Supporting Online Material for

Molecular and Genomic Data Identify the Closest Living Relative of Primates

Jan E. Janečka, Webb Miller, Thomas H. Pringle, Frank Wiens, Annette Zitzmann, Kristofer M. Helgen, Mark S. Springer, William J. Murphy*

*To whom correspondence should be addressed. E-mail: wmurphy@cvm.tamu.edu

This PDF file includes:

Materials and Methods
Figs. S1 to S14
Tables S1 to S5
References

SUPPORTING ONLINE MATERIAL
Janecka *et al*. "Molecular and Genomic Data Identify the Closest Living Relative of Primates."

Materials and Methods,
Screening for Candidate Indels and Analysis.
General Strategy:

We used an approach similar to (*S1*) for discovering indels in protein-coding regions that provide evidence for alternative phylogenetic hypotheses for Euarchonta: 1.) basal Dermoptera (colugos) within Euarchonta, 2.) Primatomorpha (Dermoptera+Primates), and 3.) Sundatheria (colugos and treeshrews as a sister clade to Primates). Specifically, an indel shared by treeshrew and primates, but not with colugo and other eutherians (say, rodents or carnivores), would support a basal position for Dermoptera within Euarchonta. Similarly, an indel shared by colugo and primates relative to treeshrew would support the Primatomorpha hypothesis, while an indel shared by colugos and treeshrews relative to Primates and outgroup species would support Sundatheria. The screen for indels supporting the first case (treeshrew and Primates) would also recover evidence supporting monophyly of Euarchonta, i.e., cases where the colugo has the same fragment size as Primates and treeshrews and different from all other placental mammals. In addition, we performed a search for indels supporting alternative hypotheses (i.e., treeshrews+Glires, or treeshrews basal among Euarchontoglires). The search for potentially informative indels was complicated by the lack of genome sequence data for colugo. Therefore, we used PCR-based methods (described below) to determine whether the interval in colugo either supported Primatomorpha (if the colugo sequence agreed with primates by having the indel) or was non-informative (i.e., agreed with treeshrew and confirmed monophyly of Primates). Similar reasoning was used to screen for the two alternate hypotheses.

Genome Alignments:

We screened a non-redundant set of 197,522 protein coding exons from the UCSC KnownGenes track whole-genome coding sequence alignments for potentially informative indels in primate, treeshrew, and dog and mouse as outgroup genomes, and then performed computational and PCR-based approaches (*S1*) with colugo DNA to evaluate indels and their phylogenetic distribution. During the initial search for candidate indels, the screening was performed with a program available as part of a package called "Phylogenomic Tools", which is freely available at http://www.bx.psu.edu/miller_lab/. First, we used a human-galago (*Otolemur garnetti*) pairwise alignment, a human-treeshrew (*Tupaia belangeri*) pairwise alignment, and a multi-species alignment of 17 vertebrate genomes (available at the UCSC Genome Browser Gateway: http://genome-test.cse.ucsc.edu/index.html). Later we used a 28-way alignment (*S2*) that included the treeshrew sequence, and searched it without reference to the pairwise alignments. In general, gaps in multiple alignments tend to be more consistently placed (e.g., an insertion in human relative to galago and an insertion in human relative to treeshrew occur at precisely the same position in the human sequence) than gaps in independently computed pairwise alignments, because the multi-alignment process forces them to agree. Thus with high-stringency criteria that require agreement among a number of species, such as those we used for our initial search, use of pairwise alignments tends to produce fewer candidates.

Prior to primer design, we prescreened all candidate indels and their flanking sequences in the alignments to identify cross-matched paralogs (arising from incomplete 2X genome assemblies), cases of indels with simple repeats, highly diverged sequences, and examples that exhibited inconsistent phylogeny with current superordinal phylogenies (e.g., *S3*). Whole genome shotgun (WGS), high throughput genomic sequences (HTGS), and genomic survery sequences (GSS) are divisions of GenBank nr nucleotide early assemblies lacking genome browsers at time of article submission. WGS, despite the name consists of small to large contigs assembled from trace reads. The highest status assembly (Table S1) was used initially to recover the orthologous

exon sequence so as not to second-guess mature trace assembly software that optimally exploits read quality data. However for the critical set of species flanking the divergence node in question and for any assembly yielding an anomolous sequence in terms of the overall alignment or indels, each applicable trace match in the archive was further examined for consistent status of the indel, degree of local coverage, supporting evidence for orthology, additional individual animal data, and the possibility of polymorphisms, heterozygosity, lineage-specific duplications, and other trace tiling choices that might not be apparent in the assembly. Data for species such as guinea pig, elephant, and armadillo was split evenly between older assemblies and millions of newly arrived trace reads on 15 June 2007. Other species had singleton traces not included in their assembly; ESTs and cDNA sequences were sometimes available even when genomic data was missing. These were consulted if the exon could be reliably found in the highest status assembly.

Candidate Screening Guidelines/Quality Control
The issue of paralogs must be taken very seriously because data is largely derived from incomplete early assemblies, low coverage trace read compilations, and non-exhaustive cDNA surveys. Without additional quality assurances, best-blast matches alone might pull in paralogs from species where the ortholog was not yet under coverage. These measures are discussed to a certain extent in the text and supplemental material in our recent paper on Atlantogenata-defining indels (*S1*). We are concerned here only with indels of individual exons, so paralogous in this context means extra copies of that exon rather than say its whole PFAM domain (if any) or full-length enveloping gene. These extra exon copies might include internal repeats, partial tandem duplication or debris, segmental duplications, processed retrogenes, and processed pseudogenes in addition to conventional gene family members. With so many species (total branch lengths exceed a billion years for each informative exon) lineage-specific contractions and expansions could reasonably occur. Further, these extra copies might interact confusingly with the actual ortholog via gene conversion or non-orthologous recombination, possibly repeatedly over the time spans involved here. Fortunately, paralogous copies generally replicate the indel status at the time of duplication and so, even if mixed in, do not necessarily affect phylogenetic scoring. While no set of checks on incomplete genomes can be infallibile, we expect that our implementation of the following quality measures greatly reduces the likelihood of paralogous contamination of our multi-species ortholog compilations. Our screening methodology requires sequential use of large-scale comparative genomics (the 28-species alignment), a stringent algorithmic screening protocol, and extensive hand-curation of the resulting (manageable) list of surviving candidates. Each case raises its own narrow curational issues; for reasons of space and lack of generality we do not provide details for each situation encountered on an individual basis.

a. Blat and Blast clustering. For each candidate indel, we first examined the number and quality of protein Blat matches, next at GeneSorter (precomputed Blastp alignments of every gene against the entire human proteome), then at alignment in the 28-species genomic alignment, all features provided by the UCSC genome browser, and finally at tBlastn cross-species matches at the nr, htgs, gss, est and wgs database divisions of GenBank (the last database contains species with 2x or better assemblies). Secondary Blastn matches at the trace archive are also examined when that is the best available data source. Our candidate exons are already filtered by the initial algorithm to require high degree of conservation and exclude compositional simplicity. A numbered gene like *NCOA4* implies previously recognized family members. A high Blast score might result at the level of a full length gene without necessarily giving match quality for the indel-containing exon at issue. Therefore we defined match quality to include score, percent identity, residue similarity, invariant residues, and signature motif conservation. We also required disjoint clustering (i.e. the *NCOA4* exon ortholog from any mammal in the compilation scores better against all other *NCOA4* exons than against any paralogous NCOAx human exon). As a practical matter (since not all sequences reside in any single querable GenBank division), a

compiled database of orthologous and paralogous exons in fasta format was queried at http://www.proweb.org/proweb/Tools/WU-blast.html.

b. Stabilizing effect of established orthologs: Some species, by virtue of stronger assemblies (e.g., rhesus, mouse, dog, opossum) or phylogenetic proximity to human (hence minimal sequence separation, e.g. chimp, orangutan), seed the overall alignment with near-certain orthologs, placing strong expectations on residual intercalated species (e.g. marmoset, rat, cat, elephant). The effect is that inadvertent paralogous sequences are readily recognized as inappropriate. Comparative genomics at the level of a few dozen species is thus somewhat self-correcting. That applies as well to individual aa residue anomalies that can arise from trace read frameshift errors or peak compression. These required additional manual curation effort.

c. Splice position and phase: An internal indel-containing exon has two flanking introns of specific phases, a leading exon and initial methionine and a single downstream intron, a terminal exon a single upstream intron and a stop codon. We required these features to be strictly conserved in all putative orthologs. This excludes processed retropositioned paralogs and older members of some gene families. This test is available for trace, contig, and assembly data but not for (intronless) transcripts where only the methionine and stop codon can be checked.

d. Synteny: The putative orthologs were required to be in homologous chromosome positions; more restrictedly, the left- and right-flanking genes of the gene containing the exon were required to correspond in other species (with ≥6X assemblies) in both position and orientation to those of human. This excludes some rearrangement mechanisms leading to paralogous copies (multi-gene block duplications need best-blast for follow-up). We did not pursue candidates located at extreme telomere and centromere positions in human as these are prone to duplication. The restricted synteny test is implemented by tBlastn of the three gene concatenate against the wgs division of GenBank. This test is not available for trace and transcript-only species and works spottily when intergenic distances are long relative to assembled contig lengths.

PCR-validation:

Prior to PCR-based validation, we excluded candidate indels that had insufficient flanking sequence to design appropriately conserved primers for PCR analysis. Remaining candidate primer pairs were first tested for amplification in colugo (*Cynocephalus volans*) with capillary-based fragment-size analysis (ABI 3730, Foster City, CA), with appropriate control DNAs of known fragment size (e.g., *Tupaia*), with fluorescent-labeled primers and/or direct-sequence-based determination of indel sizes. Those PCR primers that amplified in colugo (and were therefore potentially informative) were also examined in other treeshrews (*Urogale* and *Ptilocercus*) and the second extant colugo (*Galeopterus variegatus*). Exonic regions were then sequenced and aligned to existing coding alignments with CLUSTAL X (*S4*). These sequences have been deposited in GenBANK under accession numbers EU142217-EU142251. Additional eutherian species were added from whole genome assemblies and sequence traces deposited in public databases (Table S1), followed by electronic validation and curation to rule out paralogs and pseudogenes, as described above.

Nuclear DNA Phylogenetic Analysis and Topology Tests.

We compiled a 14,085-bp DNA alignment from 19 nuclear gene segments for 16 euarchontoglires, including 3 rodents (Eastern Chipmunk *Tamias striatus*, American Beaver *Castor canadensis*, South African Springhare *Pedetes capensis*), 2 lagomorphs (Eastern Cottontail Rabbit *Sylvilagus floridanus*, Northern Pika *Ochotona hyperborea*), 2 colugos (Sunda Colugo *Galeopterus variegatus*, Philippine Colugo *Cynocephalus volans*), 6 primates (Ring-Tailed Lemur *Lemur catta*, Gray Mouse Lemur *Microcebus murinus*, Bushbaby *Otolemur garnettii*, Black-Headed Spider Monkey *Ateles fusciceps*, Rhesus Macaque *Macaca mulatta*, Human *Homo sapiens*), and 3 treeshrews (Lesser Treeshrew *Tupaia minor*, Mindanao Treeshrew *Urogale everetti*, Pen-Tailed Treeshrew *Ptilocercus lowii*). The *P. lowii* specimen analyzed in

this study originated from Malaysia (Telok Senangin, Manjung District, Perak, West Malaysia, 4°18'N, 100°34' E; permits UPE 40/200/19 SJ. 179, 1074, 1290 to A. Zitzmann and F. Wiens). The 19 nuclear gene segments included *ADORA3, ADRA2B, ADRB2, APP, ATP7A, BCHE, BDNF, BMI1, CNR1, CREM, EDG1, FBN1, PLCB4, PNOC, RAG1, RAG2,* two segments from *TTN*, and *TYR*. This dataset expands upon smaller published matrices (*S5-S7*). To generate a larger and nearly complete data set that includes additional taxon sampling from all major lineages of Euarchonta, we added five new nuclear gene segments (*ADRA2B*: 1,218-bp, *FBN1*: 736-bp, *BCHE*: 989-bp, and two segments from exon 45 of *TTN*: 1,219-bp and 1,691-bp) with previously described amplification and sequencing protocols (*S5, S8*). A list of all new and published sequences used in our analyses is in Table S2. Primers (Table S3) were designed to include degenerate sites and were used to amplify *ADRA2B* in *U. everetti* and *Cynocephalus volans*, *CREM* in *Ptilocercus lowii*, *PNOC* in *Cynocephalus volans, RAG1* in *Cynocephalus volans,* and *TYR* in *Ptilocercus lowii*. Degenerate primers were used to amplify *FBN1* in *Cynocephalus volans*, *U. everetti*, *Ptilocercus lowii*, and *Microcebus murinus*. All *TTN* segments were amplified exclusively with degenerate primers modified from (*S7*). Missing data was supplemented with whole genome shotgun reads deposited in the trace archive database (Table S1).

