins is highly significant (*P* = 0.008). Given that the same result was obtained from two independently selected groups of genes, gene selection variations are unlikely to affect the result.

Indeed, the conclusion of genetic non-equidistance as defined by the method here is highly resistant to variations in the random selection of genes. None of the 10 groups of randomly selected fast evolving genes was able to produce an artifactual non-equidistance result as shown in Table 5. This suggests that it is highly unlikely (*P* < 0.05) to produce an artifactual non-equidistance result using the method here. The large scale analysis of nearly 20% of all known genes from orangutans as shown in Table 5 effectively established beyond any reasonable doubt that orangutans are genetically closer to chimpanzees than to humans.

The method here may not reveal a real non-equidistance if the number of genes enrolled in a test is insufficient to reach statistical significance. For example, as shown in Table 5, test groups 2, 8, 9, 10 of slow evolving genes did not show significant non-equidistance. But in all these cases, simply adding more genes to the test would easily produce a statistically significant result. All these groups showed that the number of genes with greater similarity between orangutans and chimpanzees is at least 2 fold greater than the number of genes with greater similarity between orangutans and humans. Therefore, all these groups would be expected to show statistically significant non-equidistance if the number of genes analyzed is increased to 81 (54 vs. 27, P < 0.05). Thus the method may not always reveal a real non-equidistance due to selection variation in the number of genes enrolled in a test. But when a non-equidistance is scored, it is almost always real (*P* < 0.05). If it is not real, it may not score as statistically significant. True equidistance would not be falsely scored as non-equidistance by the method, as indicated by the fact that none of the 10 groups of fast evolving genes in Table 5 showed statistically significant non-equidistance. In contrast, for the non-equidistance in slow evolving genes, even when the number of slow evolving genes enrolled is smaller than that of fast evolving genes, 6 of 10 groups scored statistically significant non-equidistance (Table 5).

Insufficient number of genes cannot account for some of the equidistance results shown here, such as the equidistance of chimpanzees and gorillas to orangutans in both slow and fast evolving genes (Table 4), because the same set of genes did reveal the real non-equidistance of gorillas and humans to orangutans. However, for some equidistance results here that have no positive controls for the sufficiency of gene numbers, such as the equidistance of gibbons to hominoids (Supplementary Table S12), it remains possible although unlikely that the result could be altered when more genes become available for analysis in the future. But these relatively weaker results of equidistance do not affect in any way the certainty of the main results of this study regarding the non-equidistance of orangutans to humans and African apes or the non-equidistance of lorises to humans and tarsiers.

The primate phylogeny as revealed by the new molecular approach is shown in Figure 4. The hominid lineage emerged 17.3 Myr ago, likely from an orangutan-like ancestor given the fossil record and the remarkable similarities between humans and orangutans [30,32,34]. The orangutan-like lineage subsequently gave rise to a gorilla-like lineage 12 Myr ago that next produced the chimpanzees at 4.5 Myr ago, given the fossil record and the closer morphological similarity between gorillas and orangutans than between chimpanzees and orangutans [39,59].

Independent verification of the internal coherence of the primate phylogeny

Given that macroevolution is not accessible to direct experimental testing, internal coherence of all facts within the whole becomes the only criterion for truth, much like the logical coherence of a mathematical proof. It is therefore imperative to verify a conclusion from several independent strategies. Unfortunately, very few past studies have attempted to support a molecular phylogeny from several independent ways, and most anyway would fail such internal coherence testing. Indeed, it is the rule rather than the exception that molecular dating from different studies using different dataset often gave conflicting results especially for macroevolution. This is of course to be expected if the molecular clock paradigm is completely mistaken for macroevolution.

In contrast, the primate phylogeny here represents a coherent picture of several independent facts. The 17.3 Myr divergence time between human and pongid was independently derived three times from different kinds of dataset. It was first calculated from sequence comparison among humans, orangutans, and gorillas using the fossil time of gorilla as calibration for the slow clock (Table 4). Next it was calculated from sequence comparison among humans, orangutans, and chimpanzees (Supplementary Table S10). Finally, it was calculated from sequence comparison among tarsiers, loris, new world monkeys, old world monkeys, orangutans, and humans using the fossil time of tarsiers as calibration for the slow clock. Since few gene sequences are presently known for tarsiers and lorises, the study here in Table 7 used all informative genes available from NCBI database. Therefore, there can be no possibility of purposely manipulating the choice of genes in order to match a result. The independent arrival at the same exact number of 17.3 was a pure coincidence. The same also applies to the data set for kangaroos as in Table 9.

By deducing, from the key results of this study (the human-pongid split at 17.3 Myr ago and the inclusion of tarsiers in the prosimian clade), the molecular time for mammal radiation and for Eutheria-Metatheria split that are actually consistent with well established fossil records, the study here shows for the first time a remarkable and unprecedented concordance between fossil/phenotype and molecule/genotype, as well as the remarkable internal consistency of the primate phylogeny with other molecular and fossil dating results of mammals, such as the split times of mouse-rat and kangaroo-opossum. The results also independently confirmed the validity of fossil dates that are less than definitive such as the 12 myr old gorilla fossil, since such dating is fully consistent with the more well established fossil dating such as mammal radiation at the K/T boundary.

Previous molecular studies all fail to find such complete consistency due to the simple fact that the molecular clock paradigm was false or doomed from the beginning. People were too easily fooled or satisfied by partial consistency but the criterion for truth can only be complete internal consistency for every fact of the whole.

The primate phylogeny is supported by ample fossil data in the literature

There is little morphological support for a human-chimpanzee group [32,34]. However, because of the seeming certainty of the molecular view, most paleoanthropologists began to go along with this view in the 1980s. By downplaying the significance of morphological features that are traditionally viewed important and baselessly regarding them as results of parallel or convergent evolution, they shifted the position of *Ramapithecus/Sivapithecus* from being the ancestor or close sister of humans to that of orangutans. However, some researchers recently suggest that the ancestor or closest sister of orangutans is a fossil from Thailand, *Khoratpithecus,* that lived in the same Miocene period as *Sivapithecus* [63]. *Sivapithecus* differs from *Khoratpithecus* and orangutans in dental characteristics and postcranial skeleton. There is little or no evidence of adaptations for suspension in *Sivapithecus*, and this has caused some anthropologists to doubt the orangutan affinities [64]. Thus, if *Ramapithecus/Sivapithecus* is not a human ancestor or a close sister of that ancestor, then it would have no close relationship with any living primate.

But, as paleoanthropologist Simons put it: “If the immunological dates of divergence devised by Sarich are correct, then paleontologists have not yet found a single fossil related to the ancestry of any living primate and the whole host of species which they have found are all parallelistic imitations of modern higher primates. I find this impossible to believe. [as] it is not presently acceptable to assume that all the fossil primates resembling modern forms are only parallelisms, that highly arboreal apes wandered hundreds of miles out of Africa across the Pontian steppes of Eurasia in search of tropical rain forests, or that Australopithecus sprang full-blown five million years ago, as Minerva did from Jupiter, from the head of a chimpanzee or a gorilla.”[65] The new molecular result here strongly supports the original view of paleoanthropologists on *Ramapithecus* [29,30,31]. It also easily accommodates the 7 Myr old *Sahelanthropus* suggested by some to be the oldest hominid [38,66], which would otherwise be difficult to reconcile with the 5 Myr split time of human and apes.

The fossil literature on humans and apes shows a coherent picture much more consistent with the human-pongid split at 17.3 Myr ago than with any other schemes. There are a number of different fossil apes around 17-15 Myr ago in Africa that can be divided roughly into two major groups according to some authors [64,67]. Group one consists of *Turkanpithecus* and *Kenyapithecus*, and the other group of *Afropithecus*, *Equatorius,* and *Nacholapithecus*. A speculative story that is most consistent with existing data is as follows. Group one has no suspension adaptation in locomotion and may have migrated to Eurasia around 15-14 Myr ago and given rise to one of the two types of *Griphopithecus* and *Sivapithecus* who later may have moved back to Africa around 8-10 Myr ago due to climate change to a temperate one in Eurasia. Group two also may have moved to Eurasia around 15-14 Myr ago and given rise to the other type *Griphopithecus (G. alpani)* and *Dryopithecus*. Change to temperate climate in late Miocene in Eurasia may have caused some *Dryopithecus* to move back to Africa around 12 Myr ago leading to African apes and some (*D. laietanus*) to tropical South East Asia leading to *Khoratpithecus* and orangutans. The African ape ancestors may be more sensitive to temperate climates and disappearance of forests than human ancestors and thus moved back to Africa earlier, at the beginning period of climate cooling.

The two groups of fossil apes at 14-10 Myr ago are more distinctly different than their earlier African ancestors [64]. According to Stringer and Andrews: “Group one has robust jaws, enlarged molar teeth with thick enamel, and some buttressing of the face to accommodate chewing stresses caused by the large teeth and a hard fruit diet. They lived in seasonal woodland to open forest environments and were adapted to some extent to ground living.” [64]. They, I suggest, were the ancestors of humans and later developed bipedalism. To some authors, walking on two legs may arose more likely from a terrestrial form of locomotion on all fours (with on twos occasionally) rather than arboreal climbing and suspension [68,69]. “The other group inhabited wetter, less seasonal forests and lived in trees employing a form of locomotion that involves some degree of suspension from overhead branches. Their jaws were more lightly built and their teeth not enlarged, so that their diet must have been soft fruits.” [64]. They are obviously the best candidates for the ancestors of pongids.

