Science
AAAS

Supplementary Materials for

Evolutionary constraint and innovation across hundreds of placental mammals

Authors: Matthew J. Christmas[1]†, Irene M. Kaplow[2,3]†, Diane P. Genereux[4], Michael X. Dong[1], Graham M. Hughes[5], Xue Li[4,6,7], Patrick F. Sullivan[8,9], Allyson G. Hindle[10], Gregory Andrews[7], Joel C. Armstrong[11], Matteo Bianchi[1], Ana M. Breit[12], Mark Diekhans[11], Cornelia Fanter[10], Nicole M. Foley[13], Daniel B. Goodman, Linda Goodman[14], Kathleen C. Keough[14,15,16], Bogdan Kirilenko[17,18,19], Amanda Kowalczyk[2,3], Colleen Lawless[5], Abigail L. Lind[15,16], Jennifer R. S. Meadows[1], Lucas R. Moreira[4,7], Ruby W. Redlich[20], Louise Ryan[5], Ross Swofford[4], Alejandro Valenzuela[21], Franziska Wagner[22], Ola Wallerman[1], Ashley R. Brown[2,3], Joana Damas[23], Kaili Fan[7], John Gatesy[24], Jenna Grimshaw[25], Jeremy Johnson[4], Sergey V. Kozyrev[1], Alyssa J. Lawler[3,4,20], Voichita D. Marinescu[1], Kathleen M. Morrill[4,6,7], Austin Osmanski[25], Nicole S. Paulat[25], BaDoi N. Phan[2,3,26], Steven K. Reilly[27], Daniel E. Schäffer[2], Cynthia Steiner[28], Megan A. Supple[29], Aryn P. Wilder[28], Morgan E. Wirthlin[2,3,30], James R. Xue[4,31], Zoonomia Consortium§, Bruce W. Birren[4], Steven Gazal[32], Robert M. Hubley[33], Klaus-Peter Koepfli[34,35,36], Tomas Marques-Bonet[37,38,39,40], Wynn K. Meyer[41], Martin Nweeia[42,43,44,45], Pardis C. Sabeti[4,31,46], Beth Shapiro[29,47], Arian F. A. Smit[33], Mark Springer[48], Emma Teeling[5], Zhiping Weng[7], Michael Hiller[17,18,19], Danielle L. Levesque[12], Harris A. Lewin[23,49,50], William J. Murphy[13], Arcadi Navarro[37,39,51,52], Benedict Paten[11], Katherine S. Pollard[15,16,53], David A. Ray[25], Irina Ruf[54], Oliver A. Ryder[28,55], Andreas R. Pfenning[2,3], Kerstin Lindblad-Toh[1,4]‡*, Elinor K. Karlsson[4,7,56]‡*

Correspondence to: kersli@broadinstitute.org (KLT) and elinor.karlsson@umassmed.edu (EKK)

This PDF file includes:

Other Supplementary Materials for this manuscript include the following:

Materials and Methods

Alignment update

After inferring the initial 242-way alignment used in (*11*), we realized that the tarsier (*Carlito syrichta*) genome we had included was a mislabeled kangaroo rat. We removed the mislabeled genome from the alignment using the halRemoveGenome command in the HAL toolkit (*181*). Due to the progressive alignment process, the misplaced genome likely negatively influenced the reconstruction of the ancestors above it in the guide tree and thus the alignment from the rest of the primates (the clade it was basal to) to the other clades. To remedy this, we re-aligned the affected part of the alignment (the parts involved in inferring all ancestors above tarsier) to remove the influence of the mislabeled genome without requiring re-alignment of unaffected parts of the tree. Since subtrees not below tarsier were not affected (due to their alignments not including tarsier as a descendant or outgroup genome), they did not require re-inference. We constructed a re-inference alignment with guide tree:

((((((fullTreeAnc110:0.1523,Galeopterus_variegatus:0.12765)fullTreeAnc112:0.01,fullTreeAnc113:0.15442)fullTreeAnc114:0.04,fullTreeAnc69:0.04)fullTreeAnc115:0.0406,fullTreeAnc237:0.0212)fullTreeAnc238:0.0237,fullTreeAnc14:0.02)fullTreeAnc239;

where we obtained the *fullTreeAnc* ancestors used as leaves from the original HAL. We used platypus as an outgroup genome. We used the relationships and ancestral sequences from the re-inferred "supertree" alignment and merged them with the existing, original alignments between species in each subtree (using halReplaceGenome from the HAL toolkit). In the process, we removed the now-redundant ancestor above tarsier (*fullTreeAnc111*). The alignment improved after removal of the duplicate genome and re-inference: coverage of the human genome from non-primate species increased by an average of 2.2% of the human genome, which was an average 6% relative increase from the coverage in the original alignment.

Identifying genes in Zoonomia genomes using halLiftover

For the *halLiftover*-based ortholog identification, we began with human protein-coding and exon sequence annotation from ENSEMBL (BioMart v99; Downloaded March 2020) (*182*). For each gene investigated, we chose the longest transcript. We used halLiftover in conjunction with Zoonomia Cactus alignment (*4*) to identify sequences orthologous to the protein-coding sequence of each exon across each of the 241 assemblies. These sequences were prone to frameshift mutations, internal stop codons, and likely premature truncations. Those issues were likely due to missing sequence or small errors in the sequence alignments and amplified by the large evolutionary distance. We therefore performed two variations of post-processing steps.

For transcripts marked “Pfenning,” we first choose the ortholog of the annotations from the most closely related of human, goat, or mouse, a set of high-quality genomes and annotations that span a large segment of the mammalian evolutionary tree. We mapped the set of ENSEMBL exons for the reference sequence to the target species using halLiftover on the Zoonomia Cactus alignment (*11*, *181*). We matched each lifted over exon sequence fragment to reference exons by translating it into an amino acid sequence and shifting start and end points to match exon boundaries. We then merged exon boundaries only if they did not create a frameshift mutation, did not create an internal stop codon, or did not increase the size of the exon by greater than or equal to three-fold.

For transcripts marked “Broad,” we use the halLiftover outputs from mapping the human gene coordinates to each species. We smoothed the halLiftover output for each protein-coding gene/species/contig combination by making a single interval from the first and last coordinates for

each gene ortholog in the halLiftover output. We padded both ends with 500bp and then, for each species, extracted the genome sequence in that interval. We then applied Exonerate protein2genome (*183*) to translate the original human gene into protein sequences and then to predict exons and introns in each species by finding sequences matching the human sequences within the smoothed, padded halLiftover outputs while accounting for splice site sequences.

For both methods, we considered a transcript to be "valid" if the predicted protein sequence started with methionine, was contained on a single contig, and was within 90-110% of the length of the human reference protein. If a transcript was not "valid," we repeated this process for the next-longest transcript for the gene; if there were no additional transcripts in a species, we did not report an annotation for that gene, species combination. If "valid" transcripts were found for multiple contigs, we reported only the first valid transcript found. We reported transcripts along with their identity and similarity scores (from Exonerate) on both the exon and transcript level. We also reported insertions and deletions (from Exonerate) at the exon level. For our final annotations, we excluded all annotations of genes that do not have human orthologs.

Given that genes are sporadically missing from each genome due to genome quality issues, cataloging "essential" genes found in all placental mammals is not feasible. We annotated just 116 genes in all 240 Zoonomia species. When we consider only the highest-quality assembly in each order, this increases to 2,718 genes, still far fewer than the 9,226 included in the BUSCO version odb10 gene set of mammalian single-copy orthologs (*184*).

We have released our data on the following website: http://genome.ucsc.edu/cgi-bin/hgGateway?genome=Homo_sapiens&hubUrl=http://cgl.gi.ucsc.edu/data/cactus/241-mammalian-2020v2-hub/hub.txt. We have labeled each entry with the corresponding human gene name, the corresponding human transcript name, the protein ENSEMBL ids, the gene name in the species if it was available, and the method used to generate the annotation.

Identification of *CMAH* gene loss

Determining gene loss from genome assemblies can be confounded by mis-assemblies and by assembly gaps. To avoid these confounding factors, we combined two approaches, one using short reads and one using genome assemblies, to determine whether the *CMAH* gene had been lost in mammalian genomes. We selected a reference genome with an intact, well-annotated *CMAH* gene for each of the species. For the short-read approach, we obtained publicly available Illumina paired-end reads obtained from the NCBI Short Read Archive (*185*) for each species and aligned them to their assigned reference genomes with bowtie2 using the very-sensitive-local parameter (*186*). We determined read coverage of the *CMAH* coding sequence using bedtools (*187*). For the genome assembly analysis approach, we aligned the protein sequence of the *CMAH* gene from the assigned reference genome to each genome assembly using exonerate (*183*). We evaluated results from the exonerate analysis by comparing them to the Cactus multiple genome alignment. We considered putative *CMAH* gene loss events to be cases where both the short-read alignment coverage and the exonerate alignment indicated gene loss and where both approaches indicated loss of the same part of the gene. Custom python scripts used for processing outputs from bedtools and exonerate are available on the github repository.

Mammalian neutral model for constraint scoring

We used the alignment of 241 mammals (*4*) in HAL-format (HAL Tools v2.1) (*181*) as the input to generate three different neutral models. The first was a general model, used for autosomes, and the other two were for each of the sex chromosomes (chrX and chrY). We used

the resulting nucleotide substitution rate matrices to generate conservation scores. For this process, we first identified the ancestral repeats by running RepeatMasker Open-4.0. 2013-2015 (*188*) on the ancestral sequence of the mammal alignment. We used sequence from the second most ancestral branch (fullTreeAnc238) instead of the most ancestral sequence (fullTreeAnc239), as interspersed repeats on fullTreeAnc238 had better reconstruction of the eutherian ancestral form than in fullTreeAnc239 (personal communication A. Smit) (*188*). We converted repeat coordinates to fullTreeAnc239 using halLiftover (*181*). Then, we filtered ancestral sequence repeat sets to exclude i) non-mammalian repeats shared with birds and reptiles; ii) repeat sequences annotated as structural RNA copies, satellites, tandem repeats, and low complexity annotations; and iii) regions where synteny was not present across the four major branches of the mammalian tree: Xenarthra, Afrotheria, Laurasiatheria, and Euarchontoglires (reconstruction using halLiftover).

The input for the neutral evolution model calculation was a random set of ancestral repeat positions (100kb total) selected from the outputs of halLiftover. The use of ancestral repeats as representative of neutrally evolving sequences is supported by evidence for a lack of purifying selection acting on ancestral repeats (*38*). We extracted an ancestor-referenced multiple alignment formatted file (MAF) alignment (hal2maf) and used PhyloFit from Phast v1.5, with default parameters (--subst-mod REV --EM) and with the corrected tree, to estimate branch lengths while fixing the tree topology (*181*, *189*). We processed the resultant alignment (mafDuplicateFilter) to filter sequences that aligned multiple times to the same region (*190*). This step avoided potential biases due to species over-alignment. For sex chromosome-specific models, we converted ancestral repeat coordinates from the HAL alignment to human hg38 coordinates (halLiftover) (*181*). We selected random sets of repeat positions from either chromosome X or Y of the human outputs of halLiftover (i.e., 100Kb as above) to calculate the neutral models specific to each sex chromosome. We constructed those models separately from the general model because sex chromosomes evolve at different rates to autosomes (*191*). We later used all these models to compute phyloP scores.

We used the same method as above to estimate primate-neutral models, with the difference being that ancestral branch reconstruction was based on the 43 primates from the alignment. For these, we used human as the reference. We evaluated synteny with six of the reconstructed branches in the primates tree (fullTreeAnc112, fullTreeAnc107, fullTreeAnc103, fullTreeAnc88, fullTreeAnc78, fullTreeAnc70). We used these primate-neutral models to compute primate-specific PhastCons scores.

<u>Alignment pre-processing, and alignment depth, nucleotide distribution, and branch length calculations</u>

We converted the HAL alignment to a species-referenced MAF-format alignment using hal2maf (*181*), then filtered out species duplicates in the alignment using mafDuplicateFilter (https://github.com/dentearl/mafTools) in order to avoid scoring biases due to over-alignment (*190*). We allowed one sequence per species (i.e. the best match) in each alignment block to remain, which we chose by comparing the sequence to the consensus for the block (*190*). We repeated this for multiple reference species, ultimately creating human, chimpanzee, house mouse, dog, and little brown bat reference-based alignments.

For the primate-specific data, we filtered out non-primate species from the alignment using the mafSpeciesSubset command from mafTools (*190*). We collected the alignment depth and list of species aligned across the alignment using the Bio.Align package from BioPython

(https://biopython.org/wiki/Multiple_Alignment_Format) (*192*). We calculated total branch lengths for each of the alignment blocks in the MAF alignment files using the tree_doctor command with the branch length (–branchlen) option from the PHAST software package (*193*).
PhyloP constraint score calculation: We used phyloP (part of the PHAST v1.5 package: https://github.com/CshlSiepelLab/phast) to calculate per-base constraint and acceleration p-values (*194*). We presented scores as -$\log_{10}$ p-values under a null hypothesis of neutral evolution, where computation involved performing a likelihood ratio test at each alignment column (--method LRT) with constraint and acceleration scores outputted (--mode CONACC, negative values indicate acceleration). We calculated phyloP scores on the human-referenced, 241-way, MAF-formatted, duplicate-filtered alignment. Scores ranged from -20.0 to 8.903 for autosomes, -20.0 to 7.765 for chromosome X, and -20.0 to 9.280 for chromosome Y; the different ranges of scores between the autosomes, chromosome X, and chromosome Y are likely due to the differences in the models used for each. Scores from the dog and little brown bat-referenced 241-way MAF-formatted alignments ranged from only -20 to 8.903, as their assemblies did not include chromosome Y. Scores from the human-referenced, primates-only alignment (43-way) ranged from -20 to 1.264. Scores calculated on the primate subgroup had lower ranges than mammals, as the total branch lengths of clade-specific trees were relatively low compared to the entire mammalian tree. All associated scripts for calculating phyloP scores are available on Github (https://github.com/michaeldong1/ZOONOMIA).

We used phyloP because it allows us to use all our genomes to quantify the constraint of individual nucleotides of interest. Thus, our phyloP scores can be used for deciphering which of multiple near-by nucleotides is likely to be functional, such as identifying which of multiple genetic variants associated with a disease in linkage disequilibrium is likely to be causal. However, a limitation of phyloP is that it does not compute confidence intervals around estimates of constraint. Computing confidence intervals would require re-computing phyloP scores with hundreds or more different subsets of filtered ancestral repeats, which is computationally intractable.

PhastCons constraint score calculation: We also used PhastCons, another part of the PHAST package, which uses a phylogenetic hidden Markov model (phylo-HMM) to identify evolutionarily conserved elements (*25*, *195*). In contrast to phyloP single-base constraint, PhastCons metrics incorporate the columns of flanking bases. We calculated two outputs: a per-base constraint score and a set of conserved elements coordinates (--viterbi option). We set the model parameters to match those used to generate the PhastCons output for the UCSC 100 Species Vertebrate Multiz Alignment & Conservation (expected-length=45, target-coverage=0.3, rho= 0.31, http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=cons100way)(*84*).

Constraint score thresholds:

We computed a mammalian phyloP threshold by converting the p-values corresponding to the phyloP scores into q-values using a false discovery rate (FDR) correction (*196*) (R function *qvalue*). We classified outputs as constrained or accelerated based on the sign of the score; we considered any column with a resulting q-value ≤ 0.05 to be significantly evolutionarily constrained or accelerated (5% FDR, 240 mammal phyloP constraint score ≥ 2.27, 3.528% of the human genome). We took a different approach for primate constraint given the fewer species (n=43) and shorter branch lengths, as a phyloP score across primate species would be underpowered to discriminate the highly constrained bases from the background (*12*). We identified the PhastCons threshold that yielded a similar fraction of the genome under constraint

as for all mammals (i.e., PhastCons base score ≥ 0.961, 3.54% of the human genome). The locations of significantly constrained bases in mammals and primates overlap, which is not surprising because the 43 primates we used are a subset of the 241 mammalian assemblies. We evaluated the mammalian phyloP threshold, the use of different scores to measure constraint in mammals and primates, and the base pair resolution of mammalian phyloP scores by comparing our results to expectations of amino acid and transcription factor motif conservation (**Figs. 2 and 3**) and by using heritability analyses of human diseases and complex traits (*24*).

Lower bound estimates of genome-wide constraint: We estimated the lower bound for the fraction of sites under purifying selection across the human, chimpanzee, dog, house mouse, and little brown bat genomes (π) by comparing the empirical cumulative distribution functions (ECDF) of phyloP scores across each genome to the ECDF of ancestral repeats, following the same method detailed in (*12*) where:

$\pi = 1 - \min_s F(s)/G(s)$.

Here, F is the ECDF for all sites across the genome (a function of scores s), and G is the ECDF for ancestral repeats (i.e., sites not under any evolutionary pressure). We extracted coordinates of ancestral repeats from the UCSC repeatmasker track and filtered these to retain a set of ancestral repeats present in the four clades of the mammalian tree (Xenarthra, Afrotheria, Laurasiatheria, and Euarchontoglires) (*81*, *188*). To account for the lack of certainty in the neutrality of ancestral repeats we generated 500 bootstrapped samples of ancestral repeats for each genome using the 'equate' package in R v.4.0.4. We calculated the ECDFs of phyloP values for all positions across each genome as well as for each set of bootstrapped ancestral repeats (500 sets per genome) using the ecdf function in R v.4.0.4 (*197*). We excluded phyloP values below -1.5 so that the extreme left tail of the ECDFs did not heavily influence estimates of constraint. We took the median values of π as our lower bound estimates of constraint due to skewed distributions and generated 95% confidence intervals around our estimates using the 'wilcox.test' function in R stats package, setting 'conf.int = TRUE'. Values of π represent the proportion of bases under constraint in our genome-wide phyloP score sets and therefore do not include any unaligned positions. To achieve a value for the percentage of the whole genome that is under constraint, we first calculated the number of bases under constraint by multiplying the total number of positions with phyloP scores by π. The number of bases under constraint divided by the genome size is then the lower bound estimate for the proportion of the genome under constraint. Median values for lower bound estimates of the number of bases under constraint and their 95% confidence intervals are as follows: human 331.69 Mbp (331.61, 332.81), chimpanzee 358.77 Mbp (358.34, 360.73), mouse 238.54 Mbp (237.91, 239.54), dog 244.81 Mbp (244.18, 245.48), and bat 367.01 Mbp (367.00, 367.02).

