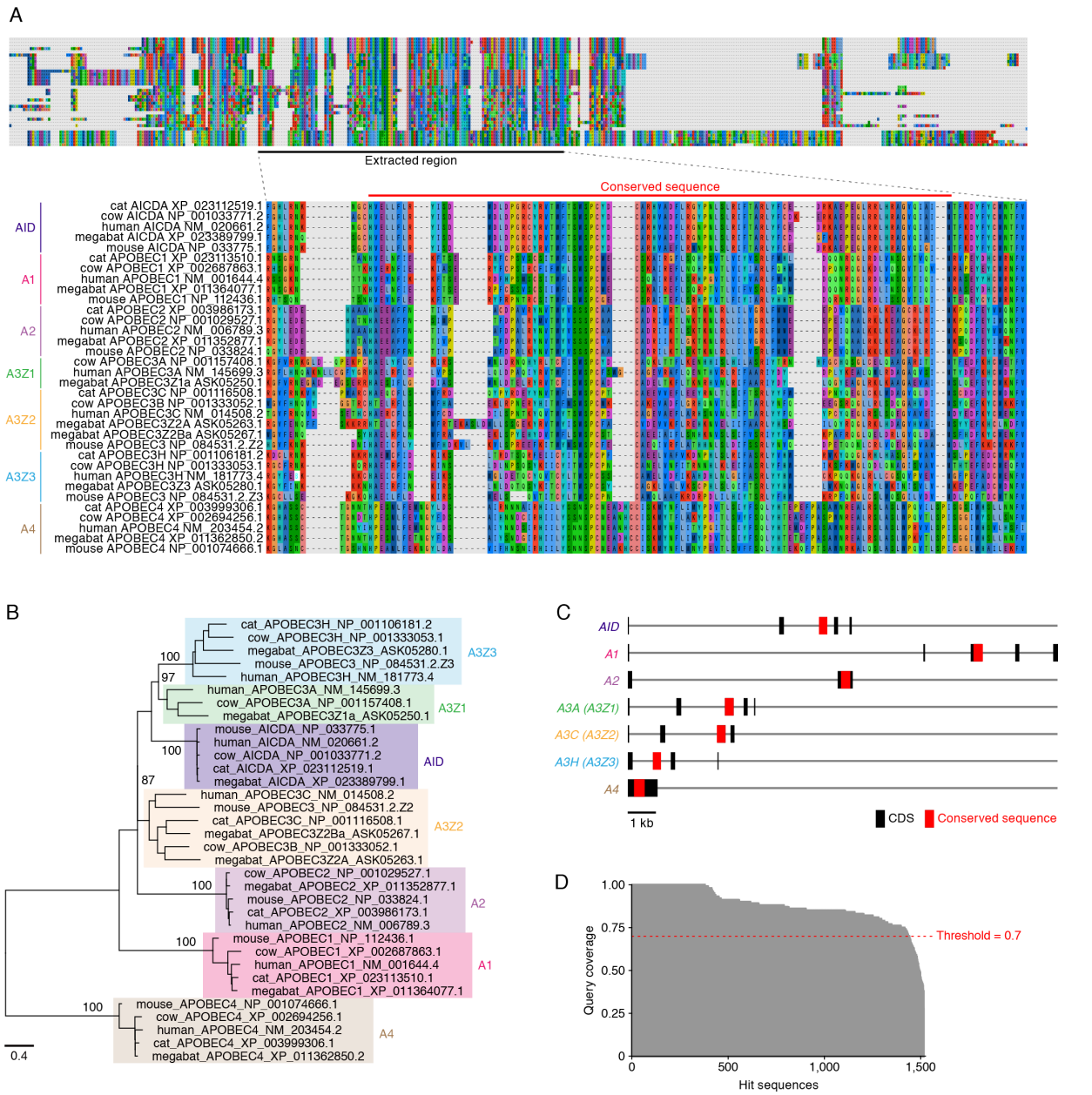


Fig. S1. Sequence conservation in *AID/APOBEC* family genes. (*A*) Amino acid sequences of *AID/APOBEC* family genes used as queries for *in silico* genome screening are shown. Sequences disclosing homology to the conserved sequences of *AID/APOBEC* family genes ("Conserved region") were extracted. (*B*) Phylogenetic tree of *AID/APOBEC* family genes. The tree was constructed using maximum likelihood (ML). Bootstrap values are indicated at the nodes. (*C*) Gene structure of *AID/APOBEC* family genes in the human genome. Black and red boxes indicate regions corresponding to the coding domain sequence (CDS) and the conserved sequence, respectively. Note that the conserved sequence is located on a single exon in all types of *AID/APOBEC* family genes. (*D*) Coverage of sequence alignment among sequences recovered via *in silico* genome screening. Y-axis indicates the proportion of the length of the query sequence covered by the alignment (referred to as coverage). Sequences that covered < 70% of the conserved region were discarded.