Sequences for each gene segment were aligned individually in CLUSTAL X (*S4*) with default parameters and manually corrected to minimize the number of indels. After exclusion of ambiguously aligned regions, the final data matrix included 13,080-bp. The Akaiki Information Criterion in MODELTEST, vers. 3.07 (*S9*) was used to select the most appropriate evolutionary model (general-time-reversible + gamma + invariants, GTR + Γ + I) and parameters. These starting parameters were used to reconstruct a ML tree with a full heuristic search with a neighbor-joining (NJ) starting tree and tree-bisection-reconnection (TBR) branch swapping in PAUP* (*S10*). Model parameters were then reestimated from the ML tree, and incorporated in subsequent iterative ML searches with re-estimation, until the model parameters stabilized. The final parameters were used to reconstruct an ML phylogeny, and perform an ML bootstrap evaluation with 100 heuristic replicates with TBR. The phylogeny of these sequences was also estimated with MR. BAYES version 3.1.2 (*S11*). The character matrix was divided by gene segment, where each gene was assigned its own model of evolution. Model selection was estimated on the basis of MODELTEST results. Two independent runs were performed with 4 independent chains, sampled every 1,000th generation for 5 million generations. The first 1 million generations were discarded as burn-in. We determined convergence between the two runs when the average standard deviation of split-frequencies was less than 0.01. An additional MR.BAYES analysis was performed on the basis of a codon site-specific model for the nuclear protein coding genes, versus the non-coding 3'-untranslated region gene segments, under the same run conditions as the gene-segment partitioned analysis. This analysis produced identical results with the gene-segment based analysis.

Statistical Tests. A statistical test developed by (*S12*) for analysis of SINEs and other indels was applied to our indel data set. Specifically, we found 7, 1, and 0 indels supporting Primatomorpha (colugos+primates), treeshrews+primates, and Sundatheria (treeshrews+colugos), respectively. Under this test , a score of 4, 1, 0, or better, is statistically significant (*P*=0.025). Therefore our results reject the other two hypotheses in favor of the Primatomorpha hypothesis. Shimodaira-Hasegawa (S-H) tests *(S13)* of alternative euarchontan topological hypotheses were performed in PAUP with RELL optimization.

Molecular Dating. Divergence dates were obtained with the programs ESTBRANCHES and MULTIDIVTIME (*S13-14*), assuming the final ML topology (Fig. S12). Domestic cat sequences were included for each gene segment to root the tree. Branch lengths and variance-covariance matrices were estimated under the Felsenstein (*S15*) model in ESTBRANCHES, and these branch

lengths were used to estimate divergence times in MULTIDIVTIME. We used the following 8 fossil-based calibration points: (1) 23 My for the minimum age of the hominoidea/ cercopithecoidea split (*S6*), (2) 36 My for the minimum age and (3) 50 My for the maximum age of Anthropoidea (*S6*), (4) 63 My for the minimum age and (5) 90 for the maximum age of Primates [following *S7*], (6) minimum of 43 My for the base of Scandentia (*S16*), (7) minimum of 55 My for the base of Rodentia (*S16*), and (8) a minimum of 37 My for the base of Lagomorpha (*S16*). The prior for the root was set at 91 My, on the basis of previous molecular estimates for the age of Euarchontoglires (e.g. *S1*). Analyses with a 65 My prior, and those iteratively removing each constraint, showed minimal differences in age estimates (Table S4). The prior is not a constraint and is only used in the initial starting parameters.

Table S1 Taxa examined for indels with species codes used, species names and Orders, and source of data.

Species Code	Species	Common Name	Order	Data Source (Assembly Version), Sequencing Center
homSap	*Homo sapiens*	Human	Primates	Genome sequence (hg18)
panTro	*Pan troglodytes*	Common Chimpanzee	Primates	Genome sequence (panTro2)
gorGor	*Gorilla gorilla*	Western Gorilla	Primates	Trace archive, SC
ponPyg	*Pongo pygmaeus*	Bornean Orangutan	Primates	HTGS, BCM-HGSC, WUGSC
macMul	*Macaca mulatta*	Rhesus Macaque	Primates	Genome sequence (rheMac2)
calJac	*Callithrix jacchus*	Common Marmoset	Primates	WGS, WUGSC
tarSyr	*Tarsius syrrichta*	Philippine Tarsier	Primates	Trace archive, WUGSC
otoGar	*Otolemur garnettii*	Bushbaby	Primates	Genome sequence (otoGar1), BI-MIT
micMur	*Microcebus murinus*	Gray Mouse Lemur	Primates	Trace archive, BI-MIT
cynVol	*Cynocephalus volans*	Philippine Colugo	Dermoptera	Genomic PCR
galVar	*Galeopterus variegatus*	Sunda Colugo	Dermoptera	Genomic PCR
tupBel	*Tupaia belangeri*	Northern Treeshrew	Scandentia	Genome sequence (tupBel1), BI-MIT
uroEve	*Urogale everetti*	Mindanao Treeshrew	Scandentia	Genomic PCR
ptiLow	*Ptilocercus lowii*	Pen-Tailed Treeshrew	Scandentia	Genomic PCR
musMus	*Mus musculus*	House Mouse	Rodentia	Genome sequence (mm8)
ratNor	*Rattus norvegicus*	Brown Rat	Rodentia	Genome sequence (rn4)
speTri	*Spermophilus tridecemlineatus*	Thirteen-Lined Ground Squirrel	Rodentia	WGS, BI-MIT
dipOrd	*Dipodomys ordii*	Ord's Kangaroo Rat	Rodentia	Trace archive, BIM-HGSC
cavPor	*Cavia porcellus*	Domestic Guinea Pig	Rodentia	WGS, BI-MIT
oryCun	*Oryctolagus cuniculus*	European Rabbit	Lagomorpha	Genome sequence (oryCun1), BI-MIT
ochPri	*Ochotona princeps*	American Pika	Lagomorpha	Trace archive, BI-MIT
canFam	*Canis familiaris*	Domestic Dog	Carnivora	Genome sequence (canFam2)
felCat	*Felis catus*	Domestic Cat	Carnivora	Genome sequence (felCat3)
eriEur	*Erinaceus europaeus*	Western European Hedgehog	Eulipotyphla	WGS, BI-MIT
sorAra	*Sorex araneus*	Common Shrew	Eulipotyphla	WGS, BI-MIT
turTru	*Tursiops truncatus*	Bottlenose Dolphin	Cetartiodactyla	Trace archive, BCM-HGSC
susScr	*Sus scrofa*	Domestic Pig	Cetartiodactyla	HTGS/EST database
bosTau	*Bos taurus*	Domestic Cow	Cetartiodactyla	Genome sequence (bosTau2), BCM-HGSC
equCab	*Equus caballus*	Horse	Perissodactyla	Genome sequence (equCab1), BI-MIT
myoLuc	*Myotis lucifugus*	Little Brown Bat	Chiroptera	WGS, BI-MIT
pteVam	*Pteropus vampyrus*	Large Flying Fox	Chiroptera	Trace archive, BCM-HGSC
dasNov	*Dasypus novemcinctus*	Nine-Banded Armadillo	Xenarthra	Genome sequence (dasNov1), BI-MIT
choHof	*Choloepus hoffmanni*	Hoffmann's Two-Toed Sloth	Xenarthra	Trace archive, WUGSC
echTel	*Echinops telfairi*	Lesser Hedgehog Tenrec	Afrosoricida	Genome sequence (echTel1), BI-MIT
loxAfr	*Loxodonta africana*	African Bush Elephent	Proboscidea	Genome sequence (loxAfr1), BI-MIT
proCap	*Procavia capensis*	Rock Hyrax	Hyracoidea	Trace archive, BCM-HGSC

SC = Sanger Center

BC-HGSC = Baylor College of Medicine - Human Genome Sequencing Center

BI-MIT = Broad Institute - Massachusetts Institute of Technology

WUGSC = Washington University Genome Sequencing Center

Table S2 List of taxa included in the 14-kb nuclear gene matrix, with GenBank (NCBI) accession numbers for 19 nuclear gene segments. Some segments were obtained from whole genome reference sequences and trace archives (in italics, Broad Institute – Massachusetts Institute of Technology) in GenBank.

				Accession Number
				ADRA2B
Order	**Family**	**Species**	**Common Name**	**1218 bp**
Glires				
Rodentia	Sciuridae	*Tamias striatus*	Eastern Chipmunk	AJ315942[1]
Rodentia	Castoridae	*Castor canadensis*	American Beaver	AJ427260
Rodentia	Pedetidae	*Pedetes capensis*	South African Springhare	AM407920
Lagomorpha	Leporidae	*Sylvilagus floridanus*	Eastern Cottontail Rabbit	Y15946[2]
Lagomorpha	Ochotonidae	*Ochotona hyperborea*	Northern Pika	AJ427253[3]
Euarchonta				
Dermoptera	Cynocephalidae	*Galeopterus variegatus*	Sunda Colugo	AJ251182
Dermoptera	Cynocephalidae	*Cynocephalus volans*	Philippine Colugo	EU142140
Primate	Lemuridae	*Lemur catta*	Ring-Tailed Lemur	AJ891067
Primate	Cheirogaleoidae	*Microcebus murinus*	Gray Mouse Lemur	AM050858
Primate	Galagidae	*Otolemur garnettii*	Bushbaby	AJ251186[4]
Primate	Atelidae	*Ateles fusciceps*	Black-Headed Spider Monkey	AM050855[5]
Primate	Cercopithecidae	*Macaca mulatta*	Rhesus Macaque	AM050852
Primate	Hominidae	*Homo sapiens*	Human	M34041
Scandentia	Tupaiidae	*Tupaia minor*	Pygmy Treeshrew	AJ251187[6]
Scandentia	Tupaiidae	*Urogale everetti*	Mindanao Treeshrew	EU142141
Scandentia	Ptilocercidae	*Ptilocercus lowii*	Pen-Tailed Treeshrew	EU142142

1. *Sciurus vulgaris* used for *Tamias striatus*

2. *Oryctolagus cuniculus* used for *Sylvilagus floridanus*

3. *Ochotona princeps* used for *Ochotona hyperborea*

4. *Nycticebus coucang* used for *Otolemur garnetti*

5. *Ateles paniscus* used for *Ateles fusciceps*

6. *Tupaia tana* used for *Tupaia minor*

Table S2 continued

Species	ADORA3 324 bp	ADRB2 833 bp	APP 699 bp	ATP7A 681 bp	BCHE 989 bp	BDNF 582 bp
T. striatus	AY011207	AY011267	AY011330	AY011393	EU142149	AY011457
C. canadensis	AY011208	AY011268	AY011331	AY011394	EU142150	AY011458
P. capensis	AY011210	MISSING	AY011333	AY011396	MISSING	AY011460
S. floridanus	AY011221	AY011282	AY011345	AY011408	EU142151	AY011472
O. hyperborea	AY011222	AY011283	AY011346	AY011409	EU142152	AY011473
G. variegatus	AY011223	AY011284	AY011347	AY011410	EU142153	AYO11474
C. volans	AF535031	AF535036	AF535039	AF535043	EU142154	AF535048
L. catta	AY011225	AY011286	AY011349	AY011412	EU142155	AY011476
M. murinus	EU142143	EU142144	EU142146	*1557720828 1551872979*	*1568929540 1582760635*	EU142161
O. garnettii	AF535032	AF535037	AF535040	AF535044	*1119599817 1088989074 1078871794*	AF535049
A. fusciceps	AY011227	AY011287	AY011351	AY011414	EU142156	AY011477
M. mulatta	AY011228	AY011288	AY011352	AY011415	NW001112559	AY011478
H. sapiens	AYO11231	AY011291	AY011354	AY011418	EU142157	AY011481
T. minor	AY011224	AY011285	AY011348	AY011411	EU142158	AY011475
U. everetti	AF535030	AF535035	AF535038	AF535042	EU142159	AF535047
P. lowii	EU213052	EU142145	EU142147	EU142148	EU142160	EU142162