These estimates may be influenced by any differences in base composition and substitution patterns between ancestral repeats and the rest of the genome (*12*). However, we note that for humans, the overall base composition of ancestral repeats is very similar to the genome-wide base composition (40.1% GC versus 41.0% GC respectively) and the distribution of ancestral repeats along each individual chromosome does not differ from a random distribution (tested using the regionR package in R, where we assessed whether the locations of ancestral repeats significantly overlaps the locations of a randomly distributed set of sequences with the same size distribution using a permutation test; the number of overlaps was significantly greater than expected by chance, $p = 0.01$).

For the human genome, our lower-bound estimate of genome-wide constraint of ~10.7% is about three times the number of bases identified as significantly under constraint at an FDR of

<0.05 (see above). The comparison of phyloP CDFs for the whole genome and for ancestral repeats provides a measure of the amount of the genome demonstrating higher conservation than observed for neutrally evolving sequences and therefore likely captures a lot of clade-specific (i.e., primates) constraint. Our stringent use of FDR=0.05 means that, whilst we are highly confident that positions below this threshold are under constraint, many constrained positions (particularly clade-specific ones) will not be captured, as our power to identify constrained positions is limited by the tree breadth and depth.

Previous attempts to estimate the proportion of the human genome under constraint have placed it at between 3 and 12% (*13*, *40*, *198*). For example, a more recent estimate of 8.2% (7.1-9.2%) (*198*) was computed using an empirical null based on a geometric distribution fit using distances between indels in pairwise alignments within ancestral repeats, and it would miss constrained nucleotides in short regions between indels. Other lower estimates used only 29 mammals (*13*) or 34 mammals but only 1% of the genome (*12*). All these previous methods used reference-based alignments that cannot handle some of the structural rearrangements that Cactus can align, such as inversions (*199*). Thus, our estimate of 10.7% is higher than previous estimates for several possible reasons, including the method used for quantifying constraint, the number of species considered, and the use of a Cactus alignment.

It has been argued that the amount of constraint cannot exceed 15% based on considerations of mutational load (*200*), although this approach is disputed (*201*). Our estimate is below this theoretical limit.

Differences between species in the proportion of genome under constraint

We found substantial differences between the proportions of constrained sequence in different species, with little brown bat having the highest proportion. To evaluate whether this may be a result of the variation in the number of closely related species in the alignment for these species rather than a true reflection of differences in constraint, we calculated the total branch length for the set of each species and each species's nine most closely related species. The correlations between these branch lengths and both the percentage ($R=-0.06$; $p=0.9$; $n=5$) and the amount of constrained sequence ($R=-0.47$; $p=0.62$; $n=5$) were not significant.

Constraint in protein-coding sequence

We obtained protein-coding annotations from GENCODE v.36 (*202*). We selected all protein-coding transcripts and checked for consistency in the cDNA, CDS, and peptide lengths (the number of amino acids in the peptide; the protein-coding length had to be 3x the peptide length, and the start codon had to be methionine.). Since most genes have many transcripts, we created a transcript hierarchy to pick one transcript per gene in this order (pickOne): MANE transcript (*203*), gnomAD canonical transcript (*64*), BUSCO transcript (*204*), protein-coding transcript, lncRNA transcript, pseudogene transcript, and other transcript. For each protein-coding transcript, we identified its structural parts: proximal promoter (500bp upstream of the TSS), TSS, 5'UTR exons, start codon, coding exons, canonical intronic splice sites, introns, stop codon, and 3'UTR exons. Some protein-coding genes do not have two UTRs, and some are intronless. We extracted phyloP scores for all positions in protein-coding genes including 5' and 3' UTRs (N=34,284,184 bp) to compare constraint between different positions within coding sequences. We tested the top 5% most accelerated and most conserved genes, measured by mean phyloP score, against a non-redundant representative set of Gene Ontology biological processes using WebGestalt (*46*, *205*) and identified the top 10 overrepresented gene sets.We ran

WebGestalt from the web interface at http://www.webgestalt.org/ and used the default parameters (minimum number of IDs in the category: 5; maximum number of IDs in the category: 2000l; FDR Method: BH) and a reference gene set of all mapped entrezgene IDs.

We defined degeneracy at each protein-coding position using International Union of Pure and Applied Chemistry (IUPAC) codes and used R v.4.1.1 to summarize mean and standard deviation phyloP scores for positions within codons, degenerate and non-degenerate positions, methionines that act as and do not act as start codons, and cysteines that form and do not form intra-peptide disulfide bridges. We obtained locations of intra-peptide disulfide bridges from the UCSC Genome Browser (*83*): https://genome.ucsc.edu/cgi-bin/hgc?hgsid=799000851_rdDAOY8aECs6ODSPVUt92mNOp7fR&c=NC_045512v2&l=14950&r=24950&o=22732&t=23137&g=unipCov2DisulfBond&i=disulf+bond.

Human variation at sites under constraint

We intersected all single nucleotide polymorphisms (SNPs) in TOPMed data freeze 8 (https://topmed.nhlbi.nih.gov/), a resource of 811 million SNPs from ~186 thousand individuals (*69*), with their phyloP scores. We compared minor allele frequencies for SNPs in constrained positions (phyloP > 2.27) to those in non-constrained positions using the Wilcoxon rank sum test in R v.4.1.1. We predicted the functional impact of all SNPs using SNPEff (*70*).

Constrained four-fold degenerate sites within transcription factor binding sites

To determine whether certain transcription factors (TFs) tend to bind to constrained coding sequence [as indicated by constraint in four-fold degenerate sites (4D sites), which otherwise would not be expected to show constraint], we used bedtools intersect (*187*) to identify overlap of 4D sites under constraint (phyloP > 2.270) and ENCODE transcription factor clusters downloaded from the UCSC Genome Browser (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/encRegTfbsClustered/) (*187*, *203*). We took all the TFBS with cluster score = 1000 (reflects strength of the ChIP-seq signal, which are the highest-confidence TFBS) for each TF with TFBS in the database and counted the total number of constrained and non-constrained 4D sites overlapping TFBS for each TF. We then calculated a relative difference between constrained and non-constrained 4D sites for each TF:

$$\left(\frac{DcTF}{Dc} - \frac{DnTF}{Dn}\right) \times 100$$

Where *DcTF* = number of 4D constrained sites for each TF, *Dc* = number 4D constrained sites, *DnTF* = number of 4D non-constrained sites for each TF, and *Dn* = number of 4D non-constrained sites, giving a percentage constraint excess (positive) or deficiency (negative) for each TF. We observed that a subset of TFs show an excess of 4D constraint in their binding sites and so we further tested whether the type of transcription factor, as well as mean phyloP for the gene and amino acid coded for, have an effect on the percent of excess constrained 4D sites in TFBS using ANOVA using the 'aov' function and 'anova_test' from the rstatix package to calculate effect sizes in R v.4.1.1.

Constraint in functional elements

We calculated constraint enrichment for several genome features (coding sequences, 5' UTRs, 3' UTRs, introns, DHS, and the five types of cCREs). We obtained coding sequences, 5' UTRs, 3' UTRs and introns from GENCODE v.36 as described above (*206*). We obtained regulatory features from ENCODE3, including DNase hypersensitive sites (DHS; 243 cell lines

and tissues) (*207*) and candidate *cis*-regulatory elements (cCREs) (*14*). cCREs include canonical promoter-like signatures (PLS; ± 200 bp of GENCODE TSS, high DHS and H3K4me3 signals), proximal enhancer-like signatures (pELS; ± 2 kb of GENCODE TSS, high DHS and H3K27ac signals, low H3K4me3 signal), distal enhancer-like signatures (dELS; > 2 kb from GENCODE TSS, high DHS and H3K27ac signals), DNase-H3K4me3 elements (promoter-like biochemical signature that are not within 200bp of an annotated TSS), and CTCF-only (high DNase and CTCF and low H3K4me3 and H3K27ac).

First, we calculated constrained fractions for each feature as the number of positions with phyloP above FDR=0.05 threshold/the total number of positions. Next, we calculated constraint enrichment as the constrained fraction of the feature divided by the constrained fraction of the genome. We assessed the significance of enrichment for each feature using a χ^2 test in R v.4.1.1.

Identifying regions of high constraint

Zoonomia ultraconserved elements (zooUCEs)

We extracted all positions in the alignment bed files where the number of species aligned was ≥ 235 and the base was the same among all species aligned at that position. We then merged neighboring positions, creating elements of ultraconserved sequences ranging in size from 2bp to 190bp. The final set of zooUCEs contains all elements ≥ 20bp. We assessed overlap between our zooUCEs and previously defined UCEs (*73*) using bedtools intersect and the "-u" flag to report all zooUCEs that overlap the original UCEs (*187*).

Regions of contiguous constraint (RoCCs)

We extracted all constraint positions (phyloP ≥ 2.270) from the phyloP bed files and grouped neighboring constraint positions to create runs of constraint ≥ 2bp. We noticed that many of these runs were separated by only a single base pair below the phyloP threshold, so we merged all runs of constraint that were separated by only a single base, creating Regions of Contiguous Constraint (RoCC).

Constraint across the genome - 100kb bins

We divided the human genome into 100kb bins, excluding chrY, any bins with < 80kb positions with phyloP values, and bins < 100kb (i.e., ends of chromosomes), for a total of 28,218 bins. We then fit a linear model of the number of constrained bases within each bin using the total number of bases with phyloP values in the bin. We included several covariates in the model to control for differences between bins that could influence estimates of constraint: the number of coding bases, the number of positions with a low number of species aligning (≤24), and a measure of mappability for each bin. For mappability, we used the "k24" UCSC Genome Browser track (*83*), the fraction of 24 x 24-mer sequences overlapping a given base (i.e., the current base ± 23 bp) that map uniquely to the genome (1 = all map uniquely, 0 = none map uniquely), and we summed the number of positions in each bin with k24 score > 0.9. We square root-transformed all values. We then calculated p-values from the absolute studentized residuals for each bin using the R *pnorm* function (using absolute values meant that low p-values represent significant excess or depletion of constraint), and we computed q-values from the p-values using the R function *qvalue*. We considered any bin with q-value ≤ 0.05 to be significant. We merged adjacent significant bins using *bedtools merge* (*187*). Analysis was carried out using R v4.1.1. We identified 54 bins as significant, 53 with excess and one lacking constraint. We performed a

gene set enrichment analysis on the 53 bins with excess constraint using GREAT version 4.0.4 (*106*) with default settings.

Constraint in gene deserts

We used the set of developmental TFs identified in (*208*), which are all TFs known to be involved in the regulation of developmental processes (2,863 genes). We defined intergenic regions as all regions between ensembl genes (GRCh38.103) and calculated length and proportion of constraint (number of positions with phyloP ≥ 2.270) for each intergenic region. We labeled all intergenic regions as either "neighboring developmental TF" or "other" if they do or do not border a developmental TF, respectively. We extracted all gene deserts, defined as the longest 5% of intergenic regions, and compared constraint between gene deserts neighboring developmental TFs and gene deserts neighboring other genes using a Wilcoxon rank sum test in R v. 4.0.4. We tested for a significant difference in the proportion of constraint in ENCODE3 cCREs located within gene deserts neighboring developmental TFs versus those in other gene deserts with a Wilcoxon rank sum test in R v. 4.0.4. We assessed whether genes neighboring gene deserts were enriched for developmental TFs using Fisher's exact test ("fisher.test" function in R v.4.0.4).

Identifying Unannotated Intergenic Constraint Regions (UNICORNs)

To investigate constraint in regions lacking a functional annotation, we used the "subtract" tool in bedtools v.2.29.2 (*187*) to remove autosomal sequences with the following annotations: GENCODE v37 exons (UTRs and exons for all protein-coding genes) and promoters (TSS +/- 1kb), introns, ENCODE3 cCREs, DHS (including TF binding sites), ChIA-PET anchors, three promoter annotation sets, and 6 enhancer annotation sets (**Table S9**). We used liftOver (*209*) to convert annotations to hg38 coordinates when the annotations were in other assemblies. We then identified groups of closely located constraint positions within the "unannotated" sequences: We identified any two unannotated constraint positions within 5bp of each other and retained all clusters of such positions greater than 10bp in length, giving a set of 423,586 unannotated intergenic constraint regions (UNICORNs). We generated a set of non-constraint clusters to compare with the UNICORNs by identifying all unannotated intergenic regions greater than 10bp that did not contain any constraint positions (phyloP less than 2.270 for all positions). This generated over 6 million clusters, which we randomly sampled to generate a set to match the number of UNICORNs.

We used the "intersect" tool in bedtools v2.30.0 (*187*) to calculate the distance of each UNICORN and non-constraint cluster to its (linearly) closest gene, cCRE, dhs, and chromatin loop anchor (**Table S9)**. We identified SNPs in each cluster from TOPMed data freeze 8 (*69*) and calculated average minor allele frequency and number of SNPs per bp (0-1). We then binned our UNICORNs and randomly sampled non-constraint clusters into five bins based on size (11-20bp, 21-50bp, 51-100bp, 101-500bp, 501-1,325bp). For each bin, we compared each of these measures between UNICORNs and matched non-constraint elements using Wilcoxon rank sum tests with a Bonferroni correction for multiple testing in R v.4.1.1.

To determine the extent to which UNICORNs overlap open chromatin regions from neurons in specific brain regions and motor cortex cell types as well as different parts of the brain at different developmental stages, we downloaded the relevant datasets and then determined their overlap with the UNICORNs. Specifically, for open chromatin in neurons from specific brain regions, we downloaded the NeuN+ ATAC-seq (*210*) data from (*102*) and processed them as

described for the neurons from primary motor cortex from this dataset in (*15*). For open chromatin from specific motor cortex neuronal cell types, we used the reproducible human motor cortex neuron open chromatin regions (*99*) processed in the same way as in our companion paper except for, in ArchR's addGroupCoverages, using an order of magnitude higher values for maxCells and maxFragments (*159*, *211*). For open chromatin from different parts of the brain at different developmental stages, we used the file GSE149268_annotation-ocr-hg38.bed.gz downloaded from GSE149268 (*100*); these open chromatin datasets came from later in development than the ENCODE fetal brain open chromatin data, and, unlike the ENCODE fetal brain open chromatin, which came from the whole embryo brain, these datasets came from specific brain regions (*14*, *212*). To compute the number of UNICORNs and nucleotides within UNICORNs covered by each of these datasets as well as by the union of these datasets, we used coverageBed from bedtools (*187*). To compute the number of UNICORNs covered by one of the datasets and not the others, we used subtractBed and coverageBed from bedtools (*187*). We found that the percentage of UNICORNs covered was close to the percentage of nucleotides within UNICORNs covered.

Curating transposable elements

The detailed methodology of TE curation efforts can be found in the companion paper (*110*). Here, we summarize the major steps of the process. First, we gathered final genome assemblies for each species under examination and identified putative TEs using RepeatModeler-4.0.9 (*46*). We assumed that most elements shared by large groups of mammals had already been described, and we focused instead on relatively new putative TEs as defined by those having hits with K2P distances (*213*) less than 4.4% (~20my or less since insertion) based on a general mammalian neutral mutation rate of 2.2×10^{-9}. Custom scripts associated with this part of the analysis are available on github (https://github.com/davidaray/bioinfo_tools).

For each iteration of manual TE curation, we generated new consensus sequences to match the top 50 blast (*51*) hits. Upon completion of the species-by-species curation step, we identified and removed duplicate TEs within the final library according to the 80-80-80 rule (*55*) and utilizing the program cd-hit-est (https://doi.org/10.1093/bioinformatics/btq003). We used blastx (*56*), RepBase (*47*), and TEclass (*57*) were to organize the library by TE type using distinguishable hallmarks of each respective TE class (*214*). We combined our library with the known vertebrate TE library (*215*) and did deduplication.

We then used the final library as an input for Repeatmasker-4.1.0 (*8*) to mask the genome assemblies with this custom library. We performed postprocessing of the outputs using a custom script, RM2Bed.py (available at https://github.com/davidaray/bioinfo_tools), which eliminates overlapping hits and converts files to .bed format.

Initial RepeatModeler output yielded 25,025 initial queries, ~101/assembly. After de novo curation and elimination of duplicates, the final library consisted of 7,707 consensus sequences. We have deposited these in the dfam TE database (https://dfam.org/home).

Constraint in repeats

We downloaded repeat annotations for the human genome from the UCSC repeatmasker (*188*) track: https://hgdownload.cse.ucsc.edu/goldenPath/hg38/database/rmsk.txt.gz. We counted the number of constraint positions in each repeat contained in this database. For each repeat class ("Simple_repeat", "Low_complexity", "rRNA", "tRNA", "DNA", "Satellite", "LINE", "RC",

"srpRNA", "SINE", "Retroposon", "snRNA", "scRNA", "LTR", "RNA", and "Unknown") we calculated an estimate of excess constraint per repeat type:

$$\left(\frac{rc}{c} - \frac{rn}{n}\right) \times 100$$

where *rc* = total constrained positions in repeat class, *c* = total constrained positions within repeats, *rn* = total non-constrained positions in repeat class, and *n* = total non-constrained positions within repeats, obtaining a percentage constraint excess (positive) or deficiency (negative) for each repeat type.