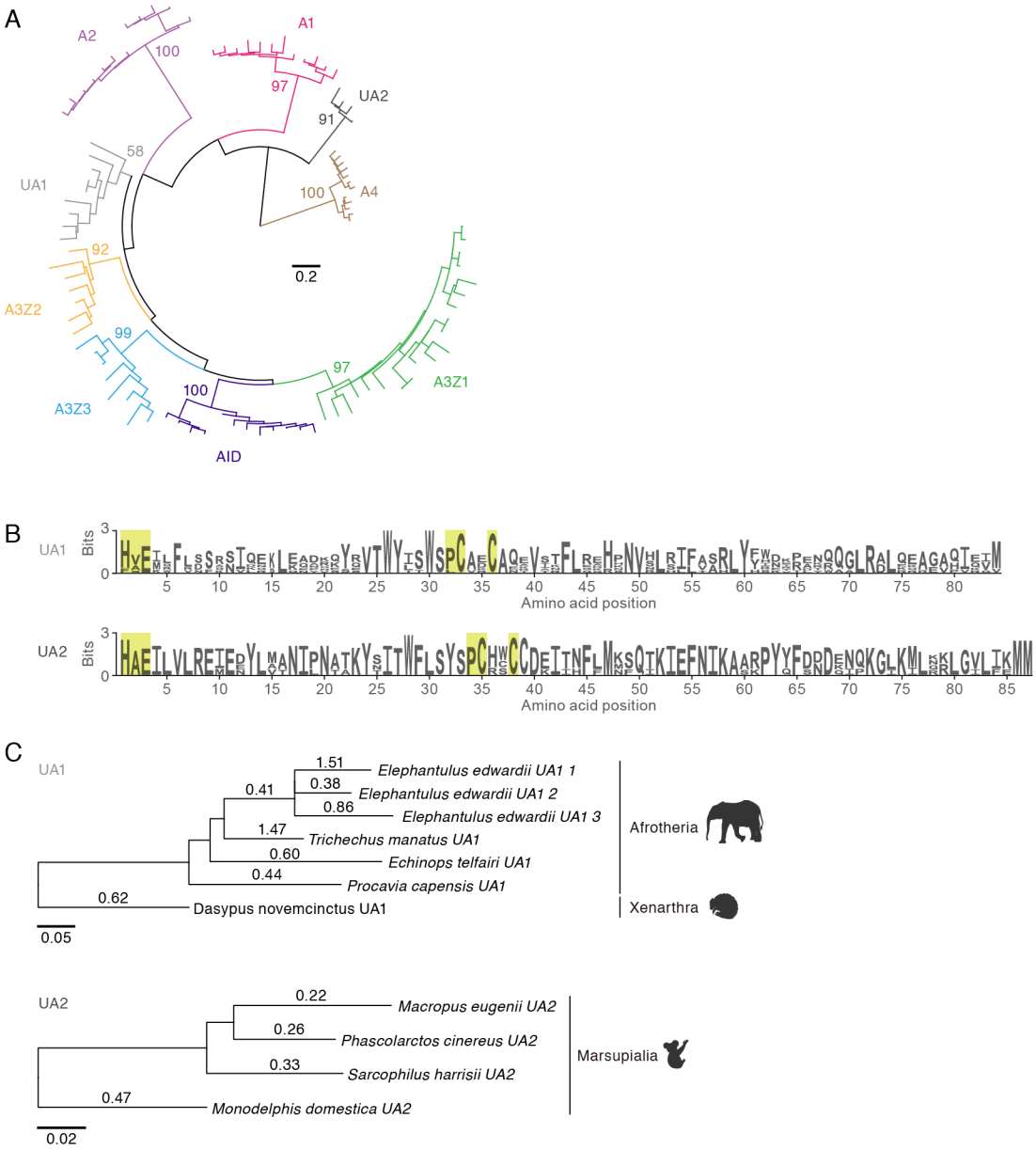


Fig. S2. Characterization of lineage-specific *AID/APOBEC* genes (*UA1* and *UA2*) found in basal eutherian mammals. (*A*) Phylogenetic tree of *AID/APOBEC* Z domains identified in Afrotheria, Xenarthra, and Marsupialia. The tree was reconstructed using a maximum likelihood (ML) approach, and is based on an alignment of nucleic acid sequences. Bootstrap values are indicated at the nodes. (*B*) Logo plots of amino acid sequences of UA1 and UA2. Yellow squares indicate the amino acid residues corresponding to the catalytic domain of AID/APOBEC proteins. (*C*) Phylogenetic tree of *UA1* and *UA2* Z domains. The tree was reconstructed by maximum likelihood (ML) method based on the nucleic acid sequences. Estimated dN/dS ratios are indicated on the branches. Note purifying selection (dN/dS < 1) was detected in the Z domains of both genes.