Table S2 continued

Species	Accession Number					
	BMI1 325 bp	**CNR1** 993 bp	**CREM** 445 bp	**EDG1** 978 bp	**FBN1** 736 bp	**PLCB4** 401 bp
T. striatus	AY011516	AY011575	AY011639	AY011702	EU142170	AY011762
C. canadensis	AYO11517	AY011576	AY011640	AY011703	EU142171	AY011763
P. capensis	AY011519	AY011578	AY011642	AY011705	EU142172	AY011765
S. floridanus	AY011529	AY011591	*639274633[C]*	AY011716	EU142173	AY011778
O. hyperborea	AY011530	AY011592	AY011655	AY011717	EU142174	AY011779
G. variegatus	AY011531	AY011593	AY011656	AY011718	EU142175	AY011780
C. volans	AF535052	MISSING	AF535058	AF535062	EU213053	AF535065
L. catta	AY011533	AY011595	AY011658	AY011720	EU142176	AY011782
M. murinus	EU142163	EU142165	*1546199914*	EU142169	*1526582216 1558942136*	EU142183
O. garnettii	AF535053	AF535056	AF535059	AF535063	EU142177	AF535066
A. fusciceps	AY011535	AY011597	AY011660	AY011721	EU142178	AY011784
M. mulatta	AY011536	AY011598	AY011661	AY011722	EU142179	AY011785
H. sapiens	AY011539	AY011601	AY011664	AY011725	EU142180	AY011788
T. minor	AY011532	AY011594	AY011657	AY011719	EU142181	AY011781
U. everetti	AF535051	EU142166	AF535057	AF535061	EU213054	AF535064
P. lowii	EU142164	EU142167	EU142168	MISSING	EU142182	EU142184

1. *Oryctolagus cuniculus* used for *S. floridanus*

Table S2 continued

Species	Accession Number					
	PNOC **327 bp**	**RAG1** **774 bp**	**RAG2** **444 bp**	**TTN2&3** **1219 bp**	**TTN6&7** **1691 bp**	**TYR** **426 bp**
T. striatus	AY011821	AY011879	AY011936	EU142191	EU142206	AY011998
C. canadensis	AY011822	AY011880	AY011937	EU142192	EU142207	AY011999
P. capensis	AY011824	AY011882	AY011939	EU142193	EU142208	AY012000
S. floridanus	AY011835	AY011895	AY011952	EU142194	EU142209	AY012010
O. hyperborea	AY011836	AY011896	AY011953	EU142195	EU142210	AY012011
G. variegatus	AY011837	AY011897	AY011954	EU142196	EU213056	AY012012
C. volans	AF535070	EU142186	AF535076	EU142197	EU142211	AF535081
L. catta	AY011839	AY01898	AY011956	EU142198	EU142212	AY012014
M. murinus	*1576058496* *1550765351*	EU142187	EU142189	EU142199	EU213057	EU142215
O. garnettii	AF535071	AF535074	AF535077	EU142200	*1094614950* *1097653575* *1077056321* *1090439925* *1108493862*	AF535082
A. fusciceps	AY011841	AY011899	AY011958	EU213055	EU213058	AY012015
M. mulatta	AY011842	AY011900	AY011959	EU142201	NW001098160	AY012016
H. sapiens	AY011845	AY011903	AY011962	EU142202	NT005403	AY02019
T. minor	AY011838	MISSING	AY011955	EU142203	EU142213	AY012013
U. everetti	AF535069	MISSING	AF535075	EU142204	EU213059	AF535080
P. lowii	EU142185	EU142188	EU142190	EU142205	EU142214	EU142216

Table S3: Nuclear primer sequences used to amplify gene segments from the 14-kb nuclear gene matrix

ADRA2B, 1,219 bp:
degenerate F-5'AGCCRATCCAGAAGAAGAACT
degenerate R-5'TCACBTTYCTCATCCTYTTCACC
CREM, 445 bp:
degenerate F-5'AGGAACTCAARGCCCTCAAA
degenerate R-5'GGGAGARCAAATGTCTTTCAA
PNOC, 327 bp:
degenerate F-5'GCATCMTDGASTGTGAAGRGMA
degenerate R-5'TGCCTCATAAAYTCACTGAAYC
RAG1, 774 bp:
degenerate F-5'AAGACMTCYTGGAAGGCRTGA
degenerate R-5'RAAGTTGCCRTTCATYCTCA
TYR, 426 bp:
degenerate F-5'TGTGGCCRGCTHTCAGGSAG
degenerate R-5'CTTCATGRGCAAARTCAATGT
BCHE, 989 bp:
F–5'TCAGAGATGTGGAACCCAAA
R–5'ATGCATCACTCCCATCCATT
FBN1, 736 bp:
F–5'AGACTACCTCAGTGGTGAACTGG
R–5'AAGCACCATTACAAACCCTCA
degenerate F-5'AGACTACCTCAGTGGTGAAYTGG
degenerate R-5'AAGCACCATTAYAAACCCTCA
*TTN*2&3, 1,219 bp:
F2–5'CACCTCTCTTGTYCTTGAHAATG
R2–5'CCTYTTGGAGGATCAGGYTTATC
F3–5'GGATGAYGTCACCAGAAACAGTG
R3–5'RCCTGGYTCTTTGTAGGGATATT
TTN6&7, 1,691 bp:
F6–5'TGTGATCCTGTVTTCAAACCT
R6–5'GCATTRCARACTTTGGATTCAGC
F7–5'TTCMCCACCAGGRAAGGT
R7–5'TGGYCCAGGYTCTTTAAATGGAT

Table S4. Divergence dates estimated in MULTIDIVTIME (*S13-14*) from a 14-kb concatenated 19 nuclear gene segment sequence matrix with the relaxed clock method. The 95% credibility intervals are given in parentheses. The table shows divergence times estimated with 91 Mya and 65 Mya priors and with each constraint removed under a 91 Mya prior. The node numbers refer to those identified in Fig. S1. Divergence times estimated in (*S17*) are based on a 58.9 kb genomic alignment that is independent from our data set.

Node	Steiper & Young 2006		Prior				Constraints (91 Mya prior)					
			91 Mya		65 Mya		no Scandentia minimum		no Rodentia minimum		no Cercopithecoidea minumum	
16: *Tupaia/Urogale*			8.3	(5.4-12.8)	8.2	(5.2-12.8)	8.3	(5.3-12.8)	8.4	(5.3-12.7)	8.2	(5.3-12.7)
17: Scandentia			63.4	(51.9-74.1)	62.2	(51.0-73.6)	63.4	(51.8-74.4)	63.5	(51.9-74.2)	63.0	(51.6-73.8)
18: Hominoidea/Cercopithecoidea	30.5	(26.9-36.4)	26.8	(23.2-32.7)	26.7	(23.2-32.9)	26.8	(23.2-32.8)	26.8	(23.2-32.9)	26.1	(20.9-32.6)
19: Anthropoidea	42.9	(37.3-52.4)	41.7	(36.4-48.6)	41.3	(36.3-48.5)	41.7	(36.5-48.7)	41.7	(36.4-48.6)	41.2	(36.2-48.5)
20: *Lemur/Microcebus*	40.9	(35.3-51.0)	40.4	(31.7-49.7)	39.8	(31.4-49.0)	40.5	(31.9-49.5)	40.5	(31.9-49.6)	40.1	(31.6-49.0)
21: Strepsirrhini	57.1	(49.4-71.4)	62.1	(51.8-71.5)	60.9	(51.1-70.9)	62.1	(51.8-71.4)	62.0	(51.9-71.4)	61.6	(51.6-71.2)
22: Primates	77.5	(67.1-97.7)	79.6	(66.5-89.5)	77.8	(65.2-89.2)	79.6	(66.6-89.4)	79.6	(66.5-89.4)	79.2	(66.5-89.4)
23: *Cynocephalus/Galeopterus*			19.8	(14.0-27.2)	19.9	(14.0-27.5)	19.9	(14.2-27.2)	19.9	(14.0-27.0)	19.6	(14.0-27.0)
24: Primatomorpha			86.2	(71.3-97.9)	84.1	(70.0-97.0)	86.2	(71.4-97.6)	86.2	(71.6-97.7)	85.8	(71.4-97.7)
25: Euarchonta			87.9	(72.6-99.9)	85.7	(71.1-98.9)	87.8	(72.7-99.7)	87.9	(73.0-99.7)	87.4	(72.7-99.7)
26: Lagomorpha			50.9	(40.8-60.8)	50.0	(40.2-60.1)	50.9	(41.0-60.5)	50.9	(40.9-60.7)	50.6	(40.8-60.5)
27: Castoridae/Pedetidae			64.6	(52.6-75.5)	63.2	(51.6-74.9)	64.6	(52.6-75.5)	64.6	(52.6-75.6)	64.2	(52.4-75.3)
28: Rodentia			73.3	(60.4-84.4)	71.7	(59.3-83.8)	73.3	(60.4-84.4)	73.3	(60.6-84.5)	72.9	(60.4-84.3)
29: Glires			85.7	(70.7-97.8)	83.5	(69.2-96.8)	85.6	(70.8-97.8)	85.7	(71.0-97.9)	85.2	(70.8-97.5)
30: Euarchontoglires			88.8	(73.2-101.0)	86.5	(71.8-100.0)	88.7	(73.3-100.9)	88.7	(73.6-100.9)	88.3	(73.3-100.8)

Node	Constraints (91 Mya prior)									
	no Anthropoidea minimum		no Anthropoidea maximum		no Lagomorpha minimum		no Primate minimum		no Primate maximum	
16: *Tupaia/Urogale*	8.2	(5.3-12.6)	8.4	(5.4-13.0)	8.3	(5.3-12.9)	8.3	(5.3-12.8)	8.8	(5.5-13.9)
17: Scandentia	62.8	(51.1-73.9)	63.6	(51.7-74.4)	63.4	(51.8-74.0)	63.3	(51.6-74.0)	67.5	(52.6-85.6)
18: Hominoidea/Cercopithecoidea	26.6	(23.1-32.6)	27.1	(23.2-34.1)	26.9	(23.2-33.0)	26.8	(23.2-32.8)	27.5	(23.3-33.4)
19: Anthropoidea	41.3	(34.8-48.4)	42.0	(36.5-50.5)	41.7	(36.4-48.8)	41.7	(36.4-48.7)	43.2	(36.6-49.6)
20: *Lemur/Microcebus*	40.1	(31.1-49.3)	40.7	(31.9-50.1)	40.4	(31.7-49.4)	40.4	(31.7-49.7)	42.8	(32.6-54.9)
21: Strepsirrhini	61.5	(50.7-71.4)	62.3	(51.9-72.0)	62.0	(51.6-71.5)	62.0	(51.6-71.5)	65.8	(52.5-81.2)
22: Primates	79.0	(65.6-89.4)	79.9	(66.6-89.6)	79.6	(66.5-89.5)	79.5	(66.0-89.4)	84.7	(67.2-105.3)
23: *Cynocephalus/Galeopterus*	19.6	(13.8-26.9)	20.0	(14.1-27.4)	19.9	(14.1-27.2)	19.9	(14.0-27.4)	21.0	(14.5-29.4)
24: Primatomorpha	85.5	(70.6-97.6)	86.4	(71.5-97.9)	86.2	(71.4-97.7)	86.1	(71.0-97.8)	91.8	(72.3-115.1)
25: Euarchonta	87.2	(72.0-99.7)	88.1	(72.7-100.0)	87.8	(72.7-99.9)	87.7	(72.2-99.7)	93.6	(73.4-117.7)
26: Lagomorpha	50.5	(40.3-60.3)	51.0	(40.8-60.8)	50.9	(40.7-60.7)	50.9	(40.7-60.5)	54.2	(41.6-69.5)
27: Castoridae/Pedetidae	64.0	(51.9-75.2)	64.8	(52.7-76.0)	64.6	(52.6-75.5)	64.5	(52.4-75.5)	68.7	(53.2-87.1)
28: Rodentia	72.7	(59.7-84.4)	73.6	(60.4-84.8)	73.3	(60.4-84.5)	73.2	(60.2-84.2)	78.1	(61.2-98.4)
29: Glires	85.0	(70.0-97.6)	85.9	(70.9-97.9)	85.6	(70.9-97.8)	85.5	(70.4-97.7)	91.2	(71.6-114.8)
30: Euarchontoglires	88.0	(72.7-100.8)	88.9	(73.5-101.1)	88.7	(73.4-101.0)	88.6	(72.9-100.8)	94.5	(74.1-119.0)

Table S5 Primers used to amplify and sequence indels informative within Euarchonta. Information provided for indel primers includes the gene in which the indel is present, the human chromosome that the gene occurs on, and the approximate location within the March 2006 genome assembly.