Since simple repeats showed an excess of constraint, we extracted all simple repeats from the dataset and calculated distance to the nearest protein-coding gene for each one using bedtools closest (*187*) and RefSeq gene annotations for GRCh38 (https://www.ncbi.nlm.nih.gov/projects/genome/guide/human/index.shtml) (*188*, *216*). We then calculated the Pearson product-moment correlation coefficient between the proportion of constraint and the distance to the nearest gene using the cor.test function in R v. 4.1.1.

Identifying olfactory receptor genes

The olfactory receptor (OR) gene family shows a high frequency of duplication events that are generally unique to each lineage. We explored the OR gene family across the Zoonomia species set, independently of alignment-based annotation. We mined all genomes for OR gene sequences using the olfactory receptor assigner (ORA) (*177*). ORA uses profile hidden Markov-models (HMMs) to identify possible OR gene sequences while also highlighting pseudogene status. We used the reference OR protein sequences from ORA to identify putative OR gene regions in each assembly with tblastn (*217*, *218*). We extracted the regions reported in the tblastn results, including an additional 500bp upstream and downstream to ensure coverage of the start and stop codon regions for which neither was reported, using a custom perl script (available on the Zoonomia github repository). We used these sequences, as well as all whole-genome assemblies independently as inputs for ORA. We classified sequences as "pseudogenes" if they contained in-frame stop codons or were shorter than 650bp and therefore not long enough to form the 7-transmembrane domain. We excluded any sequences that were less than 200bp and considered other pseudogenes to be "non-functional genes." We clustered identified OR genes in each species with 99% identity using cd-hit (*219*) to remove putative non-paralogous duplicates. We expected two types of duplicates to be present in the data: sequences found both with tblastn and ORA and sequences that occur more than once due to duplicate contigs in the assemblies (*127*). To ensure that all sequences truly represented OR genes, we mapped each sequence to a database of 139 mammalian annotations available on RefSeq (*216*). If the highest scoring hit for a query sequence was not an annotated OR gene, we excluded the sequence. Our reference-free approach avoids potentially underestimating ORs in distantly related taxa that may have occurred due to lineage-specific genome evolution and is robust to variation in genome assembly quality [OR ~ contig N50: N=236; Spearman's rho (ρ) = 0.006, $p = 0.93$; OR~ scaffold N50: N = 236; ρ = 0.013; $p = 0.84$].

Turbinal count phenotyping

We curated species-specific numbers of olfactory turbinals from both sides of the nasal cavity(**Table S12**) (*220–240*) . To identify turbinal homology and compute resulting counts, we followed established nomenclature (*220*, *221*). Our counts include frontoturbinals, ethmoturbinals, interturbinals, and the lamina semicircularis. Though the latter is not a turbinal, it exhibits a turbinal-like morphological pattern and is covered by olfactory epithelium. We did

not include nasoturbinals and the septum nasi, both of which are also partly covered by olfactory epithelium in several species (*222*, *223*, *241*), because large parts of these structures are not involved in olfaction. To date, turbinal numbers are available for 67 species of our sample. We used morphological data from only adult stages because the number and complexity of olfactory turbinals increase during postnatal growth in many species (*220*, *224*, *242*). Although the given data are available for only a small sample of species, they serve as a basis for future studies on possible associations between OR genes, phenotypes, and olfactory performance.

In our statistical analyses, we excluded certain species due to imprecise data on their turbinal morphology (e.g., *Dasypus novemcinctus* (*240*)). In case of turbinal number range, we used the average for our analyses. For instance, for the domestic dog (*Canis lupus familiaris*), whose turbinal number varies widely due to extreme breed morphology (20 up to 34, n = 27 specimens (*243*)), we used the average (27 turbinals) for breeds with varying facial length types (brachycephalic to dolichocephalic (*224*, *226*)). After filtering out species and combining measurements from the same species, we had turbinal counts for 64 placental mammals in Zoonomia.

Statistical analyses on olfactory structures

We associated the number of OR genes with the number of olfactory turbinals using phylolm (*136*), a faster version of phylogenetic generalized least squares (*135*), with the Zoonomia phylogenetic tree (*26*). phylolm runs a generalized least squares regression to associate the number of OR gene counts with the number of olfactory turbinal counts while having a covariance matrix that is based on the phylogenetic relationships between species, thereby directly incorporating the phylogenetic tree. The "phylolm coefficient" is the coefficient from this regression. We computed an empirical p-value by comparing our p-value to p-values from 999,999 phylogenetic permulations (*137*), permutation tests in which the general phylogenetic tree topology is preserved. We used the slightly sped-up code for continuous phylogenetic permutations described in our companion paper (*159*). We mapped both the number of OR genes and olfactory turbinals onto the phylogenetic tree using the countMap function for continuous traits from the R package phytools (*244*). The state of each internal node was estimated using a fast maximum likelihood algorithm and interpolated according to equation [2] of (*245*).

Associating gene conservation with hibernation

We investigated genomic differences between mammals that we defined as hibernators and as strict homeotherms (**Table S1**). The hibernator phenotype included any species exhibiting depressed body temperature (<18°C) that persists for >24 hours. We categorized species as homeothermic if experimental data showed no evidence of naturally occurring body temperature depression in the lab or field or if species did not recover from experimentally-induced hibernation or torpor (*246–326*). We removed n = 20 species categorized as daily heterotherms (body temperature declines persist for <24 hours) (*327*) from this comparison. Additionally, we removed 45 species from downstream analysis when conflicting or no evidence of hibernation had been previously described, leaving us with 22 species defined as deep hibernators and 154 species defined as strict homeotherms.

For identifying genes that are more similar to the mammalian ancestor than they are to non-hibernators through generalized least squares (GLS) forward genomics, we partitioned the exons in the 241 mammal alignment version 1 (*11*) into 100bp regions. For each element, we used

PREQUEL from the PHAST package (*192*) to compare the exonic sequences of hibernating versus strict homeotherms against the sequences of a mammalian ancestor that was constructed from the multi-way alignment containing 241 mammals in our earlier work (*4*, *11*, *25*, *192*). We generated a list of conserved exonic elements by taking the intersection of mammalian-conserved regions and exonic boundaries, where we used the conserved regions previously identified based on Siphy-ω (*328*), and the Gencode v24 exon positions. Elements shorter than 70bp were excluded. Then, grouping the conserved exonic elements within each aligned gene, we calculated the percentage of bases aligned to the ancestral reference for each species. For all species in which at least 90% of the bases aligned among all conserved elements, we calculated the percent identity (# identical / # aligned) to the ancestral reference. Note that this method does have a number of limitations–it does not account for the likely effects of a sequence difference from the common ancestor (synonymous versus nonsynonymous change or type of amino acid change if the change is nonsynonymous), and it does not account for local substitution rate (changes from the common ancestor in regions with higher substitution rates are more likely to happen by chance). We then used GLS forward genomics with the Zoonomia phylogenetic tree to identify regions conserved in hibernators relative to the placental ancestor (GLS adjusted $p < 0.05$), where p-value adjusted was done using the Benjamini-Hochberg procedure (*26*, *135*, *329*).

To determine if our GLS forward genomics method detected more conservation than it would from a random set of foreground species with a similar phenotype tree topology, we performed phylogenetic permulations (*137*). GLS forward genomics is a method for assessing phenotypic patterns in genomic data by performing linear regressions between these patterns and global percent IDs, computed values which compare extant sequences to the sequence of a common ancestor, p-values are the significance of linear regression positive slopes (*16*, *23*). For GLS permulations, foreground species were selected as the 22 species with the highest values from phylogenetic simulations in the permulation framework. We did not impose additional requirements for number or structure of foreground species on phylogenetic branches for ancestral species. For each element that was significant from GLS forward genomics (nominal $p < 0.05$), we ran GLS forward genomics for 10,000 permulations, where we allowed only permulations with 22 foreground species. We calculated permulation p-values for each element as the fraction of permulations with a p-value at least as low as the p-value for the real data and then adjusted p-values using the Benjamini-Hochberg procedure (*329*). Out of the 1,479 hibernation-associated elements detected by GLS analysis prior to FDR correction, 803 elements had adjusted permulation $p < 0.05$. All 28 significant elements from GLS forward genomics without permulations (GLS adjusted $p \leq 0.05$) associated with hibernation also had significant adjusted permulation p-values (**Table S13**).

To identify genes with significant evolutionary rate shifts in hibernating mammals versus non-hibernating mammals, we used RERconverge (*149*). Such genes are putative hibernation-related genes. Genes with low relative evolutionary rates in hibernating mammals likely have increased constraint in hibernators and therefore potential functional importance, while genes with high relative evolutionary rates in hibernating mammals may be under relaxation of evolutionary constraint (implying decreased functional importance) or positive selection (implying increased functional importance). Note that this method cannot accurately infer the ancestral trait. To perform RERconverge analyses, we first obtained filtered alignments for 16,209 gene exon amino acid sequences and used them to estimate tree branch lengths (*6*, *160*). We ran RERconverge using default parameters with the full 167 mammal species set for which hibernation phenotypes were known (21 hibernators and 146 non-hibernators). Species included

were those present in the protein-coding sequence alignment whose names matched species with available hibernation phenotypes. We ran getAllResiduals and foreground2Paths with the species we used supplied for the "useSpecies" argument and "clade=all" specified to foreground2Paths so that internal branches were assigned as foreground non-hibernators based on maximum parsimony. We performed correlation analyses using the correlateWithBinaryPhenotype function with default parameters. While this method enabled us to identify hundreds of genes with relative evolutionary rates associated with the evolution of hibernation, this method may have missed relevant genes because it used a human-referenced protein-coding sequence alignment, and humans are not hibernators.

We also conducted permulations to support RERconverge analyses. We conducted a modified complete case permulation approach by generating 10,000 null phenotypes with the following constraints: 1) all null phenotypes contained 21 foreground species, and 2) all null phenotypes had a number of total foreground branches no different from five from the observed hibernator phenotype. We ran RERconverge analyses as stated previously for all null phenotypes. We quantified the permulation p-value per gene as the number of permulated statistics more extreme than the observed statistic divided by the total number of permulations performed. We adjusted permulation p-values using the Benjamini-Hochberg procedure (*329*).

Correlated traits present a challenge for forward genomics. Many hibernators are bats, who also tend to be exceptionally long-lived given their body size. We were concerned that this might be confounding our results because we obtained genes involved in DNA repair, which is also involved in longevity. To confirm we had not unintentionally tested for associations to longevity or other traits shared by most bats but not by most other mammals, we used Bayes factor analyses to quantify the amount of signal arising from bat foreground species compared to hibernator foreground species. Using the BayesFactor package in R (*330*), we calculated per-gene linear model Bayes factors predicting RERs from either the bat phenotype or the hibernator phenotype. We then calculated the ratio of hibernator to bat Bayes factors per gene to quantify the amount of support for a gene's relationship to one phenotype over the other–we considered Bayes factors greater than five to provide notably more support for the hibernator phenotype and therefore distinguish genes associated with hibernation from those associated with being a bat. As expected, some DNA repair genes did show signal driven by bats (SLX4 Bayes factor = 0.003, TEX15 Bayes factor = 0.318), but others showed signal driven by hibernators (PARP4 Bayes factor = 67.047, PWWP3A Bayes factor = 85.282). This suggests that some DNA repair functionalities do evolve convergently in hibernators overall and not only in bats. We considered a gene to be associated with hibernation if the adjusted p-value from RERconverge was < 0.05, the adjusted permulations p-value was < 0.05, and the Bayes factor value was > 5.0.

Supplementary Text

Additional comparisons of constraint between genes and their potential regulatory elements

Investigating enhancers is challenging because cCREs do not necessarily regulate the closest genes (*331*). We used the EpiMap putative enhancer-gene linking predictions (*332*) to link ENCODE3 proximal and distal cCREs (*14*) to their likely target genes. Briefly, EpiMap predicted enhancers in more than 800 cell types (including ENCODE datasets) and linked these putative enhancers to their target genes using correlation between enhancer presence and gene expression across these cell types. Using EpiMap enhancer-gene links in all cell types, we were able to link 33% and 53% of proximal and distal cCREs, respectively, to at least one gene.

We compared these region groupings to each other and observed some expected trends that have not previously been demonstrated at this scale. There is a strong positive correlation (ρ = 0.52) between conservation of protein-coding and proximal promoter sequences. This result is consistent with the theory that, if the function of a gene in mammals requires high conservation of protein sequence, then the non-coding sequence in the proximal promoter that regulates its expression also tends to be constrained (*333*, *334*). We also observed consistent correlation of protein sequence constraint with constraint in corresponding 5'UTRs and 3'UTRs (ρ = 0.54 and 0.45, respectively), and, to a lesser extent, with constraint in introns (ρ = 0.30). To further investigate the relationship between protein-coding and regulatory conservation, we evaluated the relationship between constraint in ENCODE3 cCREs(*14*) and constraint in the genes that cCREs may regulate. We observed non-negligible correlation between distal cCRE constraint scores and constraint in associated CDSs and promoters (ρ = 0.20 and 0.23, respectively).

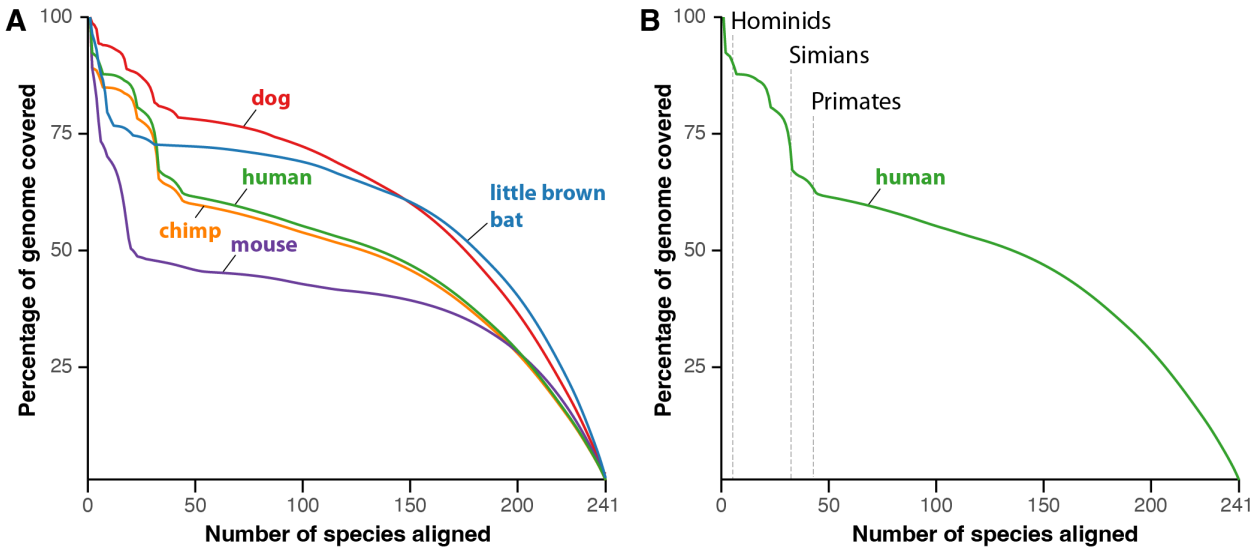

Fig. S1. Genome alignment depth varies with phylogeny among Zoonomia species.

(A) When the Cactus alignment is referenced on five different species, the depth of the alignment (number of species aligned) varies widely. Only a small percentage of the genome aligns in most species, with rapid drops correlating with longer branches of the phylogeny. **(B)** For example, in the human alignment, most of the genome is aligned in 5 other hominids, but this drops as more distantly related species are included.