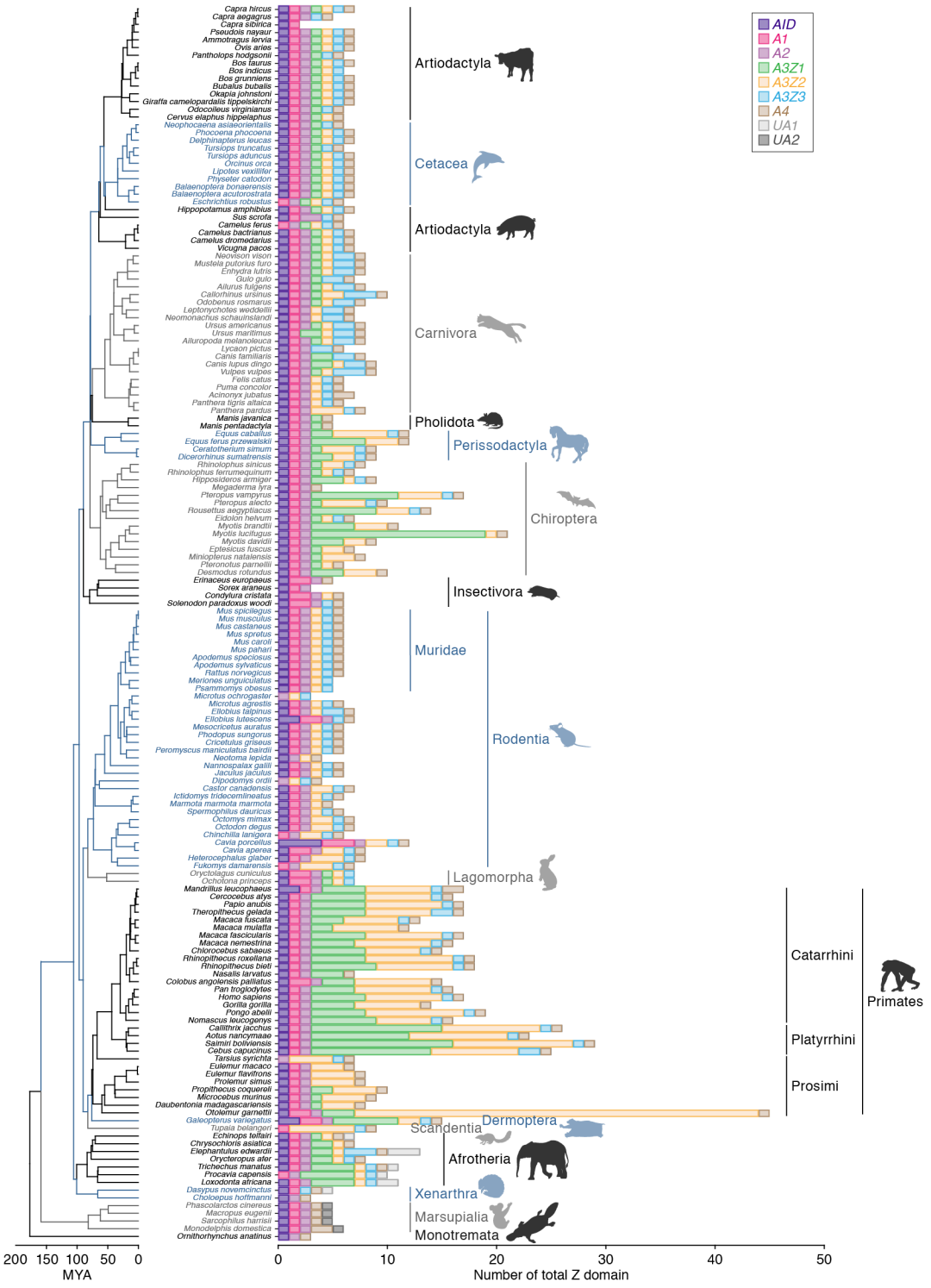


Fig. S3. Number of *AID/APOBEC Z* domains identified in each mammal species. Counts of both intact and pseudogenized *AID/APOBEC Z* domains are shown.

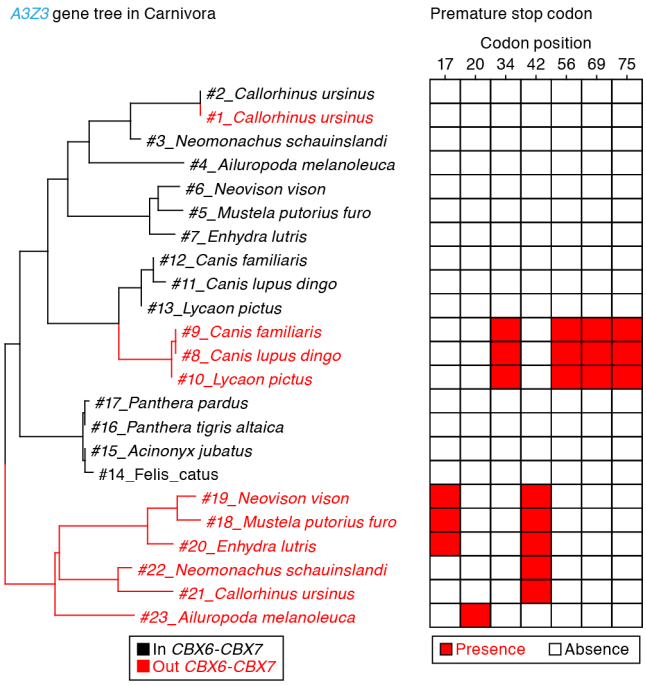


Fig. S4. Gene duplications of *A3Z3* followed by pseudogenizations in Carnivora. (*Left*) Phylogenetic tree of *A3Z3* in Carnivora. Red and black indicate *A3Z3* located in the inside and outside of the canonical *A3* gene locus. (*Right*) Presence of premature stop codons. Positions of the stop codons in the multiple sequence alignment are indicated. The phylogenetic relationships and shared patterns of premature stop codons indicate that fixation of duplicated *A3Z3* pseudogenes has occurred at least twice in carnivore evolution.

