MBE Advance Access published June 29, 2010

Research article

Mammalian evolution may not be strictly bifurcating

Björn M Hallström[*] and Axel Janke[†]

[*]Department of Cell and Organism Biology, Division of Evolutionary Molecular Systematics, University of Lund, Sölvegatan 35, S-223 62 Lund, Sweden

[†]LOEWE - Biodiversität und Klima, Forschungszentrum BiK-F, Senckenberganlage 25, D-60325 Frankfurt am Main Germany

Correspondence to: Axel Janke; email: ajanke@senckenberg.de
Phone: +49 – (0)69 7542 1842
Fax: +49 – (0)69 7542-7904

Running head: Mammalian phylogenomics

Key words: continental drift, Cretaceous warming, genome analysis, hybridization, phylogenomics, split decomposition

Abstract

The massive amount of genomic sequence data that is now available for analyzing evolutionary relationships among 31 placental mammals reduces the stochastic error in phylogenetic analyses to virtually zero. One would expect that this would make it possible to finally resolve controversial branches in the placental mammalian tree. We analyzed a 2,863,797 nucleotide-long alignment (3,364 genes) from 31 placental mammals for reconstructing their evolution**.** Most placental mammalian relationships were resolved, and a consensus of their evolution is emerging. However, certain branches remain difficult or virtually impossible to resolve. These branches are characterized by short divergence times in the order of 1-4 million years. Computer simulations based on parameters from the real data show that as little as about 12,500 amino acid sites could be sufficient to confidently resolve short branches as old as about 90 million years ago. Thus, the amount of sequence data should no longer be a limiting factor in resolving the relationships among placental mammals. The timing of the early radiation of placental mammals coincides with a period of climate warming some 100 - 80 million years ago and with continental fragmentation. These global processes may have triggered the rapid diversification of placental mammals. However, the rapid radiations of certain mammalian groups complicate phylogenetic analyses, possibly due to incomplete lineage sorting and introgression. These speciation-related processes led to a mosaic genome and conflicting phylogenetic signals. Split network methods are ideal for visualizing these problematic branches and can therefore depict data conflict and possibly the true evolutionary history better than strictly bifurcating trees. Given the timing of tectonics, of placental mammalian divergences, and the fossil record, a Laurasian rather than Gondwanan origin of placental mammals seems the most parsimonious explanation.

Introduction

As genomic sequences are the ultimate source of molecular data for evolutionary studies, the wealth of information available from the whole genome sequencing of metazoans has revolutionized evolutionary studies. After publication of the human genome (Lander et al. 2001; Venter et al. 2001), it was decided that low-coverage (2x) genome sequences from additional mammalian species would be beneficial for more accurate sequence annotation and for studying the evolution of disease genes. The data from these genome projects also enable us to study the evolution of mammalian lineages with a previously unimaginable amount of data, opening the filed of phylogenomics. The first phylogenomic study of placental mammals initially involved some 200,000 nucleotides (nt) of protein coding sequences (Nikolaev et al. 2007). Improving the sequence and taxon coverage, phylogenomic analyses of mammals included 2.2 million nucleotides (Mnt) from about 2,840 protein-coding genes (Hallström et al. 2007) or some 1,700 conserved genome loci (Wildman et al. 2007). The largest and most complete dataset so far analyzed more than 2.8 Mnt from 3,012 protein-coding genes (Hallström and Janke 2008). However, phylogenomics has failed to reliably resolve certain branches of the placental mammal tree, thereby posing new problems and questions in animal evolution (Hallström and Janke 2008).

Despite the amount of available data, some branches in the placental mammal tree are only weakly supported and phylogenomic analyses leave some branches as yet completely unresolved, due to the variable and poor support for different evolutionary scenarios (Hallström and Janke 2008). This was unexpected, because, theoretically, the sheer amount of genomic data should have easily overcome stochastic errors from single and multiple (≈20) gene analyses, a problem that vexed molecular phylogenetic studies before the wealth of information from the genomic age (Kullberg et al. 2008).

To date, three important but difficult to resolve branching points in the mammalian tree have been indentified. The first is the primary branching among placental mammals, that between the super-orders Xenarthra (in this study: sloth and armadillo), Afrotheria (elephant, tenrec and hyrax), and Boreoplacentalia (all remaining species used in this study). Previous resolutions of their divergences depended on the choice of analytical methodology and type of data, but did not enable the rejection of alternative hypotheses (Nishihara et al. 2007; Hallström and Janke 2008). Even the

analyses of retroposon insertions, otherwise generally regarded as a solid phylogenetic marker system, failed to resolve a clear bifurcation in the most basal divergences among placental mammals (Churakov et al. 2009; Nishihara et al. 2009); thus, supporting the original sequence-based findings (Hallström and Janke 2008). Retroposon insertions are, with a few exceptions (Cantrell et al. 2001; van de Lagemaat et al. 2005), free of homoplasies (Steel and Penny 2000). Therefore, the apparently contradicting results from sequence data and retroposon insertion analyses require a natural explanation. Similar observations have also been made for the other two poorly or unresolved mammalian phylogenetic branches. The detection of apparently conflicting phylogenetic signals for the position of the Scandentia (tree shrews) relative to Primates and Glires (rodents and lagomorphs) within the Euarchontoglires clade, and the position of Chiroptera (bats) relative to Artiodactyla (even-toed ungulates), Carnivora, Lipothypla (hedghog, common shrew and allies), and Perissodactyla (odd-toed ungulates) within the Laurasiaplacentalia, leave both controversial and poorly supported (Nishihara et al. 2006; Janecka et al. 2007; Kriegs et al. 2007; Hallström and Janke 2008)

A common feature of these three problematic divergences is the divergence of groups within 1-4 million years (Myr) of one another (Hallström and Janke 2008). Such time intervals may have been too short for the respective genomes to gather enough substitutions to resolve rapid divergences some 100 Myr later, or limited taxon sampling may impede such phylogenetic analyses. Another possible reason for the problems involved in clearly resolving short central branches of the placental mammalian tree by sequence and retroposon data may be speciation-associated processes, such as species hybridization and incomplete lineage sorting (Nei 1987). Species hybridization leads to introgression, the incorporation of genes from one species into the gene pool of another species, while incomplete lineage sorting produces a pattern of allele fixations from ancestral polymorphisms that does not reflect the species history. Both processes generate mosaic genomes, in which different loci support alternative mammalian relationships (Hallström and Janke 2008; Churakov et al. 2009). This type of reticulated evolution and incomplete lineage sorting can produce conflicting evolutionary signals. Split network methods can better illustrate and explore such conflicts in the data than can traditional analyses that seek a strictly bifurcating tree (Huson and Bryant 2006), and can be used to detect important processes in the evolution of the placental mammal genome.

While the exact reasons for the problematic resolution of certain branches in the mammalian tree are currently unknown, they seem to be connected to the rapid radiation of early placental mammals in the mid-Cretaceous (Hallström and Janke 2008). Such a process could have been triggered by a swift mid-Cretaceous temperature increase that reached a maximum at the Cenomanian –Turonian boundary at 93.5 Ma (Retallack 2002) and fragmentation of the super-continents Laurasia and Gondwana at about the time of the earliest divergences of placental mammals (Smith et al. 2004; Reeves 2009; www.reeves.nl).

We have built a new alignment, including recently released genome data, to further investigate problematic branches in the mammalian tree and to study the tree-likeness of the mammalian genome data. The number of placental mammals included in this study has been increased to 31, twelve (63%) more than in a previous phylogenomic analysis (Hallström and Janke 2008), and the genome data from six outgroup species were included to ensure a solid root of the tree from a maximal taxon sampling. The previously unresolved branchings inside the Laurasiaplacentalia, which include Perissodactyla, Carnivora, Artiodactyla, Cetacea, Chiroptera, and Lipotyphla (Arnason et al. 2008), may be clarified by the new genome data of a perissodactyl (horse), a second chiropteran (mega bat), a cetacean (dolphin), and an artiodactylan (alpaca). The genome data from a hyracoid (hyrax) and a second xenarthran (sloth) are invaluable additions to the hitherto long and undivided branches of Afrotheria and Xenarthra, allowing the basal divergences among placental mammals to be examined in more detail. Computer simulations were made to estimate the amount of data needed to resolve short branches, and finally, split decomposition methods (Huson and Bryant 2006) were used for visualizing conflict in the tree.

Materials and Methods

Data

Predicted cDNA sequences from all tetrapods with assemblies and gene builds in release 54 of ENSEMBL were downloaded from ftp://ftp.ensembl.org/pub/current_fasta/. In total, 37 species (table 1) were included in the data build. The taxon sampling represents 16 of the 21 extant eutherian orders. The sequence data from metatherian (marsupials), prototherian (monotremes), avian

(bird), reptile, and amphibian species were collected for rooting the placental mammal tree. For increasing the usable sequence length several outgroups were included in the analysis.