The main seeming inconsistency with this story is the intermediate thin enamel of *Dryopithecus* being unlike the intermediate thickness in orangutans and in the oldest fossil gorilla *Chororapithecus*. But in truth, enamel thickness is not an informative feature and may become thin or thick in several independent lineages. It can also vary a great deal within a species [70]. For example, while most *Australopithecus* like fossils around 4 Myr ago have thick enamel, consistent with being human, some like *Ardipithecus* has thin ones. Some *Proconsul* has thin ones while some other *Proconsul* has thick ones [70]. Besides, the enamel thickness of orangutan is really an intermediate between human and African apes [70], and its enamel deposition rate is slow like African apes rather than fast like *Sivapithecus*/humans [71]. In contrast to this trivial dental inconsistency, the human-chimpanzee grouping must assume the extremely non-parsimonious position that one of the two major groups of Miocene apes had contributed little to any living higher primates, despite the fact that it was far more abundant in number, wider in geographical distribution, more adapted to ground living, and therefore more like the situation of humans today.

Some researchers have suggested a human-orangutan clade to the exclusion of African apes based on derived shared morphologies [32,34,71]. But such analysis suffers from an inherent difficulty in cladistic analysis that deems its conclusion unreliable. Such analysis assumes each feature to be independent of each other and carries equal weight, an assumption that is more likely to be false than true and cannot be independently verified. In most cases, one major feature, such as the vertebra, is enough and can/should override numerous other features.

Convergent evolution could account for the similarity between orangutans and humans. Also can the loss of ancestor features in the African apes. Loss or displacement of ancestor features for some lineages within a clade is in fact quite common for many clades during gradual evolution (e.g., loss of limbs in snakes). It is nearly always possible to find a sister lineage to be more similar to an outgroup than other sister lineages, both in terms of morphology or DNA as shown by the genetic non-equidistance result here. I have invented the slow clock method to avoid such problems in molecular phylogeny. It is now a challenge for morphologists to develop an equivalent method to avoid the problems of convergent evolution or of loss/displacement of features as well as to separate major from minor features, which is especially important for hominoids as “parallel evolution in the jaws, teeth, and facial structure of hominoids appears to be the rule rather than the exception.” [72]. If that seems like an impossible task for them, then they would have no choice but to accept the molecular results of the slow clock method, which at least for now has no known problems or difficulties.

Chimpanzees had lived side by side with humans in the past in areas suitable for fossil formations [73]. The emergence of chimpanzees from a gorilla-like lineage was here calculated to be 4.5 Myr ago, assuming similar substitution rates for gorillas and orangutans in slow evolving genes (Supplementary Table S15). The only known ancient fossil of chimpanzees has an age of 0.5 Myr [73]. The much more recent emergence of chimpanzees easily explains the extreme rarity of chimpanzee fossils relative to that of humans (or even to gorillas). Shorter time of lineage existence and small population size would both reduce the chance for fossil formation. Chimpanzees had much less time to expand their populations.

The division between humans and great apes is obviously a fundamental one in many respects, especially in the brain or intelligence. Anyone who discounts that and considers himself just a third chimpanzee deserves to be and can only logically hope to be treated like a stupid ape by real humans [74]. If one does not take his intelligence and hence his thoughts seriously in the first place, why should anyone else? Real humans could care less about what a chimpanzee may think about evolution, regardless whether he is the third one or not.

Chimpanzees have much more phenotypes in common with other great apes than with humans. Few informative features are known that are shared by chimpanzees and humans to the exclusion of gorillas and orangutans, much less than those shared between humans and orangutans [32,34]. It is much less parsimonious to suggest that the many features shared between humans and orangutans were regained by humans after being lost in the MRCA of humans and chimpanzees. It is likely that human like anatomical features were present in the MRCA of pongids and were retained in orangutans but lost or displaced in African apes. Relative to orangutans, the closer sequence similarity of chimpanzees to humans did not translate into more phenotype similarities with humans. This highlights the point of the MGD hypothesis that the key determinant of phenotypes in complex animals is epigenetic programs. The importance of epigenetics gradually increased in a stepwise way during macroevolution. Major chromosome reorganizations such as the change from 24 pair in pongids to 23 pair of chromosomes in humans certainly qualify as major epigenetic changes rather than purely genetic. As shown by the Cambrian explosion, major divisions in forms occur prior to minor divisions [75]. If this is a real pattern, it is expected that the emergence of humans should have occurred prior to any finer differentiation further along the line of a great ape.

By the same reasoning, it is more likely for anthropoids to appear early rather than late during diversification of placental mammals. If the Cambrian radiation created vertebrates together with invertebrates, it would be inconsistent if the radiation of mammals did not produce anthropoids together with other diverse mammals. My result here provides for the first time molecular evidence for anthropoid origin around the K/T boundary, well consistent with fossil evidence such as *Altiatlasius* and *Eosmias* [50,52,53].

There are diverse opinions among paleontologists about the position of tarsiers [41,53]. Given the problem and complexity of convergent evolution and the inherent difficulty with cladistic analysis, it may be impossible to reach a firm conclusion based on morphology alone. It is true as shown by previous molecular analysis that tarsiers are closer in sequence similarity to simian primates than lemurs/lorises are. But that is likely due to convergent evolution, because Tarsiers show more features of higher epigenetic complexity than other prosimians, including long gestation time and brain at birth largest among mammals relative to body size [41]. From the fossil literature, the most likely MRCA or closest sister lineage for the prosimian clade (including tarsiers, lorises, and lemurs) is the mysterious omomyid *Rooneyia* from ~40 Myr ago that also shows lemur-like features [50]. While it is thought by some paleontologists that omomyid gave rise to tarsiers while adapids to lemurs, there are really no definitive morphological evidence, and adapids have also been thought by some as ancestors of anthropoids [76].

The pongid clade is supported by ample molecular data in the literature

I here summarize the large amount of molecular data in the literature that supports the pongid clade as found here. First, human is closer to orangutan than chimpanzee is in neutral sequences as measured by Ks but is more distant to orangutan than chimpanzee is in non-neutral sequences as measured by Ka [77]. Since neutral sequences evolve faster than non-neutral sequences, this observation is fully consistent with the result here that human is more distant to orangutan in slow evolving genes.

Second, chimpanzee is closer to orangutan than human is in gene expression pattern, suggesting a distinction between humans and pongids in epigenetic programs [78]. Also, chimpanzee is closer to gorilla than human is in gene expression pattern in the brain and fibroblasts [79,80].

Third, retrovirus insertion pattern shows sister grouping of chimpanzees and gorillas to the exclusion of humans and orangutans [81,82]. The presence or absence of certain repetitive DNA elements such as Alu can both support or contradict the sister grouping of humans and chimpanzees, and is therefore not an informative marker [83,84]. There are many other genetic differences between humans and the African apes, including cytogenetic differences, abundance and distribution of endogenous retroviruses, differences in the type and number of repetitive genomic DNA and transposable elements, the presence and extent of allelic polymorphisms, specific gene inactivation events, gene sequence differences, gene duplications, single nucleotide polymorphisms, gene expression differences, and messenger RNA splicing variations [85]. In contrast, human and chimpanzee share very few molecular features that are not also shared by gorillas, inconsistent with a human-chimpanzee clade.

Fourth, the chromosome-banding pattern of humans is more similar to orangutan than to chimpanzee or gorilla [86]. Ten chromosomes show similar patterns in human and orangutan (chromosome 5, 6, 8, 12, 13, 19, 20, 21, 22, X,), whereas only 1 (chromosome 3) does in human and chimpanzee. This is consistent with the observation that human shares more segmental duplications with orangutan than chimpanzee does [87], since duplications are likely to affect chromosome banding patterns. Also, seven chromosomes show similar patterns between chimpanzee and gorilla (chromosome 2q, 6, 7, 11, 12, 16, X), whereas none does between human and gorilla.

Fifth, consistent with low genetic diversity in humans, human specific segmented duplications show lower copy number polymorphisms in humans than chimpanzee specific segmented duplications do in chimpanzees [87]. Similarly, those duplications shared among human, chimpanzees, and orangutans, or those shared among human, chimpanzees, orangutans, and monkeys are also less polymorphic in humans than in chimpanzees, indicating clearly that duplications that are shared because of common ancestry are less polymorphic in humans than in chimpanzees. In contrast, the duplications shared between human and chimpanzees are equally polymorphic in humans and chimpanzees. This unusual result contradicts the sister grouping of humans and chimpanzees, because both the MGD and the bottleneck hypothesis would predict lower polymorphism in humans if these duplications are shared because of common ancestry. However, it is fully consistent with the interpretation that the shared duplications between humans and chimpanzees are not due to common ancestry but are due to common selection of independent duplications. Common selection leading to shared sequences is well established [88,89,90]. The MGD hypothesis interprets many of the shared sequences between humans and chimpanzees as a result of common selection rather than common ancestry. The similar selection pressure leads to similar levels of polymorphism. This result is thus one of the best that simply cannot be reconciled in any way with the sister grouping of humans and chimpanzees but fully supports the MGD hypothesis and the sister grouping of humans and pongids.