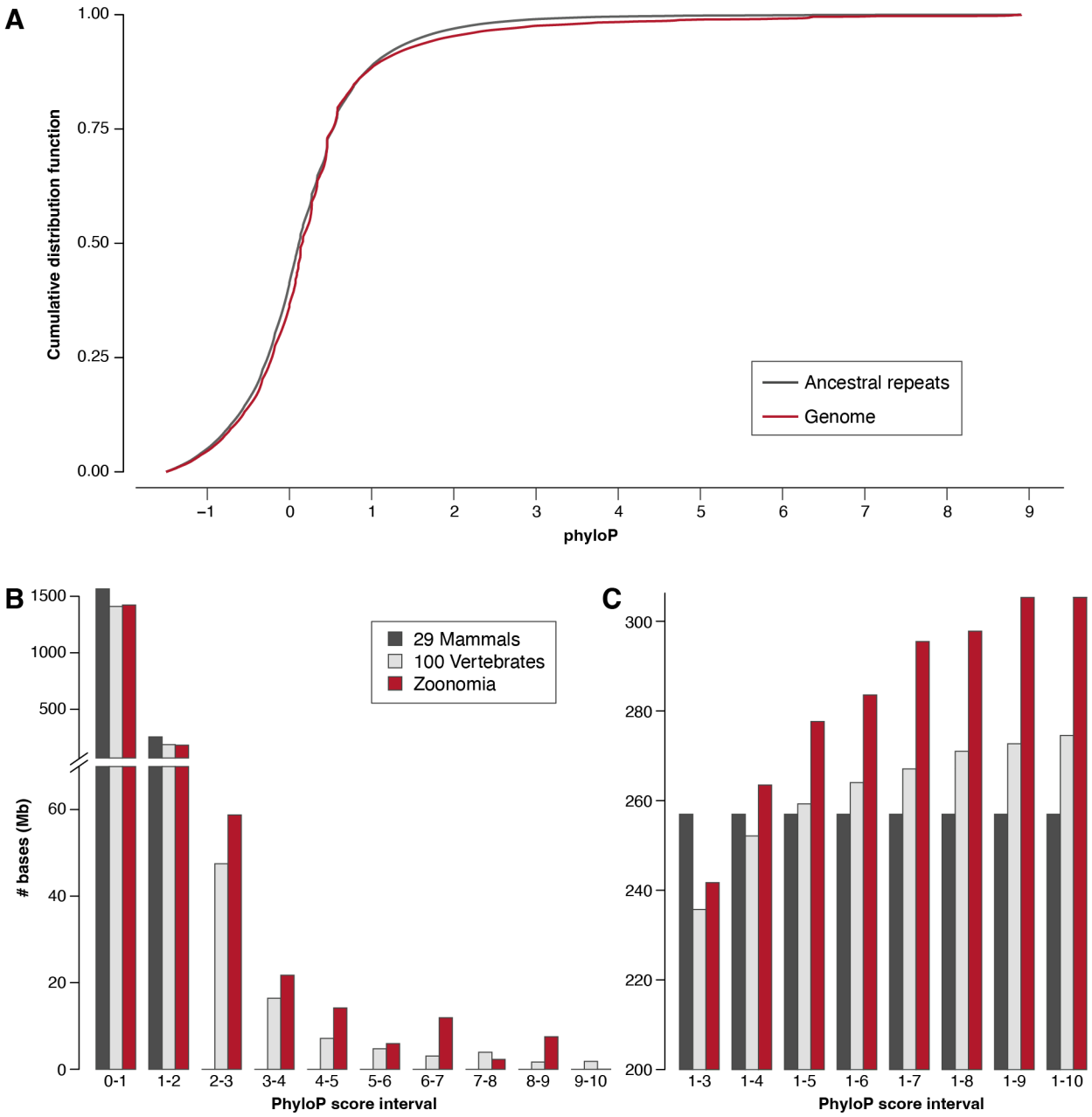

Fig. S2. PhyloP scoring of the 240 species Zoonomia alignment detects >10.7% of genome as constrained (more than in either 29 Mammals or 100 Vertebrates). (A) Cumulative distribution functions (CDFs) of phyloP scores across the human genome (red) and in ancestral repeats, representing neutrally evolving sequence (gray). We calculated our lower-bound estimate of constraint of 10.7% for the human genome by comparing these CDFs. **(B)** phyloP score bins greater than two contain more bases in Zoonomia than in 29 Mammals (which does not have any bases with scores >2) and 100 Vertebrates (which has a smaller number of high score bases). **(C)** In total, more bases in Zoonomia have high phyloP scores, indicating constraint.

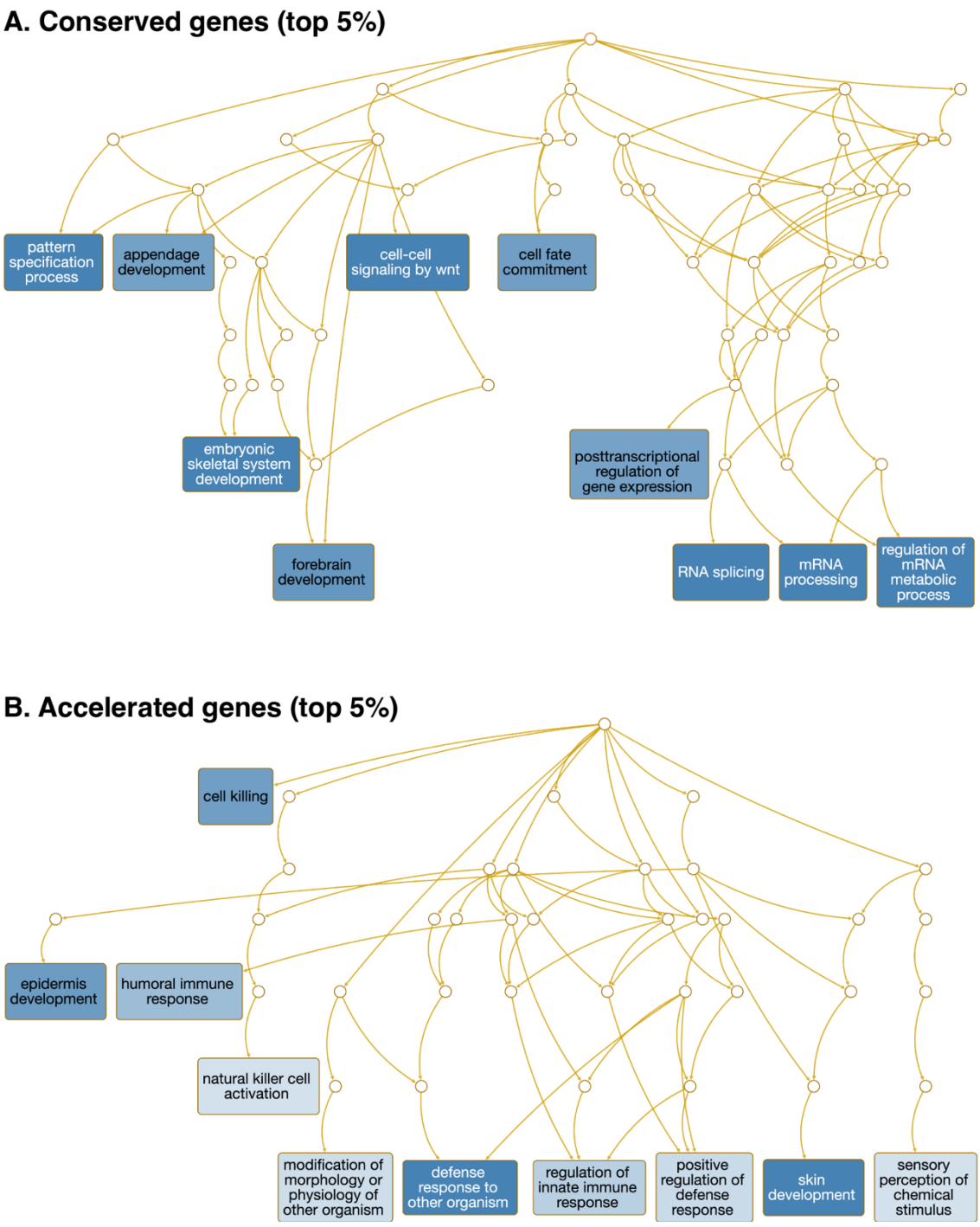

Fig. S3. Constrained and more quickly evolving genes in mammals are enriched in some gene sets. We tested both **(A)** Genes with very high average phyloP scores in protein-coding sequence and **(B)** genes with very low average phyloP scores in protein-coding sequence against a non-redundant representative set of Gene Ontology biological processes using WebGestalt (*46–48*) and plotted the top 10 overrepresented gene sets in each. All gene sets are significant after correction ($p_{FDR} < 0.0002$).

Fig. S4. Phylogenetic tree with CMAH gene loss annotated. Red lineages indicate CMAH loss. Red boxes indicate newly reported losses, and blue boxes indicate lineages previously reported to have lost *CMAH* but whose genomes contain intact CMAH coding sequence in this analysis.

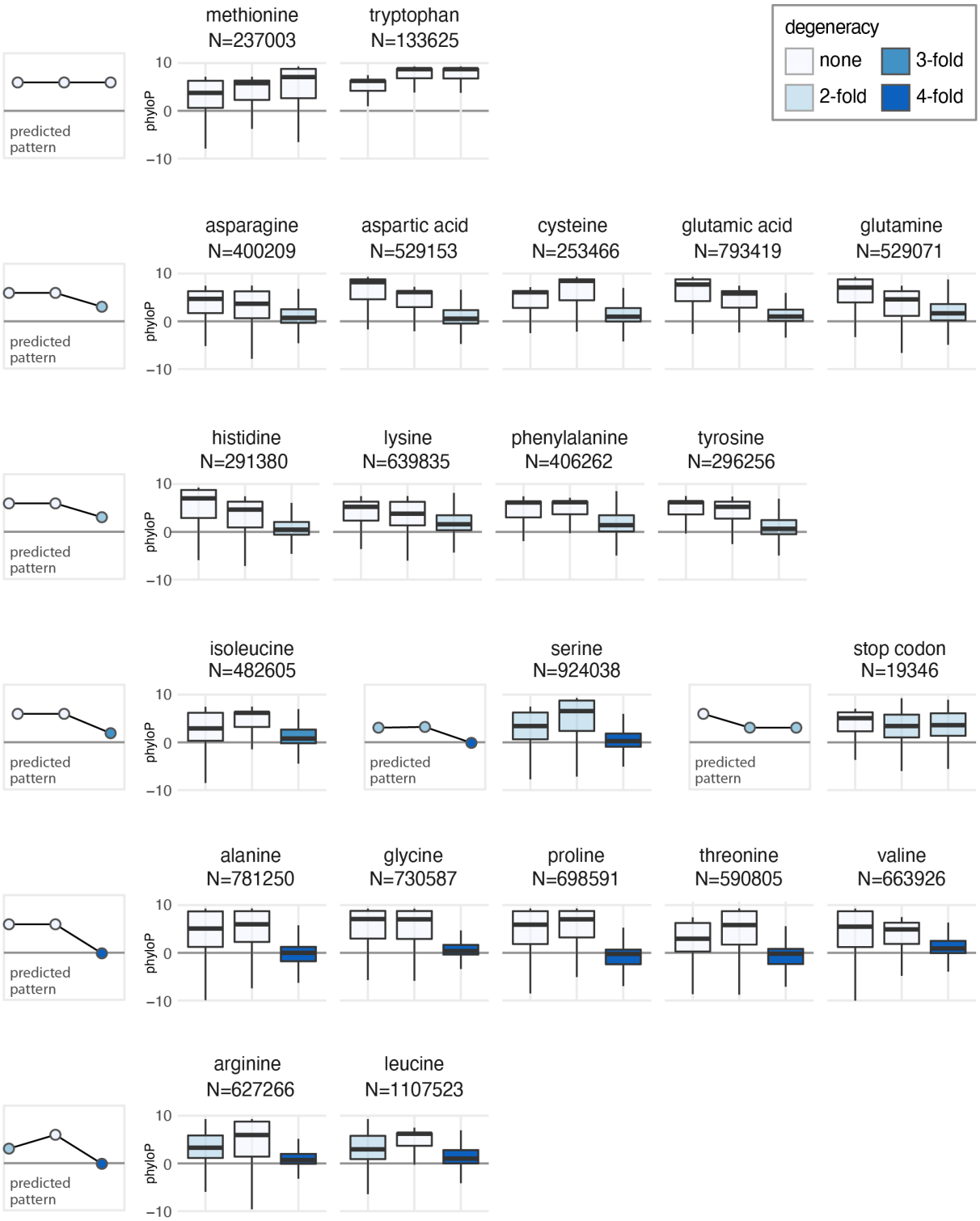

Fig. S5. Degeneracy and constraint in codon positions. Boxplots show the distribution of phyloP at each position within a codon for all 20 amino acids plus stop codon. Predicted patterns of constraint are based on the degeneracy for each amino acid [ranging from 0 (four-fold degenerate) to 4 (non-degenerate)]. They appear similar to the observed patterns of constraint.

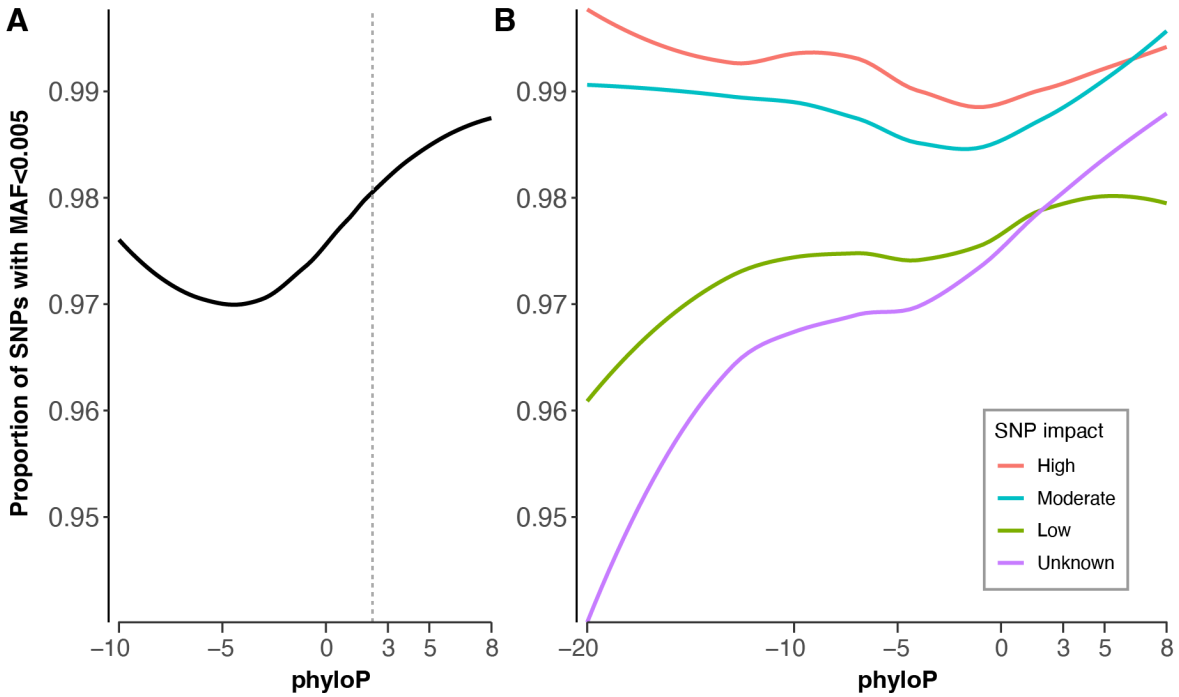

Fig. S6. phyloP scores versus percentage of single-nucleotide polymorphisms (SNPs) with minor allele frequency < 0.005. (A) phyloP score versus percentage of SNPs with minor allele frequency < 0.005 for all SNPs (N=622,177,419, ρ = 0.78, p = 0.00014). SNPs are from TOPMed data freeze 8 (*69*). **(B)** phyloP score versus percentage of SNPs with minor allele frequency < 0.005 for SNPs with different predicted impact levels by SNPEff: high (N=501,807, ρ = -0.05, p = 0.81), moderate (N=5,688,677, ρ = 0.39, p = 0.18), low (N=7,759,842, ρ = 0.60, p = 0.0018), and unknown (N=608,227,093, ρ = 0.98, p = 5.45×10^{-7}).

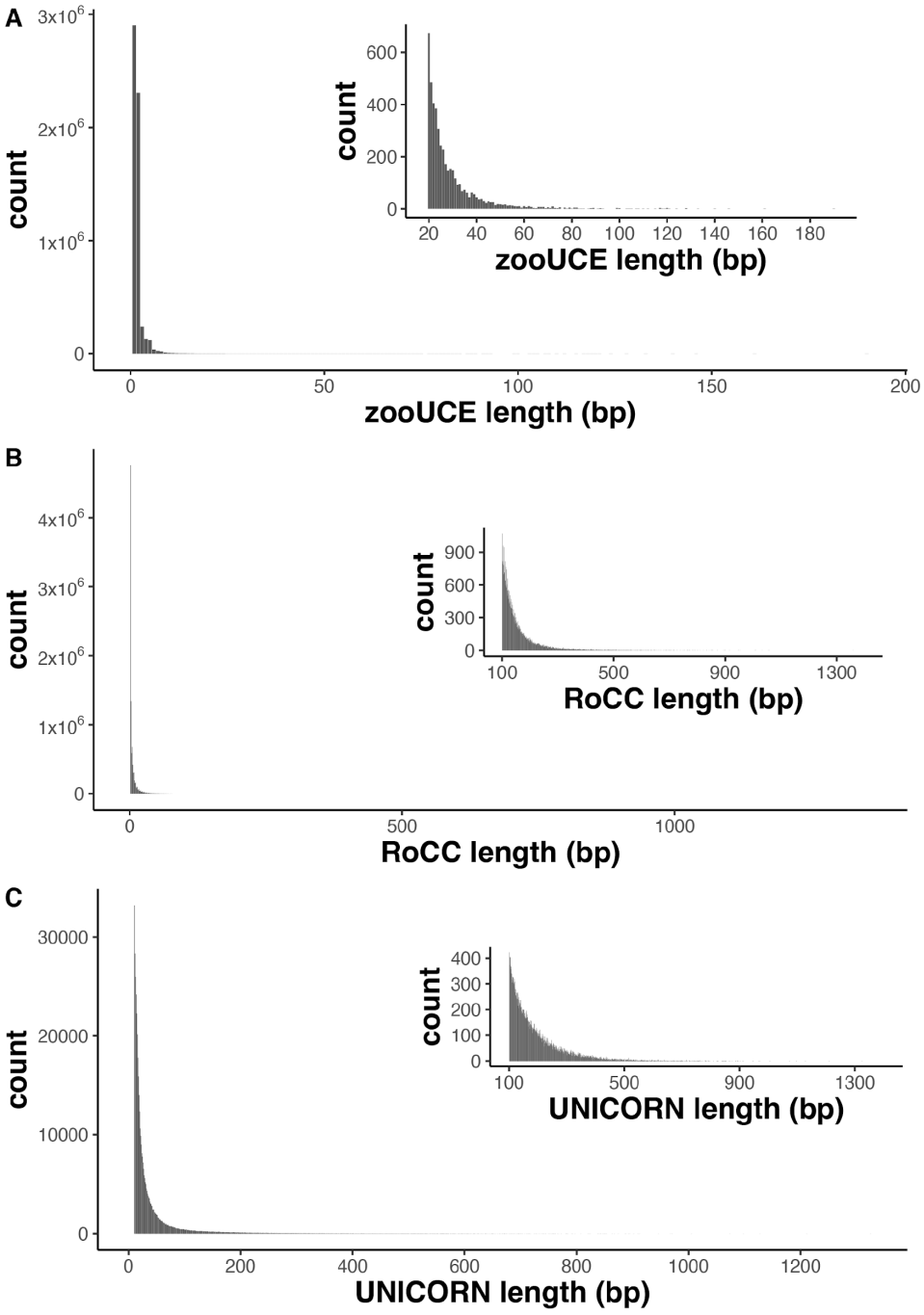

Fig. S7. Histograms of lengths of conserved regions for different definitions of conservation. (A) Histogram of lengths of Zoonomia ultraconserved elements (zooUCEs). **(B)** Histogram of lengths of regions of contiguous conservation (RoCCs). **(C)** Histogram of lengths of UNICORNs. Insets show distributions for shorter length elements.