Assembly

Orthologous cDNA sequences were detected using the recursive BLAST method (Hallström and Janke 2008) with a cutoff e-value of 10^{-12}. This approach efficiently precludes paralogous sequences from gene families being included in the analysis without assuming a phylogenetic tree. In cases where several transcript variants were present for a gene, the longest one was used. The sequences were translated to amino acids (aa), and multiple sequence alignments from the identified orthologs were constructed with the program MUSCLE (Edgar 2004) for all ortholog groups containing at least 20 species. The alignments were trimmed by removing all columns containing gaps, and alignments with an observed (p) aa distance larger than 40% were excluded from the analysis. Finally, nt alignments were constructed with the assistance of the aa alignments. A customized Perl script produced and drew a presence/absence matrix for illustrating the data density. Each row in the density map refers to a species and each column to a gene. Any gene that is present in a species is marked with a black dash. The matrix is sorted in both directions, placing the species with the best coverage at the top of the graph and the genes represented by most species to the left.

Phylogenetic analysis

ML analyses were performed as described in detail in Hallström and Janke (2008) and are therefore only briefly outlined here: Treefinder version October 2008 (Jobb et al. 2004) was used with the GTR model (Lanave et al. 1984) assuming rate heterogeneity with eight classes of gamma distributed rate categories (Yang 1994) and one class of invariable sites (8Γ + I) for nt sequences. For aa sequence analyses, the WAG2000 model, WAG, (Whelan and Goldman 2001) with a rate heterogeneity and invariable sites model (8Γ + I) was used for the ML analyses. Uncertain or controversial relationships were further analyzed by an extended ML analysis. Alternative topologies were statistically evaluated in TF using Shimodaira-Hasegawa probabilities, pSH (Shimodaira and Hasegawa 1999). Neighbor joining analyses were done in the SplitsTree4 program (Huson and Bryant 2006) on the WAG2000 + Γ + I

model using parameters estimated by TF. The complete alignment was analyzed according to these parameters. In addition, single gene alignments, selected for the 10% of the longest sequences, were evaluated for their phylogenetic content using for simplicity and speed of computation the GTR 8Γ + I model.

Dating

Divergence times were estimated from both aa and nt sequence data using the non-parametric rate smoothing method on a logarithmic scale (NPRS-LOG) implemented in TF (Jobb et al. 2004). The ten fossil-based age constraints that were used to calibrate the tree were taken from Benton et al. (2009) and are detailed in table 2. Mean values and their standard deviations were calculated from the branch lengths of 100 bootstrapped ML analyses of aa and nt sequences.

Network

The sequence data were analyzed by the SplitsTree4 program using the neighbor-net method from aa ML distances under the WAG2000 model of sequence evolution accounting for rate heterogeneity (gamma) and invariable sites as in the ML analyses. The retroposon data from Churakov et al. (2009) were recoded and simplified for the presence (1) and absence (0) of retroposons in human, elephant, armadillo, and opossum. The data matrix was then analyzed by SplitsTree4 and presented as a split network simply by plotting each retroposon insertion event onto the corresponding internal branch. A chi-2 test was performed in Excel for evaluating the uniformity of distribution of retroposon insertions between the three possible topologies..

Simulation

The amount of sequence data needed to resolve a branch that would lead to a hypothetical bifurcating Exafroplacentalia clade was estimated by computer simulation using the Seq-gen program (Rambaut and Grassley 1997). The simulations of aa sequences were performed on the tree topology and branch lengths obtained from the actual data utilizing the observed average aa composition and rate heterogeneity. The time interval from the divergence of Afrotheria and Exafroplacentalia until Exafroplacentalia itself split into Boreoplacentalia and Xenarthra was varied by changing the branch length of this internal node to 1/2, 1/3, 1/4, and 1/5th of the observed length. The amount of sequence data required for

providing pSH values below 0.05 for both of the two alternative hypotheses was recorded and plotted on a line-graph. In parallel, the effect of missing data on the tree reconstruction was investigated by removing characters from the simulated data sets. The removal was done according to the proportions of missing data in each individual species of the original data. The impact on the amount of data required for statistical significance was recorded.

Results

Dataset and quality

The total length of the filtered alignment data was 2,863,797 nt, derived from 3,364 genes of 31 placental mammal species plus 6 outgroups. The average lengths of the individual sequences were 851 ± 667 nt and the average observed distance between human and platypus was 15.7% ± 9.9%.

Figure 1 illustrates the data density. Each row represents a species and a black dash denotes the presence of a gene in the alignment. The average sequence coverage for mammalian species was 79%. The poorest sequence coverage among the placentals was observed for the common shrew, which was represented in 62% of the alignment. The non-mammalian outgroups had a sequence coverage of 49-82%, which, as expected, was generally lower than the mammalian average. However, the marsupial species, opossum, had an above average coverage of 82%, because, as one of the better-quality genomes, it has been sequenced with over 7-fold coverage. Also, compared to non-therian outgroups, the opossum's close relation to placental mammals facilitated ortholog detection.

The nt and aa frequencies appeared to be very similar among the species, but due to the large size of the dataset, a possibly overly strict chi-square test rejected compositional homogeneity for many species, even for R/Y-coded nt data. The data properties, such as distance distribution, character composition, evolutionary rates, and type of genes, resemble those of previous phylogenomic studies and are not described in detail here (Hallström et al. 2007; Hallström and Janke 2008).

Tree

The ML tree based on aa sequence and a WAG+8GI model is shown in figure 2. The tree generally conforms to that of previous phylogenomic studies (Hallström et al.

2007; Wildman et al. 2007; Hallström and Janke 2008). Most branches are supported by unit ML bootstrap and TF support values. Despite the increased taxon sampling, the same problematic branches found in previous studies were identified here. Thus, the earliest divergences of placental mammals, the Xenarthra, Afrotheria, and Boreoplacentalia, are only marginally supported. In this analysis of nt and aa sequences, Afrotheria represents the earliest split from the eutherians, but alternative topologies cannot be rejected by pSH tests based on different data types (supplemental table 1). The Epitheria hypothesis, Afrotheria and Boreoplacentalia as sistergroups, receives the least support. The grouping of the tree shrew (Scandentia) differs from previous phylogenomic studies. In this study, Scandentia group with Glires (Rodentia plus Lagomorpha), but a grouping with primates or a position outside a primate/Glires clade cannot be rejected by a pSH analysis (supplemental table 2). Finally, the phylogenetic position of the Chiroptera (bats) among the Laurasiaplacentalia receives only limited branch support from TF, 91% for NT12, 95% for aa, and in pSH analyses 3 alternative positions receive probabilities >0.05 (supplemental table 3) and cannot be formally rejected by all data types. However, the Pegasoferae hypothesis, Chiroptera as sistergroup to Perissodactyla plus Carnivora, receives only low support and can be rejected on the basis of aa and NT12 ML analyses.

The pSH analysis remained ambiguous about the best supported tree when only one outgroup, the opossum, was used for these analysis (not shown) and the amount of usable data was reduced by 12%. Therefore all analyses were done using multiple outgroups.

Dating

The estimations of divergence times shown in the chronogram of figure 3 and detailed in table 3 are based on the topology depicted in figure 2. The dating was performed solely by the NPRS-LOG method, because previous studies showed virtually no differences relative to other algorithms (Roos et al. 2007, Nilsson et al. 2010). The numerous calibrating points are marked with circles. A circle is filled when the divergence time estimate reached either the upper or lower bound and open when it stayed anywhere between the boundaries. The most influential calibration point in this study was that between dolphin and cow (Cetacea and Artiodactyla). The lower bound (minimum) is given with 52.4 Ma as the latest possible divergence time between

cetaceans from the remaining artiodactyls (Benton et al 2009). At this time, however, cetacean characters have already evolved and slightly older divergence times have been suggested by others (Bajpai and Gingerich 1998, Arnason et al. 2000, van Tuinen and Hadly 2004). Each million year this lower bound is moved back in time, almost equally affects all earlier divergences. None of the other calibration points exhibits an equally strong effect on the divergence time estimate, and most are estimated within their boundaries.

In the current phylogenomic ML tree the most basal divergence of placental mammals, between the Afrotheria and the remaining placental mammals, occurs at 90.6 Ma. Only 2.7 Myr later, at 87.9 Ma, the xenarthrans diverge from the Boreoplacentalia. Most other ordinal divergences occur between 80 and 65 Ma. The weakly supported position of the tree shrew and chiropterans are correlated with short divergence intervals of 2.1 and 2.0 Myr, respectively (table 3). Internal branches that have durations > 4 Myr or are short but very recent, are significantly supported despite compositional biases or other possible systematic errors in the data. Thus, the divergences among human, chimpanzee, and gorilla occur within 2 Myr, but are significantly resolved. The slightly older age of the divergence of Afrotheria relative to our other phylogenomic studies is a consequence of the tree topology. The divergence times are similar or in some cases younger, except for the basal divergences, than those estimated before (Hallström and Janke 2008). This is probably due to the increased taxon sampling and larger number of calibration points. When the tree is constrained to the Xenafrotheria hypothesis (Afrotheria plus Xenarthra), a probable alternative (Hallström and Janke 2008), the early divergence times become younger, but other parts of the tree are left unaffected.