Finally, the pongid clade resolves inconsistencies in the literature on the functional constraint on gene control regions in hominid genomes. Gene control regions conserved between human and chimpanzees are found in some studies to be under less selective constraint in hominids than those between mouse and rat do in murids [91,92]. This observation seems extremely anti-intuitive and against the axiom of the MGD hypothesis. In contrast, another study found that functional non-coding regions conserved among human, mouse, and dog are subject to significant selective constraint in hominids [93]. These seemingly conflicting observations in fact are completely consistent with each other if one accepts the pongid clade but cannot be reconciled under the sister grouping of human and chimpanzee. The regions studied by Keightley et al are conserved regions between human and chimpanzee, which are mostly due to common selection or convergent evolution rather than common ancestry. However, the regions studied by Bush and Lahn are conserved regions among human, mouse and dog, which are mostly due to common ancestry. Studies on segmental duplications has shown that duplications due to common ancestry show less polymorphisms in humans or chimpanzees than do duplications due to convergent evolution that are shared between human and chimpanzee [87]. So, sequences shared due to convergent evolution are subject to less selective constraint than those due to common ancestry.

Conclusions

The MGD is the only complete evolution theory that can explain all relevant facts and has not a single contradiction. The molecular clock hypothesis should never have been invented in the first place for macroevolution if people had paid attention to the overlap feature of the equidistance result. Thus, new and correct methods for molecular phylogeny analysis of macroevolution need to be invented. The MGD suggests that inferring genealogy should make use of the genetic non-equidistance to a simpler taxon as measured by slow evolving sequences. This slow clock method showed that humans are genetically more distant to orangutans than African apes are and separated from pongids 17.3 Myr ago. Also, tarsiers are genetically closer to lorises than simian primates are, suggesting a tarsier-loris clade to the exclusion of simian primates. The validity and internal coherence of the primate phylogeny here were independently verified. There exists a remarkable and unprecedented concordance between molecules and fossils that has remained hidden from view until now as revealed by the MGD hypothesis.

Acknowledgments:

I thank Wenyun Zhou for technical assistance. I thank John Grehan for reading the manuscript and for valuable discussions.

Methods

Sequence selection and alignments

Protein sequences from a specific taxon were retrieved from the NCBI protein database. For example, to retrieve all orangutan/pongo protein or cDNA sequences, I did Search for Pongo on the NCBI home page (using the word Pongo to search the Protein database). This returned 8206 items or sequences on 411 webpages. The 4330 random cDNAs (all named as hypothetical proteins) of *Pongo abelli* from the German cDNA consortium are located on webpage 21-237. Homology comparisons were performed using BLASTP on the NCBI server.

Genetic equidistance test

Genetic equidistance of taxon A and B to C can be established if the number of genes showing greater similarity between A and C than between B and C is similar to the number of genes showing less similarity between A and C than between B and C (P > 0.05). Each gene was randomly selected from the NCBI database without any intentional bias or intent to influence in any biased way the outcome of the equidistance test. The rationale of the method is straightforward. If A and B are equidistant (or non-equidistant) to C at the whole genome level, then a random sampling of a small set of the genome should show the same. Equidistance means that, while some genes may show exact equidistance, some would show approximate equidistance (non exact identity). For genes that show approximate equidistance, the number of genes with greater similarity between A and C than between B and C should be similar to the number of genes with less similarity between A and C than between B and C (P > 0.05). Thus, the informative genes in the method here are genes that show approximate equidistance.

This method of determining genetic equidistance contains no uncertain premises and is more reliable and meaningful than existing methods, such as the relative rate test, which all fail to take into account the maximum cap on genetic distance and assume incorrectly that mutations accumulate equally in the two diverging lineages in all cases regardless of the difference in epigenetic complexity of the lineages. The validity of this method, besides being self-evident, has been verified as shown in Table 5. None of the 10 independently selected groups of fast evolving genes produced artifactual violation of an expected equidistance result. The possibility of a false-positive by this method is therefore insignificant ($P < 0.05$).

The method relies on the availability of a set of randomly selected genes that is large enough for reaching statistical significance. But the exact nature of the genes (function type, reason for study, and time or order of appearance in the Genbank) is independent of their utility in the equidistance test. Thus, while the availability of a gene sequence in the Genbank has specific reasons and hence is not strictly random, none of the reasons is in anyway linked to the equidistance test. Their availability in the Genbank is therefore effectively random as far as the equidistance test is concerned. Any non-biased selection scheme of these genes would satisfy the randomness requirement of the equidistance testing method here.

A straightforward and simple scheme employed here was to select genes based on their numerical order of appearance on the NCBI webpage. Overrepresentation of genes of the same functional type was avoided when possible, although no evidence was found for such overrepresentation affecting in anyway the result of the equidistance test. The enrollment of genes for a test was stopped when the number of genes already enrolled was enough for drawing statistically significant conclusions. Each gene was enrolled prior to knowing its effect on the final result of the test. No gene was either included in or excluded from a test after knowing its effect on the test result.

The classification of a gene as fast or slow evolving was made after the enrollment of the gene for any given test. The cutoff score in percent identity was arbitrarily made for each test so that the number of fast evolving genes is approximately similar to that of slow evolving genes to ensure that each set has sufficient number of genes for statistical testing. For inferring genealogy by equidistance testing to a simpler taxon, an internal control for randomness of gene selection is that the set of fast evolving genes should give equidistance result. When the set of fast evolving genes are random enough to be able to produce an equidistance result, any result associated with the set of slow evolving genes can only be due to randomness in the selection of genes.

The complete genome of gorilla or orangutan has yet to be completed. It is not yet possible to test whether orangutan is equidistant to humans and gorillas/chimpanzees using whole genome data. However, a large set of randomly selected cDNAs of orangutan (*Pongo abelli*) have been sequenced and recently deposited in the Genbank by the German cDNA consortium. These cDNAs (4338 in total) represent nearly 20% of known genes. An analysis of all these cDNAs was performed to verify that the result of the equidistance testing method here is independent of gene selections. The cDNAs were arbitrarily divided into 10 groups (each with 433 genes). Starting from the first cDNA CAH89494, every 433 genes form a group based on the numerical order on the NCBI webpage. If all 10 groups gave the same type of result, the result would be significant (10 positive vs. 0 negative, $p < 0.05$). This large scale analysis would confirm that small scale analysis using smaller number of genes is good enough for the equidistance testing method here to give meaningful result. This has indeed been confirmed. All 10 groups showed the expected equidistance result that chimpanzees and humans are equidistant to orangutans in fast evolving genes (Table 5).

For the equidistance test, non-informative genes include those that have no orthologous Genbank sequences in one of the concerned taxa, have long alignment gaps, are identical among the taxa, show exact equidistance from the outgroup, under strong positive selection (for example, major histocompatibility complex genes), or have many polymorphisms that prevent meaningful inference of equidistance.

Calculation of divergence time

Calculation of human-orangutan divergence time based on the gorilla fossil split time of 12 Myr ago was performed using the formula: Divergence time of human and orangutan = 12 x the Poisson correction distance for any given protein between human and orangutan divided by the Poisson correction distance between gorilla and orangutan. The method of using the Poisson correction distance to infer divergence time is commonly used today, especially for distances that are less than 20% in percent identity [14,45]. To ensure randomness of gene selection, all genes used for calculation of divergence time were selected without any prior knowledge on how each gene may affect the outcome of the calculation.

Statistical methods

Statistical methods used were Student’s t test and Fisher’s exact test, 2 tailed.

References:

1. Zuckerkandl E, Pauling L (1962) Molecular disease, evolution, and genetic heterogeneity, Horizons in Biochemistry; Kasha M, Pullman B, editors. New York: Academic Press.

2. Margoliash E (1963) Primary structure and evolution of cytochrome c. Proc Natl Acad Sci 50: 672-679.

3. Kumar S (2005) Molecular clocks: four decades of evolution. Nat Rev Genet 6: 654-662.

4. Huang S (2009) Molecular evidence for the hadrosaur B. canadensis as an outgroup to a clade

containing the dinosaur T. rex and birds. Riv Biol 102: 20-22.

5. Huang S (2008) Ancient fossil specimens are genetically more distant to an outgroup than extant sister species are. Riv Biol 101: 93-108.

6. Huang S (2008) The genetic equidistance result of molecular evolution is independent of mutation rates. J Comp Sci Syst Biol 1: 092-102.

7. Huang S (2009) Inverse relationship between genetic diversity and epigenetic complexity. Submitted Preprint available at Nature Precedings <http://dxdoiorg/101038/npre200917512>

8. Pulquerio MJ, Nichols RA (2007) Dates from the molecular clock: how wrong can we be? Trends Ecol Evol 22: 180-184.

9. Laird CD, McConaughy BL, McCarthy BJ (1969) Rate of fixation of nucleotide substitutions in evolution. Nature 224: 149-154.

10. Jukes TH, Holmquist R (1972) Evolutionary clock: nonconstancy of rate in different species. Science 177: 530-532.

11. Goodman M, Moore GW, Barnabas J, Matsuda G (1974) The phylogeny of human globin genes investigated by the maximum parsimony method. J Mol Evol 3: 1-48.

