Published in final edited form as:

Mol Phylogenet Evol. 2009 September ; 52(3): 660–664. doi:10.1016/j.ympev.2009.05.002.

Phylogeny of the Ferungulata (Mammalia: Laurasiatheria) as determined from phylogenomic data

Zhuo-cheng Hou^{a,b,d}, Roberto Romero^{a,b,*}, and Derek E. Wildman^{a,b,c}

^aPerinatology Research Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development/NIH/DHHS, Detroit, MI 48201, USA

^bCenter for Molecular Medicine & Genetics, Wayne State University School of Medicine, Detroit, MI 48201 USA

^cDepartment of Obstetrics and Gynecology Wayne State University School of Medicine, Detroit, MI 48201 USA

^dDepartment of Animal Genetics, China Agricultural University, Beijing, China

Abstract

Great progress has been made toward resolving the evolutionary relationships among extant mammals, yet there are still areas of disagreement. The relationships among ferungulates that have high quality draft genome sequences available (i.e. dog, cow, horse) are unresolved, and thus we examined their phylogeny using currently known mammalian 1:1 orthologs. This dataset consists of 40 million base pairs from 2705 protein-coding genes. Maximum likelihood and Bayesian analyses of the combined and individual gene phylogenies strongly support a sister grouping of cow and horse to the exclusion of dog although topology tests could not rule out a horse and dog sister group relationship.

Keywords

ferungulata; carnivore; perissodactyla; cetartiodactyla; molecular phylogeny

INTRODUCTION

Accurate phylogenies are required to reconstruct the evolution of morphological, physiological, and behavioral characters. The advent of molecular phylogenetic techniques has led to great progress in resolving the phylogenetic relationships among extant mammals, yet there are still areas of disagreement. The clade Laurasiatheria consists of six monophyletic clades Eulipotyphla (insectivores including hedgehogs, gymnures, soricid shrews, moles, and *Solenodon*), Chiroptera (bats), Pholidota (pangolins), Carnivora (carnivores), Perissodactyla (rhinoceroses, equides, and tapirs), and Cetartiodactyla (artiodactyls and cetaceans) (Wilson and Reeder, 2005). The cetaceans are given ordinal status in current reference taxonomies (E.g., Wilson and Reeder, 2005), but molecular phylogenies demonstrate they are more closely related to artiodactyls like the hippopotamus, thus rendering the artiodactyls paraphyletic (Irwin and Arnason, 1994; Kleineidam et al., 1999; Montgelard et al., 1997; Ursing and Arnason, 1998). Molecular phylogenetic studies on mammals have identified the branching orders among most orders in the laurasiatherian

Corresponding Author: Roberto Romero, MD, Perinatology Research Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH/DHHS, Wayne State University/Hutzel Women's Hospital, 3990 John R, #4 Brush, Detroit, MI 48201.

clade (Janecka et al., 2007; Murphy et al., 2001a, b; Springer et al., 2007); the relationships among the orders in the clade Ferungulata (carnivores, pangolins, cetartiodactyls, and perissodactyls) are less clear. These previous molecular studies as well as limited morphological evidence (Shoshani and McKenna, 1998) support a grouping of carnivores and pholidotes to the exclusion of perissodactyls and cetartiodactyls; however, further resolution within this clade has been challenging. Indeed, the relationships among the orders that include the domestic dog (Carnivora), the cow (Cetartiodactyla) and horse (Perissodactyla) have defied resolution.