Network analyses

The Neighbor-Net based on aa sequence data is shown in figure 4 and a magnification of its central region with labels for the major splits in figure 5. The Neighbor-Net includes the tree in figure 1, but appears to favor the Xenafrotheria (Afrotheria plus Xenarthra) hypothesis. A neighbor-joining analysis conforms to the Xenafrotheria hypothesis and neighbor-joining bootstrap supports this grouping with 90% and most others with unit support (supplemental figure 1).

The data conflict is best exemplified for the splits of the Xenarthra, Afrotheria and Boreoplacentalia, certain splits within the Eurchontoglires, and particular

relationships among the Laurasiaplacentalia. These branches are generally also those that were identified to be problematic by individual ML analyses. Thus, the placement of the Chiroptera and Lipotyphla relative to Carnivora, Perissodactyla, and Cetartiodactyla, as well as Scandentia relative to primates and Glires are uncertain by the Neighbor-Net analysis.

Differing from the ML tree, the Neighbor-Net shows a tendency to group the Chiroptera with a Carnivora, Perissodactyla, and Cetartiodactyla (Cetacea plus Artiodactyla) clade. Deep divergences of well-defined and supported groups, like Laurasiaplacentalia, Afrotheria, or Rodentia are separated by stretched boxes, illustrating a strong signal and limited conflict in the data, in agreement with the ML analysis. However, even among clearly resolved species Neighbor-Net has the power to indicate possible conflict in the data, as exemplified in the cases of some primate, the rodent, or carnivore divergences. The reason for this conflict remains unknown, however.

The 10-percentile of the longest genes (336 sequences) exceeded lengths of 1850 nt. Even though phylogenetic analyses of some genes resolve particular branches as in the ML tree shown in figure 2, a majority-rule consensus tree of ML trees from individual genes resulted in a star-like tree at the ordinal level and above, resolving only relationships among the most closely related taxa. A consensus network using a 2% or 4% threshold value did not show more or other structures than the Neighbor-Net (supplemental figure 2a, 2b). Finally, a network analysis has been done for about 500 selected genes that are capable of rejecting two of the three possible hypotheses of a certain node, at pSH <0.10. This analysis resulted in cube-like (i.e. unresolved) structures for the three critical nodes discussed above. The same analysis has been done for some selected and well-resolved nodes. Even these nodes showed a high degree of conflict, with about 20% of the sequences supporting either of the two alternative topologies (not shown). Thus, the analyses of single genes did not allow drawing further conclusions about the unresolved nature of certain branches that were encountered in the ML analysis of all sequence data.

The split network from the retroposon insertion data (Churakov et al. 2009) illustrates the unresolved nature of the earliest placentalian divergences in an ideal way (figure 6). There are nine retroposon insertions for Epitheria (all placentals except Xenarthra) and eight events that support the Xenafrotheria hypotheses. The Exafroplacentalia (Boreoplacentalia plus Xenarthra) hypothesis, which is supported

by five retroposon insertions, cannot be excluded. Even though one may recognize a slight favor of the Epitheria and less support for the Exafroplacentalia hypothesis, a chi-2 analysis yields a probability of p=0.55, thus a uniform distribution of retroposon insertion cannot be rejected.

Computer simulation

Computer simulations were performed to determine the theoretical minimum sequence length needed to significantly resolve the deepest divergence among placental mammals, based on pSH values. In this study the aa ML analyses reconstruct Afrotheria as diverging first from the remaining placentals some 90 Ma. The following split between Xenarthra and Boreoplacentalia occurs only some 2.7 Myr later. Given this topology, the parameters used, and dating of the short branch, it is surprising that as little as about 7,000 aa sites of evenly evolving sequence data are needed to significantly resolve this early divergence some 90 Ma. Even taking the, albeit limited, patchiness of the data into account by including the observed percentage of unknown data in the simulations, the amount of aa sequence data needed to resolve a very short early branch doubles. Thus, when including missing sites as a modeling parameter, the number of aa sites needed to significantly resolve such a branch increases to ca 12,500. Under the same conditions, a time interval as short as 540 thousand years (ka) at 90 Ma could be significantly resolved with as little as 60,000 aa sites (supplemental table 4). Figure 7 illustrates the results. It shows also that, as expected, the amount of data needed for resolving increasingly shorter branches increases approximately exponentially.

Discussion

New genome data from mammalian species are becoming available at an accelerating rate. Within a year of a recent phylogenomic study (Hallström and Janke 2008) eleven new genomes were released in databases. These data enable more detailed analysis of the evolution of placental mammals. The new genome data alleviate concerns about reconstruction artifacts caused by limited taxon sampling and sequence length. The present phylogenomic analyses place the newly sequenced species into their expected positions on the mammalian tree and a consensus of their evolution emerges. Before

the genomic era, comprehensive datasets and a dense taxon sampling for phylogenetic studies were available from mitochondrial protein coding data, mitogenomics (Arnason et al. 2008), or from a few selected nuclear genes (Murphy 2001). Mitogenomics clarified major parts of the mammalian tree, which are now being confirmed by phylogenomic analyses. For example, the unexpected sistergroup relationship between Carnivora and Perissodactyla was first identified by a mitogenomic study (Xu et al. 1996) and was reconstructed in later nuclear gene and phylogenomic analyses. Similarly, the grouping of the order Cetacea within Artiodactyla into the clade Cetartiodactyla was suggested by mitochondrial cyt*b* data (Irwin and Arnason 1994) and later strongly supported by nuclear gene (Gatesy et al. 1996), retroposon insertion (Nikaido et al. 1999), and mitogenomic (Arnason et al. 2000) analyses. Finally, the strong support for monophyly of the super-order Afrotheria was shown by nuclear gene analyses (Stanhope et al. 1998) and mitogenomics (Mouchaty et al. 2000) and is reconstructed from phylogenomic analyses.

However, in the current phylogenomic analyses certain branches of the placental mammalian tree still received only limited support or remained – in effect – unresolved, despite the high data density, the authoritative amount of genome data, and the increased taxon sampling. Problematic branches involve the relationships among Xenarthra, Boreoplacentalia, and Afrotheria, the most ancient divergence of placental mammals. In addition, the position of the Chiroptera among Laurasiaplacentalia and the placement of the Scandentia (tree shrew) with the Glires or Euarchonta remain insufficiently supported or even unresolved. These branches were previously identified as problematic in early phylogenomic analyses (Nikolaev et al. 2007, Hallström and Janke 2008, Nishihara et al. 2008), and the resolution of their evolution remains vague even when new analytical approaches such as Outgroup Scoring are used (Schneider and Cannarozzi 2009).

Outgroup Scoring offers an alternative approach to traditional phylogenetic reconstruction and may overcome some of its limitations. The method analyzes the topology and the support of a tree by extracting a signal utilizing outgroups at different distances. In this way it overcomes, or at least reduces, the effect of model violations and long-branch attraction. Compared to standard ML analysis, simulations using Outgroup Scoring have demonstrated equal or better performances in resolving problematic relationships (Schneider and Cannarozzi 2009). This method, however,

identifies the same problematic branches as the current analyses and provides no certain resolution. While Outgroup Scoring favors the Xenafrotheria hypothesis (Xenarthra plus Afrotheria) as do some previous phylogenomic studies, a support value of 0.91 indicates that alternative topologies are possible. Inspections of alternative topologies by Outgroup Scoring unfortunately offer no more certainty than traditional ML analyses and provide no explanation for why some branches of the placental mammalian tree are refractory to resolution.

One might suspect that the lack of resolution in the placental mammalian tree is coupled to the amount of data available for each species. The common shrew (Lipothyphla) has the lowest sequence coverage (62%) in the alignment, yet it is still represented by 1.8 Mnt of sequence data and receives significant phylogenetic support. In contrast, the tree shrew (Scandentia) receives limited phylogenetic support despite having a similar or higher sequence coverage of 70% (2.0 Mnt). The phylogenetic position of the tree shrew is basically unresolved in ML and neighbor network analyses. As in the cases of other poorly resolved branches, the tree shrew joins other placental orders within Euarchontoglires with very short branches. This would explain the contradictory phylogenetic placements of this order and the limited number of synapomorphies from rare genomic changes that have been recovered in previous studies of scandentian evolution (Janecka et al. 2007; Kriegs et al. 2007). There simply may not have been enough time for genomes in temporally narrow divergences to accumulate sufficient numbers of informative sites. For this reason and the stochastic nature of short sequences, single gene analyses did not allow further conclusions about the nature of the conflicting signals.

Simulations

Computer simulations were made to investigate whether sequence length is a limiting factor in resolving short branches. For obvious reasons, such simulations are idealizations (simplifications) of natural processes and are not able to include all factors that shape a sequence during evolution. One potential pitfall of any simulation is that the same model is used for both simulating and analyzing the data. In our study this probably leads to an underestimation of the required sequence length. However, the strength of simulations is their ability to isolate certain parameters intentionally, by computational or practical limitations. The aim was to identify the theoretical minimum amount of sequence data that enables significant reconstruction of the

placentalian tree based on the general properties of the sequence data and their evolutionary history, that is, the tree. The present simulation includes a number of known sources of systematic errors, such as varying evolutionary rates between branches (varying branch lengths), varying rates along sequences (rate heterogeneity), multiple substitutions (absolute branch lengths), individual aa replacement probabilities (the aa replacement model), and isolated long branches from using the observed ML tree topology. In a separate simulation the fraction of missing data was added as a parameter in the simulation. Thus, the simulations are relatively realistic, even if all possible parameters were not included in the study. A model for simulating compositional bias has not yet been developed and was therefore not part of the simulation.