Gene	Position on Human Genome Flanking the Indel	Number of Placental Taxa Examined	Forward Primer	Reverse Primer
N4BP2	chr4 near 39,780,754	33	TCTCTTCTCAGYAGTTCTTYAAA	GCTTGAAAYTGAACCCTTCG
ZNF12	chr7 near 6,698,680	33	TGCATAGCTTCCATCRCTACT	GAGAGAGGTAATGTTCCTRGTAAAA
CDCA5	chr11 near 64,603,721	31	GCCTGACYTTCTTAGACATTTCC	TTTYTTGGARAAAGAAAACAACC
SPBC25	chr2 near 169,454,207	35	GGCATAAATGAATTTTGGAATAA	CCTGCAAATGYTTTRATGGA
SMPD3	chr16 near 66,953,100	33	AGGGTCTTGGAGAGTGAGGA	GCGTRCAGCATGTAGTCGATG
MTUS1	chr8 near 17,557,778	32	GTCACTGCTTCAACCACCTG	TGCGTTTTATACTTCTCWGCTTCTT
SH3RF2	chr5 near 145,373,602	32	CCAAACCTCACTGCAAGACA	CTGGGGTGCTGATTTCTACC
NCOA4	chr10 near 51,254,904	35	ATGGATCTTTCTGATTGG	CATTCAGGCACTTCAGATTGC
TEX2	chr17 near 59,624,835	32	TCCTCTTTTTCTCGGCCAGT	AAGGCTCAGACTGATAAGGAGA
SSH2	chr17 near 24,982,804	36	GATTTCTTGAGTGCGCTCCT	CAAGTGCTACCTCTGCCTCA
ADD2	chr2 near 70,753,563	32	AAGACTTGGAAGGRGAVACC	AAACCGCCCCAGAAGAAC

Figure S1-S11 Indels found to be informative for *a priori* Euarchonta and Primatomorpha hypotheses. Indels supporting Euarchonta are presented first (Figs. S1-S3) with the euarchontan taxa in blue. Indels supporting Primatomorpha follow (Figs. S4-S10), with primatomorphan taxa in green. The indel supporting monophyly of treeshrews+Primates (in orange) is shown in Fig. S11. Taxon designations follow a 6-letter nomenclature: the first 3 letters of the genus followed by the first three letters of the species (Table S1). Sequences with truncated ends were generated by PCR with conserved primers that were placed in the exon flanking the indel. HTGS=high throughput genome sequence. WGS=contig assembly from whole genome shotgun traces

Figure S12. Maximum likelihood tree (-ln L = 56234.7) reconstructed in PAUP* (*S10*) with the final (following removal of ambiguous sites) 13.1-kb concatenated sequence alignment, with a GTR + Γ + I model with the following parameters: Base frequencies, 0.2651, 0.2474, 0.2341, 0.2534; rate matrix, 1.4031, 5.3779, 0.8426, 1.4209, 6.5144; gamma shape parameter, 0.8675; proportion of invariant sites, 0.3549. Numbers used to identify nodes correspond to those in Table S4.

Figure S13. Maximum likelihood bootstrap tree (100 replicates) reconstructed in PAUP* (*S10*) with the 13.1-kb concatenated sequence alignment with *Ptilocercus* removed from the analysis. The maximum likelihood search was conducted under a GTR + Γ + I model with the following parameters: Base frequencies, 0.2668, 0.2422, 0.2343, 0.2567; rate matrix, 1.4071, 5.3038, 0.8410, 1.5736, 6.8836; gamma shape parameter, 0.7440; proportion of invariant sites, 0.3500.

Figure S14. Phylogeny of extinct and extant Euarchonta on the basis of the morphological character matrix of (*S18*). Shown is the strict consensus tree of 3 equally parsimonious trees, with all but four characters unordered (following *S18*), constrained to fit the Primatomorpha hypothesis (Scandentia(Dermoptera+Primates)). Bootstrap values >50% are shown above each branch. These results are consistent with a similar analysis reported in the Supplementary Information of (*S18*).

Figs. S1-S3: *Indels Supporting Euarchonta* **(euarchontan taxa in blue)**

Fig. S1. Gene *N4BP2*

```
homSap KELLESECVEAQFSEAPVDLDASEPQACLNLP----GLDLPGTGGDQKSTRVSDVFLP
panTro KELLESECVEAQFSEAPVDLDASEPQACLNLP----GLDLPGTGGDQKSTRVSDVFLP
ponPyg KELLESECVEAQFSEAPVDLDASEPQACLNLP----GLDLPGTGGDQKSTRVSDVFLP
macMul KELLESECVEAQFSEAPVDLDASEPQASLNLP----GLDLPGTGGDQKSTRVSDVFLP
calJac KELLDSECVEALFSRAPVDLDASEPQASLNPP----GLDLPGMGGDQKSTPVSDMFLP
tarSyr KEPLESECAKAQFSQSPVDLDAHEPQASSNLP----EFDSPDSEGDQKSASVSDEFVP
otoGar KEVLESECIAVQLPQASVNLDAREPQPPSSFP----GLDLPSTGGDQKSS-VSDVFVP
micMur KELLESECAEAQFPQASVHLDASEPQAPLSLP----GLDLPGTGGDQKSTFVSDVFVP
cynVol KELLESECVEAQFSQAPADLDASEPQALLNLP----ELDLPDTGGDQKSTSVSDVFVP
tupBel .ELLEPERTEAQSSQGPVDLGASEPQAPSNLP----GLDSAGTGGDRGSPAAADVSVP
ptiLow KELLESECIEAQFPLAPIDLDASEPQAPLNLP----GLDLPGTGGDQD-SSVSDVFVP
musMus KDLSESECPSAQHSQALADLGNSDPQAPSTHPLHNSGSDLPGTDGDQKSASAPDVFVP
ratNor KDHLESECPSVQHSQALVSVGNSEPQVPSNRSLHNSGSDLPSTDGDQKSTSAPDVFVP
dipOrd KEGLEAECVDNQLSQPSVDMEANEPQAPSTLPVQNLGLDLPGTSGDQKSTSASDVFVP
speTri .ELLESEHVEAQLSEAPVNLNANEPQASLNLPVQNSGCDLPDTGGNQRSTSVSDIFVP
cavPor KVLLDSNCVETQLSPAPVDLDANVPQGPLNLTVQNPELDLAGTGEGQKSVSISGVFVP
oryCun .ELLESECVETQFCQTLLLESDANELKAPLNLPVQNPGLDIPGTGGDQKPTSVSDVFV.
ochPri NELLQSEC-ETQFSQNPEKFDASEPQPTLNLPVQNSGPNIPGTGGDQKSTPVSDVF..
canFam KELLESECVETPFSQAPVDLDASEPQAPLSLTAQNLGLDLLGTSGDQKSASVPDAFVP
felCat KELLETEYVETPFSQAPVDLDDSEPQA-LSLTVQNLGLDYlVQGGDQKSASVPDMFVP
equCab KELLESECLQAQFSQAPVDLDVSESQAPLNLTVENPGLGLLGRGENQKSTSVPDMFVP
myoLuc .DLLESECAETQFSQAPVDLDANEPQAPSNSTVQNPGLGLLGKGGNQKSTSIPDVFVP
pteVam NELLESECVEAQFSETPVDLDANEPQAPLN--VQNPGLGLLGIGGDQKSTSIPDVFVP
turTru KELLESVCVETQFSQAPVDLDANELQAPLNLTVQNLGLGLLGTDGDQKS-SVPDVFVP
bosTau KELSESECVKTQFSQAPVDLDANELQAPLNLAVQNPGIGLIGTGGDQKSTSVPDVFVP
susScr RELLESECVETQFSQAPVDSDASELQAPLNLTVQNPGLGLLGTGGDQKSTSVSDVFVP
sorAra KECSESESAEAQFSQDPGHLDIHEPQA-LNLPVQNPGVGLQDAGGDQSSTSASDVFVP
eriEur .ELVESECVEAQFSEDPVDLDANDSS--LNLAVQNSGLDLLGAGSDQIPTSVPDVFVP
dasNov .ELLESECIEAQLFQVPVDMDTNELQAPLNLSAQNPGLDLLDIDGDQKTASASDVFVP
choHof KELLESECTEAQFFQVPVDLDASEPQSPLNLSAQNPGLDLLAADGDQKSISVPDVFVP
loxAfr EELLESECVQTQFSQAPVDLDANEPQPPLNLSVQNPELDLLVTDGDQKSTPVPNVLVP
proCap EDLAESECVETHFSEAPVDLDASEPQIPLNIFVENPELDSVHTFGDQRCTSASDVLAP
echTel ILKKRCEYVETQFSQALVDLDANEFQVPlNLSPQVPDLGLLGKGGDQKSTSAPGMFVP
```

Fig. S2 Gene *ZNF12*

```
homSap NVPGKTFDVETNPVPSRKIAYKNSLCDSCEKCLTSVS
panTro NVPGKTFDVETNPVPSRKIAYKNSLCDSCEKCLTSVS
ponPyg NVPGKTFDVETNPVPSRKIAYKNSLCDSCEKCLTSVS
macMul NVPRKTFDVETNAVPSGKTAYRNSLCDSCEKCLASVS
calJac NVPGKTFDVETNPVPSRKTSYKNSLCDSCEKCLTSVS
tarSyr NVPGKPFTVETNLDPSRKIAHKNSLCDSCEKSLTSVS
otoGar NVPSKIFNVEMNSIPLRKTAYKNSLCGSCEKSLTSIS
micMur NVPGKIFNVETNPVSSRKIAYKNSLCDSCEKSLTSIS
galVar ..........VNPVPSRKIAYKSSLCDSCEKSLTSIS
cynVol ......FKVEVNPVPSRKIACKSSLCNSCEKSLTSIS
uroEve ......SNVEMNLVPSRKIAYQNSLCDSCEKTLTSVS
tupBel NVPGKTSNVETNSVPSRKIAFQNSLCDSCEKTLTSVS
ptiLow ......FNVEMNPVPSRKIAYQNSLCDSCGKNLTSVS
musMus DASSKATDGETKPFPSQKALPQ---CNSCEKSLMCVS
ratNor DASRMAAEVAKKLAPSGKVLPK---CSSCEESLMCVS
dipOrd SVSGKAFSVEKNLVP-RKIASK---CDLCEKSLACVP
speTri NVSGKAFNVEINPVPSRKVASK---CDLCEKSLTSVS
cavPor NVSEKAFSVEMSPVASRKISSR---CDACEKSLTSVS
oryCun NGLGKTLNLDTNPVSSRKTPCQ---YDSSGMNLKYIS
canFam NVLGKTFNVETNPIPSRKMSYK---CDSCEKSLKSVS
felCat NVLGKTLNVETNPIPSRKISYK---CGSCEKSLKSIS
equCab HVPGKTFNVETNPVPSRKMSYK---CDSCGKSLKSVS
myoLuc DVLGKTFNVETNPVPSRKTSYK---CDSCEKSLKSTS
pteVam NVLSKTFNVETNPVLSRKISYK---CDSCEKSVKSI.
turTru NVFGKTFNVETNPVPSRKISCK---CDSCGKSLKSTS
bosTau NGLGKTCNVETSPVSSRKISYQ---CDSCEKSLKSIS
sorAra NVPSKTFSIETNSAPSRKMPCK---CNACGKSCKPAS
eriEur NDLGKTFHVESNPVPSNTVSYK---CGSCEKNVKAAS
dasNov NALGKTFNVETNPVPSRQIFSK---CDSCEKSLKSIS
choHof NALGKTFEVETNPVSPRKISSK---CDSCGTSLKSIS
loxAfr NILDKTLNVESSPDLSRKISYK---CDSCEKNLKSIS
proCap NVPDKTFNVETSPDPSRNISYT---CDSCEKSLKSIS
echTel NVLGKTFSVETSPDSSGKILHK---CDSCEKSLASIS
```