12. Langley CH, Fitch WM (1974) An examination of the constancy of the rate of molecular evolution. J Mol Evol 3: 161-177.

13. Li W-H (1997) Molecular evolution. Sunderland, MA: Sinauer Associates.

14. Nei M, Kumar S (2000) Molecular evolution and phylogenetics. New York: Oxford University Press.

15. Avise JC (1994) Molecular markers, natural history and evolution. New York, NY: Springer.

16. Van Valen L (1974) Molecular evolution as predicted by natural selection. J Mol Evol 3: 89-101.

17. Clarke B (1970) Darwinian evolution of proteins. Science 168: 1009-1011.

18. Richmond RC (1970) Non-Darwinian evolution: a critique. Nature 225: 1025-1028.

19. Kimura M (1968) Evolutionary rate at the molecular level. Nature 217: 624-626.

20. Kimura M, Ohta T (1971) On the rate of molecular evolution. J Mol Evol 1: 1-17.

21. King JL, Jukes TH (1964) Non-Darwinian evolution. Science 164: 788-798.

22. Ayala FJ (1999) Molecular clock mirages. BioEssays 21: 71-75.

23. Huang S (2009) Molecular clock at best explains half the story on ‘genetic equidistance’ and at worst explains none. The Golden Gnomon http://thegoldengnomon.blogspot.com/2009/04/molecular-clock-at-best-explains-half.html.

24. Huang S (2009) The overlap feature of hemoglobin. The Golden Gnomon http://thegoldengnomon.blogspot.com/2009/05/overlap-feature-of-hemoglobin.html.

25. Huang S (2008) Histone methylation and the initiation of cancer, Cancer Epigenetics; Tollefsbol T, editor. New York: CRC Press.

26. Nevo E (2001) Evolution of genome-phenome diversity under environmental stress. Proc Natl Acad Sci U S A 98: 6233-6240.

27. Halabi N, Rivoire O, Leibler S, Ranganathan R (2009) Protein sectors: evolutionary units of three-dimensional structure. Cell 138: 774-786.

28. Huang S (2009) Complexity Associated Protein Sectors (CAPS), new evidence for the MGD hypothesis. The Golden Gnomon http://thegoldengnomon.blogspot.com/2009/09/complexity-associated-protein-sectors.html.

29. Simons EL (1961) The phyletic position of Ramapithecus. . Postilla 57: 1-9.

30. Simons EL, Pilbeam DR (1965) Preliminary revision of the Dryopithecinae (Pongidae, Anthropoidea). Folia Primatol (Basel) 3: 81-152.

31. Pilbeam D (1968) The earliest hominids. Nature 219: 1335-1338.

32. Schwartz JH (1984) The evolutionary relationships of man and orang-utans. Nature 308: 501-505.

33. Lewin R (2005) Human Evolution. Malden, MA 02148, USA: Blackwell Publishing Ltd.

34. Schwartz JH (2005) The Red Ape, Orangutans and Human Origins. Cambridge, MA 02142, USA: Westview Press.

35. Goodman M (1962) Immunochemistry of the primates and primate evolution. Ann N Y Acad Sci 102: 219-234.

36. Sarich VM, Wilson AC (1967) Immunological time scale for hominid evolution. Science 158: 1200-1203.

37. Wilson AC, Sarich VM (1969) A molecular time scale for human evolution. Proc Natl Acad Sci U S A 63: 1088-1093.

38. Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, et al. (2002) A new hominid from the Upper Miocene of Chad, Central Africa. Nature 418: 145-151.

39. Suwa G, Kono RT, Katoh S, Asfaw B, Beyene Y (2007) A new species of great ape from the late Miocene epoch in Ethiopia. Nature 448: 921-924.

40. Shoshani J, Groves CP, Simons EL, Gunnell GF (1996) Primate phylogeny: morphological vs. molecular results. Mol Phylogenet Evol 5: 102-154.

41. Schwartz JH (2003) How close are the similarities between Tarsius and other primates? In Tarsiers: Past, Present, and Future; Wright PC, Simons EL, Gursky S, editors. Piscataway, N.J.: Rutgers University Press.

42. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, et al. (2007) The delayed rise of present-day mammals. Nature 446: 507-512.

43. Wible JR, Rougier GW, Novacek MJ, Asher RJ (2007) Cretaceous eutherians and Laurasian origin for placental mammals near the K/T boundary. Nature 447: 1003-1006.

44. Flynn JJ, Parrish JM, Rakotosamimanana B, Simpson WF, Wyss AR (1999) A new Middle Jurassic mammals from Madagascar. Nature 401: 57-60.

45. Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392: 917-920.

46. Luo ZX, Ji Q, Wible JR, Yuan CX (2003) An Early Cretaceous tribosphenic mammal and metatherian evolution. Science 302: 1934-1940.

47. Benton MJ, Donoghue PC (2007) Paleontological evidence to date the tree of life. Mol Biol Evol 24: 26-53.

48. Rossie JB, Ni X, Beard KC (2006) Cranial remains of an Eocene tarsier. Proc Natl Acad Sci U S A 103: 4381-4385.

49. Seiffert ER, Simons EL, Attia Y (2003) Fossil evidence for an ancient divergence of lorises and galagos. Nature 422: 421-424.

50. Bajpai S, Kay RF, Williams BA, Das DP, Kapur VV, et al. (2008) The oldest Asian record of Anthropoidea. Proc Natl Acad Sci U S A 105: 11093-11098.

51. Sige B, Jaeger JJ, Sudre J, Vianey-Liaud M (1990) Altiatlasius koulchii n. gen. et sp., primate omomyidé du Paléocène supérieur du Maroc, et les origines des euprimates. . Palaeontographica Abt A 214: 31-56.

52. Beard C (2004) The Hunt for the Dawn Monkey. Berkeley: University of California Press.

53. Kay RF, Ross C, Williams BA (1997) Anthropoid origins. Science 275: 797-804.

54. Simons EL (1995) Skulls and anterior teeth of Catopithecus (primates:Anthropoidea) from the Eocene and anthropoid origins. Science 268: 1885-1888.

55. Elango N, Thomas JW, Yi SV (2006) Variable molecular clocks in hominoids. Proc Natl Acad Sci U S A 103: 1370-1375.

56. Wildman DE, Uddin M, Liu G, Grossman LI, Goodman M (2003) Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: enlarging genus Homo. Proc Natl Acad Sci U S A 100: 7181-7188.

57. Dawkins R (2004) The Ancestor's Tale. Boston: Houghton Mifflin Company.

58. Rowe T, Rich TH, Vickers-Rich P, Springer M, Woodburne MO (2008) The oldest platypus and its bearing on divergence timing of the platypus and echidna clades. Proc Natl Acad Sci U S A 105: 1238-1242.

59. Collard M, Wood B (2000) How reliable are human phylogenetic hypotheses? Proc Natl Acad Sci U S A 97: 5003-5006.

60. Fischer A, Pollack J, Thalmann O, Nickel B, Paabo S (2006) Demographic history and genetic differentiation in apes. Curr Biol 16: 1133-1138.

61. Zhi L, Karesh WB, Janczewski DN, Frazier-Taylor H, Sajuthi D, et al. (1996) Genomic differentiation among natural populations of orang-utan (Pongo pygmaeus). Curr Biol 6: 1326-1336.

62. Xu X, Arnason U (1996) The mitochondrial DNA molecule of Sumatran orangutan and a molecular proposal for two (Bornean and Sumatran) species of orangutan. J Mol Evol 43: 431-437.

63. Chaimanee Y, Suteethorn V, Jintasakul P, Vidthayanon C, Marandat B, et al. (2004) A new orang-utan relative from the Late Miocene of Thailand. Nature 427: 439-441.

64. Stringer C, Andrews P (2005) The Completer World of Human Evolution. New York: Thames and Hudson.

65. Simons EL (1969) The origin and radiation of primates. Ann New York Acad Sci 167: 319-331.

66. Lebatard AE, Bourles DL, Duringer P, Jolivet M, Braucher R, et al. (2008) Cosmogenic nuclide dating of Sahelanthropus tchadensis and Australopithecus bahrelghazali: Mio-Pliocene hominids from Chad. Proc Natl Acad Sci U S A 105: 3226-3231.

67. Cela-Conde CJ, Ayala FJ (2007) Human Evolution: Trails from the Past. Oxford: Oxford University Press.

68. Schwartz JH (2007) The origins of human bipedalism. Science: 1065.

69. Thorpe SK, Holder RL, Crompton RH (2007) Origin of human bipedalism as an adaptation for locomotion on flexible branches. Science 316: 1328-1331.

70. Smith TM, Martin LB, Leakey MG (2003) Enamel thickness, microstructure and development in Afropithecus turkanensis. J Hum Evol 44: 283-306.

71. Grehan JR (2006) Mona Lisa smile: the morphological enigma of human and great ape evolution. Anat Rec B New Anat 289: 139-157.

72. Kay RF (1982) Sivapithecus simonsi, a new species of miocene hominoid, with comments on the phylogenetic status of the ramapithecinae. Int J Primatology 3: 113-173.