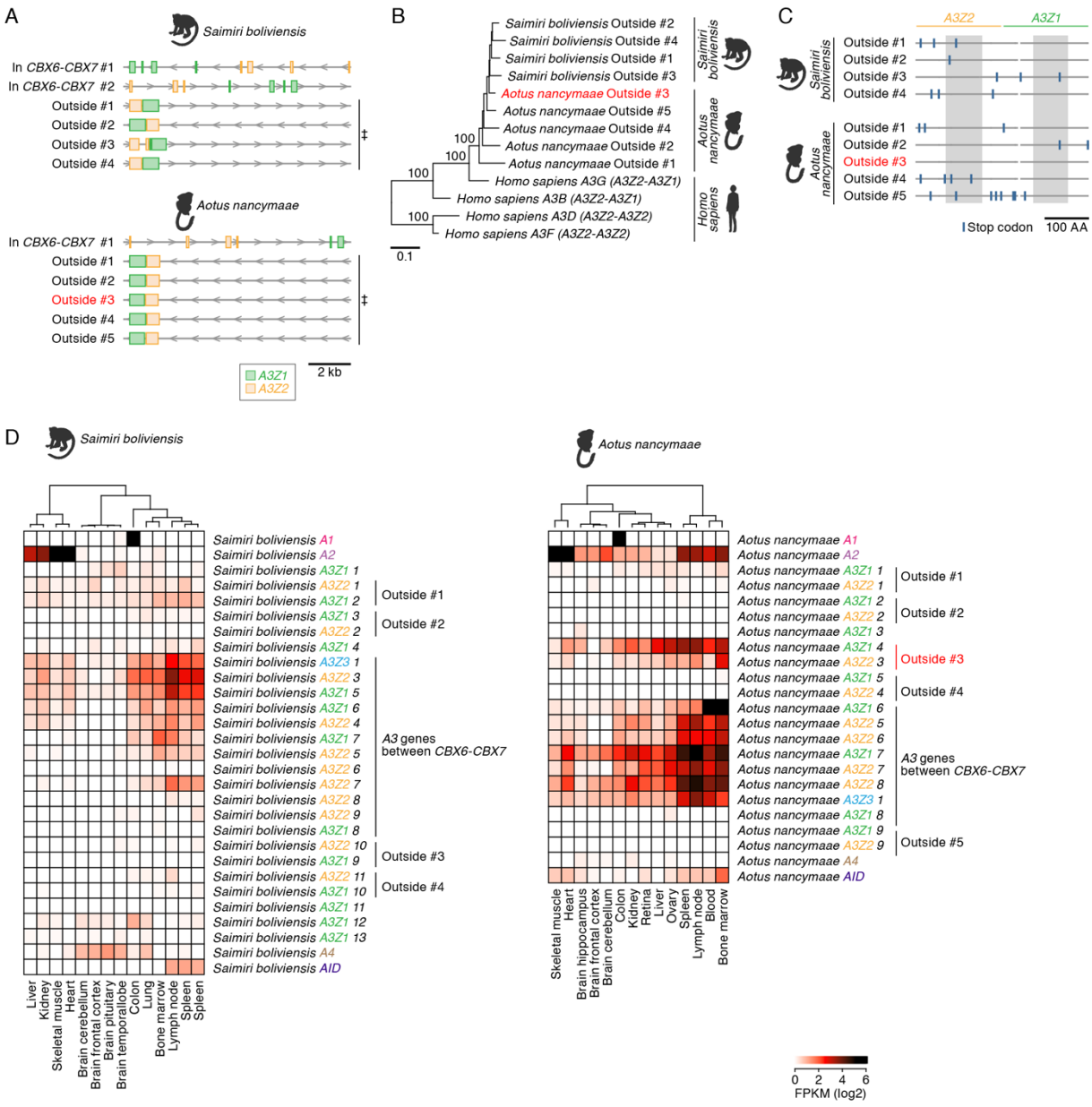


Fig. S5. Retrotransposition of *A3G*-like genes in New World monkeys. (*A*) Gene structures of *A3Z2-A3Z1* type genes in the genomes of New World monkeys. Genes within and outside the canonical *A3* gene locus ("in *CBX6-CBX7*") are shown. Arrowheads indicate the direction of the respective loci. The sequences indicated by double daggers are intron-less sequences and correspond to those in **Fig. 3*C***. (*B*) A phylogenetic tree of retrotransposed *A3Z2-A3Z1* type genes identified in New World monkeys. Note that the retrotransposed genes form a cluster with the human *A3G* (*A3Z2-A3Z1* type) gene. (*C*) Presence or absence of premature stop codons in the retrotransposed *A3Z2-A3Z1* type genes in the two New World monkeys. Regions corresponding to the conserved sequence of AID/APOBEC Z domains are boxed. (*D*) RNA expression of the retrotransposed *A3* genes in the tissues of New World monkeys. Rows indicate the sequences of *AID/APOBEC* family genes, and columns indicate the tissues of New World monkeys. Color indicates RNA expression level (log2 FPKM). Note that the "outside #1-4" genes in *Saimiri boliviensis* and "outside #1-5" genes in *Aotus nancymaae* are identical to those in panels *A-C.*