The analyses of the computer simulated sequence evolution clearly show that under the modeled conditions and in the absence of introgression or lineage sorting a few ten thousand aa of random protein coding data are capable of resolving even very short branches from lineages that existed for less than 1 Myr at 100 Ma. Yet, the problems encountered in resolving such branches, even with millions of nts, suggest that natural processes involved in speciation or significant systematic errors, other than those included in the current study, may mask the phylogenetic signal.

The simulation data show a strong inverse correlation between the data density and the amount of data needed for the resolution of short branches millions of years ago. When a parameter of missing data, represented by the same fraction missing in the original alignment, is included, the amount of data needed for significant resolution increases by more than 60%. A recent phylogenomic analysis of fundamental metazoan divergences that are at least six times as old as the mammalian radiation, utilized expressed sequence tags (EST) from several species as data sources. Consequently, the resulting data density in this study was only about 50% from 150 genes, due to the lack of overlap between EST libraries (Dunn et al. 2008). The degree to which the reliability of the tree resolution is affected by increasing amounts of missing data and taxon sampling remains to be studied.

Networks

The limited resolution of certain branches may indicate that the evolution of placental mammals did not proceed in a strictly bifurcating way. Under these circumstances current phylogenetic methods seeking a fully resolved, two-dimensional tree are not

suitable for reconstructing the history of placental mammals. Possibly, the phylogenetic signals of contradicting branches annihilate each other, which may lead to an apparently unresolved trifurcation. Collapsing problematic branches into trifurcations, however, hides invaluable evolutionary information. A case in point is best demonstrated from data of recent phylogenetic analyses of retroposon insertions on the early radiation of placental mammals (Churakov et al. 2009, Nishihara et al. 2009). The data of Churakov et al. (2009) are individually confirmed in other species by experimental sequence analysis, while the Nishihara et al. (2009) data are based on database entries only.

A split network representation of the Churakov et al. (2009) data, illustrates the complex radiation of the three early placental mammalian lineages. A traditional presentation of these data as a tree would either require a multifurcation, the presentation of three separate trees, or an extended tree-like diagram (see figure 4A in Churakov et al. 2009), all of which blur the actual evolutionary message of a complex evolutionary process that led to the three clades of placental mammals. Other problematic divergences have not yet been analyzed by retroposon insertion data in the same detail, but sequence-based analyses are suggesting several such non-bifurcating radiations in the mammalian tree. The study of retroposon data for the evolution of the Chiroptera, Perissodactyla, and Carnivora (Nishihara et al 2006), where only a single apparently contradicting signal has been described, may be such an example. The difficulty to resolve the placement of the Chiroptera via sequence analysis and conflict from the retroposon data suggest that a more detailed retroposon insertion analysis may also lead to a network-like picture of the radiation of these groups.

More complex split networks, like the phylogenomic Neighbor-Net of figures 4 and 5, depict the intricate evolutionary signals of the sequence data in a still decipherable way, whereas traditional, bifurcating tree presentations hide alternative hypotheses or conflict in the data. Split networks have not been routinely used to study deep divergences among placental mammals, because reticulate evolution events were not suspected to occur or be detectable over long time scales. Therefore significantly conflicting data have rarely been observed in single gene analyses. Such conflict has only been sporadically observed in the phylogenetic analysis of single genes that support different phylogenies (Satta et al. 2000) or by testing individual hypotheses using pSH or similar statistical tests. In this way, many controversial

branches are going unnoticed. This is especially the case when branches receive high support values, leading one not to suspect alternative hypotheses. Branches can receive high bootstrap and especially high Bayesian probabilities even in the presence of strong conflict in the data, which would leave them, in effect, unresolved by ML test statistics (Nishihara et al 2007, Hallström and Janke 2008). The individual identification of such problematic branches occurs often only by chance or when they contradict preconceived hypotheses.

In contrast, network methods objectively present most inconsistencies in a tree in a single picture. This information can then be used to further investigate the nature and extent of conflict between alternative hypotheses. Unfortunately, the extreme age of basal mammalian divergences so far preclude the analysis of single genes for signs of conflict from reticulate evolution, due to the limited amount of information each gene provides and its stochastic nature. Yet, the Neighbor-Net of the whole dataset identifies all nodes where disagreement in the data is observed by other methods. From the network graphs in figures 4 and 5 it becomes immediately visible that the evolution and radiation of placental mammals has a more complex history than previously assumed.

Dating

The analysis of divergence time estimates may provide an explanation for the poor resolution of some branches. These branches are all characterized by short internal nodes that span intervals of 1-4 Myr. The computer simulations demonstrated that the limited resolution is not due to the lack of data. As little as 60 kaa sites are sufficient to resolve branches that are much shorter. Thus, as discussed earlier (Hallström and Janke 2008), speciation-related processes, such as incomplete lineage sorting and hybridization, might explain the lack of resolution. These processes would explain the reticulate network from the reconstruction of retroposon data for the three fundamental mammalian lineages and the other contradictory results like the grouping of the Chiroptera with Perissodactyla and Carnivores in some analyses (Nishihara et al. 2006).

The molecular divergence time estimates indicate that the onset and rapid radiation of placental mammals at ca 90 Ma coincides with an increase in global temperatures in the Albian (112-99.6 Ma) of the mid-Cretaceous (Puceat et al. 2003, Wilson and Norris 2001) that reached a maximum at the Cenomanian–Turonian

boundary at 93.5 Ma (Retallack 2002). The climate change and a generally high biological turnover (Wilson and Norris 2001) are likely to have triggered the radiation of placental mammals to a similar extent as did plate tectonics, by producing more favorable and diversified environments. However, these developments may have also caused stress and extinction events by replacing ecosystems. These "cryptic" extinctions could be interpreted as rapid radiations (Crisp and Cook 2009) and may further complicate the interpretation of mammalian evolution. The cause of the true rapid radiations cannot be determined by genomic studies alone; however, careful correlations of the occurrence/disappearance of mammalian fossils, divergence times, networks, and paleoclimates may soon provide a more detailed picture.

Origin of placental mammals

When placental mammals began their radiation in the mid-Cretaceous some 90 Ma, the super-continents Laurasia and Gondwana were fragmenting. It has therefore been suggested that the early radiation of placental mammals was shaped by vicariance (Hedges et al. 1996). Recently, tectonic movements have also been used to explain the difficulties encountered in resolving the early placental mammalian radiation and distribution pattern by analyzing sequence and retroposon data (Wildman et al. 2007; Churakov et al. 2009; Nishihara et al. 2009). These studies assume a Gondwanan distribution of the earliest placental mammals and a later distribution of Boreoplacentalia to Laurasia and Xenarthra to South America, while the members of Afrotheria remained in what became Africa. However, this scenario requires a highly choreographed, rapid, and nearly simultaneous splitting or reconnection of the continents (Nishihara et al. 2009) and seems rather unlikely.

Using geological events to explain the distribution and rapid divergence of three placentalian clades may be problematic, because geological events are not points in time. They have durations of several million years, periods that exceed the short-spanned divergences of the early mammalian radiation and speciation, which occurred over a period of 2-4 Myr (Curnoe et al. 2006; van Dam et al. 2006). It is probable that during continental breakups recurring land bridges caused by ridges and sea level fluctuations frequently connected continental plates over extended period of times. As an example, the Grande Rise and the Walvis Ridge were still exposed into Maastrichtian/Palaeocene times at 70-60 Ma (Sclater et al. 1977; Reyment and Dingle 1987), providing a semi continuous south Atlantic connection between South America

and Africa. Thus, the speciation process and divergence of lineages can be approximately an order of magnitude more rapid than continental drift dynamics.

A complete separation of South America, Africa, and Laurasian continents at 120 Ma has been suggested to explain the difficulty in resolving the Xenarthra-Afrotheria-Boreoplacentalia split by retroposon and sequence data (Nishihara et al. 2009). This date, 120 Ma, is considerably older than phylogenomic divergence time estimates among placental mammals, making their distribution by vicariance improbable. Furthermore, the Laurasian-Gondwanan separation occurred considerably earlier than 120 Ma (Smith et al. 2004, Reeves 2009, see: www.reeves.nl), and was not concordant with the separation of South America and Africa.