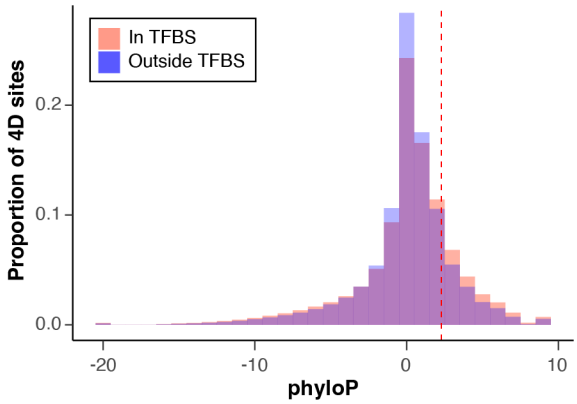

Fig. S8. Constraint and transcription factor binding at four-fold degenerate sites. (A) phyloP scores at four-fold degenerate sites inside versus outside of TFBS (N = 2,420,610; χ^2 test, $p < 2.2\text{x}10^{-16}$). Dashed vertical line indicates phyloP threshold for significant constraint (phyloP = 2.270, FDR < 0.05). ANOVA analysis of 4D sites overlapping the 100 transcription factors with the most TFBS (20,696,533 sites) suggests that the transcription factor itself has almost no effect on the percent of excess constrained 4D sites in TFBS [effect size (generalized Eta squared or ges) = 0.006; $p = 0$]. Mean phyloP for the gene explains little of the variation (ges = 0.050; $p = 0$). The amino acid the site codes for has a small effect (ges=0.097; $p = 0$).

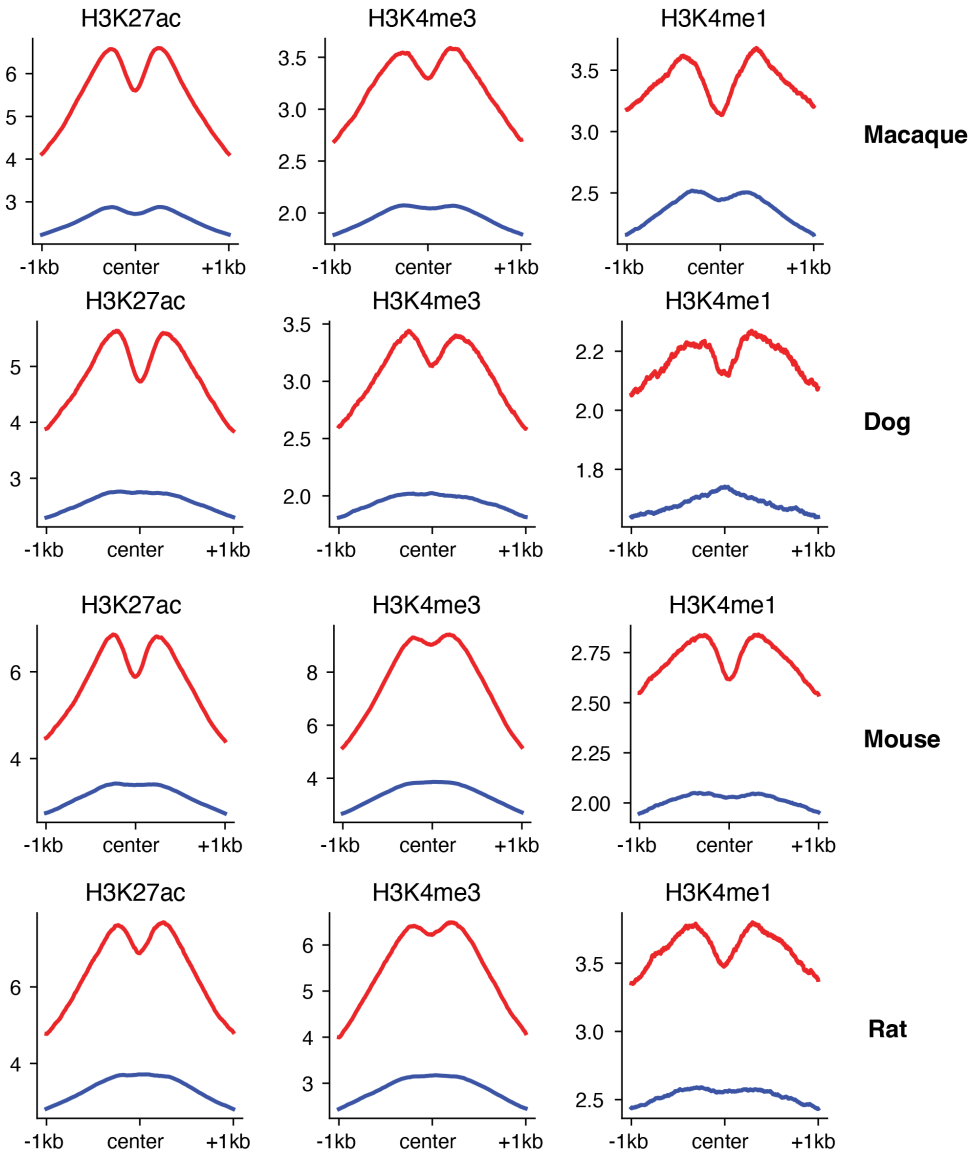

Fig. S9. Histone marks correlated with function are enriched at constrained TFBS. Histone marks H3K4me3, H3K27ac, and H3K4me1 active in non-human species are enriched at constrained human TFBS (red) but not at unconstrained human TFBS (blue). Y-axis shows enrichment.

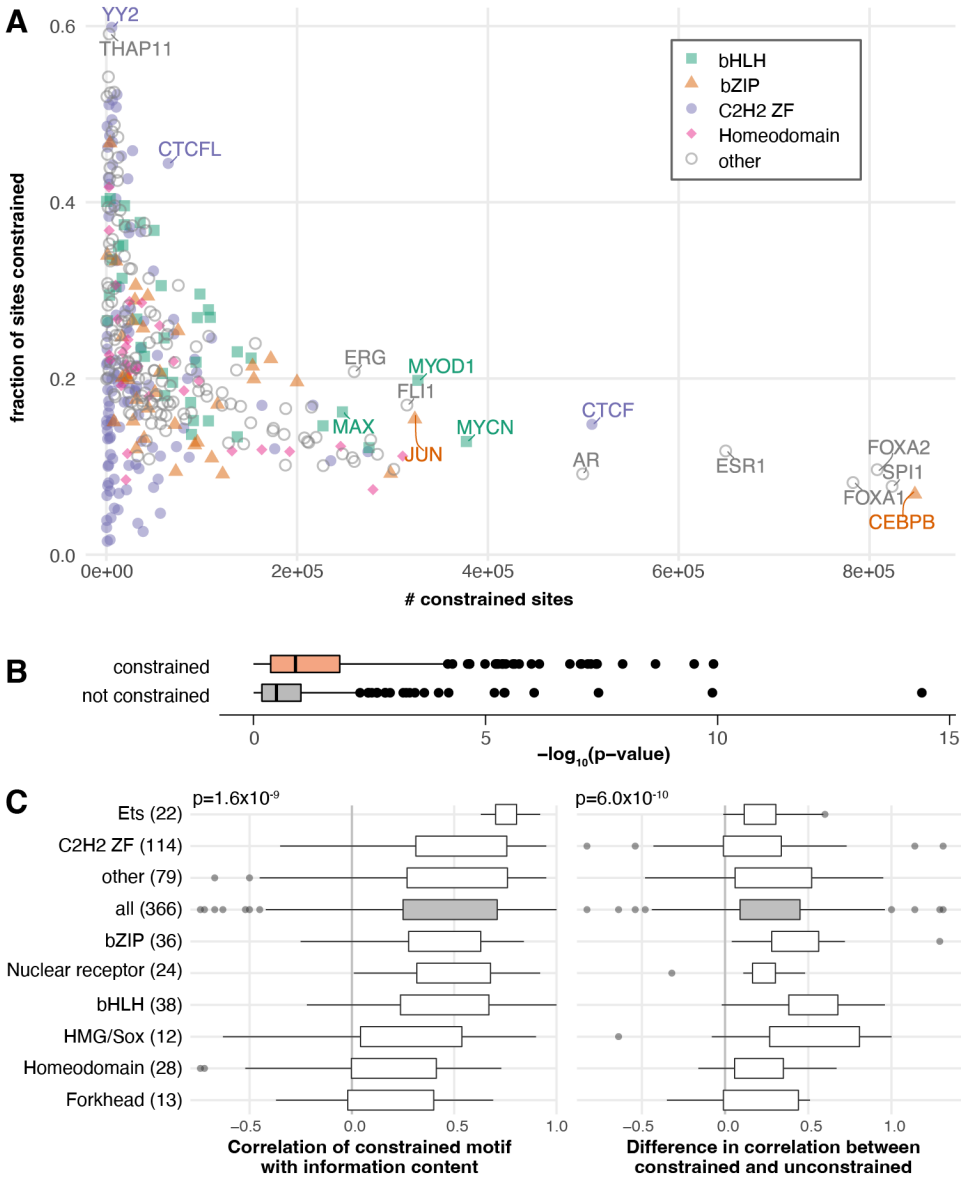

Fig. S10. Constraint in TFBS by transcription factor. (A) The fraction of TFBS sites in the human genome that are constrained varies by transcription factor. **(B)** Across all transcription factors, the correlation of phyloP with TFBS information content tends to result in lower p-values for constrained relative to unconstrained sites. **(C)** The correlation varies between different types of transcription factors, with the median for all types exceeding zero (left). The median difference is also greater than zero (right), indicating constrained sites are more strongly correlated than unconstrained sites. The significance of the effect of transcription factor family (excluding "all") on variation in each metric, measured using ANOVA, is shown in upper left.

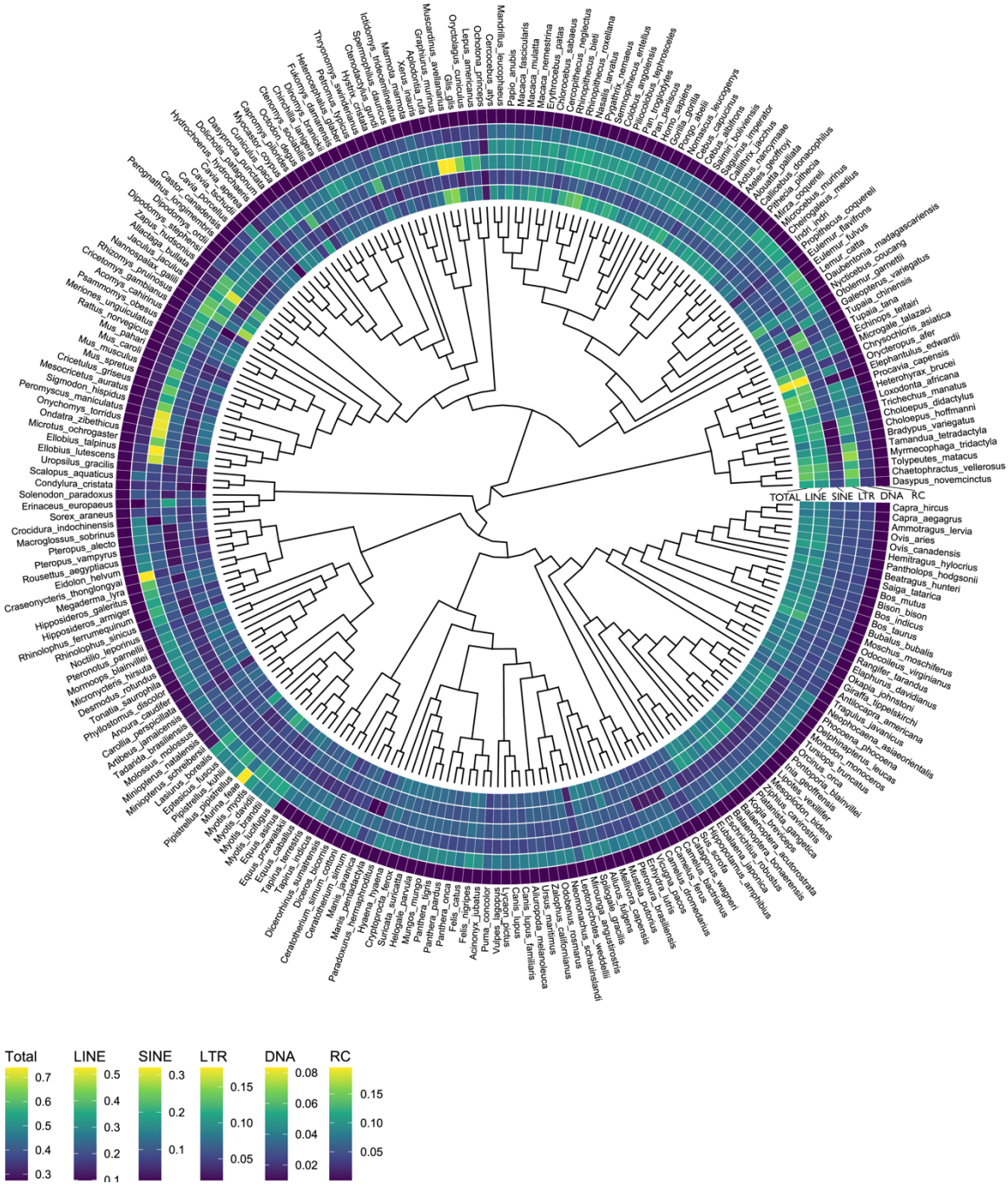

Fig. S11. Proportion of genomic content attributed to transposable element accumulation within a phylogenetic context. The rings, from inner to outer, depict the transposable element accumulation data as proportions of the total genome assembly: all transposable element content, LINEs, SINEs, LTRs, DNA transposons, rolling-circle transposons. Cladogram adapted from (*26*).

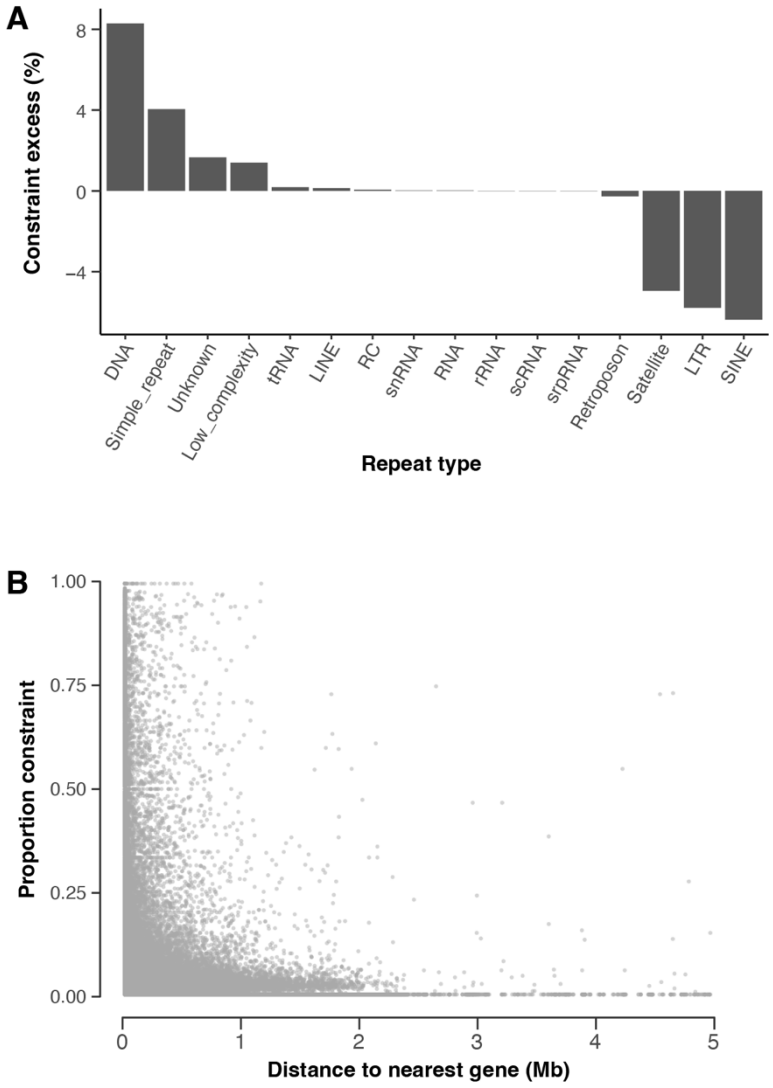

Fig. S12. Constraint of repeats. (A) Percentage of excess or depletion of constraint in each human repeat class. **(B)** Distance to nearest protein-coding transcription start site versus proportion of positions under constraint (phyloP > 2.27) for all human simple repeats.