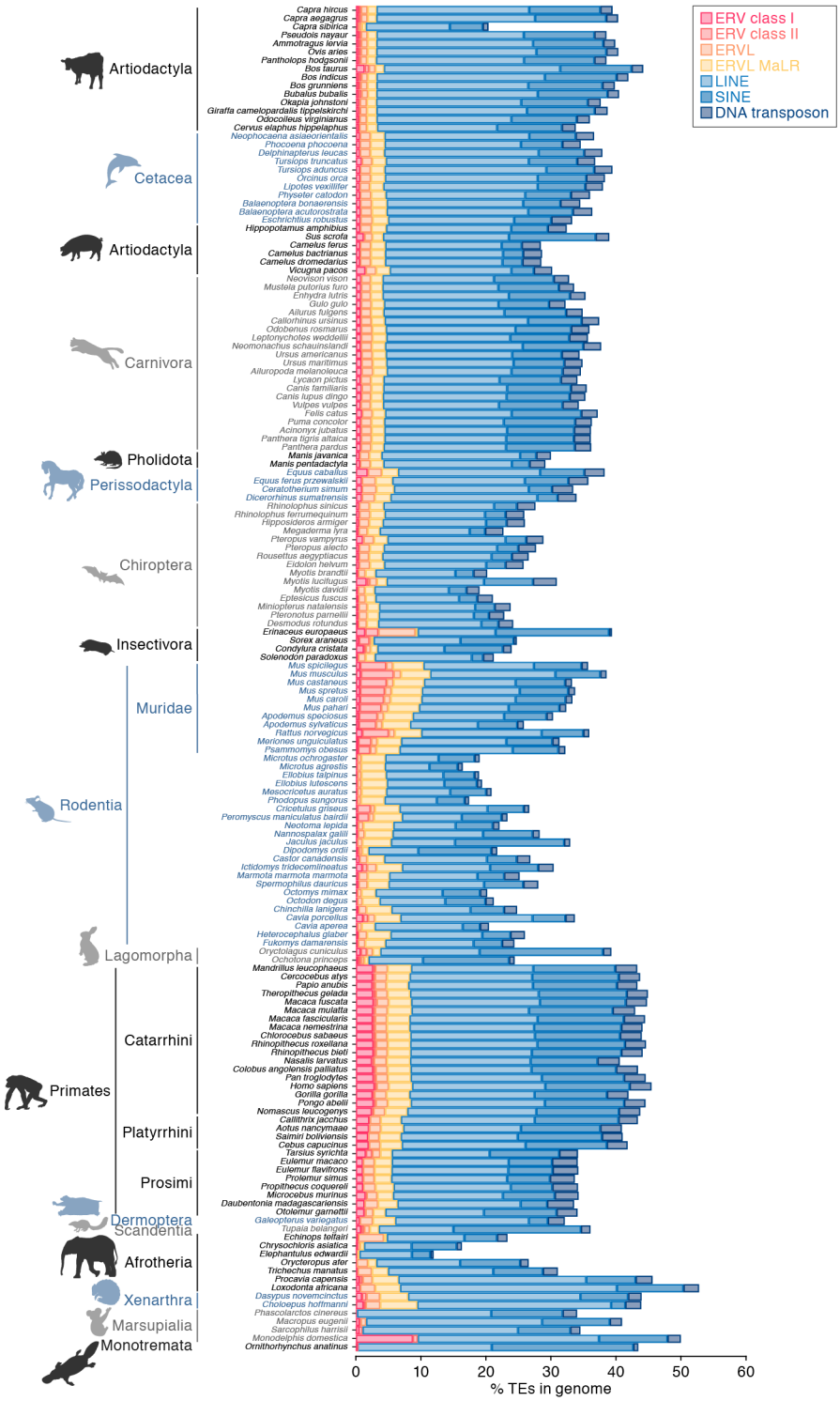


Fig. S6. Proportions of TE sequences in the genomes of the respective mammalian species. Proportions of ERV sequences in the genomes of the respective mammalian species are also shown in **Fig. 4*A***.