73. McBrearty S, Jablonski NG (2005) First fossil chimpanzee. Nature 437: 105-108.

74. Diamond J (1992) The Third Chimpanzee: The Evolution and Future of the Human Animal. New York: Harper Perennial.

75. Valentine JW, Awramik SM, Signor PW, Sadler PM (1991) The Biological Explosion at the Precambrian-Cambrian Boundary. Evolutionary Biology 25: 279-356.

76. Franzen JL, Gingerich PD, Habersetzer J, Hurum JH, von Koenigswald W, et al. (2009) Complete primate skeleton from the Middle Eocene of Messel in Germany: morphology and paleobiology. PLoS One 4: e5723.

77. Chen FC, Li WH (2001) Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 68: 444-456.

78. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, et al. (2002) Intra- and interspecific variation in primate gene expression patterns. Science 296: 340-343.

79. Uddin M, Wildman DE, Liu G, Xu W, Johnson RM, et al. (2004) Sister grouping of chimpanzees and humans as revealed by genome-wide phylogenetic analysis of brain gene expression profiles. Proc Natl Acad Sci U S A 101: 2957-2962.

80. Karaman MW, Houck ML, Chemnick LG, Nagpal S, Chawannakul D, et al. (2003) Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. Genome Res 13: 1619-1630.

81. Yohn CT, Jiang Z, McGrath SD, Hayden KE, Khaitovich P, et al. (2005) Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. PLoS Biol 3: e110.

82. Barbulescu M, Turner G, Su M, Kim R, Jensen-Seaman MI, et al. (2001) A HERV-K provirus in chimpanzees, bonobos and gorillas, but not humans. Curr Biol 11: 779-783.

83. Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, et al. (2004) Differential alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Res 14: 1068-1075.

84. Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, et al. (2003) Alu elements and hominid phylogenetics. Proc Natl Acad Sci U S A 100: 12787-12791.

85. Gagneux P, Varki A (2001) Genetic differences between humans and great apes. Mol Phylogenet Evol 18: 2-13.

86. Yunis JJ, Prakash O (1982) The origin of man: a chromosomal pictorial legacy. Science 215: 1525-1530.

87. Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, et al. (2009) A burst of segmental duplications in the genome of the African great ape ancestor. Nature 457: 877-881.

88. Bull JJ, Badgett MR, Wichman HA, Huelsenbeck JP, Hillis DM, et al. (1997) Exceptional convergent evolution in a virus. Genetics 147: 1497-1507.

89. Bollback JP, Huelsenbeck JP (2009) Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. Genetics 181: 225-234.

90. Castoe TA, de Koning AP, Kim HM, Gu W, Noonan BP, et al. (2009) Evidence for an ancient adaptive episode of convergent molecular evolution. Proc Natl Acad Sci U S A.

91. Keightley PD, Lercher MJ, Eyre-Walker A (2005) Evidence for widespread degradation of gene control regions in hominid genomes. PLoS Biol 3: e42.

92. McVicker G, Gordon D, Davis C, Green P (2009) Widespread Genomic Signatures of Natural Selection in Hominid Evolution. PLoS Genet 5: e1000471.
doi:1000410.1001371/journal.pgen.1000471.

93. Bush EC, Lahn BT (2005) Selective constraint on noncoding regions of hominid genomes. PLoS Comput Biol 1: e73.

Table 1. Human relationship with mollusks. The percent identities in protein sequence between species ***(****Octopus vulgaris, Acanthocardia tuberculatum,* and *Homo sapiens*) are shown for 10 mitochondrial proteins.

		Percent 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 2. Human relationship with brachiopods. The percent identities in protein sequence between species (*Terebratulina retusa*, *Lingula anatina,* and *Homo sapiens*) are shown for 10 mitochondrial proteins.

		Percent 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 3. Fast evolving genes reach the maximum distance faster. The percent 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.

	Percent Identity		
	D. rerio vs.		
	T. nigroviridis	*H. sapiens*	*M. musculus*
Genes reached cap			
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
Genes not yet reached cap			
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

Table 4. Orangutans are closer to gorillas or chimpanzees than to humans but are equidistant to gorillas and chimpanzees. Protein sequences from orangutans were randomly retrieved from Genbank and used to BLASTP human, chimpanzee, and gorilla protein databases at NCBI. Among the 64 informative proteins listed here, about half (30) were arbitrarily grouped as fast evolving genes based on the percent identity between orangutans and gorillas being equal to or lower than 95%. Divergence time between orangutan and human was calculated based on the fossil split time of gorilla of 12 Myr ago. The average divergence time was calculated using slow evolving genes. Four 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 gorillas. To compensate for this loss of genes showing the greatest time of split between orangutans and humans, four genes from the list showing less similarity between orangutans and gorillas are also excluded, which show the smallest distance between orangutans and humans.

	Number of identical amino acids			% Identity	Div. time (Myr)
	Or.-Hu.	Or.-Ch.	Or.-Go.	Or.-Go.	Or.-Hu.
Or.-Go. > Or.-Hu., Slow evolving, 27 genes:					
APOE	310	312	311/317	98	14.1
MBP1	228	228	229/235	97	14.1
KLK3	175	175	178/180	98	30.5
T2R38	298	298	299/310	96	13.1
ASIP	126	129	129/132	97	24.3
WNT7A	346	349	349/349	100	ni
FSHB	127	127	128/129	99	24.0
GSC	254	255	255/257	99	18.1
Myostatin	374	374	375/375	100	ni
GPR56	667	671	670/687	97	14.2
BRCA1	1098	1110	1108/1141	96	15.9
RNAseA1	149	150	151/156	96	16.7
MAOA	101	102	102/103	99	24.0
HNMT	112	112	113/117	96	15.0
SCML2	175	176	176/176	100	ni
CXCR4	346	346	347/347	100	ni
UTY	210	214	217/226	96	21.5
CFTR	1464	1465	1466/1480	99	16.0
Oxytocin receptor	283	285	284/289	98	14.4
CXCR2	340	340	342/355	96	14.0
ASPM	3393	3398	3395/3447	98	12.5
CCR5	349	351	351/352	99	36.7
FUT2	330	330	331/343	96	13.0
Prion	248	248	249/253	98	15.0
TPMT	235	237	236/245	96	13.3
Globin a2	137	138	139/141	97	24.7
COX1	494	494	497/512	97	14.3
Or.-Go. < Or.-Hu., Slow evolving, 7 genes:					
CHRM5	290	278	286/296	96	6.0ni

MET	1382	1383	1380/1390	99	9.6ni
HTR1F	362	362	359/365	98	6.0ni
CHRM3	582	582	580/590	98	9.6
FMO 2	527	527	525/535	98	9.6
A4GALT	214	212	211/218	99	6.9ni
CORTBP2	1638	1635	1633/1663	98	10.0
				Average:	17.3 ± 6.7

Or.-Go. > Or.-Hu., Fast evolving, 14 genes:

ND2	297	299	298/346	86
APOBEC3G	334	335	335/384	87
COX2	214	220	219/227	94
COX3	241	241	243/261	93
Trim5	461	465	466/493	94
ND6	164	164	166/174	94
COB	339	339	342/378	90
MCPH1	801	806	805/839	95
MAPT	454	454	455/480	94
NACA2	199	204	201/210	95

SEMG2	427	ni	428/459	93
Saitohin	119	120	121/128	94
T2R10	234	235	236/248	95
T2R48	257	255	258/280	92

Or.-Go. < Or.-Hu., Fast evolving, 16 genes:

MRGX2	316	314	313/330	95
Elafin	111	111	110/117	94
Leptin	141	141	140/146	95
T2R41	282	281	280/307	91
T2R5	286	282	284/299	94
T2R4	268	268	263/277	95
Twist	193	190	185/203	91
Rh50	388	387	385/409	94
MC1R	305	305	296/317	93
OR1D2	279	279	275/313	87

ND5	498	496	485/585	83
ND4	407	404	403/458	88
ND1	277	273	274/318	87
ATP6	188	188	181/226	80
RNAse3	131	131	130/153	85
T2R14	282	279	280/318	88

Table 5. *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. 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 percent 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).

	Genes			
	Analyzed	Number of	Number of genes	
	Start-	Informative	C-O > H-O vs. C-O < H-O	
Groups	End	genes	>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		(0.016)	(0.36)
4	CAH92408-	119	29 vs. 7	35 vs. 48
	CAH91971		(0.013)	(0.35)
5	CAH91970-	98	28 vs. 8	25 vs. 37
	CAH91540		(0.026)	(0.38)
6	CAH91539-	105	28 vs. 9	33 vs. 35
	CAH91107		(0.032)	(1.00)
7	CAH91106-	106	22 vs. 6	33 vs. 45
	CAH90673		(0.049)	(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		(0.069)	(0.55)
Total	4330	1056	247 vs. 80	335 vs. 384
			(< 0.0001)	(0.21)

Table 6. 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.