Fig. S3. Gene *CDCA5*

```
homSap SFFLEKENEPPGRELTKEDLFK---THSVPATPTSTPVPN-PEAESSSKEG-ELDARDLEMSKK
panTro SFFLEKENEPPGRELTKEDLFK---THSVPATPTSTPVPN-PEAESSSKEG-ELDTRDLEMSKK
gorGor SFFLEKENEPPGRELTKEDLFK---THSVPATPTSTPVPN-PEAESSSKEG-ELDARDLEMSKK
macMul AFFLEKENEPPGRELTKEDLFK---THIIPATPTTTPVPN-PDAESSSTEG-ELDARDLEMSKK
calJac AFFLEKENKPPCRELTKEDLSK---IHSVPDTPNTSPVLN-PEAESSSKER-ELDARDLEMFKK
tarSyr AFYLEKENNPPIREPTKEDLLK---TCDVPATPASTPVHI-LNAKSSSKDG-ELDARDLEMSKK
otoGar SFLLEKENNPPNKEPTKEECFKIHPSPVTPATPTTTPVQN-LNSDFDSGEG-DLDARDLEMSKK
micMur SFFLEKENNPPNREPTKEDLFK---IHSVPGTPATTPVHS-LSAESNSSEG-SLDARDVEMSKK
galVar ..........PRKELTKEDLFK---TCSIPVTPATTPVHN-LNIKSSSKEG-DPDARD......
cynVol ..........PRKEPTKEDLFK---TCSIPVTPATTPVHN-LNIKSNSKEG-DPDARD......
uroEve ..........PSSEPTKEDFFK---ICSVPVTPTTTPVFT-QNTEPNSSEG-DLDARD......
tupBel AFSLEKENNPPSSEPTKEDFFK---ICSVPVTPTTTPVFT-QNTEPNSSEG-DLDARDLEMSKK
musMus SFILEKENNPPLKVPTKEDLFK---TCSVPGTPSSTPVLYTQNVEPDSGEA-ELDSRDLEMSQK
ratNor SFILEKENNPPLKVPTKEDLFK---TCSVPGTPSSTPVLCAQNVKSNSREV-EPDTRDLEMSQK
dipOrd AFFLEKENNPPCQETTKEDLFN---PVPM--TPTTTPVLTTLNVIRSVHQQ-EVDARDMEMSKK
cavPor TFFLGKENNPPLQKPTKRDLFK---TPSVPVTPTTTPVLDPPEVESCSKEE-ELDIRTVEMSRK
oryCun AFFLEKENNPPKREPAREELFK---SCSVPGTPATTPVLCSLNTSSNAKEG-QLDARDLEMSQK
ochPri AFLLEKENNPPQREPAREDLFK---SCSVPTTPTTTPVLCPPSAKSTSRDG-ELDARDLEMSQK
canFam AFFLEKENNPPTKEPAREDLFK---SCSIPVTPTTTPVLYPVNVESNSREE-DLDARDLEMSKK
felCat LFFLE-ENKPSSKKPAREDLFK---TCSIPGTPTATPVLYPLNVESNSRKE-DLDARNLEMSNK
equCab AFFLEKENHLPSKEPTKEDLFK---TCSVPSTPTTTPVLQPLNVESNSGE--DLDTRDLEMSKK
myoLuc AFSLEKENNPPTKEPAKEDLFK---QCSVPA----TPVQSPD--ESNATEG-DLDTRDLEMSKK
turTru ..FLEKENNPPSKEPTREDIFP---TRSVRVTSASTPALCSLNVESDSREG-DLDARDLEMSKK
bosTau ...LEKENNPPSKKLTKEDLFH---TCSVPVTAPSTPVLCPVNAESNSWKG-DLDARDLEMSKK
susScr AFFLEKENNPPSKEPTREDLFQ---TRSVPVTPTPTPVLCSLNVESNSGEG-DLDARDLEMSKK
sorAra ...LEKENCPPGPEPTWEDLFK---PCEAPGTPASAPLVPPLSADPATSAS-ALDLRDLEMSKK
eriEur AFFLEKENKPPYKELTRGDLFK---TRDVPVTPASSPVPNAPSSADPNSGYREPDARDLEMSRK
dasNov AFFLEKGNNPPRKETTKEDLFQ---TSSVPVTTATTPVLYPLEVKSNSKKG-HPDPRDLEMSTK
choHof AFFLEKENNPPIKEPTKEDLFK---TCSVPITPATTPVLNSLKVESSSKEG-HLDARDLEMSKK
loxAfr AFLLEKENKPPSKELTKEDLFK---TCSTPVTPSTTPVLLPLTVESSPRSG-DLDTRDLEMSKK
proCap AFLLEKENNPPSKELTKADLFK---THPTPVTPASSPVLVPLNIESRSRSG-DLDPRDLEMSKK
```

Figs. S4-S10 Indels Supporting Primatomorpha (primatomorphan taxa in green)

Fig. S4 Gene *SPBC25*

```
homSap MVEDELALFDKSINEFWNKFKST--DTSCQMAGLRDTYKDSIKAFA
panTro MVEDELALFDKSINEFWNKFKST--DTSCQMAGLRDTYKDSIKAFA
ponPyg MVEDELALFDKSINEFWNKFKST--DTSCQMAGLRDTYKDSIKAFA
macMul MVEDELALFDKSINEFWNKFKST--DTSCQMAGLRDTYKDSIKAFA
calJac MVEDELALFDKSLNEFWNKFKST--DTTFQMAGLRDTYKDSLKAFA
tarSyr MVEDELTLFDKSINEFWNKFKST--DTANQMMGLRDTYKDSVKAFA
otoGar MVEDQLALLDKNINEFWNKFKST--DTAGQMAGLRDTYKDSIKTFA
micMur MVEDELVLFDKTVNEFWNKFKST--DTSCHMVGLRDTYKDSLKAFA
galVar ...................FTST--DTDCQMMGLR............
cynVol ...................FTST--DTSCQMMGLRGTNK.......
tupBel MVEDELALFDKGINEFWNKFRSTVSDTSCQMVGLRDAYKDSIKAFA
uroEve ...................FRSTVSDTSCQMVGLRDAYK.......
musMus MGEDELALLNQSINEFGDKFRNRLDDNHSQVLGLRDAFKDSMKAFS
ratNor MGEDELAAFEKSINEFGDKFRYRLSDNRSQVLGLKDAFKDSIRALS
cavPor MVEDELALFDKSINEFGNKFRNTLSDTPCQMLGLRDACKDSIKTLA
speTri MMEDELARFDKSINEFGNKFRNTFSDTRCQMVGLRDVFKDSIEALA
dipOrd MVEDELAHFDKSISEFGSKFRNTLSDTPSQTVGLRDAYKDSIKALS
oryCun MVEDELALFDKSINEFGSKFRSTLSDAPCQMVGLRDAYKDSVKSLT
ochPri MVEDELALFDKSINEFGSKFRSTLSDTPCQMVGLREACKDSVRLLT
canFam MIDDELAQFDKSISEFWSKFKGTVSDTSSQMVGLRETYKDSIKACA
felCat MIEDELALFDKSINEFWNKFKSTLSDTSCQMMGLRDTYKDSIKALT
equCab MVEDELALFDKSINEFWNKFKNTVSDTSCQMVGLRDAYKDSIKAFA
myoLuc MVEDELALLDKNINEFWNKFKSNVNDTSCQMVGLRDNYKDISKAFT
pteVam MVEDELALLDKSINEFWNKFKSSVSDTSCQMMALRDSYKDINKAFT
bosTau MVEDELALFDKSINEFWNKFKSTVSDTSCQMVGLRETYKDSIKAFA
turTru MVEDELALFDKSINEFWNKFRSTVSDTSCQMVGLRDTYKDSIKAFA
susScr MVEDELALFDKSINEFWNRFKSTVSDTSCQMVGLRENYKDSLKAFA
oviAri MVEDELALFDKSLNEFWNKFKSTVNDTSCQMVGLREAYKDSIKAFA
eriEur MVEDELALFDKSINEFWNKFKGTVSDTSFQMVGLRDTYKDSIKIFT
sorAra MVEDELVLFEKSINEFVNEFESTASDTTCQVVGPRDADKDSIKALA
dasNov MIEDELALFDKSINEFWNKFKGTVSDNSCQMVGLRDTYKDSIKAFA
choHof MIEDELALFDKSINEFWNKFKSAVSDTSCQMVGLRDTYKDSIKAFA
loxAfr MIEDELVQFDKSINEFWNKFINTASDTSCQMVGLRDAYKDSMKAFA
proCap MIEDELRQFDKSINEFWNKFINTTSDTSCQMAGLRDAYKDSMKAFA
echTel MIEDELLQFDKSMNEFRNKHFNTLNDTSGQMMGLRDTYRDSMKAFA
```

Fig. S5. Gene *SMPD3*

```
homSap  VLESEEGRREYLAFPTSKSSG--QKGRKELLKGNGRRIDYMLHAEEGLCPDWKA
panTro  VLESEEGRREYLAFPTSKSSG--QKGRKELLKGNGRRIDYMLHAEEGLCPDWKA
ponPyg  VLESEEGRREYLAFPTSKSSG--QKGRKELLKGNGRRIDYMLHAEEGLCPDCNA
macMul  VLESEEGRREYLAFPTSKSSG--QKGRKELLKGNGRRIDYMLHAEEGLCPDWKA
calJac  VLESEEGRREYLAFPTSKSSG--QKGRKELLKGNGRRIDYMLHAEEGLCPDWKA
otoGar  VLESEEGRKEYLAFPTSKSPG--QKGRKDLLKGNGRRIDYMLHAEEGLCPDWKA
micMur  VLESEEGRREYLAFPTSKSPG--QKGRKDLLKGNGRRIDYMLHAEEGLCPDWKA
galVar  ......GRREYLAFPTSKSPG--QKGRKDLDKGNGR.................
cynVol  ......GRREYLAFPTSKSPG--QKGRKDLLKGNGR.................
tupBel  VLESEEGRREYLAFPTSKSPGAGQKGRKDLLKGNGRRIDYMLYAEEGLCPDWKA
uroEve  ......GRREYLAFPTSKSPGAGQKGRKDLLKGNGR.................
ptiLow  ......GRREYLAFPTSKSPGSGQKGRKDLLKGNGR.................
oryCun  VLESEERRREYLACPISKSQGGCQYGRKDLLKGSGRPLDNMLHAEDGLGPDWKA
ochPri  VLESEEGRREYLAFPTSKSPGGGQKGRKDLLKGNGRRIDYMLHAEDGLCPDWKA
cavPor  VLESEEGRREYLAFPTSKSPGGGQKGRKDLLKGNGRRIDYMLHAEEGLCPDWKA
speTri  VLESEEGRREYLAFPTSKSSGSGQKGRKDLLKGNGRRIDYMLHAEEGLCPDWKA
dipOrd  VLESEEGRREYLAFPTSKSPGPDQKGRKDLLKGNGRRIDYMLHAEEGLCPDWKA
musMus  VLESEEGRREYLAFPTSKSPGAGQKGRKDLLKGNGRRIDYMLHAEEGLCPDWKA
ratNor  VLESEEGRREYLAFPTSKSPGAGQKGRKDLLKGNGRRIDYMLHAEEGLCPDWKA
canFam  VLESEEGRREYLAFPTSKSPGAGQKGRKDLLKGNGRRIDYILYAEEGLWPDWKA
felCat  VLETEEGRREYLAFPTSKSPGAGQKGRKDLLKGNGRRIDYMLYAEEGLCPDWKA
equCab  VLESEEGRREYLAFPTSKSAGAGQKGRKDLLKGNGRRIDYMLYTEEGLCLDWKA
myoLuc  VLESEEGRREYLAFPTSKNSGAGQKGRKDVLKGNGRRIDYMLYAEEGLCPDWKA
pteVam  VLESEEGRREYLAFPASKSPGACQKGRKDVLKGNGRRIDYMLYAEEGLCLDWKA
bosTau  VLESEEGRREYLAFPTSKSPGGGQKGRKELLKGNGRRIDYMLHGEEGLYPDWKA
oviAri  VLESEEGRREYLAFPTSKSPGGGQKGRKELLKGNGRRIDYMLHGEEGLCPDWKA
turTru  VLESEEGRREYLAFPSSKSPGGGQKGRKELLKGNGRRIDYMLHGEEGLCPDWKA
susScr  VLESEEGRREYLAFPTSKSPGGGQKGRKDLLKGNGRRIDYMLHGEEGLCPDWKA
eriEur  ILESEEGRREYLAYPTSKSSGGGQKGRKDLLKGNGRRIDYLLHAEEGLAPDWKA
sorAra  VLESEEGRREYLAFPSSKSPGGCQKGRKDLLKGNGRRIDYLLHAEEGLGPDCKA
dasNov  VLESEEGRREYLAFPTSKSPGAGQKGRKDLLKGNGRRIDYMLFTEEGLCPDWKa
loxAfr  VLESEEGRKEYLAFPTSKSSGAGQKGRKDLLKGNGRRIDYILHGEEGLGPDWKA
proCap  VLENEEGRREYLVFSTSKSSGAGQKGRKGLLKGNGRRIDYILHGEEGLGPDWNA
```