Correlating the divergence of placental mammals in this way with continental drift is problematic, because it requires speculations about the origin and distribution of living groups based on their current biogeography and ignores the mammalian fossil record. Such correlations seem to be intuitively convincing, but they presuppose a Gondwanan origin of placental mammals to parsimoniously explain their current distribution by continental drift (Wildman et al. 2007, Churakov et al. 2009, Nishihara et al. 2009). However, the existing fossil record does not support such a scenario. The most parsimonious interpretation of the actual fossil record suggests that basal divergences among placental mammals took place in Laurasia and not Gondwana (Archibald 2003; Hunter and Janis 2006; Wible et al. 2007). The current phylogenomic analyses estimate the origin of the Xenarthra at about 103 Ma. While some placentalian orders may have already been present in the Cretaceous (Asher et al. 2005; Benton and Donoghue 2007), no xenarthran fossils have been identified in South America over a time span of about 30 Myr from about this period (Albian/Cenomanian) until about 70 Ma (Maastrichtian). However, South America has an otherwise rich fossil record from archaic mammals, such as multituberculates, gondwanatheres, and sudamericids, prior to the Cretaceous–Tertiary (K/T) boundary (Flynn and Wyss 1998; Pascual and Ortiz-Jaureguizar 2007), suggesting that the lack of placentalian fossils is real and not an artifact. Therian mammals populated South America and replaced the archaic ones close to the K/T boundary (Pascual and Ortiz-Jaureguizar 2007). Molecular estimates of the origin and diversification time of the two major South American mammalian groups, marsupials at 68.5 Ma and xenarthrans at 65 Ma (Delsuc et al. 2004; Nilsson et al. 2004), correlate well with the fossil-based findings. Thus, molecular- and fossil-based data support a colonization of

South America by xenarthrans and other therian mammals from the north (Laurasia) in the late Cretaceous and not via Africa in the mid-Cretaceous as some schemes suggest (Wildman et al. 2007; Churakov et al. 2009; Nishihara et al. 2009; Murphy et al. 2001; Waddell et al. 1999).

Likewise, the depauperate fossil record of African mammals does not indicate the presence of Xenarthra or members of the Afrotheria on the African continent during a period of 95-80 Ma, when, according to molecular-based divergence time estimates, major afrotherian radiations should have occurred. In fact, the oldest members of the crown group Afrotheria are of Laurasian origin (Asher et al. 2003; Zack et al. 2005; Tabuce et al. 2007). This makes the hypothesis of a Laurasian origin and diversification of placental mammals currently the best supported by the fossil record.

Taking the fossil, tectonic, and molecular data into account yields a simple scenario for the early radiation of placental mammals. The first divergences, possibly triggered by climate change rather than plate tectonics, occurred in Laurasia. These rapid splits left no time for fixation of polymorphisms, but allowed for introgression, and led to mosaic genomes. This radiation was followed immediately by dispersal to different geographic regions, with the Xenarthra reaching South America at a later point in time. This scenario is not unreasonable given that xenarthrans and marsupials reached South America at approximately the same time, with marsupials obviously coming via North America, a Laurasian continent (Nilsson et al. 2004). Thus, the dispersal routes of xenarthrans and marsupials contradict speculations of a Gondwanan origin of placental mammals.

Summary

The advancement of sequencing technology enabled the generation of large quantities of genome data from numerous placental mammals, and eventually the genomes of most species will be sequenced. This gargantuan amount of data already makes it possible to resolve mammalian evolution and phylogeny in a detail that was unimaginable not too long ago (Novacek 1992). Exact divergence time estimates enable detailed studies about whether and how historic events such as plate tectonics and climate changes correlate with the evolution of placental mammals. However, current phylogenomic analyses already show that the evolution of placentalian genomes may not have been strictly bifurcating. Instead, incomplete lineage sorting,

species hybridization, and possibly other as yet unknown processes led to mosaic genomes, with different parts having different phylogenetic histories (Ebersberger et al. 2007, Hallström and Janke 2008, Churakov et al. 2009). The current study as well as previous phylogenomic and retroposon analyses clearly show that certain branches of the mammalian tree, and possibly that of other species, cannot be simply resolved as strictly bifurcating by genome data, and are in many cases best viewed and interpreted as network-like processes.

Acknowledgements

We are grateful to Drs. Maria Nilsson and Adrian Schneider for critical comments on the study and the manuscript and Collin Reeves for comments on the plate tectonics. The Carl-Trygger, Nilsson-Ehle Foundations, and LOEWE, supported the work.

References

Arnason U, Gullberg A, Gretarsdottir S, Ursing B, Janke A. 2000. The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. J. Mol. Evol. 50:569-578.

Archibald JD. 2003. Timing and biogeography of the eutherian radiation: fossils and molecules compared. Mol. Phylogenet. Evol. 28:350–359.

Arnason U, Adegoke JA, Gullberg A, Harley EH, Janke A, Kullberg M. 2008. Mitogenomic relationships of placental mammals and molecular estimates of their divergences. Gene 421:37-51.

Asher RJ, Novacek MJ, Geisler JH. 2003. Relationships of endemic African mammals and their fossil relatives based on morphological and molecular evidence. J. Mamm. Evol. 10:131-162.

Asher, RJ. 2005. Insectivoran-grade placental mammals: character evolution and fossil history. In: Rose KD, Archibald JD editors, The Rise of Placental Mammals. Baltimore: The Johns Hopkins University Press. p. 50–70.

Bajpai S, Gingerich PD (1998) A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. Proc Natl Acad Sci USA 95:15464–15468

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

Benton M, Donoghue PCJ, Asher RJ. 2009. Calibration and constraining molecular clocks. In: Hedges SB and Kumar S, editors. The timetree of life. Oxford University Press. p. 35-86.

Cantrell MA, Filanoski BJ, Ingermann AR, Olsson K, DiLuglio N, Lister Z, Wichman HA. 2001. An ancient retrovirus-like element contains hot spots for SINE insertion. Genetics 158:769-777.

Churakov G, Kriegs JO, Baertsch R, Zemann A, Brosius J, Schmitz J. 2009. Mosaic retroposon insertion patterns in placental mammals. Genome Res. 19:868-875.

Crisp MD, Cook LG. 2009.Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. Evolution 63:2257-2265.

Delsuc F, Vizcaino SF, Douzery EJ. 2004. Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. BMC Evol. Biol. 4:11.

Dunn CW, Hejnol A, Matus DQ, et al. (18 co-authors) 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452:745-749.

Ebersberger I, Galgoczy P, Taudien S, Taenzer S, Platzer M, von Haeseler A. 2007. Mapping human genetic ancestry. Mol. Biol. Evol. 24:2266-2276.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792-1797.

Flynn JJ, Wyss AR. 1998. Recent advances in South American mammalian paleontology. Trends Ecol. Evol. 13:449-454.

Gatesy J, Hayashi C, Cronin MA, Arctander P. 1996. Evidence from milk casein genes that cetaceans are close relatives of hippopotamid artiodactyls. Mol. Biol. Evol. 13:954-963.

Hallström BM, Kullberg M, Nilsson MA, Janke A. 2007. Phylogenomic data analyses provide evidence that Xenarthra and Afrotheria are sister groups. Mol. Biol. Evol. 24:2059-2068.

Hallström BM, Janke A. 2008. Resolution among major placental mammal interordinal relationships with genome data imply that speciation influenced their earliest radiations. BMC Evol. Biol. 8:162.

Hedges SB, Parker PH, Sibley CG, Kumar S. 1996. Continental breakup and the ordinal diversification of birds and mammals. Nature 381:226-229.

Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254-267.

Hunter PJ, Janis CM. 2006. Spiny Norman in the garden of eden? J. Mamm. Evol. 13:89-123.

Irwin DM, Arnason U. 1994. Cytochrome b gene of marine mammals: phylogeny and evolution. J. Mamm. Evol. 2, 37-55.

Janecka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ. 2007. Molecular and genomic data identify the closest living relative of primates. Science 318:792-794.

Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4:18.

Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J. 2007. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. Trends Genet. 23:158-161.

Kullberg M, Hallström BM, Arnason U, Janke A. 2008. Phylogenetic analysis of 1.5 Mbp and platypus EST data refute the Marsupionta hypothesis and unequivocally support Monotremata as sister group to Marsupialia/Placentalia. Zool. Scr. 37:115-127.

Lanave C, Preparata G, Saccone C, Serio G. 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20:86-93.

Lander ES, Linton LM, Birren B Nusbaum C, et al. (254 co-authors) 2001. Initial se-

quencing and analysis of the human genome. Nature 409:860-921.

Murphy WJ, Eizirik E, O'Brien SJ, et al. (11 co-authors) 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294:2348-2351.

Mouchaty SK, Gullberg A, Janke A, Arnason U. 2000. Phylogenetic position of the Tenrecs (Mammalia: Tenrecidae) of Madagascar based on analysis of the complete mitochondrial genome sequence of Echinops telfairi. Zool. Scr. 29:307-317.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nikolaev S, Montoya-Burgos JI, Margulies EH, Program NCS, Rougemont J, Nyffeler B, Antonarakis S. 2007. Early history of mammals is elucidated with the ENCODE multiple species sequencing data. PLoS Genet. 3:1.

Nikaido M, Rooney AP, Okada N. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. USA 96:10261-10266.

Nilsson MA, Arnason U, Spencer PB, Janke A. 2004. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. Gene. 340:189-196.

Nilsson MA, Härlid A, Kullberg M, Janke A. 2010. The impact of fossil calibrations, codon positions and relaxed clocks on the divergence time estimates of the native Australian rodents (Conilurini). Gene. 455:22-31.

Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. Proc. Natl. Acad. Sci. USA 103:9929-9934.

Nishihara H, Okada N, Hasegawa M. 2007. Rooting the eutherian tree: the power and pitfalls of phylogenomics. Genome Biol. 8:R199

Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. Proc. Natl. Acad. Sci. USA 106:5235-5240.