	No. identical amino acid		% identity	No. id. a.a.	% identity
	Lo-Hu	Lo-Ta	Lo-Ta	Lo-NWM	Lo-NWM
Lo.-Ta. > Lo.-Hu.,	*Slow evolving, 19 genes:*				
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 hormone	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
Lo.-Ta. < Lo.-Hu.,	*Slow evolving, 3 genes:*				
EDG1	151	144/158	91	ni	
HBB	135	130/146	89	133/146	91
AAR2B	336	321/368	87	324/368	88
Lo.-Ta. > Lo.-Hu.,	*Fast evolving, 10 genes:*				
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

Lo.-Ta. < Lo.-Hu., Fast evolving, 8 genes:

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

Table 7. Calculation of divergence time among primates. The slow evolving genes from Table 6 were used except two genes that are non informative for new world monkeys (NWM or N). Calculation of divergence time between NWM and lorises/prosimians was calibrated using the fossil split time of 40 Myr between tarsier and loris. This gave rise to a molecular split time of 66.7 Myr between NWM and prosimians, which was next used as calibration to calculate the divergence time between OWM and NWM. Such calculation gave rise to a molecular split time of 47.8 Myr between OWM and NWM, which was next used as calibration to calculate the divergence time between orangutan (Or) and OWM. This gave rise to a molecular split time of 29.7 Myr between orangutan and OWM, which was next used as calibration to calculate the divergence time between human and orangutan. Also, the divergence time between loris and cattle (*Bos taurus*) was calculated using the fossil split time of 40 Myr between tarsier and loris.

	Number of identical aa		Div. time	Number of identical aa			
	Loris-Tasier	Loris-NWM	Lo-NWM	N-OW	OW-Or	Or-Hu	Lo-Bos
Lo.-Ta. > Lo.-NWM, Slow evolving, 17 genes:							
PAX9	338	333/341	106.7	334	338	340	336
COX1	487	469/512	68.8	465	471	491	489
Cyt c	99	90/105	100.0	96	104	105/ni	94
ISP	190	181/202	70.0	181	192	197	187
HBA	132	130/141	48.9	136	137	140	121
COX5A	127	123/136	57.8	132	131	130	126
Epsilon-globin	62	61/67	48.0	65	66	65	61
COX3	234	197/261	94.8	210	222	239	224
LHB	108	85/119	123.6	89	100	107	
	121						110/136
IRBP	267	254/301	55.3	277	289	297	
	228						189/275
COX4I1	120	112/137	58.8	118	125	132	108
Tyr	122	118/140	48.9	131	136	140/ni	102
Growth hormone	150	116/174	96.7	154	169	174/ni	152
COX6c	64	60/75	54.5	59	66	74	59
COB	322	301/379	55.7	280	301	334	305
COX2	185	163/220	65.1	148	192	206	179
COX8A	58	50/68	72.0	52	55	67	63
Lo.-Ta. < Lo.-NWM, Slow evolving, 3 genes:							
HBB	130	133/146	32.5	137	139	144	117
AAR2B	321	324/368	37.4	ni	ni	ni	316
Cnr1	283	285/299	35.0	298	299/ni	ni	ni
			Divergence time average				
			66.7	47.8	29.7	17.3	63.6
			±25.4	±23.0	±12.6	±13.9	±35.8

Table 8. Divergence time between human and mouse. Slow evolving genes were randomly selected from the German pongo cDNA project as shown in Table 5 that show 99% identity between human and orangutan. Among these, some show lineage specific rate acceleration with a distance between mouse and rat that is 2 fold more than that between human and orangutan and were therefore excluded as non neutral clock genes. Divergence time between human and mouse was calculated for each gene as shown by using human mutation rate for both lineages (Hu/17.3), mouse mutation rate for both lineages (Mus/12.3), or using human mutation rate only for the lineage leading to human and mouse mutation rate only for the lineage leading mouse (Hu/Mus).

	Number of identical amino acids			Divergence time of human-mus (Myr)		
	Hu-Or	Hu-Mus	Mus-rat	Hu/17.3	Mus/12.3	Hu/Mus
WNT7A	346	344	348/349	28.8	61.5	39.4
Wnt1	367	366	367/370	23.1	16.4	19.1
CAH93506	738	725	738/739	242.2	172.2	200.0
CAH90891	531	521	532/535	60.6	57.4	58.8
CAH90590	216	211	215/217	103.8	36.9	54.5
CAH93476	416	402	417/420	77.9	73.8	75.6
CAH93429	206	204	206/207	51.9	36.9	43.1
CAH93390	470	465	470/471	149.6	73.8	86.2
CAH93367	471	450	468/473	199.0	56.6	88.1
CAH93330	325	317	324/327	86.5	41.0	55.6
CAH93284	544	543	545/546	13.0	36.9	30.5
CAH93155	674	667	673/679	41.5	26.7	30.9
CAH93143	322	320	323/325	28.8	30.8	29.8
CAH92769	336	332	337/338	51.9	73.8	60.6
CAH92767	1138	1136	1136/1140	34.6	12.3	18.2
CAH92738	365	358	364/366	138.4	49.2	72.6
CAH92650	224	196	224/226	259.5	184.5	215.2
CAH92595	1223	1223	1220/1230	17.2	8.6	11.5
CAH92324	906	893	904/911	61.9	31.6	41.9
CAH92088	813	800	815/819	54.5	58.4	56.4
CAH92076	513	505	512/515	86.0	41.0	55.6
CAH92050	336	321	335/338	146.2	69.7	94.4
CAH92747	342	338	342/343	86.5	61.5	71.8
Divergence time average (Myr):				89.0±69.9	57.1±43.0	65.7±50.2

Table 9. Opossum and human divergence time. Slow evolving genes with greater than 90% identity between kangaroo (*Macropus eugenii*) and opossum (*Monodelphis domestica*) and between human and mouse were randomly selected from the NCBI database. All informative genes available from the database were included in the Table. Genes showing lineage specific mutation rate acceleration were non informative and excluded. Divergence time between human and opossum was calculated for each gene as shown by using opossum mutation rate for both lineages (Opo/66.4), human mutation rate for both lineages (Hu/65.7), or using opossum mutation rate only for the lineage leading to opossum and human mutation rate only for the lineage leading human (Opo/Hu).

	Number of identical amino acids			Divergence time of Opo-Human (Myr)		
	Kan-Opo	Mus-Hu	Opo-Hu	Opo/66.4	Hu/65.7	Opo/Hu
Capza2	284	281	277/286	298.8	118.3	169.8
AAA62345	184	185	184/186	66.4	131.3	88.1
ACG50801	236	233	211/240	481.4	272.2	347.7
Mkrn1	419	406	376/428	383.6	167.2	221.3
G6PD	500	481	476/515	172.6	75.4	105.4
GAPDH	220	216	217/228	91.3	60.2	72.4
ACM88712/Rag1	174	176	171/181	94.9	131.4	109.9
PR	172	170	157/180	190.9	151.1	169.1
Pgk1	390	407	383/416	84.3	240.9	125.0
UBE1y1	141	149	144/152	48.3	175.2	75.5
Cav1	165	169	160/178	91.9	131.4	108.4
PRDX1	182	189	180/198	74.7	131.4	95.2
ABW82472	182	189	183/198	62.3	109.5	79.4
Cox1	493	466	459/512	185.2	75.7	107.5
CytoC	101	96	95/105	166.0	73.0	101.4
Divergence time average (Myr):				166.1±128.6	136.1±60.0	131.7±72.5

Figure legends:

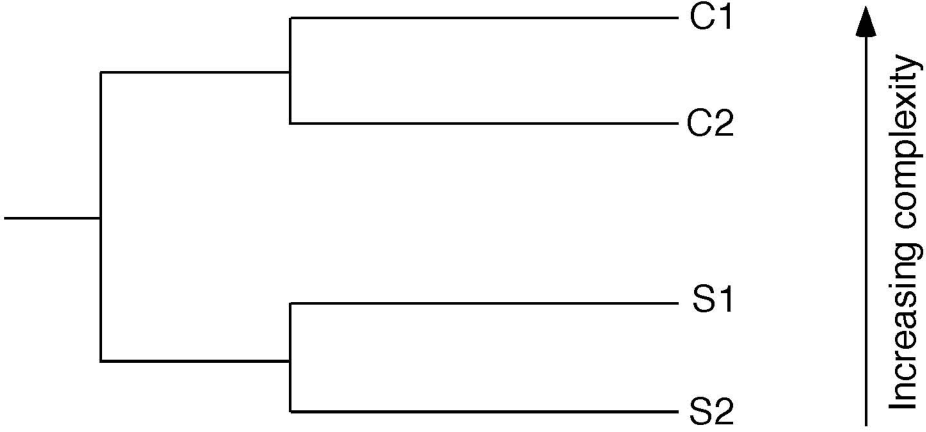
Figure 1. Genetic equidistance and non-equidistance. For any two clades of organisms, with one being more complex than the other, any individual species from one clade is equidistant in time to all species of the other clade. Within each clade, there are also variations in degree of epigenetic complexity among different species as indicated by the arrow. Both the molecular clock and the MGD hypothesis can make 4 predictions on genetic distance as shown. The two hypotheses differ only in predictions 2 and 4. Also see Figure 2 for details on predictions 3 and 4 on the difference between fast and slow evolving genes.

Figure 2. Inferring genealogy from sequence similarity in slow evolving genes. For any given three species A, B, and C, with A having low maximum genetic diversity (say, 5% protein dissimilarity for a given protein) and B higher (10%) and C still higher (20%), there are two possible phylogenetic models as shown. **A.** Slow evolving genes can distinguish the two models. The two models predict different results for slow evolving genes at a time (T1) when the genetic distances in these genes have not yet reached the maximum. However, the two models predict the same results when analysis is done at a time (T2) when the genetic distances have already reached the maximum. **B.** Fast evolving genes cannot distinguish the two models. The two models predict the same results for fast evolving genes.