There are three competing hypotheses regarding the phylogenetic relationships among Cetartiodactyla (cow), Carnivora (dog) and Perissodactyla (horse) (Fig. 1A–C). Early molecular studies (Montgelard et al., 1997) based on the mitochondrial genes 12S rRNA and cytochrome b support the ((cow, horse), dog) hypothesis (Fig. 1A). Similarly, Murphy et al. (2001a) using a concatenation of 15 nuclear genes and three mtDNA provided support for the relationships depicted in Fig. 1A. In contrast to this pattern, maximum likelihood (ML)¹ and Bayesian methods analyzing slightly more nuclear genes and mtDNA genes weakly supported the ((horse, dog), cow) hypothesis (Fig. 1B) (Murphy et al., 2001b). However, a recent expansion of this dataset (Springer et al., 2007) using 20 different nuclear genes and a broad sample of placental mammals again favored the ((cow, horse), dog) relationship although posterior probability was low (48%). A phylogeny based on housekeeping genes supported the ((horse, dog), cow) arrangement (Fig. 1B) with 52% bootstrap support (Kullberg et al., 2006). A recent phylogenomic study (Prasad et al., 2008) based on ENCODE data from 44 mammalian species also addressed this issue. When protein coding sequences were analyzed this study recovered the ((cow, horse), dog) topology, albeit with very low ML bootstrap support (18%). However, we note that this dataset did not recover a monophyletic ferungulate clade, and instead had the Chiroptera as the sister order to the Perissodactyla + Cetartiodactyla clade, although with low ML bootstrap support (20%). Conversely, the topology ((cow, dog), horse) (Fig. 1C) was weakly supported (ML bootstrap 64%, Bayesian posterior probability = 1.0) when a larger dataset (including all the coding and non-coding sequences) was used. Moreover, retroposon data (Nishihara et al., 2006) support the clade Pegasoferae in which bats are sister to a dog-horse clade. Inferring the relationships among laurasiatherian mammals will be a necessary first step in unraveling the evolution of morphological features and adaptations in this diverse clade.

The release of high quality (6.8 x fold coverage) horse genome data (<http://www.broad.mit.edu/mammals/horse/>) along with the Ensembl annotated version that was released in April, 2008, combined with the already available dog (Lindblad-Toh et al., 2005) and cow (The Bovine Genome Sequencing and Analysis Consortium et al., 2009) assemblies provide an unprecedented opportunity to elucidate the relationships among these three taxa, and we took advantage of these data to address the question. In order to avoid problems associated with orthology assessment and complex evolution of gene families, we used a large, but relatively conserved phylogenomic dataset in an attempt to infer the relationships among dog, cow, and horse. We inferred well-supported relationships within the clade, and we discuss the implications of our findings.

MATERIALS AND METHODS

Genome data sources

All sequences were obtained from the Ensembl comparative database (v. 49, April, 2008). We used Ensembl comparative homology databases (Ensembl protein homology, <http://www.ensembl.org>). The protein homology database is limited to only the longest transcripts for each gene and relies solely on Ensembl predicted gene datasets (Ensembl help desk). When genome sequence coverage is good, inferences of protein homology are robust.

However, it is not possible to accurately infer gene duplication and loss using genomes with coverage lower than 2x. Therefore, we decided to use Ensembl protein homology and only those laurasiatherian genomes with fold coverage >6x for initial orthology assessment. Because of gaps, assembly and sequence errors in 2x genomes (Green, 2007) we did not analyze the low coverage laurasiatherian assemblies from bat, cat, hedgehog, and sorcid shrew. Instead, we only used the high fold coverage assemblies from human, orangutan (<http://genome.wustl.edu/>), mouse, rat, cow (<http://www.hgsc.bcm.tmc.edu/projects/bovine/>), dog, horse (<http://www.broad.mit.edu/mammals/horse/>), and opossum for the purposes of this study. Genome coverage, taxon names, and sequence length used are listed in Table 1. We only used established single copy orthologs as determined by Ensembl. The rationale for this choice is that lineage specific gene duplications resulting in sub- and/or neo-functionalization in new gene copies along with processes such as gene conversion complicate interpretations of branch specific adaptive evolution. Therefore, we took the conservative approach and examined only those genes with relatively simple evolutionary histories. All gene families with multiple copies in a given genome were excluded from further analysis.