Novacek MJ. 1992. Mammalian phylogeny: shaking the tree. Nature 356:121-125.

Pascual R, Ortiz-Jaureguizar E. 2007. The Gondwanan and South American episodes: two major and unrelated moments in the history of the South American mammals. J. Mamm. Evol. 14:75-137.

Puceat E, Lecuyer C, Sheppard SMF, Dromart G, Reboulet S, Grandjean P. 2003. Thermal evolution of Cretaceous Tethyan marine waters inferred from oxygen isotope composition of fish tooth enamels. Paleoceanography 18:1029.

Rambaut A, Grassley NC. 1997. Seq-Gen: an application for the Monte Carlo

simulation of DNA sequence evolution along phylogenetic trees. Bioinformatics. 13:235-238.

Reeves C. 2009. Re-examining the evidence from plate-tectonics for the initiation of Africa's passive margins. London: Geological Society of Huston/petroleum Exploration Society of Great Britain.

Retallack GJ. 2002. Carbon dioxide and climate over the past 300 Myr. Phil. Trans. R. Soc. Lond. A 360:659-673.

Reyment R, Dingle RV. 1987. Palaeogeography of Africa during the Cretaceous period. Palaeogography, Palaeoclimatology, Palaeoecology 59:93-116.

Roos J, Aggarwal RK, Janke A. 2007. Extended mitogenomic phylogenetic analyses yield new insight into crocodylian evolution and their survival of the Cretaceous-Tertiary boundary. Mol. Phylogenet. Evol. 45:663-673.

Satta, Y, Klein J, Takahata N. 2000. DNA archives and our nearest relative: the tri-chotomy problem revisited. Mol. Phylogenet. Evol. 14:259-275.

Schneider A, Cannarozzi GM. 2009. Support patterns from different outgroups provide a strong phylogenetic signal. Mol. Biol. Evol. 26:1259-1272.

Sclater JG, Hellinger S, Tapscott, C. 1977. The paleobathymetry of the Atlantic Ocean from the Jurassic to the present. J. Geol. 85:509-552.

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114-1116.

Smith AG, Smith DG, Funnell BM. 2004. Atlas of Mesozoic and Cenozoic Coastlines. Cambridge: Cambridge University Press).

Stanhope MJ, Waddell VG, Madsen O, de Jong W, Hedges SB, Cleven GC, Kao D, Springer MS. 1998. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. Proc. Natl. Acad. Sci. USA 95:9967-9972.

Steel, M, Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol. Biol. Evol. 17:839-850.

Tabuce R, Asher RJ, Lehmann T. 2008. Afrotherian mammals: a review of current data. Mammalia 72:2-14.

van de Lagemaat, LN, Gagnier, L, Medstrand P, Mager DL. 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 15:1243-1249.

van Tuinen M, Hadly EA: Calibration and error in placental molecular clocks: a conservative approach using the cetartiodactyl fossil record. J Hered 95: 200–208 (2004).

Venter JC, Adams MD, Myers EW, Li PW, et al. (273 co-authors). 2001. The Sequence of the Human Genome. Science 291:1304-1351.

Waddell PJ, Cao Y, Hasegawa M, Mindell DP. 1999. Assessing the cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. Syst. Biol. 48:119-137

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum likelihood approach. Mol. Biol. Evol. 18:691-699.

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

Wildman DE, Uddin M, Opazo JC, Liu G, Lefort V, Guindon S, Gascuel O, Grossman LI, Romero R, Goodman M. 2007. Genomics, biogeography, and the diversification of placental mammals. Proc. Natl. Acad. Sci. USA 104:14395-14400

Wilson PA, Norris RD. 2001. Warm tropical ocean surface and global anoxia during the mid-Cretaceous period. Nature 412.425-429

Xu X, Janke A, Arnson U. 1996. The complete mitochondrial DNA sequence of the Greater Indian Rhinoceros, *Rhinoceros unicornis*, and the phylogenetic relationship among Carnivora, Perissodactyla and Artiodactyla (+Cetacea). Mol. Biol. Evol. 13:1167-1173.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. J. Mol. Evol. 39:306-314.

Zack SP, Penkrot TA, Bloch JI, Rose KD. 2005. Affinities of 'hyopsodontids' to elephant shrews and a Holarctic origin of Afrotheria. Nature 434:497-501.

Figure Legends

Figure 1 – Presence/absence matrix for illustrating sequence density. Each dash indicates the presence of a given gene (x-axis) in a given species (y-axis). The matrix is sorted for maximum sequence density along both axes, placing the species with the best coverage at the top of the plot and the genes represented by the most species at the left.

Figure 2 – ML tree reconstructed from an alignment length of 954k aa from 37 species under the WAG2000 +8GI of sequence evolution.

Figure 3 – Chronogram of mammalian divergences. Open and filled circles indicate calibration points: open circles represent those estimated divergence times anywhere between the boundaries, filled circles those reaching either the upper or lower boundary.

Figure 4 – Neighbor-Net based on the 954k aa alignment from 37 species. All intra-ordinal, ordinal, and most super-ordinal relationships are clearly defined in the Neighbor-Net by stretched boxes that are longer than they are wide, indicating limited conflict in the data. The clades that are poorly supported by ML aa sequence data analysis are characterized by boxed nodes that are nearly square or by negligibly short branch lengths.

Figure 5 – Close up of the Neighbor-Net highlighting major splits.

Figure 6 – Retroposon split network.

Figure 7 – Simulation of aa sequence lengths required for resolving temporally tight divergences at 100 Ma. The black line is reconstructed from the simulation of a dataset free of gaps, while the grey line represents a simulation where gaps were included according to their frequency in the original data.

Supplemental figure 1 – Neighbor-joining bootstrap analysis.

Supplemental figure 2a – Consensus network of 10% longest alignments, threshold = 0.02

Supplemental figure 2b – Consensus network of 10% longest alignments, threshold = 0.04

Table 1: Scientific and common names of species included in the phylogenomic analyses

Species	**Common name**	**Species**	**Common name**
Homo sapiens	Human	*Vicugna pacos*	Alpaca
Pan troglodytes	Chimpanzee	*Bos taurus*	Cow
Gorilla gorilla	Gorilla	*Tursiops truncatus*	Dolphin
Pongo pygmaeus	Orangutan	*Myotis lucifugus*	Little brown bat
Macaca mulatta	Macaque	*Pteropus vampyrus*	Large flying fox
Otolemur garnettii	Galago	*Sorex araneus*	Common shrew
Microcebus murinus	Mouse lemur	*Erinaceus europaeus*	Western European hedgehog
Tarsius syrichta	Tarsier	*Loxodonta africana*	African elephant
Tupaia belangeri	Treeshrew	*Echinops telfairi*	Tenrec
Rattus norvegicus	Common rat	*Procavia capensis*	Hyrax
Mus musculus	House mouse	*Dasypus novemcinctus*	Armadillo
Dipodomys ordii	Kangaroo rat	*Choloepus hoffmanni*	Sloth
Spermophilus tridecemlineatus	Ground squirrel	*Monodelphis domestica*	Opossum
Cavia porcellus	Guinea pig	*Ornithorhynchus anatinus*	Platypus
Oryctolagus cuniculus	Rabbit	*Gallus gallus*	Chicken
Ochotona princeps	Pika	*Taeniopygia guttata*	Zebra finch
Felis catus	Cat	*Anolis carolinensis*	Anole lizard
Canis familiaris	Dog	*Xenopus tropicalis*	Frog
Equus caballus	Horse		

Table 2 - Calibration points used for dating mammalian divergences.

Split		Minimum age (Ma)	Maximum age (Ma)
Eutheria	Metatheria	124	138.4
Boreoplacentalia	Exafrotheriaplacentalia	61.5	113
Euarchontoglires	Laurasiaplacentalia	61.5	113
Primates	Glires	61.5	100.5
Lagomorpha	Rodentia	61.5	100.5
Cetacea	Cow	70	52.4
Caniformia	Feliformia	39.7	63.8
Apes	Old World monkeys	23.5	33.7
Rat	Mouse	10.4	12.3
Human	Chimpanzee	6.5	10

Table 3: Divergence times and their standard deviation for splits shown in figure 3.

Split	Divergence time (Ma)
1	161.7 ± 0.35
2	138.4[a]
3	90.60 ± 0.11
4	87.89 ± 0.12
5	78.01 ± 0.21
6	81.48 ± 0.32
7	65.08 ± 0.37
8	64.98 ± 0.38
9	74.28 ± 0.15
10	76.22 ± 0.30
11	72.18 ± 0.16
12	67.34 ± 0.27
13	63.64 ± 0.42
14	73.47 ± 0.21
15	67.72 ± 0.21
16	52.45 ± 0.33
17	62.47 ± 0.26
18	70.31 ± 0.26
19	71.75 ± 0.25
20	41.58 ± 0.17
21	60.54 ± 0.22
22	25.95 ± 0.17
23	48.48 ± 0.16
24	59.27 ± 0.26
25	60.90 ± 0.26
26	57.50 ± 0.21
27	18.10 ± 0.14
28	52.4[a]
29	51.33 ± 0.20
30	11.76 ± 0.08
31	12.30[a]
32	9.376[a]

[a] Estimates that have reached a limit set by the calibration point.