Figure 3. The concept of common ancestor. B1 and B2 are extant individuals of taxon B and shared a common ancestor at time T1. Taxon A is the sister taxon of B and shared a common ancestor with a fraction of B (B1) at time T1. The difference in time between T1 and T2 can be from zero to any size. B-like lineage is represented by solid line while A-like lineage by dashed line. Between T1 and T2, the line leading to A is still part of the B-like lineage. The predictions by the molecular clock and the MGD hypothesis are shown. Only predictions by the

MGD hypothesis conform to factual observations. The only way for the fact to accommodate the molecular clock is to assume that the time difference between T1 and T2 is extremely small or zero.

Figure 4. A phylogeny of primates. The relationships of selected major primates are shown, based on results of this study. The shorter vertical distance between the MRCA and one of the sister taxa indicates that the ancestor lineage of that taxon is also the ancestor of the MRCA. For example, the ancestor lineage of gorillas is also the ancestor of the MRCA shared by all extant chimpanzees and a fraction of extant gorillas. Divergence times calculated by the slow clock method are indicated and those in bold represent fossil times used as calibration for the slow clock. Organisms are listed from top to bottom based on epigenetic complexity.



Predictions by the molecular clock hypothesis:

1. Genetic equidistance to a simpler outgroup:
   Distance C1-S2 = Distance C2-S2

2. Genetic equidistance to a complex outgroup:
   Distance S2-C1 = Distance S1-C1

3. Genetic non-equidistance to a simpler taxon in slow evolving genes given non-equidistance in time:
   Distance C1-S2 > Distance S1-S2

4. Genetic non-equidistance to a simpler taxon in fast evolving genes given non-equidistance in time:
   Distance C1-S2 > Distance S1-S2

Predictions by the MGD hypothesis:

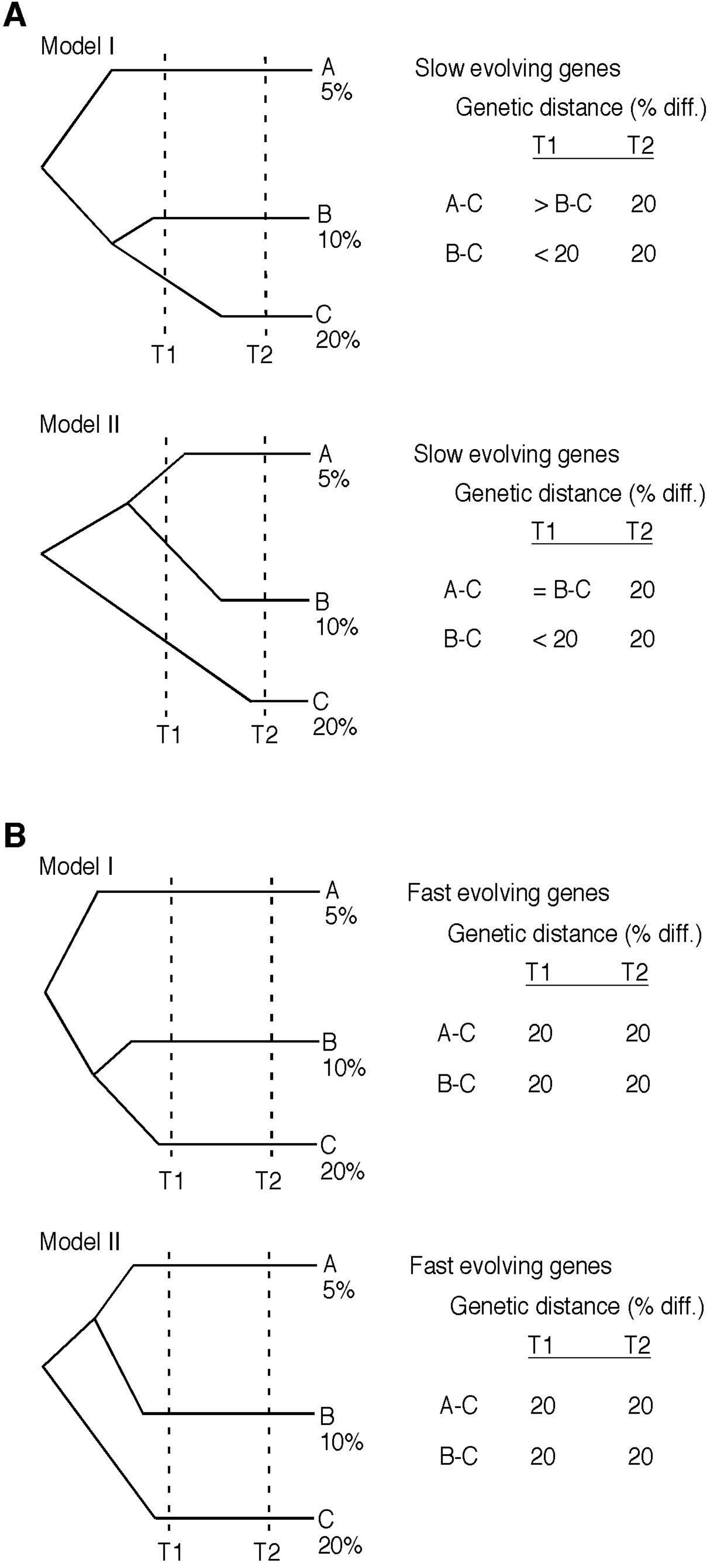
1. Same as above.

2. Genetic non-equidistance to a complex outgroup:
   Distance S2-C1 > Distance S1-C1

3. Same as above.

4. Genetic equidistance to a simpler taxon in fast evolving genes despite non-equidistance in time:
   Distance C1-S2 = Distance S1-S2

Figure 1



A Model I

Slow evolving genes

Genetic distance (% diff.)

	T1	T2
A-C	> B-C	20
B-C	< 20	20

Model II

Slow evolving genes

Genetic distance (% diff.)

	T1	T2
A-C	= B-C	20
B-C	< 20	20

B Model I

Fast evolving genes

Genetic distance (% diff.)

	T1	T2
A-C	20	20
B-C	20	20

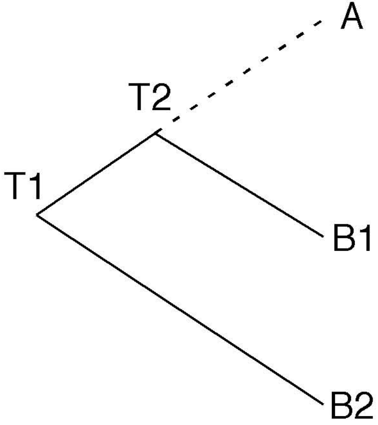
Model II

Fast evolving genes

Genetic distance (% diff.)

	T1	T2
A-C	20	20
B-C	20	20

Figure 2



Predictions by the molecular clock:

Distance B1-B2 > Distance A-B1
Distance A-B1 < Distance A-B2

Predictions by the MGD hypothesis
for slow evolving genes:

Same as above

Predictions by the MGD hypothesis
given long enough time or
for fast evolving genes:

Distance B1-B2 = Distance A-B1
Distance A-B1 = Distance A-B2

Figure 3

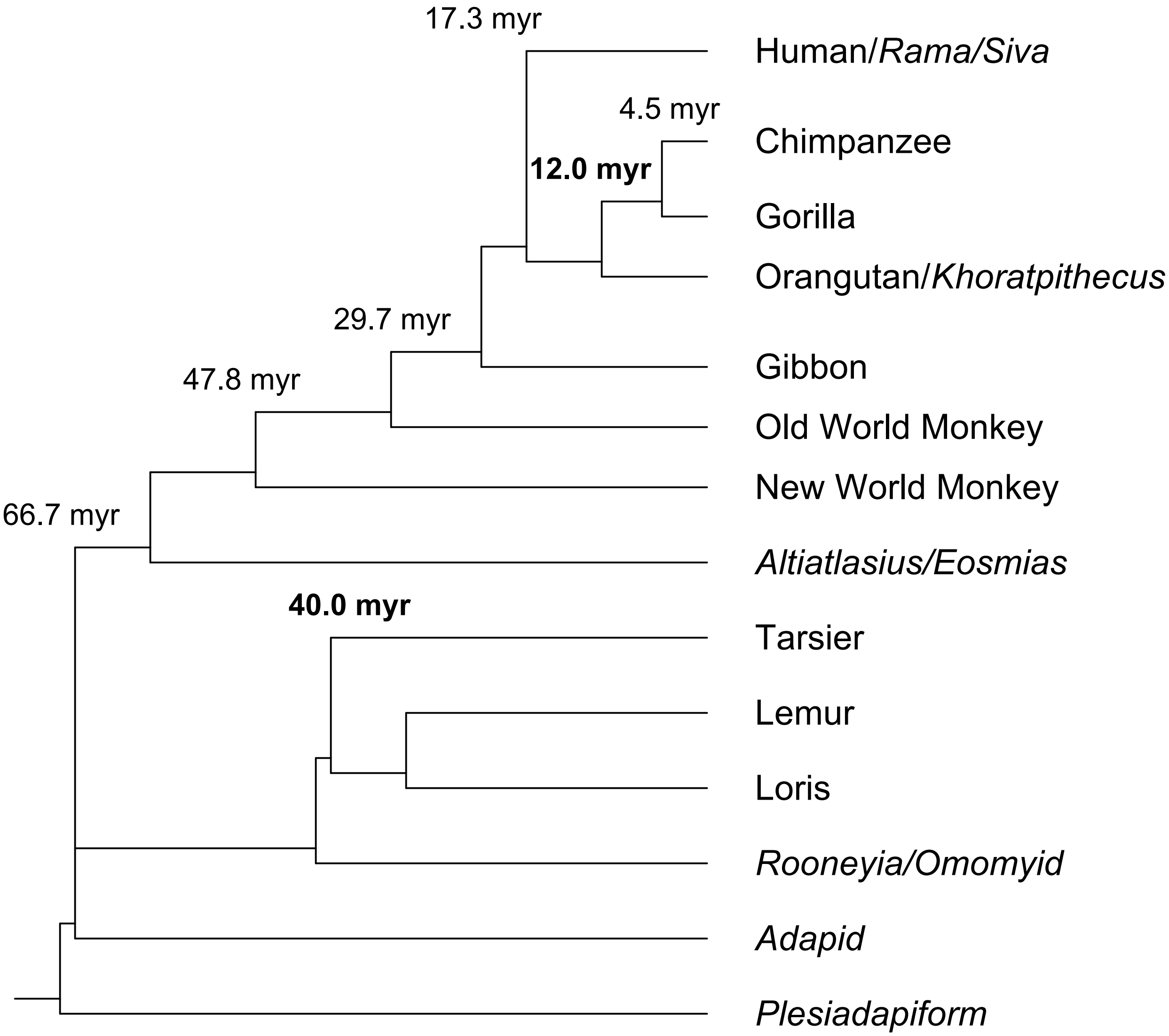

17.3 myr
Human/*Rama/Siva*
4.5 myr
Chimpanzee
12.0 myr
Gorilla
Orangutan/*Khoratpithecus*
29.7 myr
Gibbon
47.8 myr
Old World Monkey
New World Monkey
66.7 myr
Altiatlasius/Eosmias
40.0 myr
Tarsier
Lemur
Loris
Rooneyia/Omomyid
Adapid
Plesiadapiform

Figure 4

Supplementary Information to accompany

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

Shi Huang

State Key Laboratory of Medical Genetics
Xiangya Medical School
Central South University
110 Xiangya Road
Changsha, Hunan 410078, China

shuangtheman at yahoo.com

Table of contents

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

Table S1. The reptile clade (including birds): human is closer to birds than to snakes. The percent 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).

		Percent 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 S2. The amphibian group. The percent 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.

		Percent 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 S3. The teleost fish group: human is closer to the loach than to the three spined frogfish. The percent 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.

		Percent 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 S4. The echinoderm phylum. The percent 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.

		Percent 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 S5. The arthropod phylum: human is closer to the dragonfly than to the louse. The percent 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.

		Percent 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 S6. The nematode phylum. The percent 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.

	Percent 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 S7. The porifera phylum: human is closer to the chicken liver sponge than to *H. lachne.* The percent 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).

		Percent 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 S8. The fungi kingdom: human is closer to the corn smut than to yeast. The percent 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).

		Percent 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 S9. The protist alveolates superphylum. The percent 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.

		Percent 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	90.60
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 have. However, complex plants do not have more in common with complex vertebrates than simple plants have.

The bacteria kingdom

Using COX1 of humans as query, the bacterium *Magnetospirillm magnetotactilum* was identified among the closest to humans (59% identity), while *Gemmta 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. magnetotactilum* (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. *Pongo abelli* is closer to *Pan troglodytes* than to *Homo sapiens*

Table S10. *Pongo abelli* is closer to *Pan troglodytes* than to *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.

	Number of identical amino acids		% identity	Div. time (Myr)
	P.a-H.s	P.a.-P.t.	P.a.-P.t.	P.a.-H.s.
P.a.-P.t. > P.a.-H.s., Slow evolving, 53 genes:				
CAI29661	356	357/359	99	18.0
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	18.0
CAH91922	299	300/302	99	18.2
CAH91891	569	571/574	99	20.2
CAH91877	906	907/910	99	16.0
CAH91875	241	242/243	99	24.2
CAH91872	402	403/407	99	15.0
CAH91839	398	399/402	99	16.0
CAH91837	463	465/469	99	18.0
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	24.0
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	18.0
CAH90484	1383	1386/1393	99	17.3
CAH90800	879	881/888	99	15.5
CAH90788	342	345/347	99	30.0
CAH90758	407	409/412	99	20.1
CAH90750	263	264/266	99	18.0
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	16.0
CAH90500	907	908/910	99	18.0
CAH90495	1524	1526/1539	99	20.0
CAH90473	703	704/707	99	16.0
CAH90462	332	333/333	100	ni
CAH90449	434	436/440	99	18.1

P.a.-P.t. < P.a.-H.s., Slow evolving, 15 genes:

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	6.0
CAH90823	914	912/918	99	8.0
CAH90785	289	288/289	99	ni
CAH90589	1021	1020/1023	99	8.0
CAH90585	933	932/941	99	10.7
CAH90494	1060	1059/1066	99	10.3
			Average:	17.3+/-5.1

P.a.-P.t. > P.a.-H.s., Fast evolving, 66 genes:

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

3. Gorillas are closer to chimpanzees than to humans.

Table S11. Gorillas are closer to chimpanzees than to humans. Of the 69 informative gorilla proteins listed here, 35 have greater than 97% identity between gorillas and chimpanzees 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).

	No. of identical a.a.	% identity

	Go.-Hu.	Go.-Chimp.	Go.-Chimp.
Go.-Ch. > Go.-Hu., Slow evolving, 27 genes:			
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
Go.-Ch. < Go.-Hu., Slow evolving, 8 genes:			
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
Go.-Ch. > Go.-Hu., Fast evolving, 18 genes:			
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
AMAC1L2	314	321/338	94
APOBEC3G	363	367/384	95
COX2	218	222/227	97
Cob	350	353/379	92
Go.-Ch. < Go.-Hu., Fast evolving, 16 genes:			
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

4. Gibbons are the ougroup to a pongid-hominid clade.

Table S12. 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 and orangutans 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* >> 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* >> 0.05). The data show that gibbons are equidistant to orangutans and humans in both slow and fast evolving genes.

	No. of identical a.a.		% identity
	Gi.-Hu.	Gi.-Orang.	Gi.-Orang.
Gi.-Or. > Gi.-Hu., Slow evolving, 12 genes:			
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
Gi.-Or. < Gi.-Hu., Slow evolving, 7 genes:			
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
Gi.-Or. > Gi.-Hu., Fast evolving, 13 genes:			
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
Gi.-Or. < Gi.-Hu., Fast evolving, 21 genes:			
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

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

Table S13. 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 and gibbons 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* >> 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* >> 0.05). The data show that macaques are equidistant to gibbons and humans in both slow and fast evolving genes.

	No. of identical a.a.		% identity
	Ma.-Hu.	Ma.-Gi.	Ma.-Gi.
Ma.-Gi. > Ma.-Hu., Slow evolving, 7 genes:			
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
Ma.-Gi. < Ma.-Hu., Slow evolving, 11 genes:			
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
Ma.-Gi. > Ma.-Hu., Fast evolving, 8 genes:			
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
Ma.-Gi. < Ma.-Hu., Fast evolving, 8 genes:			
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

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

Table S14. 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 and macaque 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.

	No. of identical a.a.		% identity
	Sa.-Hu.	Sa.-Ma.	Sa.-Ma.
Sa.-Ma. > Sa.-Hu., Slow evolving, 8 genes:			
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
Sa.-Ma. < Sa.-Hu., Slow evolving, 9 genes:			

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
Sa.-Ma. > Sa.-Hu., Fast evolving, 9 genes:			
ND4	171	175/234	74
MC1R	268	281/317	88
Interferon a	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
Sa.-Ma. < Sa.-Hu., Fast evolving, 13 genes:			
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

7. Calculation of the divergence time between chimpanzees and gorillas

Table S15. 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.

	Number of identical a.a.		Chimp.-Go.
	Chimp.-Go.	Go.-Orang.	Div. time (Myr)
APOE	314	311/317	6.0
MBP1	234	229/235	2.0
KLK3	ni		
T2R38	330	325/333	4.5
ASIP	ni		
WNT7A	ni		
FSHB	ni		
GSC	ni		
Myostatin	ni		
GPR56	677	670/687	7.0
BRCA1	1125	1108/1141	5.8
RNAseA1	153	151/156	7.2
MAOA	ni		
HNMT	116	113/117	3.2
SCML2	ni		
CXCR4	ni		
UTY	223	217/226	4.0
MBL2	234	227/235	1.5
Oxytocin receptor	287	284/289	4.8
CXCR2	349	342/355	5.5
ASPM	3426	3403/3477	8.3
CCR5	ni		
FUT2	341	331/343	2.0
Prion	252	249/253	3.0
TPMT	244	236/245	1.3
Globin a2	ni		
COX1	504	497/512	6.4
Average of 16 informative genes:			4.5 ± 2.2