Sequences and alignment quality control

Sequences with a proportion of “Ns” greater than 10% across the length of the gene sequence were removed from analysis. Further, any gene sequence with total length <300 bp was excluded from further analysis. Alignment of multiple DNA sequences based on predicted protein sequences was performed with CLUSTAL W (Thompson et al., 1994) using Bioperl modules (<http://www.bioperl.org>). Some sequences, especially partial sequences of the full coding sequence, will significantly reduce the alignment quality and affect phylogenetic analyses (Liu et al., 2007). We used a custom perl script to concatenate the aligned cDNA sequences that were curated using Gblock v0.91 (Castresana, 2000) to retain only high quality alignment regions.

ML, MP and Bayes phylogenetic trees

MP trees from aligned nucleotide sequences were inferred with MEGA4.0 (Tamura et al., 2007) using random addition sequence replicates and 1000 branch and bound bootstrap replicates (500 replicates for amino acid MP analysis). Because substitution saturation is more likely to occur at synonymous sites, and because synonymous sites are much more frequent at 3rd codon positions, we did separate MP analyses using (1) all codon positions, and (2) 1st + 2nd codon positions. ModelTest (Posada and Crandall, 1998) was used to select the best nucleotide substitution model for the concatenated datasets and also for each gene according to the Akaike Information Criterion (AIC). ML tree searches on the concatenated dataset were performed with RAxML-7.04 (Stamatakis, 2006) using both the GTR + Γ + I model suggested by ModelTest and the GTR + Γ model that represented the most common model selected for the individual genes. The best-known likelihood (BKL) tree was inferred after conducting 100 RAxML runs using f-d option for thorough searching. Bootstrap replicates were performed using the parallelized MPI-enabled version of RAxML7.04. Opossum was used as the outgroup to the placental ingroup. We used the WAG model for analysis of deduced protein sequences MrBayes-3.1.2 (recompiled as an MPI version) was used to construct the Bayes tree with a GTR + Γ model and 500,000 MCMC generations. The parallelized MPI version of RAxML and MrBayes were run on the Wayne State University (Detroit, MI, USA) high performance computing grid (<http://www.grid.wayne.edu>). Individual gene trees were reconstructed using PhyML-2.4 (Guindon and Gascuel, 2003) according to the best model suggested by ModelTest (Posada and Crandall, 1998). UpdownDistance was used to compare the gene and species trees (<http://web.njit.edu/~wangj/updown.htm>).

RESULTS AND DISCUSSION

Data composition

A total of 2705 single copy genes (1:1 orthologs) for the eight species were obtained for comparative genomics analysis. The total evidence, combined, coding sequence dataset consists of 41,924,938 bases (Table 1), and the conserved, GBlock curated, coding sequence dataset includes 33,991,128 bases (Table 1). ModelTest analysis of each of the 275 multiple sequence alignments selected a GTR + Γ model (Fig. 2) as the best fit for the largest proportion of these single orthologs ($1859/2705 = 66.27\%$). HKY + Γ , GTR + Γ + I and K2P + Γ and SYM + Γ models were also often chosen as the best model for the analyzed genes (Fig. 2). In contrast to the GTR + Γ model that was most commonly optimal for individual genes; the best model for the concatenated, Gblock curated, total evidence dataset is the GTR + Γ model. Because of this discrepancy, we analyzed the concatenated data twice. We first used the best model for concatenated data: GTR + Γ + I, and we also analyzed the combined dataset according to model recovered as best for the largest proportion of single copy genes (GTR + Γ).

In some situations nucleotide composition bias causes incorrect tree inference (Nei and Kumar, 2000). We found no significant base difference among species. The G + C nucleotide composition ranged from 48.1% (opossum) to 52.6% (cow) (Table S1). Based on this finding, we did not consider base composition as a factor affecting the different phylogenetic analyses in this study.