Figure 1

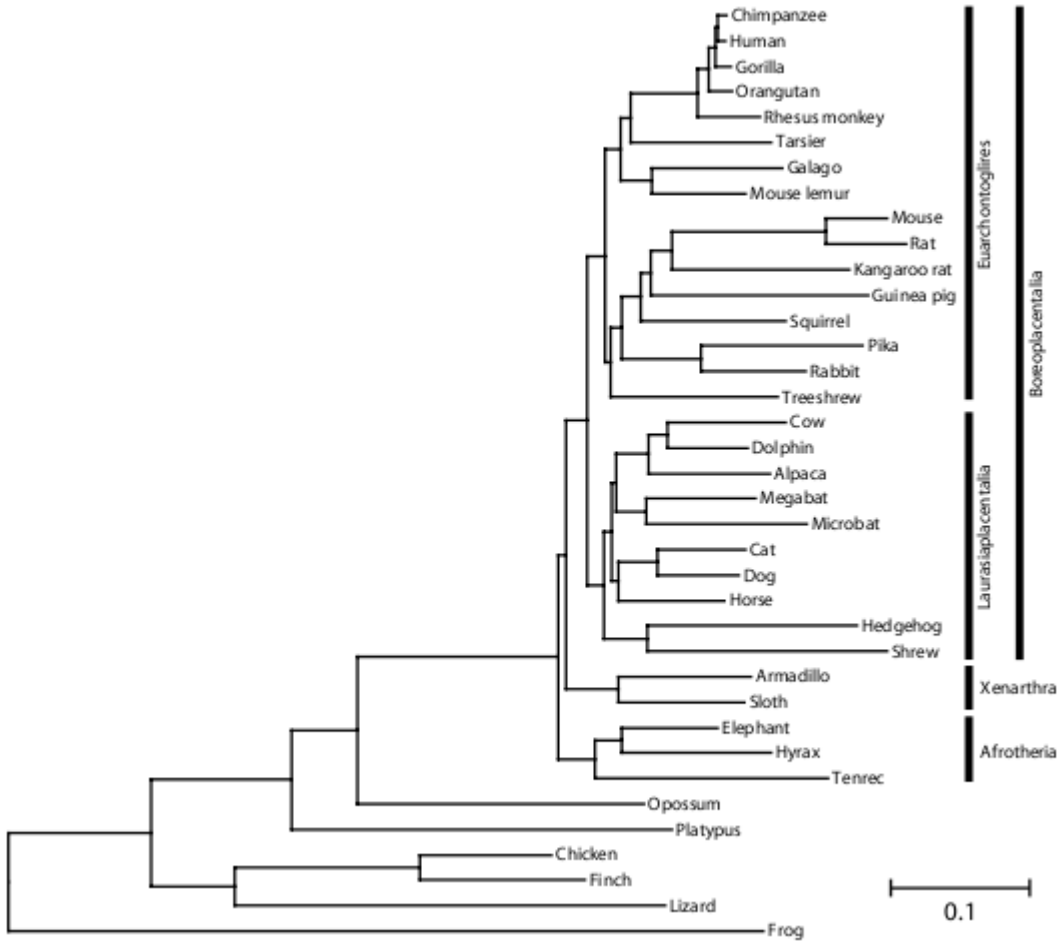
Chimpanzee
Human
Gorilla
Orangutan
Rhesus monkey
Tarsier
Galago
Mouse lemur
Mouse
Rat
Kangaroo rat
Guinea pig
Squirrel
Pika
Rabbit
Treeshrew
Cow
Dolphin
Alpaca
Megabat
Microbat
Cat
Dog
Horse
Hedgehog
Shrew
Armadillo
Sloth
Elephant
Hyrax
Tenrec
Opossum
Platypus
Chicken
Finch
Lizard
Frog
Euarchontoglires
Laurasiaplacentalia
Boreoplacentalia
Xenarthra
Afrotheria
0.1

Figure 2

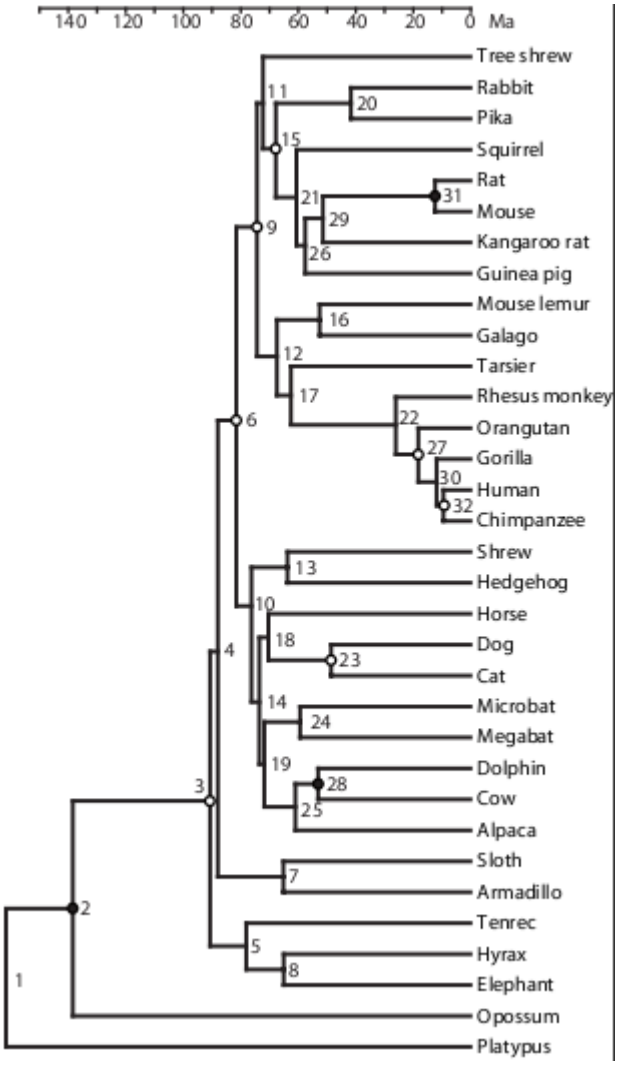
140 120 100 80 60 40 20 0 Ma
Tree shrew
Rabbit
Pika
Squirrel
Rat
Mouse
Kangaroo rat
Guinea pig
Mouse lemur
Galago
Tarsier
Rhesus monkey
Orangutan
Gorilla
Human
Chimpanzee
Shrew
Hedgehog
Horse
Dog
Cat
Microbat
Megabat
Dolphin
Cow
Alpaca
Sloth
Armadillo
Tenrec
Hyrax
Elephant
Opossum
Platypus
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

Figure 3

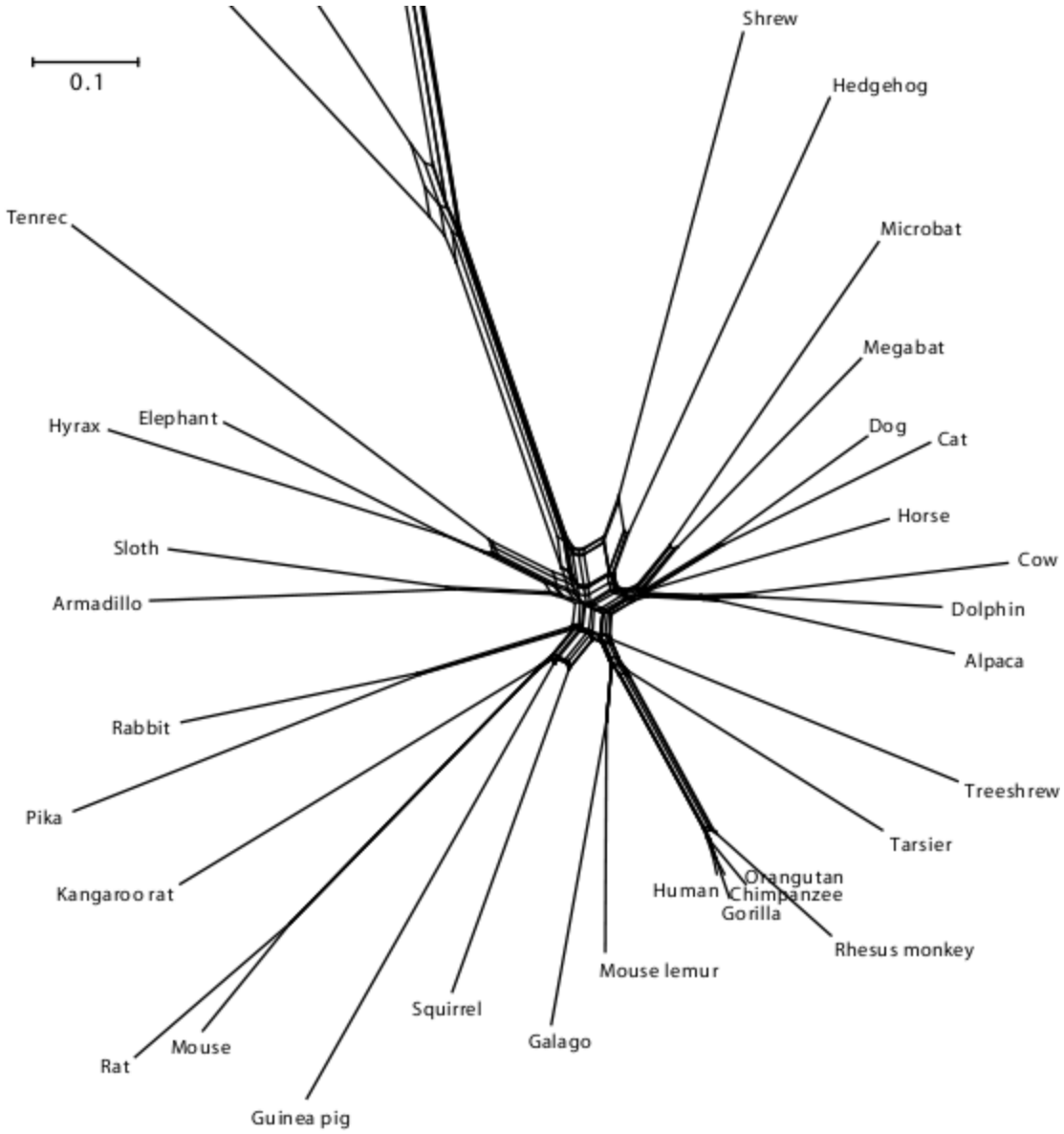
0.1
Shrew
Hedgehog
Tenrec
Microbat
Megabat
Hyrax
Elephant
Dog
Cat
Horse
Sloth
Cow
Armadillo
Dolphin
Alpaca
Rabbit
Treeshrew
Pika
Tarsier
Kangaroo rat
Orangutan
Human
Chimpanzee
Gorilla
Rhesus monkey
Mouse lemur
Squirrel
Galago
Mouse
Rat
Guinea pig