Fig. S6. Gene *MTUS1*

```
homSap   VTASTTCEKLEKARNELQTVYEAFVQ----QHQAEKTERENRLKEFYTREYEKLRDTYIEEAEKYKMQLQEQ
panTro   VTASTTCEKLEKARNELQTAYEAFVQ----QHQAEKTERENRLKEFYTREYEKLRDTYIEEAEKYKMQLQEQ
ponPyg   VTASTTCEKLEEARNELQTAYEAFVQ----QHQAEKTERENRLKEFYTREYEKLRDTYIEEAEKYKMQLQEQ
macMul   VTASTTCEKLEKARNELQTAYEAFVQ----QHQAEKTERENRLKEFYTREYEKLRDTYIEEAEKYKMQLQEQ
calJac   VTASTTCEKLEKARNELQISYEEFVQ----QHQAEKTERENRLKEFYTREYEKLRDTYIEEAEKYKMQLQEQ
tarSyr   VTASATCEKLEKARNELQTAYEGFVQ----QHQAEKTERENRLKEFYTREYEKLQNTYIEEAEKYKVQLQEQ
otoGar   VTASTTCEKLEKSRNELQLAYEGFVR----QHQADRTERENQLKEFYTREYEKLRDTYVEEAEKYKTQLQEQ
micMur   DSASTTCEELEKAKNELETAYEGFVQ----QHQADKTERENRLKEFYTREYEKLRDAYIEEAEKYKTQLQEQ
galVar   ..........EKARNELQIAYEGFVQ----QHRADKIERENQLKEFYTREFEKLQSTYI..............
cynVol   ..........EKAKDELQIAYEGFVQ----QHRADKIERENQLKEFYTREFEKLQSTYI..............
tupBel   VTASTTCDKLQKAKDELQIAYEGFVQKLNQQHQADLSELENRLKEFYTGECEKLQNIYIEEAEKYKTQPQEQ
ureEve   ............AKDELQIAYEGFVQKLNQQHQADLSELENRLKDFYTEECEKLQNMYI..............
oryCun   ATASTTCEKLEKARNELQIAYESVVQKLNQQHQTDRTELESRLKEFYTKECEKLQNIYIEEAEKYKTQKQEQ
ochPri   ATASTTCEKLEKARSELQIAYESFVQKLKQQHQTDRTELEDRLKEFYTKECEKLQSIYIEEAEKYKSQLQEQ
cavPor   VAASTTCEKLEKARNELQTAYEAFVQKLNQQHQTDQTELESRLKEFYTGECEKLQNIYIEEAEKYKAQLQEQ
speTri   VAASTTCEKLEKARNELQTAYEGFVQKLNQQHQTDQAELENRLKEFYTGECEKLQNIYIEEAEKYKIQLQEQ
dipOrd   VAASNTCEKLEIARNELQTAYEGFVQKLNQQHQTDRTELENRLKEFYTGECEKLQNIYIEEAEKYKTQLQEQ
musMus   VAASSACEKLEKARTDLQTAYQEFVQKLNQQHQTDRTELENRLKDLYTAECEKLQSIYIEEAEKYKTQLQEQ
ratNor   VAASSTCEKLEKARNDLQTAYEGFVQKLNQQHQTDQTELENRLKEFYTAECEKLQSIYIEEAEKYKTQLQEQ
canFam   VTASSTCEKLEKAKNELQIAYEGFVQKLNQQHQTDLSELENRLKEFYTGECEKFQNIYIEEAEKYKTQLQEQ
felCat   VTASSTCEKLEKARNELQIAYEGFVQKLNQQHQTDLTELENRLKEFYTGECEKLQNIYVEEAEKYKNQLQEQ
equCab   VTASTTCEKLEKARNELQIAYEGFVQKLNQQHQTDLTELENRLKEFYTGECEKLQNIYIEEAEKYQSQLQEQ
myoLuc   VTASTACEKLEKARNELQAAYEGFVQKLNQQHQTDLTELENRLKEFYTEECEKLQNIYIEEAEKYKTQLQEQ
pteVam   VTASTTCEKLEKARNELQIAYEGFVQKLNQQHQTDLTELESRLKEFYTGECEKLQNIYIEEAEKYKTQLQEQ
bosTau   VTVSTTCEKLEKARNELQIAYEGFVQKLNQQHQTDLTELENRLKEFYTGECEKLQNIYIEEAEKYKTQLQEQ
turTru   VTASTTCEKLEKARNELQIAYEGFVQKLNQQHQTDLTELENRLKEFYTGECEKLQNIYIEEAEKYKTQLQEQ
susScr   VTASTTCEKLEKARNELQIAYEGFVQKLNQQHQTDLTELENRLKEFYTGECEKLQNIYIEEAEKYKTQLQEQ
eriEur   ATASTTCEKLEKARNELQLAYEGFVQKLNQQHQTDLTELENRLKEFYTRECEKLQDIYIKEAENYKTQLQEQ
sorAra   VTASTTCEKLEKARNELQITYEGFVQKLNQQHQTDLTELENRLKEFYTGECEKLQNIYIEEAEKYKTQLQEQ
choHof   VTASTTCEKLEKARNELQIAYEGFVQKLNQQHQTDITELENRLKEFYTGECEKLQNIYIEEAEKYKTQLQEQ
loxAfr   VTASSTCEKLEKARNELQIAYEGFVQKLNQQHQTDLTELENRLKEFYTEECEKLQNIYIEEAEKYKTQLQEQ
proCap   VTASTTCEKLEKARSELQLAYEGFVQKLNQQHQTDLTELENRLKEFYTGECEKLQHIYIEEAEKYKMQLQEQ
```

Fig. S7. Gene *SH3RF2*

```
homSap SRTKNLSLVSSSSRG---NTSTL-RRGPGSRRK
panTro SRTKNLSLVSSSSRG---NTSTL-RRGPGSRRK
gorGor SRTKNLSLVSSSSRG---NTSTL-RRGPGSRRK
macMul SRTKNLSLVSSSSRG---NTSTL-RRGPGSRRK
calJac SNTKNLSLVSSSSRG---NTPTL-RRGPGSRRK
tarSyr ..TKNLSLVFSSSRG---HTPTF-*RGPGSKRRK
otoGar SRTKNLSLASSSSRG---NTPTI-RRGPGSRRK
galVar SCTKNLSLVSSSSRG---NTPTL-RRGPGSRRK
cynVol SCTKNLSLVSSSSRG---NTPTL-RRGPASRRK
uroEve SHSKSLSLVPSPSRGKATNIPTL-RRGPGSRRK
tupBel SHSKSLSLVPSPSRGKATNIPTL-RRGPGSRRK
ptiLow SHTKSLSLVSS-SRGKVTNTPTL-RRGPGSRRK
musMus SRTRHLSLMSSPSRGKATNTSSL-RKSPGSRRK
ratNor SRTKHLSLMSSPSRGKATNTSTL-RKSPGSRRK
cavPor SCTKTLSLASSSPRSKTANTPTL-RRVPGSRRK
oryCun SHNKSLSLMSSPSRGKATSTPTL-RRGPGSRRK
ochPri SRNKSLSLVSSSFRGKPTSTPTL-RRGPGSRRK
canFam ARTKTLSLVPSSSRGKANNTPTL-RRGPGSRRK
felCat VRTKNLSLVPSCSRSKAANTPTL-RRGPGSRRK
equCab SRTKNLSLVASPSRGKAVNTPTL-RRGPGSRRK
myoLuc SRTKSLSLVSS-SRGKATNTPSL-RRGPGSRRK
pteVam SCTKSMSLVSSSSRGKAANTPTL-RRGPGSRRK
turTru SHTKGMSLVSS-SRGKATNAPTL-RRGPGSRRK
bosTau SRTKNLSLVSSPSRGKATSTPTL-RRGPGSRRK
susScr ..TKSLSLVPSPARGKATHAPAL-RRGQGSGRK
sorAra ..TKSLSLMSPSSRGKETSMATL-RRGPGFRRK
eriEur .HPKASPWCPHPPRGKETNTPTL-RRGLGS...
dasNov SHTKSLPLASSSSRGKATNSPTL-RRGPGSRRK
choHof SRTKSLSLGSSFSRGKATNSPTL-RRGPGSRRK
loxAfr PHKKTLPLGPSSPRGKAINSPVL-RRGPGSRRK
proCap SHTKNPSLRSSSPRGKATNWPAL-RRGPGSRRK
echTel SRTKTLSLESPSPRGKATNSPTLQKRGPGSRRK
```

Fig. S8. Gene *NCOA4*

```
homSap   PENGSRET-SEKFKLLFQ----SYNVNDWLVKTDSCTNCQGNQPKGVEIE
panTro   PENGSRET-SEKFKLLFQ----SYNVNDWLVKTDSCTNCQGNQPKGVEIE
gorGor   PENGSRET-SEKFKLLFQ----SYSVSDWLVKPDSCTNCRGNQPKGVEIE
ponPyg   PENGSRET-SEKFKLLFQ----SYSVSDWLVKPDSCTNCRGNQPKGVEIE
macMul   PENDSRET-SEKFKLLFQ----SYNVNDWLVKTDSCTNCQGNQPKGVEIE
calJac   PENGSHEI-SEKFKLLFQ----SYNVNDWLVKTDSCTNCQGNQPKGVEIE
tarSyr   PENGSRET-REKFKLFFQ----SYSVSDWLVKPDSCTNCRGNQPKGVEIE
otoGar   PENDSHEA-SEKYKLLFQ----SYSVSDWLVKPDSCTNCRGNQPKGVEIE
micMur   LENGSCET-SEKFKIFFQ----SYSVSDWLVKPDSCTNCRGNQPKGVEIE
galVar   PENGSHEP-SEKFKHLFQ----SFNVSDWLLKPDSCTNCQGSQPKGVEIE
cynVol   PENGSHEP-SEKFKHLFQ----SFNVSDWLLKPDSCTNCQGSQPKGVEIE
uroEve   PENGSCET-SDKFKLLFQAFQESYSVNDWLVKPDSCTNCQGNQPKGVEIE
tupBel   PENGSCET-SDKFKLLFQAFQESYSVNDWLVKPDSCTNCQGNQPRGVEIE
musMus   PVDGSWET-SEKFKLLFQVFREPYNVSDWLVKPDSCTNCQGNQPRGVEIE
ratNor   PVNGRSDT-SDKFKLLFQVFREPYNVNDWLVKPDSCTNCQGNQPRGVEIE
dipOrd   PENGSHET-SEKIKLLFQVFQEPYSVNDWLVKPDSCTNCQGNQPKGVEIE
cavPor   PESSSWEA-SEKAKRLVQVFQQPYSVNDWLAKPDSCSNCGGNQPKGVEIE
oryCun   PESSSWEA-SEKAKRLVQVFQQPYSVNDWLAKPDSCSNCGGNQPKGVEIE
ochPri   PENGSCEA-SEKLKLLFQAFQEPYSVSDWLAKPDSCTNCGGNQPKGVEIE
canFam   PGNGSCET-SGKFKLLLQVFQEPYSVNDWLVKPDSCTNCRGNQPKGVEIE
felCat   PGKGSCEA-SEKFKLLFQVFQEPYSVSDWLLKPDSCTNCRGNQPKGVEIE
equCab   SRNDSCDT-SEKFKLLFQVFQESYNVNDWLAKPDSCTSCQGNQPKGVEIE
myoLuc   PGNGSCET-SEKFKRLFQVFHESYNVNDWLVKSDSCTNCQGNQPKGVEIE
pteVam   PGNGSCET-TEKFKLLFQVFQESYSVNDWLFKPDSCTNCRGNQPKGVEIE
turTru   PGNGSCET-SEKFKLLFQVFQESYGVNDWLIKPDSCTSCQGKQPKGVEIE
bosTau   PENGGCETTSEKFKLLFQVFHESYNVNDWLVKSDSCTYCQGNQPKGVEIE
oviAri   PENGGCETTSEKFKLLFQVFQESYNVNDWLVRSDSCTCCQGNQPKGVEIE
susScr   PRNGSYET-SEKFKLLFQVFQESYSVNDWLVKPDSCTSCQGNQPKGVEIE
sorAra   PGSSSCET-SEKFKSLFQVFQESYNVNDWLVKPDSCTNCQGSLPKGVEIE
eriEur   PGNGSYET-PEKFKSLFQVFQEPYNVSDWLVKSDSCTNCQGNQPKGVEIE
dasNov   PENGSCET-SEKLKLLFQVFQESYNVSDWLVKPDSCTNVR-QQPKGVEIE
choHof   ..SGSCET-SEKFKLLFQIFQESYSVSDWLVKPDSCSSCQGNQPKGVEIE
loxAfr   ..NGSCET-NEKFKLLFQVFQESYNVSDWLVKPSSCTNCQGNQPKGVEIE
proCap   ..NGSCET-SEKFKLLFQVFQESYNVNDWLVKPDSCTICQGNQPKGVEIE
echTel   ..NGSCET-SGKSQSWFQVFQEPFNISDWLVKPDSCINCQGIQPKAMEIE
```