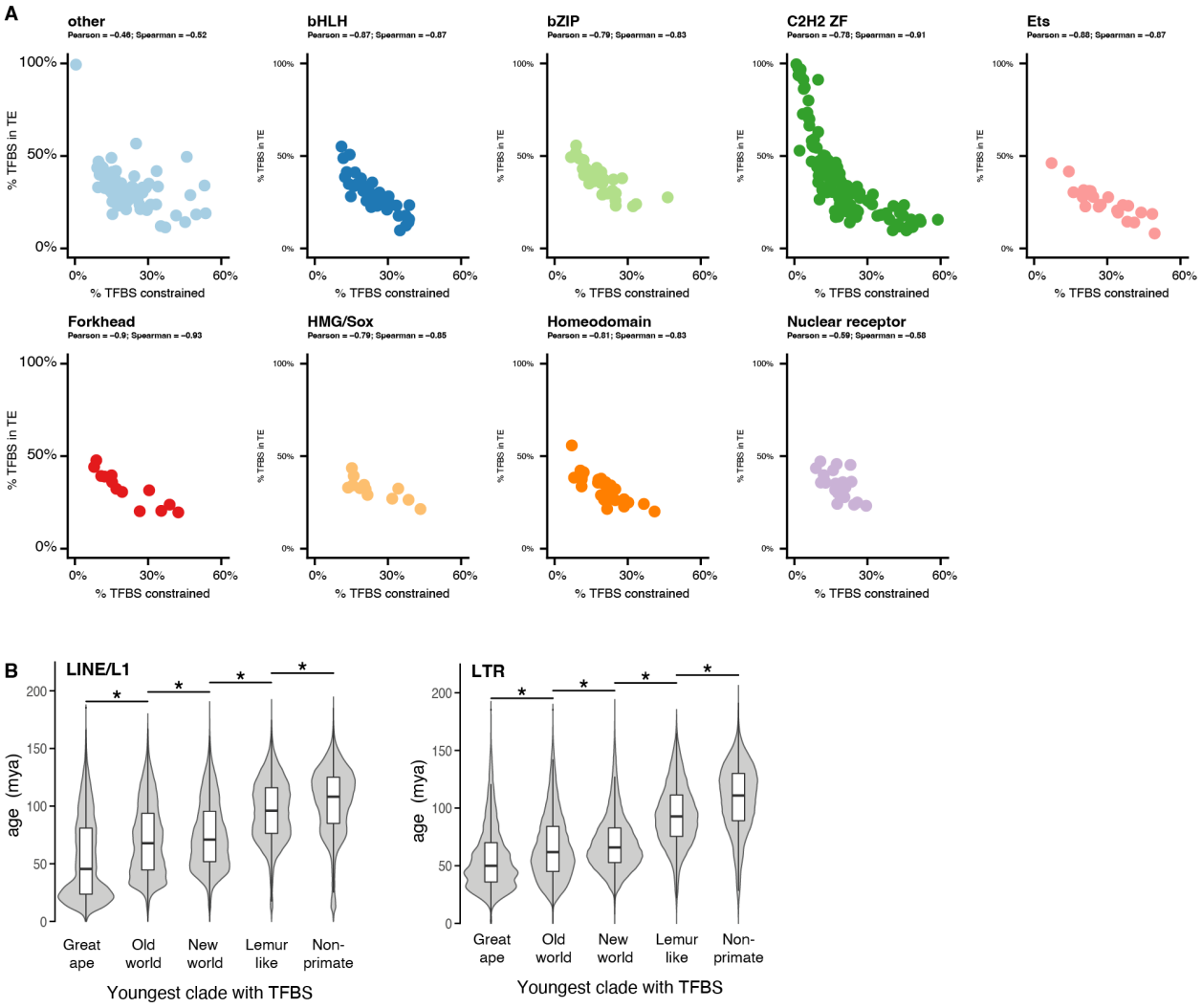

Fig. S13. Transcription factor binding sites (TFBS) in transposable elements. (A) Percent of TFBS found in a transposable element versus the percent found to be evolutionary constrained for 368 transcription factors, separated by family. **(B)** TFBS shared in only more closely related species tend to be on younger transposable elements, and vice versa.

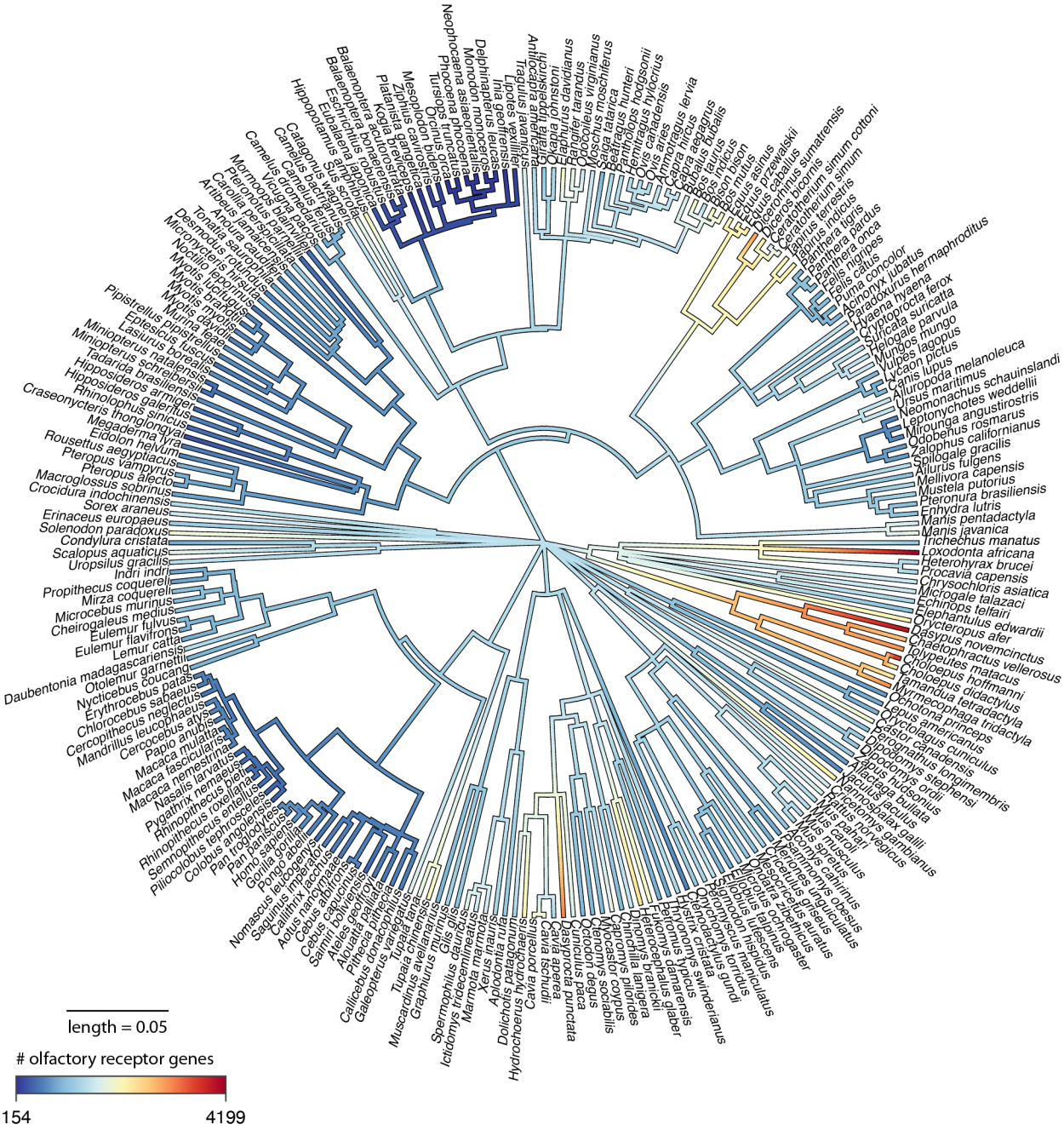

Fig. S14. Phylogenetic tree annotated with the number of olfactory receptor genes in each species. Branches are colored by the total number of olfactory receptor genes (functional and non-functional). States at internal nodes were estimated with the R package phytools which uses a maximum likelihood algorithm (*244*). "length = 0.05" shows the branch length scale of the phylogenetic tree.

Fig. S15. Olfactory receptor gene number for each species grouped by mammalian clade. **(A)** The number of olfactory receptor genes for each species in mammalian clades (gray dots), with boxplots showing the distribution for all clades with more than three species surveyed. The box represents 25%/75% quartile with a horizontal line at the median, and the vertical line demarkating 5-95% quartiles. Red text labels the top and bottom species in each order. **(B)** There is no significant correlation between the completeness of a genome assembly (measured as percent of BUSCO (*129*) genes that are complete) and the number of olfactory receptor genes detected. Estimates for incomplete assemblies that are unusually low compared to other species in the clade, as for the brown-throated sloth (starred) compared to other xenarthrans, should be validated.

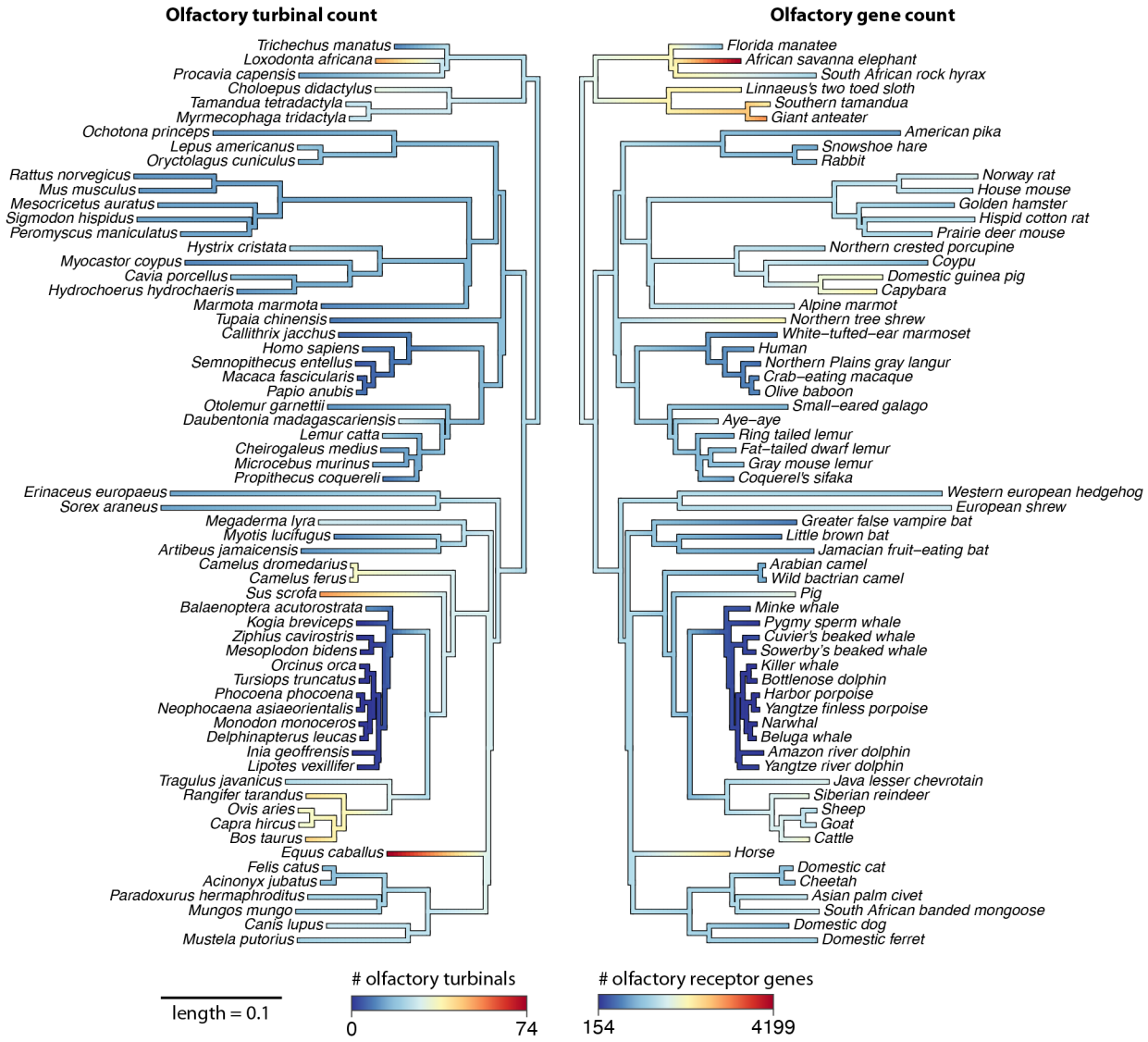

Fig. S16. Phylogenetic trees annotated with number of olfactory turbinals (left) and olfactory receptor gene counts (right). Species included are the subset of species with both olfactory receptor gene counts and number of olfactory turbinal annotations. Species' binomial names are listed on the left, and species' common names are listed on the right. States at internal nodes were estimated with the R package phytools which uses a maximum likelihood algorithm (*244*). "length = 0.1" shows the branch length scale of the phylogenetic tree**.**

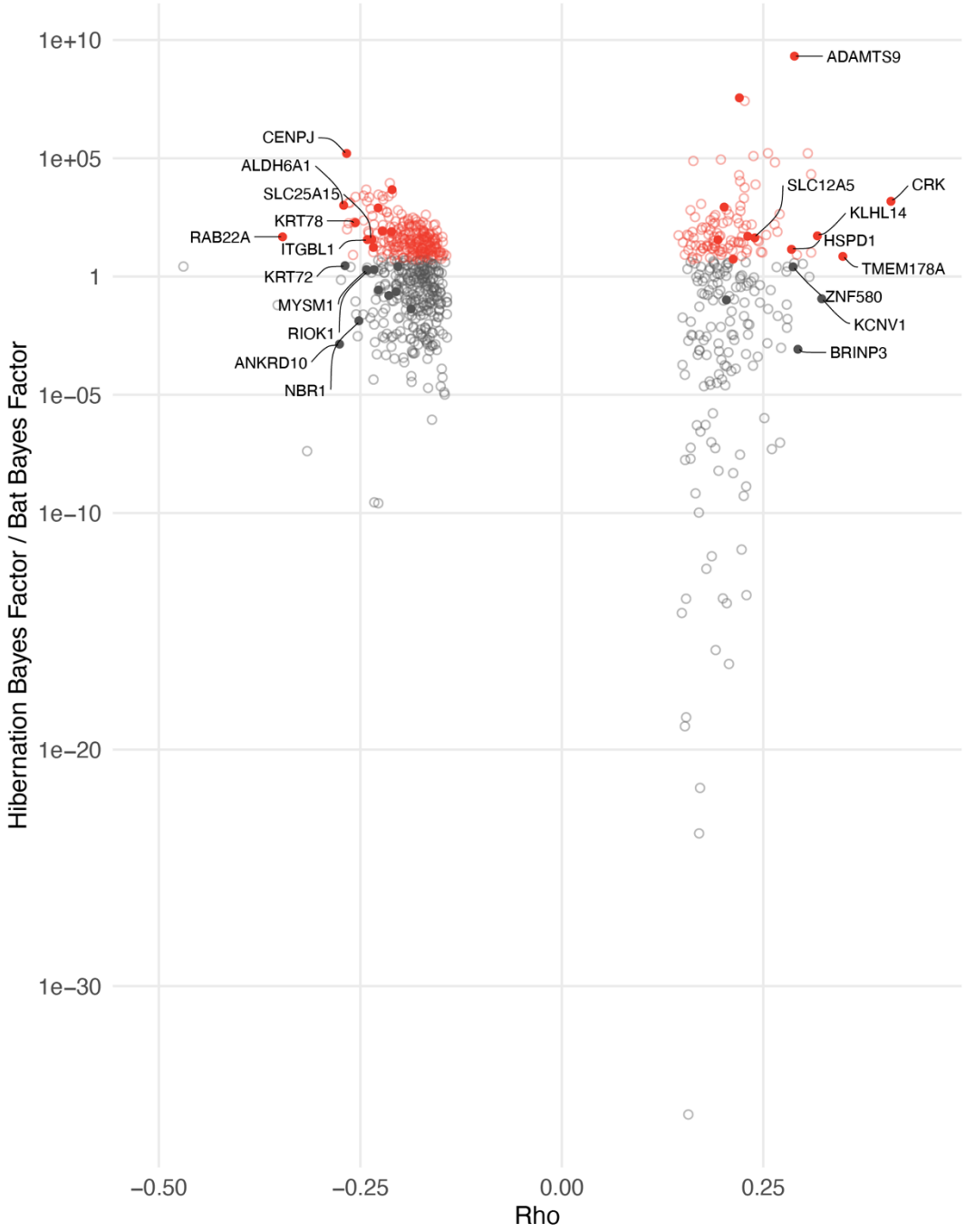

Fig. S17. Hibernation RERconverge Bayes factor analysis. Per-gene associations between the relative evolutionary rate and hibernation across mammal species (RERconverge Rho; x-axis) compared to the ratio of Bayes factors testing the hibernation phenotype and the bat phenotype (y-axis) for all genes with RERconverge $p_{FDR} < 0.05$. Filled circles represent genes that are significant after testing through permulations (permulation $p_{FDR} < 0.05$). The 20 genes with the highest absolute Rho are labeled. For some genes, the association may be confounded by high frequency of hibernation phenotype in bats (gray, ratio < 5), while others are more confidently associated with hibernation (red).