Figure 4

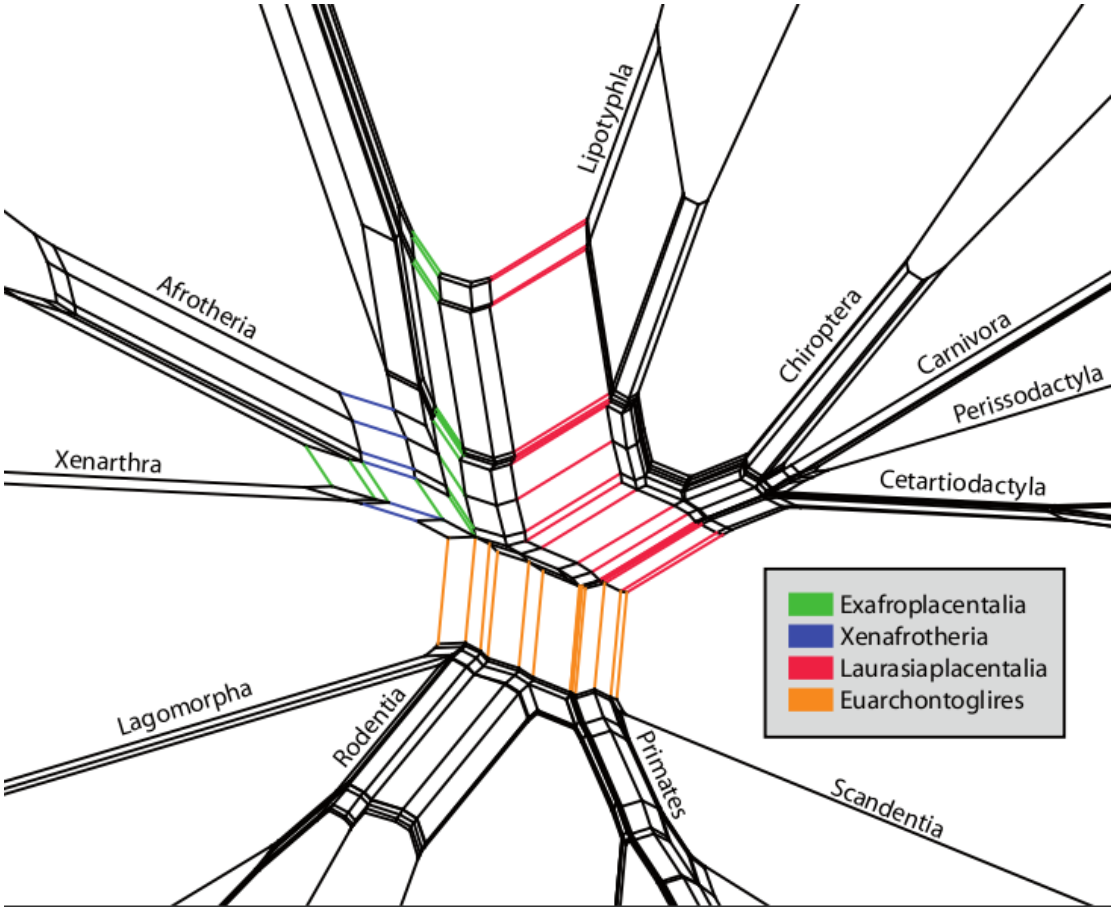
Lipotyphla
Afrotheria
Chiroptera
Carnivora
Perissodactyla
Xenarthra
Cetartiodactyla
Exafroplacentalia
Xenafrotheria
Laurasiaplacentalia
Euarchontoglires
Lagomorpha
Rodentia
Primates
Scandentia

Figure 5

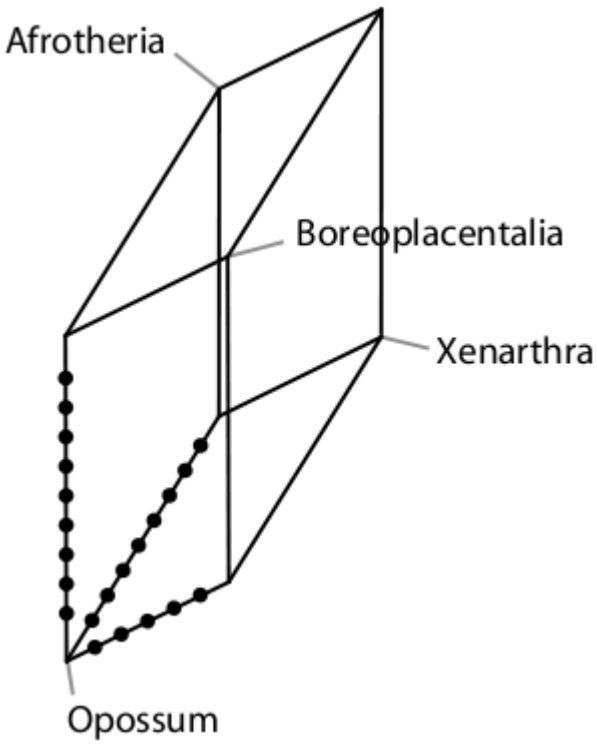

Figure 6

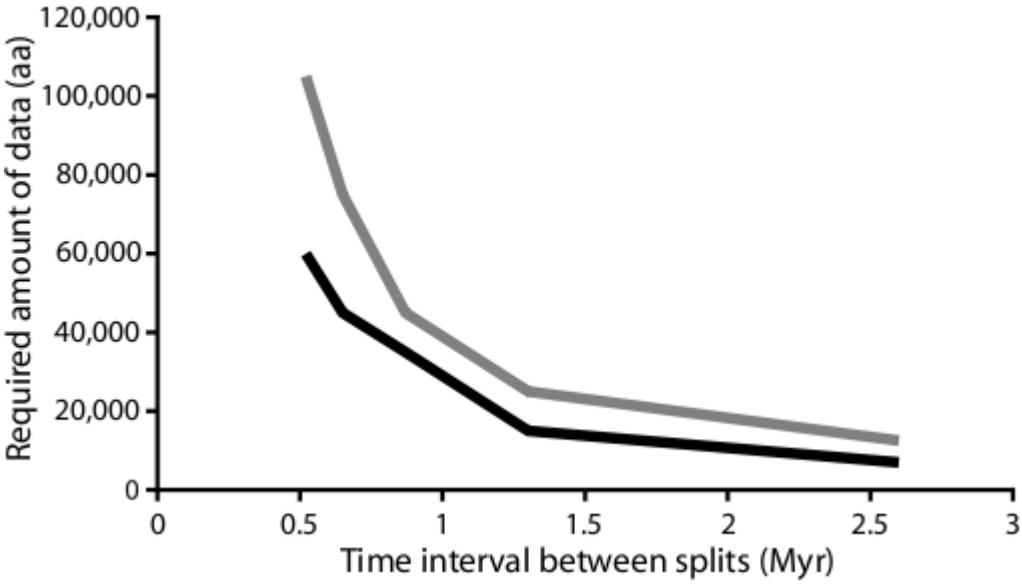
120,000
100,000
80,000
60,000
40,000
20,000
0
Required amount of data (aa)
0
0.5
1
1.5
2
2.5
3
Time interval between splits (Myr)

Figure 7