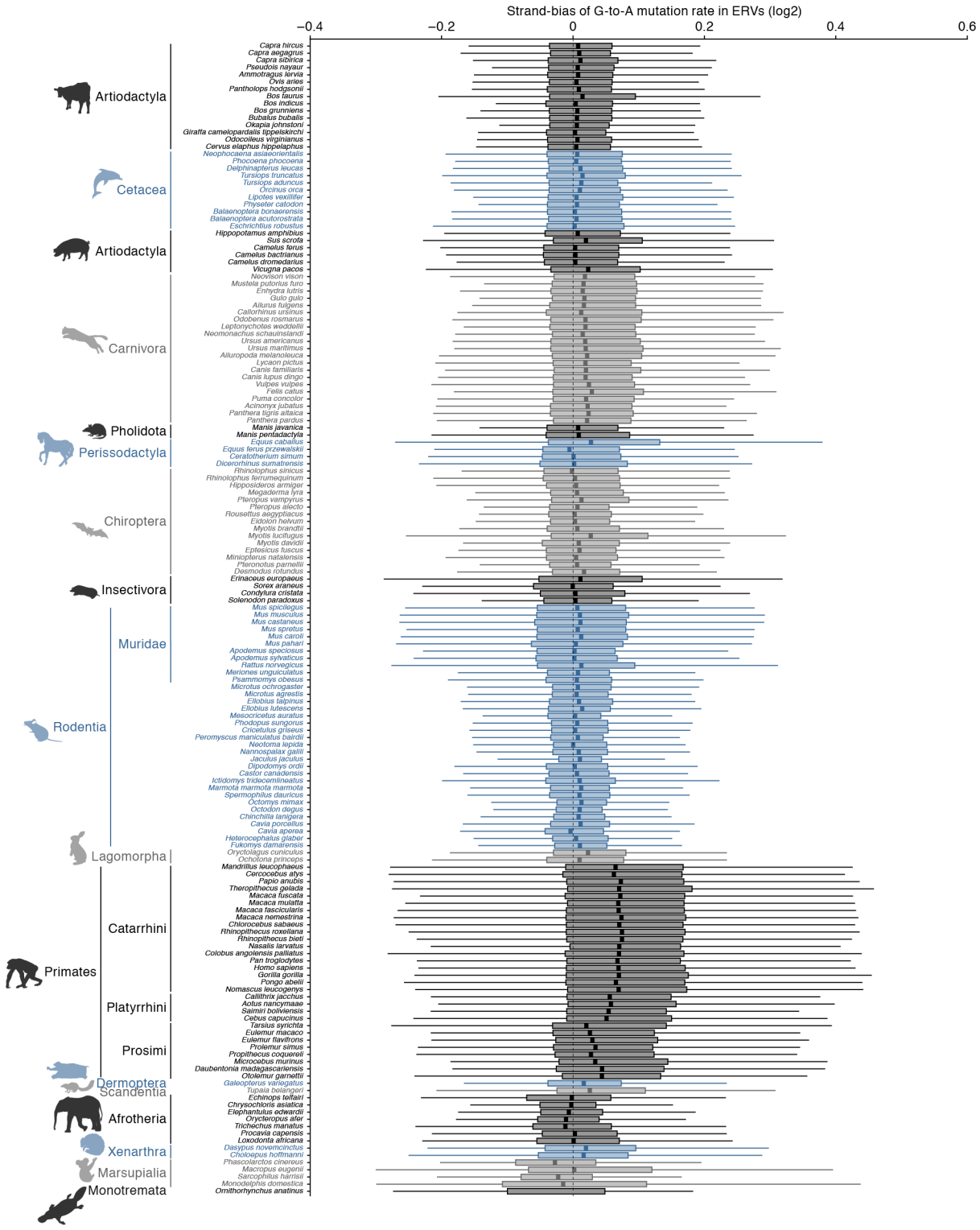


Fig. S7. Difference in the strand biases of G-to-A mutation rates in ERVs among mammals. The log2-transformed strand bias score of G-to-A mutations is shown (i.e., the mutation rate in the positive strand divided by the one in the negative strand).