Phylogenetic inference based on combined data

Fig. 3 depicts the results of the phylogenetic analyses. The BKL ML tree topology was identical whether or not we partitioned the data according to the best model per gene. All 100 runs recovered the same BKL tree (Fig. 3) for the two different models used; GTR + Γ and GTR + Γ + I. The BKL tree supports the ferungulate branching order depicted in Fig. 1A for both the GTR + Γ and GTR + Γ + I models. ML analysis using the best nucleotide substitution model (GTR + Γ + I) recovered a well-resolved topology ($\ln L = -16,809,034.51$) for the eight studied species. The percent bootstrap support was 97.5% for the grouping of cow and horse to the exclusion of dog, and 100% at all other nodes. ML analysis of the concatenated dataset using the model selected for the greatest proportion of individual genes (GTR + Γ) recovered the identical topology ($\ln L = -16,809,379.72$); however, percent bootstrap values were lower (79%) for the cow + horse clade. The Bayesian analysis also recovered this topology ($\ln L = -16,810,712.650$) and all nodes had posterior probabilities of 1.0. Because our dataset was based exclusively on sequences from protein-coding genes, we were also able to conduct further analyses by using the predicted protein sequences to construct phylogenetic relationships. Results of ML (WAG model) analyses recovered an identical topology as depicted in Fig. 3 ($\ln L = -7,454,513.83$).

MP analyses recovered different tree topologies compared to the ML and Bayesian analyses. The tree derived from 1st and 2nd positions had an identical topology to the tree that included all codon positions. The single MP tree (tree length = 2,496,212; all codon positions included) supports a sister relationship between Primates and Ferungulata, rather than a Primates plus Rodentia clade. Moreover, the nucleotide MP tree supported a cow-dog grouping instead of the horse-cow clade seen in ML and Bayesian analyses. Bootstrap proportions for these MP groupings are also strong (Fig. S1). MP analysis of deduced protein sequences (tree length = 1,114,173) disagreed with the MP results from nucleotide sequences by weakly supporting (58% bootstrap) a dog-horse grouping (Fig. S2).

Topology tests of combined data

To compare the likelihood of the data supporting each of the three computing hypotheses depicted in Fig. 1, we conducted an expected-likelihood weights (ELW) test (Strimmer and Rambaut, 2002) as implemented in RAxML-7.0.4. The comparison is between the best known-likelihood tree (BKL; Fig. 3) and the two possible alternative groupings among dog, horse, and cow (i.e., those depicted in Fig. 1B and C). In this comparison, the trees depicted in Fig. 1A and B are more likely than the tree Fig. 1C. The ELW test revealed that the hypotheses depicted in Fig. 1A and B are both significantly better than Fig. 1C ($p = 0.0006$ and $p = 0.03$, respectively). There is no significant difference between the topologies depicted in Fig. 1A and B ML trees ($p = 0.2366$). Possibly, denser taxon sampling in future studies may help to better resolve this issue.

We did an additional comparison of the optimal topology recovered by ML and Bayesian analyses (Fig. 3) to the best MP tree (Fig. S1). The MP topology did not recover a monophyletic Eurachontoglires clade; instead the primates were sister to laurasiatherians, and rodents next join this group. In this comparison, the EWL test found the ML and Bayesian tree to be significantly better than the MP tree. Moreover, we found that the placement of rodents outside of the Primates + Laurasiatheria group, significantly worsened the likelihood regardless of the branching order among dog, horse, and cow ($p = 0.0000$ for all three comparisons). The inconsistency of parsimony has been the strongest challenge to its use. Long branch attraction adversely affects the performance of parsimony and thus, parsimony will work best for recently diverged species (Felsenstein, 2004). Indeed, simulation studies concluded that ML is more accurate than MP under realistic combinations of both heterotachy and evolutionary rate variation across lineages (Philippe et al., 2005). We note that the generation times of mouse, rat, and opossum (14–15 days (Fadem and Rayve, 1985)) are much shorter than other species in this study, and generation-time effects (Hartl and Clark, 2007) may have resulted in long branch attraction among these species that consequently affected the MP tree topology (Fig. S1).