Fig. S9. Gene *TEX2*

```
homSap MSKAQTDKETSEEKPPAE---GSEDPKKPPRPQEGTRSSQRDQILYLFGR
panTro MSKAQTDKETSEEKPPAE---GSEDPKKPPRPQEGTRSSQRDQILYLFGR
ponPyg MSKAQTDKETSEEKPPAE---GSEDPKKPPRPQEGTRSSQRDQILY....
macMul MSKAQTDKETSEEKPPAE---GSEDPKKPPRPQEGTRSTQRDQILYLFGR
calJac MSKAQTDKETSEEKPSAE---GSEDPKKPPRPQEGTRSSQRDQILYLFGR
tarSyr MSKAQ-DKETSEEKPPPE---GSEDPQKPPPPQEGTRSGQREQILYLFGR
otoGar MSKAQTDKETLEEKLPAE---GSEDPKKPPHPQEGTRSGQRDQILYLFGR
micMur MSKAQTDKETLEEKLPVE---GSEDPKKPPHPQEGARSGQRDQILYLFGR
cynVol ..........VEEKLPAE---GSEDPKKPPVPQEGTRYGQRDQILYLFGR
uroEve ..........SEDKQPAEREQGSEDPKKPPHSQEGTRSGPRDQILYLFGR
tupBel MAKAQTDKETSEDKPPAERELGSEDPKKPPHSQEGTRSGPRDQILYLFGR
ptiLow ..........SEEKPPAEREPGSEDPKKPPHSQEG-RSGQRDQILYLFGR
musMus MSKAQSDKEATEEKPPPEKELPSEDLKKPPQPQEGTKSSQRDPILYLFGR
ratNor MSKAQSDKEATEEKPPPEKELPSEDLKKPPQPQEGTKSSQRDPILYLFGR
dipOrd MAKAQTDKETSEDKPPAERELGSEDPKKPPHSQEGTRSGPRDQILYLFGR
cavPor MSKAQADKETSEEKPPAEKELGSEDPKKPSHPQEGTRSVHRDQILYLFGR
oryCun MSKAQTDKETSEEKPPAERELASEDPKKPPQPQEGTRAGQRDQILYLFGR
ochPri MSKAQTDKETCEEKPAAERELANEDPKKPLQPPEGTRAGQRDQILYLFGR
canFam MSKAQTDKETSEEKPPAERELGGEDPKKPPHPQEGTRSGQRDQILYLFGR
felCat MSKAQTDKETSEEKPPAERELGGEDPKKPPHPQEGTRSGQRDQILYLFGR
equCab MSKAQSDKETSEEKPPTEKEQGVEDPKKPSPPQEGTRSSQRDQILYLFGR
myoLuc MSKAQTDKETSEEKPPAERDLGVEDPKKPPHPQEGTRSGQRDQILYLFGR
pteVam MSKAQTDKETSEEKLPAEREVGGEDPKKPP-PQEGTRSGQRDQILYLFGR
turTru MSKAQTDKETSEEKPPAERELGGEDPKKPTHPQEGTRSGQRDQILYLFGR
bosTau MSKAQTDKETCEEKPPAERELGGEDPKKPPHPQEGTRSGHRDQILYLFGR
susScr MSKAQPDKETSEEKPPAERELGGEDPKKPPHPQERTRSGQRDQILYLFGR
sorAra MAKAQTDKESSEEKLPAEKELGAEDPKKPAHPQEGTRA-QRDQILYLFGR
eriEur MSKAETDKELPAEKPPAERELGGEDPKRLPHSQEGTRSGQRDQILYLFGR
dasNov MSKTQTDKETSEEKPSAERELGSEDSKKPPHSQEGTRSGQRDQILYLFGR
choHof MSKTQTDKETSEEKPAAERELGNEDPKKPPHPQEGTRSVQRDQILYLFGR
loxAfr MSKAQTDKETSEEKPPAERELAGEDPKKPP--LEGTRSGQRDQILYLFGR
proCap MSKAQNGKETSEEKPPVERELAGEDPKKPP--QEATRSGHRDQILYLFGR
```

Fig. S10. Gene *SSH2*

```
homSap SLDYLHPQTMVHLEGFTEQSSTTD-EPSAEQVSWEESQESPLSSGSEVPY
panTro SLDYLHPQIMVHLEGFTEQSSTTD-EPSAEQVSWEESQESPLSSGGEVPY
gorGor SLDYLHPQTMVHLEGFTEQSSTTD-EPSAEQVSWEESQESPLSSGSEVP.
ponPyg SLDYLHPQTMVHLEGFTEQSSTTD-EPSADQVSWEESQESPLSSGSE...
macMul SLDYLHPQTMVHLEGFIEQSSTTD-EPSAEQVSWEESQDSPLSSGSEVPY
calJac ..DYLHPQTMVHLEGFTKQSSTTD-EPSTEQVSWEESQESPLFSGNEVPY
tarSyr NLDYLHSQAVVHLEGFTEQSSTTD-EPSAEQVSWEENQEGPLSSGNDVPH
otoGar NLDYLHPQTLVHLEGFTEQSSTTD-DPSTEQVSWEESQEGP-PSGNE...
micMur NLDYLHPQTVVHLEGFTEQSSTTD-EPSTEQVSWEESQEGPLWSGSEVPH
galVar NLDYLHPQTMVHLKGFTEQSSTTD-EPSAEQPSWEESQEGPLCSGSEVPY
cynVol NLDYLHPQTMVHLKGCTEQSSTTD-EPSAEQPSWEESQEGPLCSGSEVPY
uroEve .LDYLHPQTVVHLEGFTEQSSTTDSEPSAEQLSWKESREGPLPSGGEATH
tupBel NLDYLHPQTVVHLEGFTEQSSTTDSEPSAEQLSWKESQAGPLPSGGEATH
ptiLow ....LHPWAVLHLEGFTKQSSTTDSELSAEQVSWEDGQEG..........
musMus SLDYLLPHSVVHLEGCTEQSSATDNELSPEQASWEDSRGHFLSSGSGMAH
ratNor SLDCSHRHSVLHLEGCTEQSSTTDSRLSSEHMNWEDSQGDFLSSSTGMAH
dipOrd ..........VHLEGFTEQSSTTDNEPSAEQVSWEESAEGSLSV......
speTri .LDYLHPHTVVHLEGFTEQSSTTDSEPSTEQVSWEERQEDPLSRGYEVTH
cavPor NLDYMHPQTVVHLEGCTKQSSTTDSEPSTEQVNWAETQKGHLSGGDEVPH
oryCun SLDFLHPQTVVHLEGCTEQSSTTDNEPSAEQVGWEETQEVPLSCGSEAPH
ochPri .LDFLHPQTVVHLKDVTEHSSTTDHEPSTAQLSWEGSPLVPLSGGGPVP.
canFam NLDYAPPQTVVHLEGFTEQTSTTDNEPSSEQGSWEESQEGPFSGGSE...
felCat NLDYAHSRTIVHLEGFTEQSSTTDNEPSLEQGGWEEGLDGALSSGSEVPY
equCab SLDYLHPQTVVHLQGCTEQSSTTDSEPSAEQGSWEGSQEGPLSSG.....
myoLuc NLDYLHPQTMVHLEGFTEQSSTTDSEPCAEPGRWEESQEGPLSSGNEVPY
pteVam SLDYLNPQTVVHLEGFTEQSSTTDSEPSAEQGSWEENQEG..........
turTru SLAYLHPQTVVHLEGFTEQSSTTDNEPSAEKGSWEESQEGPLSRGSGVPY
bosTau ..DHLHPQTVVHLEGFTGQSSTTDSEPSTEQGSWEESQEGPLSRASEVPY
susScr SLDCLPPQTVVHLEGFTEQSSTTD-EPSAEQGSWEESQEGPAPRRSEVPY
sorAra SLDYLHPQTVVHLEGFTEQSSTTDSEPPTEPG.FREGRHEGLLSGSKET.
eriEur SLDYLHPQTVVHLEGLTEQSSTTDMESSAEQDSCGNSQEVALSSGNE...
dasNov NLDYLHPQTVVQLEGFTEQSSTTDTESSAEQVSWEESQEGLLSIGNDVPH
choHof NLDYLHPQTVVQLEGFTEQSSTTDDEPSVEQVSWEESQEGPLSVGSEVPY
loxAfr NVDYLHSQTVVHLEGFTEQSSSTDNEPSTQQGRWEESQEGPLCSGDEVPH
proCap SLDCLHSPKMLPLEGVTERSSSTDREPSPQQGSWEESQEGCLSGGTEVPH
echTel DLDSPQSQTVVHLEGSTEQSSSTDPEPSTWQDSWEQSQGGPLCRATKVPH
```

Indel Supporting Scandentia+Primates (treeshrews and primates in orange)