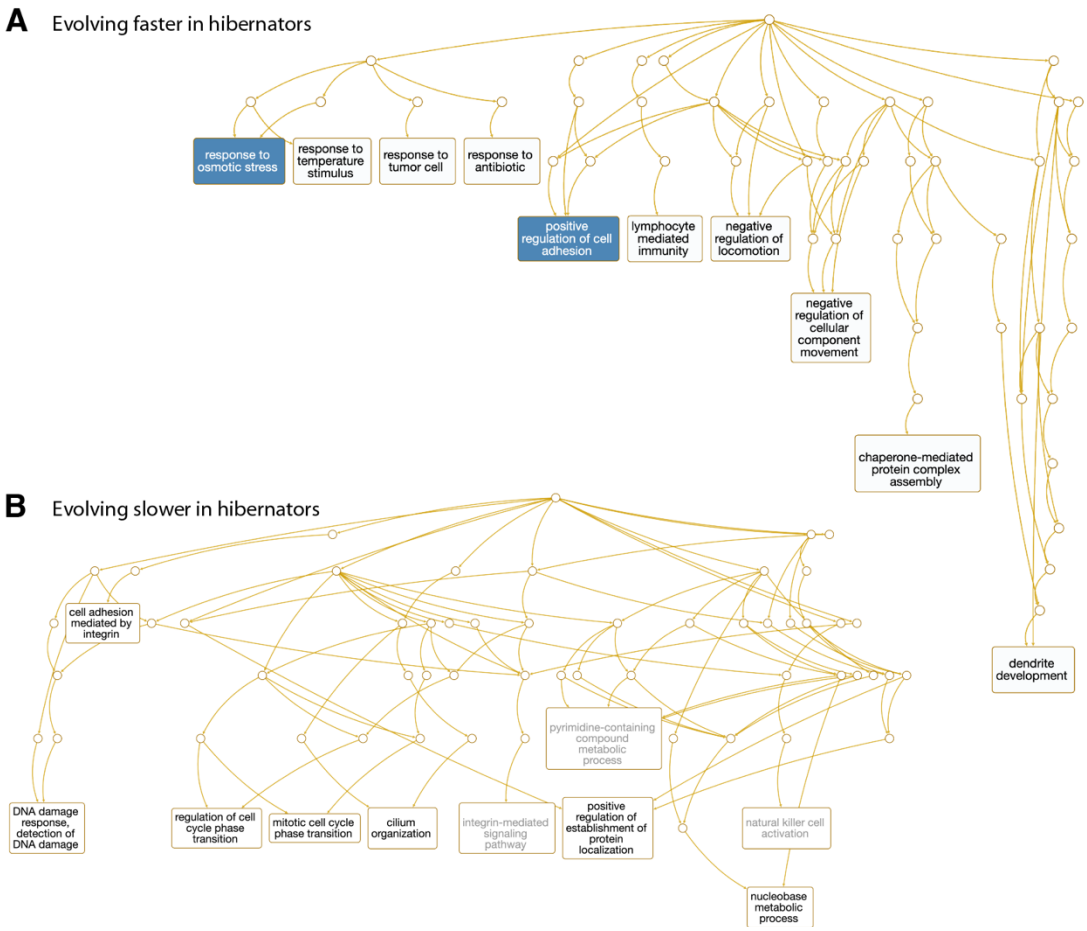

Fig. S18. Genesets enriched for hibernation-associated genes. We tested all 22 genes significantly associated with hibernation against a non-redundant representative set of Gene Ontology biological processes using WebGestalt (*46–48*) and plotted the top 10 gene sets for **(A)** 11 genes evolving faster in hibernators and **(B)** 11 genes evolving more slowly than expected in hibernators. No gene sets were significant after correction (all had $p_{FDR} > 0.05$). Seventeen gene sets were nominally significant ($p < 0.05$; black text) (**Table S15**). The top two sets (blue shading) are response to osmotic stress ($p = 9.1 \times 10^{-4}$; $p_{FDR} = 0.58$; genes SCN2A and SLC12A5) and positive regulation of cell adhesion ($p = 0.0014$; $p_{FDR} = 0.58$; genes DENND6A, HSPD1, and CRK).

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Acinonyx jubatus	cheetah	VU	Yes	Yes	Yes	No	Yes	Yes	No	-0.1698	No
Acomys cahirinus	cairo spiny mouse	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.0135	No
Acomys russatus	golden spiny mouse	LC	No	Yes	No	No	No	No	NA	NA	NA
Aeorestes cinereus	hoary bat	NA	No	Yes	No	No	No	No	NA	NA	NA
Aepyceros melampus	impala	LC	No	Yes	No	No	No	No	NA	-0.0674	NA
Ailuropoda melanoleuca	giant panda	VU	Yes	Yes	Yes	No	Yes	Yes	No	-0.1348	No
Ailurus fulgens	lesser panda	EN	Yes	Yes	Yes	No	Yes	Yes	No	0.1196	No
Alces alces	eurasian elk	LC	No	Yes	No	No	No	No	NA	-0.0763	NA
Allactaga bullata	gobi jerboa	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	NA	No
Allenopithecus nigroviridis	allen’s swamp monkey	LC	No	Yes	No	No	No	No	NA	0.2726	NA
Alligator sinensis	chinese alligator	NA	No	No	No	No	No	No	NA	NA	NA
Alouatta palliata	mantled howler	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.1283	No
Ammotragus lervia	aoudad	VU	Yes	Yes	Yes	No	Yes	Yes	No	-0.201	No
Anoura caudifer	tailed tailless bat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	0.1199	NA
Antidorcas marsupialis	springbok	LC	No	Yes	No	No	No	No	NA	-0.0316	NA
Antilocapra americana	pronghorn	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.1434	No
Antrozous pallidus	pallid bat	LC	No	Yes	No	No	No	No	NA	NA	NA
Aonyx cinereus	Asian small-clawed otter	NA	No	Yes	No	No	No	No	NA	NA	NA
Aotus nancymaae	nancy ma’s night monkey	VU	Yes	Yes	Yes	Yes	Yes	Yes	No	NA	No
Aplodontia rufa	mountain beaver	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.0722	No
Apodemus sylvaticus	long-tailed field mouse	LC	No	Yes	No	No	No	No	NA	0.0534	NA
Arctocephalus gazella	antarctic fur seal	LC	No	Yes	No	No	No	No	NA	0.0527	NA
Artibeus jamaicensis	jamacian fruit-eating bat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	0.0553	NA
Arvicanthis niloticus	african arvicanthis	LC	No	Yes	No	No	No	No	NA	-0.0613	NA
Arvicola amphibius	eurasian water vole	LC	No	Yes	No	No	No	No	NA	-0.1948	NA
Ateles geoffroyi	geoffroy's spider monkey	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.4599	No
Axis porcinus	hog deer	EN	No	Yes	No	No	No	No	NA	0.0544	NA
Babyrousa celebensis	North Sulawesi babirusa	NA	No	Yes	No	No	No	No	NA	NA	NA
Balaena mysticetus	bowhead	LC	No	Yes	No	No	No	No	NA	-1.2719	NA
Balaenoptera acutorostrata scammoni	minke whale	LC	Yes	Yes	Yes	No	Yes	No	No	-0.3456	NA
Balaenoptera bonaerensis	antarctic minke whale	NT	Yes	Yes	Yes	No	Yes	Yes	No	NA	NA
Balaenoptera edeni	bryde’s whale	LC	No	Yes	No	No	No	No	NA	NA	NA
Balaenoptera musculus	blue whale	EN	No	Yes	No	No	No	No	NA	-0.9595	NA
Balaenoptera physalus	fin whale	VU	No	Yes	No	No	No	No	NA	-0.8794	NA
Bassariscus astutus	ringtail	NA	No	Yes	No	No	No	No	NA	NA	NA
Bassariscus sumichrasti	cacomistle	LC	No	Yes	No	No	No	No	NA	0.306	NA
Beatragus hunteri	hirola	CR	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Bison bison bison	american bison	NT	Yes	Yes	Yes	No	Yes	Yes	No	-0.3054	No
Bos frontalis	gaur	NA	No	Yes	No	No	No	No	NA	NA	NA
Bos gaurus	gaur	VU	No	Yes	No	No	No	No	NA	NA	NA
Bos grunniens	yak	NA	No	Yes	No	No	No	No	NA	-0.2648	NA
Bos indicus	zebu cattle	LC	Yes	Yes	Yes	No	No	Yes	No	NA	No
Bos mutus	wild yak	VU	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Bos taurus	cattle	LC	Yes	Yes	Yes	No	Yes	Previous	No	0.4181	No
Bradypus variegatus	brown-throated sloth	LC	No	No	No	No	No	Yes	NA	NA	NA
Bubalus bubalis	water buffalo	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Callithrix jacchus	common marmoset	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.2425	NA
Callithrix pygmaea	pygmy marmoset	NA	No	Yes	No	No	No	No	NA	0.3293	NA
Callorhinus ursinus	northern fur seal	VU	No	Yes	No	No	No	No	NA	-0.0993	NA
Camelus bactrianus	bactrian camel	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.1819	NA
Camelus dromedarius	arabian camel	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.0597	NA
Camelus ferus	wild bactrian camel	CR	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Canis lupus dingo	dingo	LC	No	Yes	No	No	No	No	NA	NA	NA
Canis lupus familiaris	domestic dog	LC	Yes	Yes	Yes	No	Yes	Previous	No	0.0043	NA
Canis lupus familiaris	domestic dog	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	NA
Capra aegagrus	wild goat	NT	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Capra hircus	goat	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.0107	NA
Capra ibex	alpine ibex	LC	No	Yes	No	No	No	No	NA	NA	NA
Capra sibirica	siberian ibex	NT	No	Yes	No	No	No	No	NA	NA	NA
Capreolus pygargus	siberian roe	LC	No	Yes	No	No	No	No	NA	NA	NA
Capromys pilorides	desmarest's hutia	LC	Yes	No	Yes	No	Yes	Yes	NA	NA	NA
Carlito syrichta	philippine tarsier	NT	No	Yes	No	No	No	No	No	0.2122	NA
Carollia perspicillata	seba's short-tailed bat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	0.0422	Yes
Castor canadensis	north american beaver	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.3538	No
Catagonus wagneri	chacoan peccary	EN	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Cavia aperea	brazilian guinea pig	LC	Yes	Yes	Yes	No	No	No	No	0.0016	No
Cavia porcellus	domestic guinea pig	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.0643	No
Cavia tschudii	montane guinea pig	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Cebus albifrons	white-fronted capuchin	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.6518	No
Cebus capucinus imitator	white-faced sapajou	VU	Yes	Yes	Yes	Yes	Yes	No	No	0.6135	No
Cephalophus harveyi	harvey's red duiker	LC	No	Yes	No	No	No	No	NA	NA	NA
Cephalorhynchus commersonii	Commerson's dolphin	NA	No	Yes	No	No	No	No	NA	NA	NA
Ceratotherium simum cottoni	northern white rhinoceros	CR	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Ceratotherium simum simum	southern white rhinoceros	NT	Yes	Yes	Yes	No	Yes	Yes	No	-0.5415	No
Cercocebus atys	sooty mangabey	VU	Yes	Yes	Yes	Yes	Yes	Yes	No	0.3542	No
Cercopithecus mona	mona monkey	NT	No	Yes	No	No	No	No	NA	0.4781	NA
Cercopithecus neglectus	de brazza's monkey	LC	Yes	Yes	Yes	Yes	No	Yes	No	0.3326	No
Cervus elaphus hippelaphus	central european red deer	LC	No	Yes	No	No	No	No	NA	NA	NA
Cervus hanglu yarkandensis	yarkand deer	LC	No	Yes	No	No	No	No	NA	NA	NA
Chaetophractus vellerosus	screaming hairy armadillo	LC	Yes	No	Yes	No	No	Yes	No	NA	NA
Cheirogaleus medius	fat-tailed dwarf lemur	VU	Yes	Yes	Yes	Yes	Yes	Yes	Yes	0.0929	No
Chinchilla lanigera	long-tailed chinchilla	EN	Yes	Yes	Yes	No	Yes	Yes	NA	0.1646	No
Chlorocebus sabaeus	green monkey	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.3463	No
Choloepus didactylus	linnaeus’s two-toed sloth	LC	Yes	Yes	Yes	No	No	Yes	NA	-0.0686	NA
Choloepus hoffmanni	hoffmann’s two-toed sloth	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.1275	NA
Chrysochloris asiatica	cape golden mole	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.1626	NA
Chrysocyon brachyurus	maned wolf	NA	No	Yes	No	No	No	No	NA	NA	NA
Coendou prehensilis	brazilian porcupine	LC	No	Yes	No	No	No	No	NA	-0.2491	NA
Colobus angolensis palliatus	angolan colobus	VU	Yes	Yes	Yes	Yes	Yes	Yes	No	0.1425	No
Condylura cristata	star-nosed mole	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1225	No
Connochaetes taurinus	blue wildebeest	LC	No	Yes	No	No	No	No	NA	-0.0888	NA
Craseonycteris thonglongyai	bumblebee bat	NT	Yes	Yes	Yes	No	Yes	Yes	NA	-0.1067	NA
Cricetomys gambianus	gambian pouched rat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.218	No
Cricetulus griseus	chinese hamster	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Crocidura indochinensis	indochinese shrew	LC	Yes	No	Yes	No	No	Yes	Yes	NA	NA
Crocuta crocuta	spotted hyena	LC	No	Yes	No	No	No	No	NA	-0.0933	NA
Cryptoprocta ferox	fossa	VU	Yes	Yes	Yes	No	Yes	Yes	No	-0.1067	No
Ctenodactylus gundi	common gundi	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Ctenomys sociabilis	social tuco-tuco	CR	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Cuniculus paca	lowland paca	LC	Yes	No	Yes	No	Yes	Yes	NA	NA	NA
Cynomys gunnisoni	gunnison's prairie dog	LC	No	Yes	No	No	No	No	NA	-0.1349	NA
Cynopterus brachyotis	lesser short-nosed fruit bat	LC	No	Yes	No	No	No	No	NA	0.109	NA
Damaliscus lunatus	common tsessebe	LC	No	Yes	No	No	No	No	NA	-0.0608	NA
Dasyprocta punctata	central american agouti	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.1026	No

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Dasypus novemcinctus	nine-banded armadillo	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.3894	NA
Daubentonia madagascariensis	aye-aye	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.3292	No
Delphinapterus leucas	beluga whale	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.2704	Yes
Desmodus rotundus	common vampire bat	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0804	Yes
Dicerorhinus sumatrensis sumatrensis	sumatran rhinoceros	CR	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Diceros bicornis	black rhinoceros	CR	Yes	Yes	Yes	No	Yes	Yes	No	-0.3368	No
Dinomys branickii	pacarana	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.4191	No
Dipodomys ordii	ord's kangaroo rat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	0.1039	No
Dipodomys stephensi	stephen's kangaroo rat	VU	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Dolichotis patagonum	patagonian mara	NT	Yes	Yes	Yes	No	Yes	Yes	NA	-0.1831	No
Dugong dugon	dugong	NA	No	Yes	No	No	No	No	NA	NA	NA
Echinops telfairi	lesser hedgehog tenrec	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.4042	NA
Eidolon dupreanum	malagasy straw-colored fruit bat	VU	No	Yes	No	No	No	No	NA	NA	NA
Eidolon helvum	straw-colored fruit bat	NT	Yes	Yes	Yes	No	Yes	Yes	No	0.087	NA
Elaphurus davidianus	pere david's deer	CR	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Elephantulus edwardii	cape elephant shrew	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	NA
Elephas maximus	asian elephant	EN	No	Yes	No	No	No	No	NA	0.1455	NA
Ellobius lutescens	transcaucasian mole vole	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Ellobius talpinus	northern mole vole	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	NA	No
Enhydra lutris kenyoni	sea otter	EN	Yes	Yes	Yes	No	Yes	Yes	No	0.0558	No
Enhydra lutris nereis	sea otter	EN	No	Yes	No	No	No	No	NA	NA	No
Eonycteris spelaea	lesser dawn bat	LC	No	Yes	No	No	No	No	NA	0.0559	NA
Eptesicus fuscus	big brown bat	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	-0.1967	NA
Equus asinus asinus	donkey	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.1295	NA
Equus burchellii boehmi	quagga	NA	No	Yes	No	No	No	No	NA	-0.0572	NA
Equus burchellii quagga	quagga	NA	No	Yes	No	No	No	No	NA	NA	NA
Equus caballus	horse	LC	Yes	Yes	Yes	No	Yes	Previous	No	-0.0486	NA
Equus przewalskii	przewalski's horse	EN	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Equus zebra	mountain zebra	NA	No	Yes	No	No	No	No	NA	NA	NA
Erethizon dorsatum	north american porcupine	LC	No	Yes	No	No	No	No	NA	-0.1482	NA
Erignathus barbatus	bearded seal	LC	No	Yes	No	No	No	No	NA	-0.1638	NA
Erinaceus europaeus	western european hedgehog	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.3186	No
Erythrocebus patas	patas monkey	NT	Yes	Yes	Yes	Yes	Yes	Yes	No	0.36	No
Eschrichtius robustus	grey whale	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.7128	NA
Eubalaena australis	Southern right whale	NA	No	Yes	No	No	No	No	NA	NA	NA
Eubalaena glacialis	north atlantic right whale	CR	No	Yes	No	No	No	No	NA	NA	NA
Eubalaena japonica	north pacific right whale	EN	Yes	Yes	Yes	No	Yes	Yes	No	NA	Yes
Eudorcas thomsonii	thomson's gazelle	LC	No	Yes	No	No	No	No	NA	-0.0674	NA
Eulemur flavifrons	sclater's lemur	CR	Yes	Yes	Yes	Yes	Yes	Yes	No	NA	No
Eulemur fulvus	common brown lemur	VU	Yes	Yes	Yes	Yes	Yes	Yes	No	0.1514	No
Eulemur macaco	black lemur	EN	No	Yes	No	No	No	No	NA	0.1245	NA
Eulemur mongoz	mongoose lemur	CR	No	Yes	No	No	No	No	NA	0.246	NA
Eumetopias jubatus	steller sea lion	NT	No	Yes	No	No	No	No	NA	-0.2762	NA
Felis catus	domestic cat	LC	Yes	Yes	Yes	No	Yes	Previous	No	0.0293	NA
Felis nigripes	black-footed cat	VU	Yes	Yes	Yes	No	No	Yes	No	0.2821	No
Fukomys damarensis	damara mole-rat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Galeopterus variegatus	sunda flying lemur	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	NA
Gallus gallus	red junglefowl	NA	No	No	No	No	No	No	NA	NA	NA
Giraffa camelopardalis	giraffe	VU	No	Yes	No	No	No	No	NA	-0.4568	NA
Giraffa tippelskirchi	giraffe	VU	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Glis glis	edible dormouse	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.089	No
Globicephala melas	long-finned pilot whale	LC	No	Yes	No	No	No	No	NA	0.1676	NA