Because some studies have supported a nesting of bats within the ferungulate taxa included in the present study (Nishihara et al., 2006; Prasad et al., 2008), we examined a dataset that included 1.7 million base pairs of sequence from the low coverage *Myotis lucifugus* genome assembly. These data were derived from 1255 genes. RAxML analysis of the multiple sequence alignment of this data with the other eight taxa resulted in a laurasiatherian tree topology ($\ln L = -7176519.36$) in which a cow – bat clade was sister to a horse – dog clade, with both clades enjoying 100% bootstrap support (Fig. S3). Our finding disagrees with Nishihara et al. (2006) who had a (((Perissodactyla, Carnivora), Chiroptera), Cetartiodactyla) grouping, and with Prasad et al. (2008) who had a (((Perissodactyla, Artiodactyla), Chiroptera), Carnivora) grouping. We thus attribute little confidence to this result that has not been reported elsewhere, and we treat it with caution. Clearly, more work needs to be done regarding the phylogenetic relationships among laurasiatherians, and scrotiferans in particular.

Individual gene tree analyses

In addition to analyses of the concatenated, Gblock curated, multiple sequence alignments, we also used PhyML (Guindon and Gascuel, 2003) to produce a gene tree for each individual gene using the best substitution model as determined by ModelTest. ModelTest results determined that only 158 genes best fit the GTR + Γ + I model while 1859 genes best fit the GTR + Γ model (Fig. 2). The resultant gene trees were then compared to the topologies in Fig. 1. None of the three competing hypotheses emerged as a clear favorite based on these analyses. The total number of gene trees that were consistent with the dog-horse-cow grouping is 68.5% (1852/2705; Fig. S4). The total number of gene trees that were

consistent with the horse-cow grouping is 27.1% (734/2705); the horse-dog grouping is supported by 28.5% (771/2705) of the genes, and the cow-dog grouping is supported by 26.9% (728/2705) of the genes. Only 568 of the 1852 gene trees that recovered a monophyletic Ferungulata are consistent in terms of the branching order of the remaining mammal taxa. This finding could explain why the *a priori* selected genes used in previously published studies (Murphy et al., 2001a; Prasad et al., 2008) also apparently do not have enough power to resolve the phylogenetic relationships among these three species. We propose that only by including all the available single orthologs were we able to obtain well-supported topologies for the branching order among our study taxa. However, we note that high bootstrap values are somewhat meaningless in large-scale phylogenomic studies because bootstrap analysis is in effect a test of whether the dataset is of adequate size to find a well-supported solution. That the incongruent MP vs. ML/Bayesian results both have high bootstrap supports and/or posterior probabilities further highlights the inadequacy of traditional measure of branch support in phylogenomic studies.

Implications of this study

Given all of the results presented in the current study, we can conclude that massive amounts of genomic data can provide insight into difficult phylogenetic problems. Transposable element and indel data also have demonstrated applicability to difficult phylogenetic problems (Kriegs et al., 2006; Nishihara et al., 2006; Murphy et al., 2007). In the current study, even massive amounts of data led to different results from parsimony compared with those from ML and Bayesian approaches. Based on known problems with the MP such as long branch attraction (Felsenstein, 2004), we provisionally accept the branching arrangement shown in Fig. 3 with the caveat that ML and Bayesian models of sequence evolution developed for studies of one or a few genes may not be effective when analyzing large datasets.