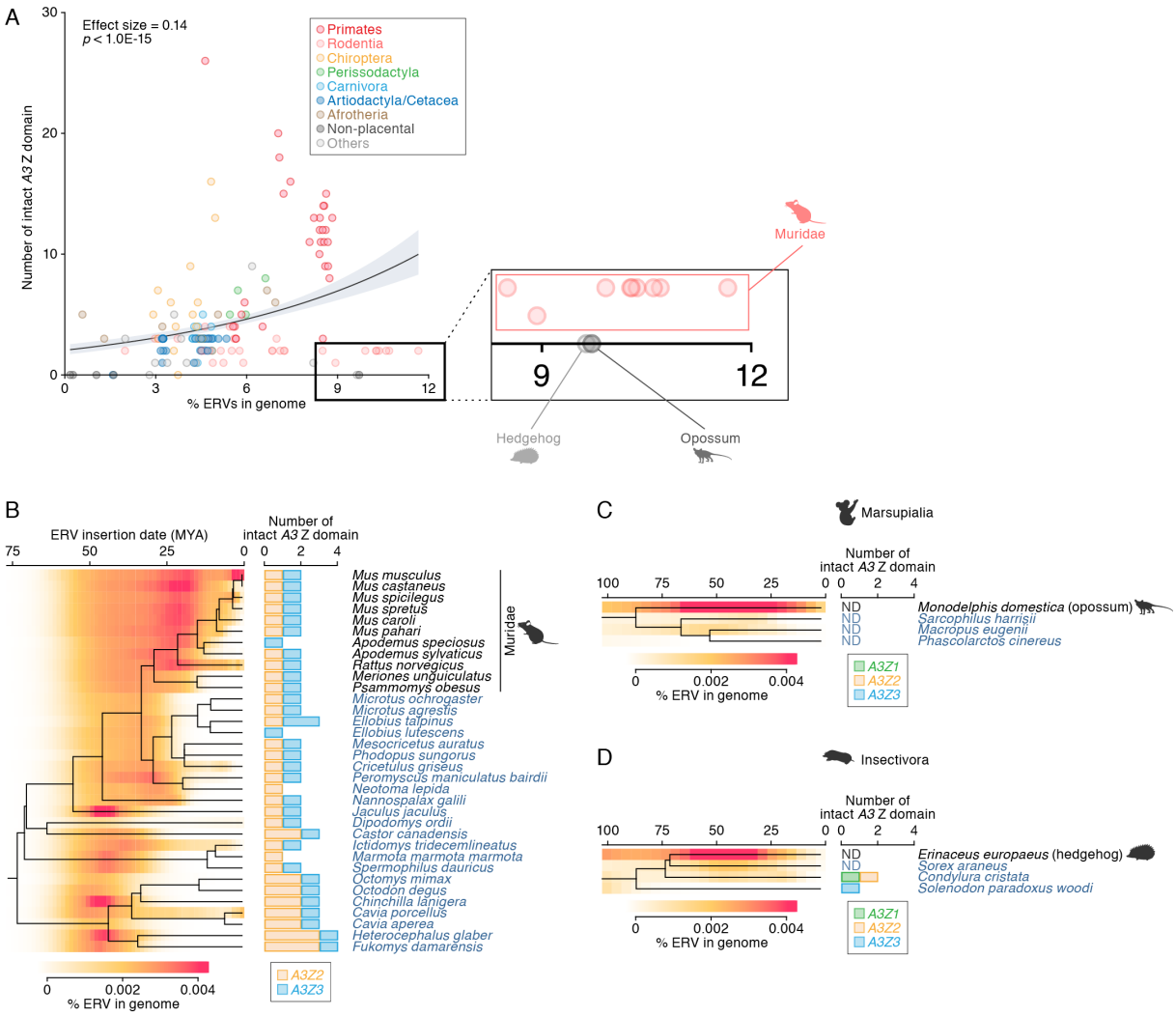
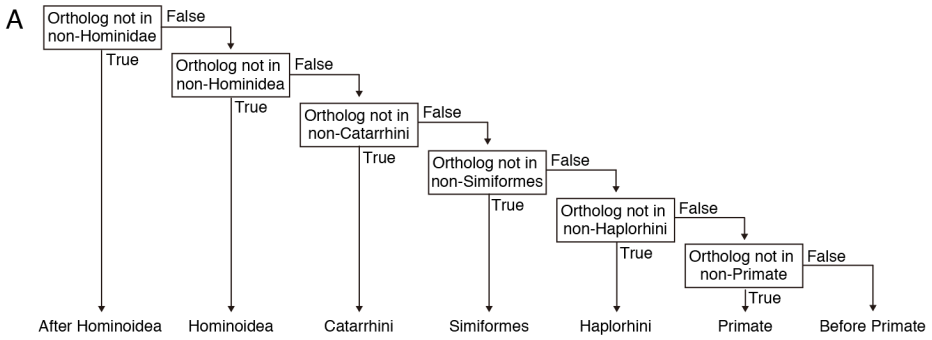
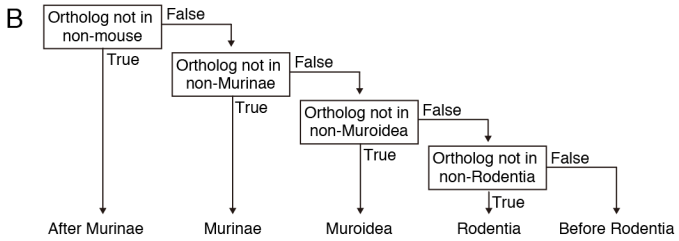
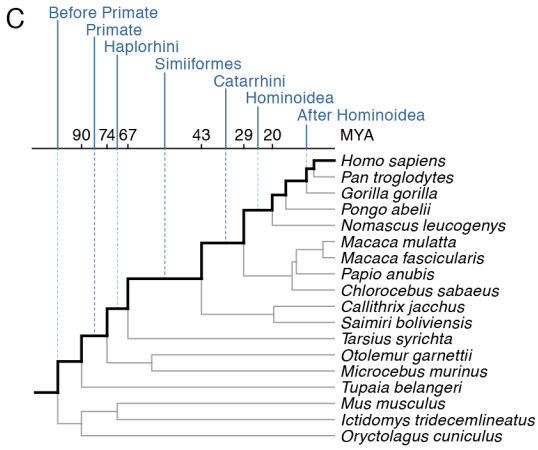
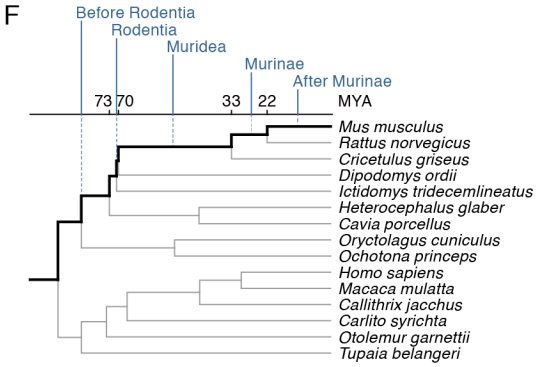


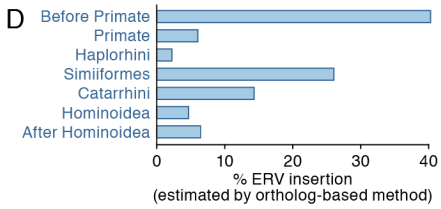
Fig. S8. "Outlier" species with a high number of ERV invasions but only few or no *A3* genes. (*A*) Relationship between the number of *A3* Z domains and the amount of ERV insertions in the genome. "Outlier" species (Muridae, hedgehog, and opossum), are indicated by enlarged dots. (*B-D*) Insertion dates of ERVs and the numbers of *A3* genes in mammals. The insertion dates of ERVs and numbers of *A3* genes in Rodents (*B*), Marsupialia (*C*), and Insectivora (*D*) are shown. (*Left*) Amounts of ERV insertions at each age in the respective mammalian species. ERV insertion date was estimated according to the genetic distance of each ERV integrant from the consensus sequence under the molecular clock assumptions (*B,* 7 × 10^{-9} mutation/site/year; *C* and *D,* 2.2 × 10^{-9} mutation/site/year). (*Right*) Number of intact *A3* Z domains.