Fig. S11. Gene *ADD2*

```
homSap GEKETAPEEPGSPAKSAPASPVQSPAKEAETKSPLVSPSKSLE
panTro GEKETAPEEPGSPAKSAPASPVQSPAKEAETKSPLVSPSKSLE
ponPyg GEKETAPEEPGSPAKSAPASPVQSPAKEAETKSPLVSPSKSLE
macMul GEKETAPEEPGSPAKSAPASPVQSPAKEAETKSPLVSPSKSLE
calJac GEKETAPEEPGSPAKSAPPSPVQSPVKEAETKSPVVSPSKSLE
tarSyr GEKETTTEEPGSPAKSAPASPVQSPAKEAETKSPVVSPSRSSE
otoGar GEKETTAEEPGSPLKSAPASPVQSPAKKAETKSPVVSPSKSSE
micMur GEKETATEEPSSPVKSAPASPAKGPAKEAETKSPVVSPSKSSE
cynVol ....TAPEEPSSPVKSAPASPAQSPVK-AETKSPVVSPSKSSE
uroEve ....TAPEEPGSPVKSAPASPAQSPAKEAETKSPVVSPSKSSE
tupBel GEKETAPEEPGSPVKSAPASPAQSPAKEAETKSPVVSPSKSSE
ptiLow ....TAPEEPGSPVKSAPASPAQSPAKETETKSPVVSPSKSSE
musMus GEKDIATEKPGSPVKSTPASPVQSPSK-AGTKSPAVSPSKTSE
ratNor GEKDAATEEPGSPVKSTPASPVQSPTR-AGTKSPAVSPSKASE
dipOrd GEKDTATEEAGSPVKSTPASPVQSPTK-ADTKSPTISPTKSVE
speTri GEKEAATEEPGSPEKSAPASPAQSPVK-AETKSPVVSPSKSAE
cavPor GEKDTTTEEPGSPAKSAPASPAQSPTK-TETKSPAVSPSKSTE
oryCun GEKETATEEPGSPVKSAPASPAQSPTK-AETKSPAVSPSKSAE
ochPri GEKETAPEEPGSPAKSAPASPAESPTK-AETKSPAVSPSK...
canFam GEKETPAEGPGSPG-SAPASAAQSPAR-SETKSPAVSPSRSAD
felCat GEKETTAEEPGSPVRSAPASPAQSPAK-SETKSPLVSPSKSLD
equCab GEKETATEEPGSPVKSAPASPAQSPAK-SETKSPVVSPSKSLD
myoLuc GEKETAAEEPGSPVKSAPASPAQSPVK-SETKSPVVSPSKSLD
pteVam GEKETATEEPGSPVKSAPASPAQSPAK-SETKSPVVSPSKS..
turTru GEKEPAPEEPGSPVKSAPASPAQSPAK-SETKSPVGSPSKS..
bosTau GEKEPVPEEPGSPVKSAPASPAQSPAK-SEPKSPVGSPSKSVD
sorAra GEKETAPAEPGSPVKSAPASPAQSPVK-PDTESPVVSPAKSVE
eriEur GEKEAAPEEPGSPAKSVPASPAQSPVK-SETKSPVVSPSKSL.
dasNov GEKETTTEEPGSPVKSAPASPVQSPVK-SEAKSPVVSPAKSSE
loxAfr GEKEIATEGPESPVKSAPTSPVKSPSK-SETKSPMVSPSTTLD
proCap GDKEIATEGPESPVKSAPTSPVKSPSK-SETKSPPVSPSK...
echTel GEKEIAPEGPGSPTKSAPTSPVQSPAK-SEAKSPVVSPSKTVE
```
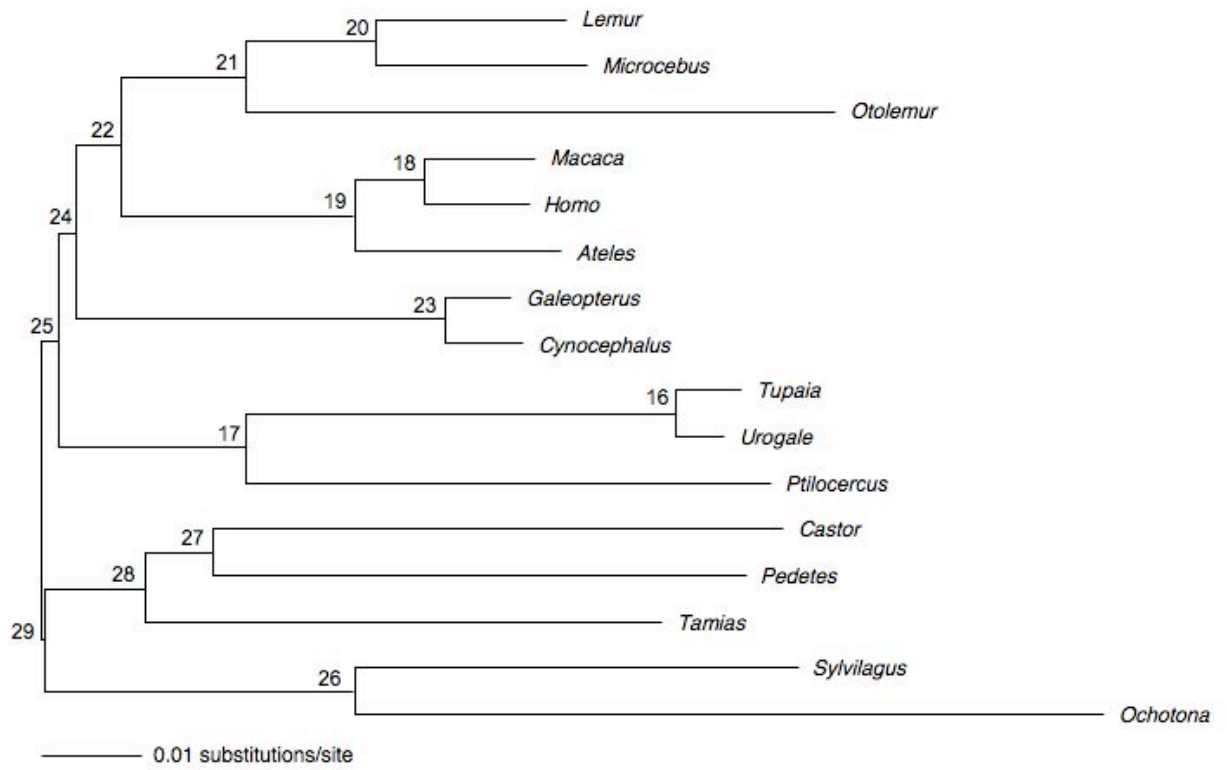
Fig. S12

0.01 substitutions/site

Fig. S13
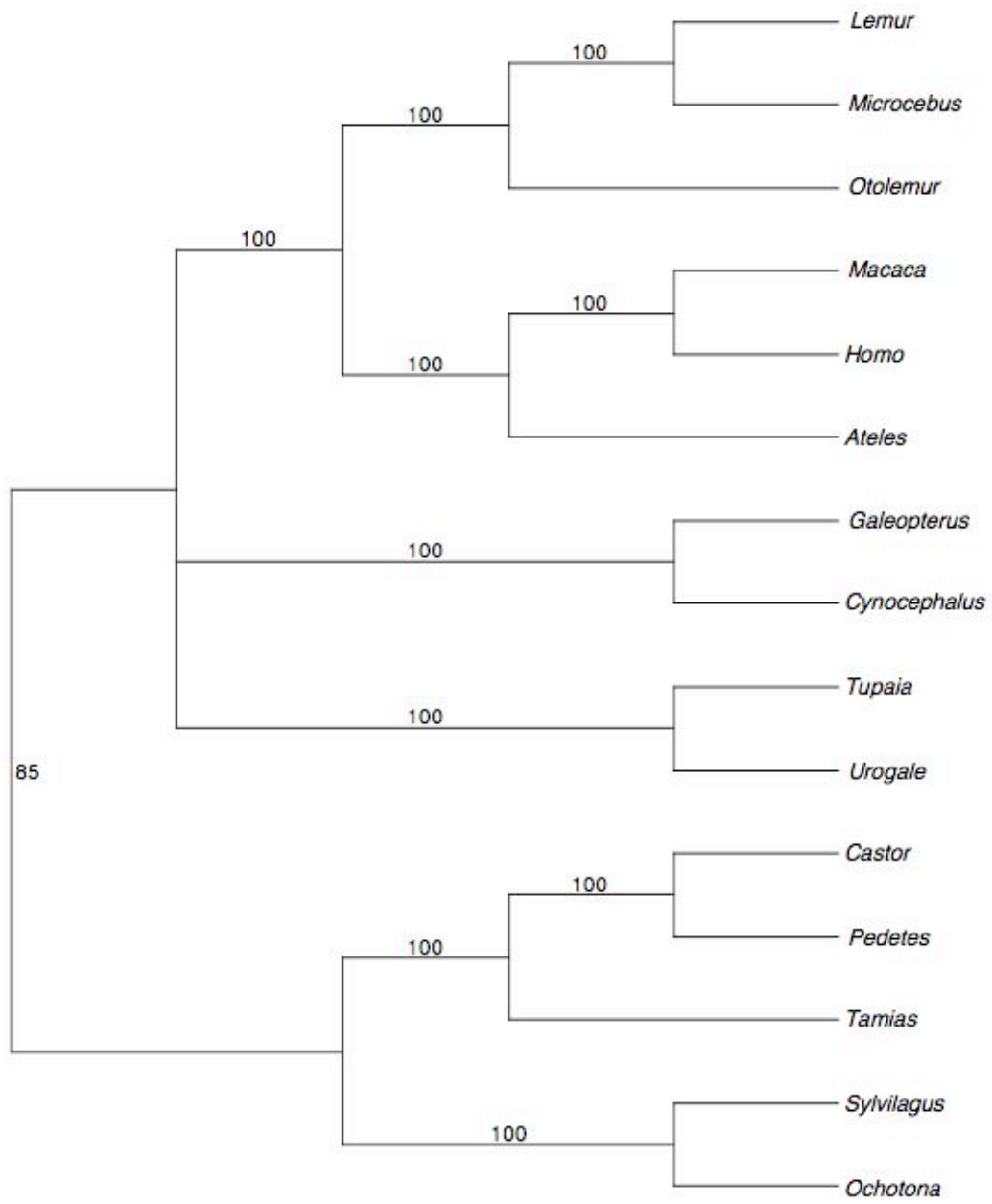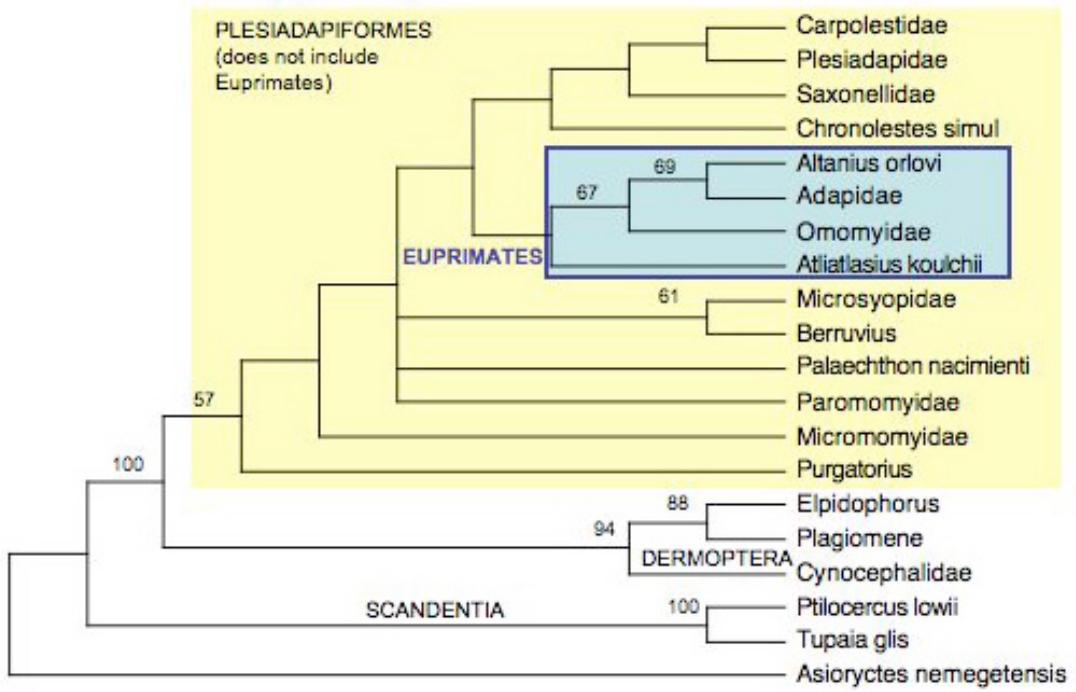
Fig. S14

Supplemental References:
S1. Murphy WJ, et al. (2007) *Genome Res* 17: 413-420.
S2. Miller W, et al. (2007) *Genome Res*, in press.
S3. Murphy WJ, et al. (2001b) *Science* 294:2348–2351.
S4. Thompson, JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) *Nucleic Acids Res* 25:4876–4882.
S5. Murphy WJ, et al. (2001a) *Nature* 409: 614-618.
S6. Eizirik E, Murphy WJ, Springer MS, O'Brien SJ (2004) in *Anthropoid Origins: New Visions*, eds Ross CF, Kay RF (Kluwer Academic/Plenum Publishers, New York), pp 45– 64.
S7. Poux C, Chevret P, Huchon D, De Jong WW, Douzery EJP (2006) *Syst Biol* 55:228–44.
S8. Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ (2005) *Science* 307:580–584.
S9. Posada D, Crandall KA (1998) *Bioinformatics* 19:301–302.
S10. Swofford DL (2003) *PAUP** (Sinauer Associates, Sunderland, MA).
S11. Ronquist F, Huelsenbeck JP (2003) *Bioinformatics* 19:1572–1574.
S12. Waddell, PJ, Kishino, H, Ota, R. *Genome Inform*. **12**, 141 (2001).
S13. Shimodaira H, Hasegawa M (1999) *Mol. Biol Evol*. 16:1114–1116.
S14. Thorne JL, Kishino H, Painter IS (1998) *Mol Biol Evol* 15:1647–1657.
S15. Thorne JL, Kishino H (2002) *Syst Biol* 51:689–702.
S16. Felsenstein J (1984) *Evolution* 38:16–24.
S17. McKenna MC, Bell SK (1997) *Classification of Mammals Above the Species Level*. (Columbia University Press).
S18. Steiper ME & Young NM (2006) *Mol. Phylogenet. Evol*. **41:** 384–394.
S19. Bloch JI, et al. (2007) *Proc. Natl. Acad. Sci. USA*. **104:** 1159-1164.