If we accept the topology depicted in Fig. 3, a number of interesting observations regarding the evolution of morphological features within the Ferungulata emerge. Both perissodactyls and artiodactyls are ungulates, while carnivores have claws on their appendages. The closest outgroups to the ferungulates not included in our study are the chiropterans followed by the eulipotyphlans (Murphy et al., 2001a). Species in these orders have claws rather than hoofs, and thus, we can parsimoniously infer that proto-hoofs characteristic of early ungulates such as *Hyracotherium* had emerged at the time of the last common ancestor of horses and cows rather than independently in each of these lineages. Indeed, our study supports the existence of a monophyletic perissodactyl – cetartiodactyl clade. Most carnivores have an invasive endotheliochorial placental interface while the placentas of horses and cows are of the less invasive epitheliochorial form (Mossman, 1987). Cetaceans, like perissodactyls and artiodactyls, also have epitheliochorial placentas (Mossman, 1987). Phylogenetic reconstructions show that the degree of placental invasion into the uterus at the time of the last common ancestor of dog, cow and horse is ambiguous (Mess and Carter, 2007; Wildman et al., 2006); however, our phylogenetic results would suggest that the epitheliochorial placenta was present at the time of the last common ancestor of horse and cows. Finally, carnivores often have multiple offspring per pregnancy while artiodactyls and perissodactyls usually have single offspring. Most bats are monotocous (single offspring per pregnancy) and most eulipotyphlans are polytocous (multiple offspring per pregnancy). The Pholidota are ferungulates who were not included in our study, are closely related to carnivores (Murphy et al., 2001b), and are usually monotocous. Thus, we can infer that the last common ancestor of dog, cow, and horse, had a single offspring per pregnancy, and that carnivores gained the ability to produce multiple offspring per pregnancy. Unraveling the evolutionary history of these and other key features requires an accurate resolution of the

phylogeny of the Ferungulata combined with careful morphological studies, high quality genomic sequence from additional species, and experimental studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by the Intramural Research program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development/NIH/DHHS. We thank the Wayne State University (Detroit, MI, USA) high performance grid for access to computing resources.

References

1. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 2000; 17:540–552. [PubMed: 10742046]
2. Elsik CG, Tellam RL, Worley KC. The Bovine Genome Sequencing Analysis Consortium. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science.* 2009; 324:522–528. [PubMed: 19390049]
3. Fadem BH, Rayve RS. Characteristics of the oestrous cycle and influence of social factors in grey short tailed opossums (*Monodelphis domestica*). *J Reprod Fertil.* 1985; 73:337–342. [PubMed: 4039362]
4. Felsenstein, J. *Inferring Phylogenies*. Sinauer Associates, Inc; Sunderland, Massachusetts: 2004. Statistical Properties of Parsimony.
5. Green P. 2_ genomes – does depth matter? *Genome Res.* 2007; 17:1547–1549. [PubMed: 17975171]
6. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003; 52:696–704. [PubMed: 14530136]
7. Hartl, DL.; Clark, AG. *Principles of Population Genetics*. Sinauer Associates, Inc; Sunderland, Massachusetts: 2007.
8. Irwin DM, Arnason U. Cytochrome b gene of marine mammals: phylogeny and evolution. *J Mamm Evol.* 1994; 2:37–55.
9. Janecka JE, Miller W, Pringle TH, Wiens F, Zitzmann A, Helgen KM, Springer MS, Murphy WJ. Molecular and genomic data identify the closest living relative of primates. *Science.* 2007; 318:792–794. [PubMed: 17975064]
10. Kleineidam RG, Pesole G, Breukelman HJ, Beintema JJ, Kastelein RA. Inclusion of cetaceans within the order Artiodactyla based on phylogenetic analysis of pancreatic ribonuclease genes. *J Mol Evol.* 1999; 48:360–368. [PubMed: 10093226]
11. Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* 2006; 4:e91. [PubMed: 16515367]
12. Kullberg M, Nilsson MA, Arnason U, Harley EH, Janke A. Housekeeping genes for phylogenetic analysis of eutherian relationships. *Mol Biol Evol.* 2006; 23:1493–1503. [PubMed: 16751257]
13. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ 3rd, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander EA, Ponting CP, Galibert F, Smith DR, DeJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin CW, Cook A, Cuff J, Daly MJ, DeCaprio D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli KP, Parker HG, Pollinger JP, Searle SM, Sutter NB, Thomas R, Webber C, Baldwin J, Abebe A, Abouelleil A, Aftuck L, Ait-Zahra M, Aldredge T, Allen N, An P, Anderson S, Antoine C, Arachchi H, Aslam A, Ayotte L, Bachantsang P, Barry A, Bayul T, Benamara M, Berlin A, Bessette D, Blitshteyn B, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Brown A, Cahill P, Calixte N, Camarata J, Cheshtsang Y, Chu J, Citroen M, Collymore A, Cooke P, Dawoe T, Daza R, Decktor K, DeGray S, Dhargay N, Dooley K, Dorje P, Dorjee K, Dorris L, Duffey N, Dupes A, Egbiremolen O, Elong