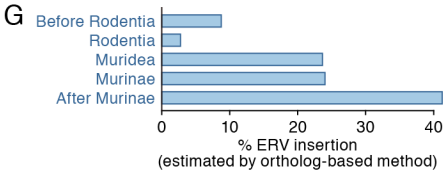

A
Ortholog not in non-Hominidae
False
True
After Hominoidea
Ortholog not in non-Hominidea
False
True
Hominoidea
Ortholog not in non-Catarrhini
False
True
Catarrhini
Ortholog not in non-Simiformes
False
True
Simiformes
Ortholog not in non-Haplorhini
False
True
Haplorhini
Ortholog not in non-Primate
False
True
Primate
Before Primate

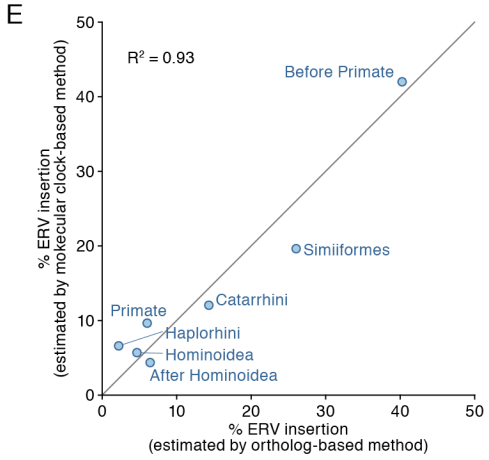

B
Ortholog not in non-mouse
False
True
After Murinae
Ortholog not in non-Murinae
False
True
Murinae
Ortholog not in non-Muroidea
False
True
Muroidea
Ortholog not in non-Rodentia
False
True
Rodentia
Before Rodentia


C
Before Primate
Primate
Haplorhini
Simiiformes
Catarrhini
Hominoidea
After Hominoidea
90 74 67 43 29 20
MYA
Homo sapiens
Pan troglodytes
Gorilla gorilla
Pongo abelii
Nomascus leucogenys
Macaca mulatta
Macaca fascicularis
Papio anubis
Chlorocebus sabaeus
Callithrix jacchus
Saimiri boliviensis
Tarsius syrichta
Otolemur garnettii
Microcebus murinus
Tupaia belangeri
Mus musculus
Ictidomys tridecemlineatus
Oryctolagus cuniculus


F
Before Rodentia
Rodentia
Muridea
Murinae
After Murinae
73 70 33 22
MYA
Mus musculus
Rattus norvegicus
Cricetulus griseus
Dipodomys ordii
Ictidomys tridecemlineatus
Heterocephalus glaber
Cavia porcellus
Oryctolagus cuniculus
Ochotona princeps
Homo sapiens
Macaca mulatta
Callithrix jacchus
Carlito syrichta
Otolemur garnettii
Tupaia belangeri


D
Before Primate
Primate
Haplorhini
Simiiformes
Catarrhini
Hominoidea
After Hominoidea
0 10 20 30 40
% ERV insertion
(estimated by ortholog-based method)


G
Before Rodentia
Rodentia
Muridea
Murinae
After Murinae
0 10 20 30 40
% ERV insertion
(estimated by ortholog-based method)


E
R² = 0.93
Before Primate
Simiiformes
Catarrhini
Primate
Haplorhini
Hominoidea
After Hominoidea
0 10 20 30 40 50
% ERV insertion
(estimated by mokecular clock-based method)
% ERV insertion
(estimated by ortholog-based method)

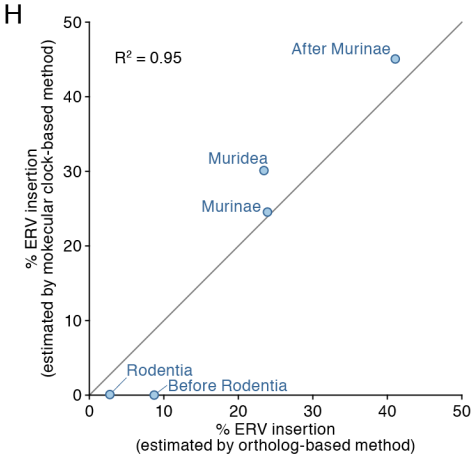

H
R² = 0.95
After Murinae
Muridea
Murinae
Rodentia
Before Rodentia
0 10 20 30 40 50
% ERV insertion
(estimated by mokecular clock-based method)
% ERV insertion
(estimated by ortholog-based method)

Fig. S9. Estimation of insertion dates of ERVs using the ortholog distribution-based and genetic distance-based methods. (*A* and *B*). Flowchart showing the procedure used to estimate the insertion dates of ERVs based on ortholog distributions. The procedures used for ERVs in the genomes of humans (*A*) and mice (*B*) are shown. (*C*-*E*) Comparison of ERV insertion dates estimated by the ortholog distribution-based and genetic distance-based methods (human ERVs). (*C*) Stratification of human ERVs according to estimated insertion dates. (*D*) Amount of ERV insertions acquired in distinct time periods as estimated using an ortholog distribution-based approach. (*E*) Comparison of ERV insertion amounts estimated by the two methods. In the genetic distance-based method, 2.2×10^{-9} mutation/site/year was used as the mutation rate. The line y = x and coefficient of determination (R^2) are shown. (*F*-*H*) Comparison of ERV insertion dates estimated by ortholog distribution and genetic distance-based methods in mouse ERVs. (*F*) Stratification of mouse ERVs according to the estimated insertion dates. (*G*) Amount of ERV insertions acquired in distinct time periods as estimated using the ortholog distribution-based methods. (*H*) Comparison of ERV insertion amounts estimated by the two methods. In the genetic distance-based method, 7×10^{-9} mutation/site/year was used as the mutation rate. The line y = x and coefficient of determination (R^2) are shown.