- R, Falk J, Farina A, Faro S, Ferguson D, Ferreira P, Fisher S, FitzGerald M, Foley K, Foley C, Franke A, Friedrich D, Gage D, Garber M, Gearin G, Giannoukos G, Goode T, Goyette A, Graham J, Grandbois E, Gyaltzen K, Hafez N, Hagopian D, Hagos B, Hall J, Healy C, Hegarty R, Honan T, Horn A, Houde N, Hughes L, Hunnicutt L, Husby M, Jester B, Jones C, Kamat A, Kanga B, Kells C, Khazanovich D, Kieu AC, Kisner P, Kumar M, Lance K, Landers T, Lara M, Lee W, Leger JP, Lennon N, Leuper L, LeVine S, Liu J, Liu X, Lokyitsang Y, Lokyitsang T, Lui A, Macdonald J, Major J, Marabella R, Maru K, Matthews C, McDonough S, Mehta T, Meldrim J, Melnikov A, Meneus L, Mihalev A, Mihova T, Miller K, Mittelman R, Mlenga V, Mulrain L, Munson G, Navidi A, Naylor J, Nguyen T, Nguyen N, Nguyen C, Nicol R, Norbu N, Norbu C, Novod N, Nyima T, Olandt P, O'Neill B, O'Neill K, Osman S, Oyono L, Patti C, Perrin D, Phunkhang P, Pierre F, Priest M, Rachupka A, Raghuraman S, Rameau R, Ray V, Raymond C, Rege F, Rise C, Rogers J, Rogov P, Sahalie J, Settippalli S, Sharpe T, Shea T, Sheehan M, Sherpa N, Shi J, Shih D, Sloan J, Smith C, Sparrow T, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Stone S, Sykes S, Tchuinga P, Tenzing P, Tesfaye S, Thoulutsang D, Thoulutsang Y, Topham K, Topping I, Tsamla T, Vassiliev H, Venkataraman V, Vo A, Wangchuk T, Wangdi T, Weiland M, Wilkinson J, Wilson A, Yadav S, Yang S, Yang X, Young G, Yu Q, Zainoun J, Zembek L, Zimmer A, Lander ES. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*. 2005; 438:803–819. [PubMed: 16341006]
14. Liu G, Uddin M, Islam M, Goodman M, Grossman LI, Romero R, Wildman DE. OCPAT: an online codon-preserved alignment tool for evolutionary genomic analysis of protein coding sequences. *Source Code Biol Med*. 2007; 2:5. [PubMed: 17877817]
 15. Mess A, Carter AM. Evolution of the placenta during the early radiation of placental mammals. *Comp Biochem Physiol A Mol Integr Physiol*. 2007; 148:769–779. [PubMed: 17347003]
 16. Montgelard C, Catzeflis FM, Douzery E. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome b and 12S rRNA mitochondrial sequences. *Mol Biol Evol*. 1997; 14:550–559. [PubMed: 9159933]
 17. Mossman, HW. *Vertebrate Fetal Membranes*. Rutgers University Press; New Brunswick: 1987.
 18. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. Molecular phylogenetics and the origins of placental mammals. *Nature*. 2001a; 409:614–618. [PubMed: 11214319]
 19. Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS. Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*. 2001b; 294:2348–2351. [PubMed: 11743200]
 20. Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res*. 2007; 17:413–421. [PubMed: 17322288]
 21. Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*. Oxford University Press; New York: 2000.
 22. Nishihara H, Hasegawa M, Okada N. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc Natl Acad Sci USA*. 2006; 103:9929–9934. [PubMed: 16785431]
 23. Philippe H, Zhou Y, Brinkmann H, Rodrigue N, Delsuc F. Heterotachy and long-branch attraction in phylogenetics. *BMC Evol Biol*. 2005; 5:50. [PubMed: 16209710]
 24. Posada D, Crandall KA. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 1998; 14:817–818. [PubMed: 9918953]
 25. Prasad AB, Allard MW, Green ED. Confirming the Phylogeny of Mammals by Use of Large Comparative Sequence Datasets. *Mol Biol Evol*. 2008; 25:1795–1808. [PubMed: 18453548]
 26. Shoshani J, McKenna MC. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Mol Phylogenet Evol*. 1998; 9:572–584. [PubMed: 9668007]
 27. Springer MS, Burk-Herrick A, Meredith R, Eizirik E, Teeling E, O'Brien SJ, Murphy WJ. The adequacy of morphology for reconstructing the early history of placental mammals. *Syst Biol*. 2007; 56:673–684. [PubMed: 17661234]
 28. Stamatakis A. RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22:2688–2690. [PubMed: 16928733]

29. Strimmer K, Rambaut A. Inferring confidence sets of possibly misspecified gene trees. *Proc Biol Sci.* 2002; 269:137–142. [PubMed: 11798428]
30. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol.* 2007; 24:1596–1599. [PubMed: 17488738]
31. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22:4673–4680. [PubMed: 7984417]
32. Ursing BM, Arnason U. Analyses of mitochondrial genomes strongly support a hippopotamus–whale clade. *Proc Biol Sci.* 1998; 265:2251–2255. [PubMed: 9881471]
33. Wildman DE, Chen C, Erez O, Grossman LI, Goodman M, Romero R. Evolution of the mammalian placenta revealed by phylogenetic analysis. *Proc Natl Acad Sci USA.* 2006; 103:3203–3208. [PubMed: 16492730]
34. Wilson, DE.; Reeder, DM. *Mammal Species of the World: A Taxonomic and Geographic Reference.* Johns Hopkins University Press; Baltimore: 2005.

Figure 1.
Competing hypotheses for the phylogenetic relationships among Cetartiodactyla (cow), Carnivora (dog) and Perissodactyla (horse).

Figure 2. Distribution of the best nucleotide substitution models. ModelTest (Posada and Crandall, 1998) was used to determine the optimal substitution model for each of 2705 1:1 orthologs.

Figure 3.

Maximum likelihood and Bayesian tree derived from the analysis of the conserved coding sequences. Branch lengths indicate Bayesian-inferred substitutions per site with a GTR + Γ model. Maximum likelihood bootstrap proportions are listed above Bayesian posterior probabilities for all branches as follows: ML (GTR + Γ)/ML (GTR + Γ + I)/Bayesian posterior probabilities (GTR + Γ). Opossum was the outgroup used to root the tree.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Phylogenomic data used in this study

Clade	Scientific name	Common name	Coding sequence	Genome coverage	Conserved coding sequence (Gblock)
Primates	<i>Homo sapiens</i>	Human	5,416,500	Finished	4,248,891
	<i>Pongo abelii</i>	Sumatran orangutan	5,210,688	6x	4,248,891
Rodentia	<i>Rattus norvegicus</i>	Brown Norway rat	5,1199,122	11.9x	4,248,891
	<i>Mus musculus</i>	Mouse	5,364,882	Finished	4,248,891
Cetartiodactyla	<i>Bos taurus</i>	Cow	5,167,082	7x	4,248,891
Perissodactyla	<i>Equus caballus</i>	Horse	5,151,942	6.79x	4,248,891
Canivora	<i>Canis lupus familiaris</i>	Dog	5,257,251	7.6x	4,248,891
Marsupialia	<i>Monodelphis domestica</i>	Gray short-tailed opossum	5,157,471	6.8x	4,248,891
Total			41,924,938		33,991,128