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Gorilla gorilla gorilla	western lowland gorilla	CR	Yes	Yes	Yes	Yes	Yes	Previous	No	0.1055	No
Grammomys surdaster	common grammomys	NA	No	Yes	No	No	No	No	NA	NA	NA
Grampus griseus	Risso's dolphin	NA	No	Yes	No	No	No	No	NA	NA	NA
Graphiurus murinus	woodland doormouse	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Gulo gulo	wolverine	LC	No	Yes	No	No	No	No	NA	0.1013	NA
Halichoerus grypus	gray seal	LC	No	Yes	No	No	No	No	NA	-0.2181	NA
Helarctos malayanus	Malayan sun bear	NA	No	Yes	No	No	No	No	NA	NA	NA
Helogale parvula	dwarf mongoose	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1213	No
Hemitragus hylocrius	nilgiri tahr	EN	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Heterocephalus glaber	naked mole-rat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.362	NA
Heterohyrax brucei	yellow- spotted rock hyrax	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.2155	NA
Hippopotamus amphibius	hippopotamus	VU	Yes	Yes	Yes	No	Yes	Yes	No	-0.4807	NA
Hipposideros armiger	great roundleaf bat	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	-0.1481	Yes
Hipposideros galeritus	cantor's leaf-nosed bat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.0572	Yes
Hippotragus equinus	roan antelope	NA	No	Yes	No	No	No	No	NA	NA	NA
Hippotragus niger niger	sable antelope	LC	No	Yes	No	No	No	No	NA	-0.2029	NA
Homo sapiens	human	LC	Yes	Yes	Yes	Yes	Yes	Previous	No	0.7466	Yes
Hyaena hyaena	striped hyena	NT	Yes	Yes	Yes	No	Yes	Yes	No	-0.171	No
Hydrochoerus hydrochaeris	capybara	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Hydrodamalis gigas	Steller's sea cow	NA	No	Yes	No	No	No	No	NA	NA	NA
Hydropotes inermis	chinese water deer	VU	No	Yes	No	No	No	No	NA	-0.0639	NA
Hylobates moloch	silvery javan gibbon	EN	No	Yes	No	No	No	No	NA	0.4093	NA
Hystrix cristata	crested porcupine	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.3616	No
Ictidomys tridecemlineatus	thirteen-lined ground squirrel	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	NA	No
Indri indri	indri	CR	Yes	Yes	Yes	Yes	Yes	Yes	No	-0.0038	NA
Inia geoffrensis	amazon river dolphin	EN	Yes	Yes	Yes	No	Yes	Yes	No	0.3389	Yes
Jaculus jaculus	lesser egyptian jerboa	LC	Yes	Yes	Yes	No	Yes	No	NA	0.089	No
Kobus ellipsiprymnus	waterbuck	LC	No	Yes	No	No	No	No	NA	-0.2336	NA
Kobus leche leche	lechwe	NT	No	Yes	No	No	No	No	NA	-0.2098	NA
Kogia breviceps	pygmy sperm whale	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1253	NA
Lagenorhynchus obliquidens	pacific white-sided dolphin	LC	No	Yes	No	No	No	No	NA	0.6008	NA
Lama glama chaku	guanaco	NA	No	Yes	No	No	No	No	NA	NA	NA
Lama guanicoe cacsilensis	guanaco	LC	No	Yes	No	No	No	No	NA	NA	NA
Lasiurus borealis	eastern red bat	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	-0.1744	NA
Lemur catta	ring tailed lemur	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.1049	No
Leptonychotes weddellii	weddell seal	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.1723	Yes
Leptonycteris yerbabuenae	lesser long-nosed bat	NT	No	Yes	No	No	No	No	NA	NA	NA
Lepus americanus	snowshoe hare	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.1344	No
Lepus timidus	mountain hare	LC	No	Yes	No	No	No	No	NA	-0.1974	NA
Lipotes vexillifer	yangtze river dolphin	CR	Yes	Yes	Yes	No	Yes	Yes	No	0.0651	NA
Litocranius walleri	gerenuk	NT	No	Yes	No	No	No	No	NA	-0.1195	NA
Lontra canadensis	north american river otter	LC	No	Yes	No	No	No	No	NA	0.0421	NA
Loxodonta africana	african savanna elephant	EN	Yes	Yes	Yes	No	Yes	Yes	No	0.02	NA
Lutra lutra	european otter	NT	No	Yes	No	No	No	No	NA	-0.021	NA
Lycaon pictus	african hunting dog	EN	Yes	Yes	Yes	No	Yes	Yes	No	0.113	No
Lynx canadensis	canadian lynx	LC	No	Yes	No	No	No	No	NA	-0.0295	NA
Lynx pardinus	iberian lynx	EN	No	Yes	No	No	No	No	NA	NA	NA
Macaca fascicularis	crab-eating macaque	VU	Yes	Yes	Yes	Yes	Yes	Yes	No	0.4661	No
Macaca fuscata	japanese macaque	LC	No	Yes	No	No	No	No	NA	0.3173	NA
Macaca mulatta	rhesus monkey	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.4564	No
Macaca nemestrina	southern pig-tailed macaque	VU	Yes	Yes	Yes	Yes	Yes	Yes	No	0.5646	No
Macroglossus sobrinus	long-tongued fruit bat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	0.0977	NA

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Macrotus californicus	california leaf-nosed bat	LC	No	Yes	No	No	No	No	NA	0.0156	NA
Madoqua kirkii	kirk's dikdik	LC	No	Yes	No	No	No	No	NA	0.0832	NA
Mandrillus leucophaeus	drill	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.2981	No
Mandrillus sphinx	mandrill	VU	No	Yes	No	No	No	No	NA	0.4164	NA
Manis javanica	malayan pangolin	CR	Yes	Yes	Yes	No	Yes	No	NA	-0.3872	No
Manis pentadactyla	chinese pangolin	CR	Yes	Yes	Yes	No	Yes	Yes	NA	-0.2459	No
Manis tricuspis	tree pangolin	NA	No	Yes	No	No	No	No	NA	-0.2066	NA
Marmota flaviventris	yellow-bellied marmot	LC	No	Yes	No	No	No	No	NA	-0.295	NA
Marmota himalayana	himalayan marmot	LC	No	Yes	No	No	No	No	NA	-0.3821	NA
Marmota marmota marmota	alpine marmot	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	-0.2722	No
Marmota monax	woodchuck	LC	No	Yes	No	No	No	No	NA	-0.1741	NA
Marmota vancouverensis	vancouver island marmot	CR	No	Yes	No	No	No	No	NA	-0.3253	NA
Martes zibellina	sable	LC	No	Yes	No	No	No	No	NA	0.2373	NA
Mastacomys fuscus	broad-toothed mouse	NA	No	Yes	No	No	No	No	NA	NA	NA
Mastomys coucha	southern african mastomys	LC	No	Yes	No	No	No	No	NA	NA	NA
Megaderma lyra	greater false vampire bat	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0502	Yes
Megaptera novaeangliae	humpback whale	LC	No	Yes	No	No	No	No	NA	-0.5631	NA
Mellivora capensis	honey badger	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1925	No
Meriones unguiculatus	mongolian jird	LC	Yes	Yes	Yes	No	No	Yes	NA	-0.0061	No
Mesocricetus auratus	golden hamster	VU	Yes	Yes	Yes	No	Yes	Yes	NA	-0.2994	No
Mesoplodon bidens	sowerby’s beaked whale	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	NA
Mesoplodon densirostris	Blainville's beaked whale	NA	No	Yes	No	No	No	No	NA	NA	NA
Mesoplodon europaeus	Gervais' beaked whale	NA	No	Yes	No	No	No	No	NA	NA	NA
Mesoplodon stejnegeri	Stejneger's beaked whale	NA	No	Yes	No	No	No	No	NA	NA	NA
Microcebus murinus	gray mouse lemur	LC	Yes	Yes	Yes	Yes	Yes	Yes	Yes	0.2022	No
Microcebus sp. 3 GT-2019	mouse lemur	NA	No	Yes	No	No	No	No	NA	NA	NA
Microcebus tavaratra	northern rufous mouse lemur	VU	No	Yes	No	Yes	Yes	Yes	NA	NA	NA
Microgale talazaci	talazac's shrew tenrec	LC	Yes	Yes	Yes	No	No	No	NA	-0.1192	NA
Micronycteris hirsuta	hairy big-eared bat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	NA
Microtus agrestis	field vole	LC	No	Yes	No	No	No	No	NA	-0.0037	NA
Microtus arvalis	common vole	LC	No	Yes	No	No	No	No	NA	-0.1118	NA
Microtus fortis	reed vole	NA	No	Yes	No	No	No	No	NA	NA	NA
Microtus ochrogaster	prairie vole	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.1199	No
Microtus oeconomus	root vole	LC	No	Yes	No	No	No	No	NA	NA	NA
Miniopterus natalensis	atal long-fingered bat	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	NA	NA
Miniopterus schreibersii	common bent-wing bat	VU	Yes	Yes	Yes	No	Yes	Yes	Yes	-0.1115	NA
Mirounga angustirostris	northern elephant seal	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.2978	Yes
Mirounga leonina	southern elephant seal	LC	No	Yes	No	No	No	No	NA	-0.3859	NA
Mirza coquereli	coquerel's giant mouse lemur	EN	Yes	Yes	Yes	Yes	Yes	Yes	NA	0.153	No
Mirza zaza	coquerel's giant mouse lemur	VU	No	Yes	No	No	No	No	NA	NA	NA
Molossus molossus	pallas's mastiff bat	LC	No	Yes	No	No	No	Previous	NA	-0.1269	NA
Monodelphis domestica	gray short-tailed opossum	LC	No	No	No	No	No	No	NA	NA	NA
Monodon monoceros	narwhal	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0875	NA
Mormoops blainvillei	ghost-faced bat	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	NA
Moschus berezovskii	forest musk deer	EN	No	Yes	No	No	No	No	NA	NA	NA
Moschus chrysogaster	alpine musk deer	EN	No	Yes	No	No	No	No	NA	NA	NA
Moschus moschiferus	siberian musk deer	VU	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Mungos mungo	south african banded mongoose	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0803	No
Muntiacus crinifrons	black muntjac	VU	No	Yes	No	No	No	No	NA	NA	NA
Muntiacus muntjak	red muntjac	LC	No	Yes	No	No	No	No	NA	0.19	NA
Muntiacus reevesi	reeves' muntjac	LC	No	Yes	No	No	No	No	NA	NA	NA
Murina aurata feae	ashy-gray tube-nosed bat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	Yes

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Mus caroli	ryukyu mouse	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Mus musculus	house mouse	LC	Yes	Yes	Yes	No	Yes	Previous	NA	-0.0222	No
Mus pahari	shrew mouse	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Mus spicilegus	mound-building mouse	LC	No	Yes	No	No	No	No	NA	NA	NA
Mus spretus	western mediterranean mouse	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Muscardinus avellanarius	hazel dormouse	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Mustela erminea	ermine	LC	No	Yes	No	No	No	No	NA	0.2741	NA
Mustela putorius putorius	domestic ferret	LC	Yes	Yes	No	No	No	No	No	-0.0651	NA
Mustela putorius furo	domestic ferret	LC	No	Yes	Yes	No	Yes	Yes	NA	NA	NA
Myocastor coypus	coypu	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.3935	No
Myodes glareolus	bank vole	LC	No	Yes	No	No	No	No	NA	0.0363	NA
Myotis brandtii	brandt's bat	LC	Yes	Yes	Yes	No	Yes	Previous	Yes	NA	Yes
Myotis davidii	david's myotis	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	NA	NA
Myotis lucifugus	little brown bat	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	-0.2096	NA
Myotis myotis	greater mouse-eared bat	LC	Yes	Yes	Yes	No	Yes	Previous	Yes	-0.1352	Yes
Myotis septentrionalis	northern myotis	NT	No	Yes	No	No	No	No	NA	NA	NA
Myrmecophaga tridactyla	giant anteater	VU	Yes	Yes	Yes	No	Yes	Yes	NA	-0.0238	NA
Nanger granti	grant's gazelle	LC	No	Yes	No	No	No	No	NA	-0.0526	NA
Nannospalax galili	upper galilee mountains blind mole rat	DD	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Nasalis larvatus	proboscis monkey	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.1921	No
Nasua narica	white-nosed coati	LC	No	Yes	No	No	No	No	NA	0.0582	NA
Neofelis nebulosa	clouded leopard	VU	No	Yes	No	No	No	No	NA	-0.1205	NA
Neomonachus schauinslandi	hawaiian monk seal	EN	Yes	Yes	Yes	No	Yes	Yes	No	NA	NA
Neophocaena asiaeorientalis asiaeorientalis	yangtze finless porpoise	EN	Yes	Yes	Yes	No	Yes	Yes	No	NA	NA
Neotoma lepida	desert woodrat	LC	No	Yes	No	No	No	No	NA	NA	NA
Neotragus moschatus	suni	LC	No	Yes	No	No	No	No	NA	0.1312	NA
Neotragus pygmaeus	royal antelope	LC	No	Yes	No	No	No	No	NA	NA	NA
Neovison vison	american mink	LC	No	Yes	No	No	No	No	NA	-0.0021	NA
Noctilio leporinus	greater bulldog bat	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0043	NA
Nomascus leucogenys	northern white-cheeked gibbon	CR	Yes	Yes	Yes	Yes	Yes	Yes	No	NA	No
Nycticebus coucang	sunda slow loris	EN	Yes	Yes	Yes	Yes	Yes	Yes	NA	0.2527	No
Nycticeius humeralis	egyptian slit-faced bat	LC	No	Yes	No	No	No	No	NA	NA	NA
Ochotona princeps	american pika	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.0326	No
Octodon degus	degu	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.1595	No
Odobenus rosmarus rosmarus	pacific walrus	VU	No	Yes	No	No	No	No	No	-0.0979	Yes
Odobenus rosmarus divergens	pacific walrus	VU	Yes	Yes	Yes	No	Yes	Yes	NA	NA	Yes
Odocoileus hemionus hemionus	mule deer	LC	No	Yes	No	No	No	No	NA	-0.041	NA
Odocoileus virginianus virginianus	white-tailed deer	LC	Yes	Yes	No	No	No	No	No	-0.028	No
Odocoileus virginianus texanus	white-tailed deer	LC	No	Yes	Yes	No	Yes	No	NA	NA	NA
Okapia johnstoni	okapi	EN	Yes	Yes	Yes	No	Yes	Yes	No	-0.0754	No
Ondatra zibethicus	muskrat	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.3301	No
Onychomys torridus	scorpion mouse	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0547	No
Orcinus orca	killer whale	DD	Yes	Yes	Yes	No	Yes	Yes	No	0.1556	Yes
Oreamnos americanus	mountain goat	LC	No	Yes	No	No	No	No	NA	NA	NA
Oreotragus oreotragus	klipspringer	LC	No	Yes	No	No	No	No	NA	0.0027	NA
Ornithorhynchus anatinus	platypus	NT	No	No	No	No	No	No	NA	NA	NA
Orycteropus afer afer	aardvark	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.3271	NA
Oryctolagus cuniculus cuniculus	rabbit	EN	Yes	Yes	Yes	No	Yes	Yes	No	-0.1956	No
Oryx dammah	scimitar-horned oryx	EW	No	Yes	No	No	No	No	NA	NA	NA
Oryx gazella	gemsbok	LC	No	Yes	No	No	No	No	NA	-0.1878	NA
Otolemur garnettii	small-eared galago	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.1582	No

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Ovis ammon	argali	NT	No	Yes	No	No	No	No	NA	-0.1492	NA
Ovis aries	sheep	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.0552	No
Ovis canadensis canadensis	peninsular bighorn sheep	EN	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Ovis nivicola lydekkeri	snow sheep	NA	No	Yes	No	No	No	No	NA	NA	NA
Ovis orientalis	mouflon	NA	No	Yes	No	No	No	No	NA	NA	NA
Pachyuromys duprasi	fat-tailed gerbil	NA	No	Yes	No	No	No	No	NA	NA	NA
Pan paniscus	bonobo	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.329	NA
Pan troglodytes	chimpanzee	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.224	NA
Panthera leo	lion	VU	No	Yes	No	No	No	No	NA	-0.0191	NA
Panthera onca	jaguar	NT	Yes	Yes	Yes	No	Yes	Yes	No	-0.2347	No
Panthera pardus	leopard	VU	Yes	Yes	Yes	No	Yes	Yes	No	-0.1207	No
Panthera tigris altaica	amur tiger	EN	Yes	Yes	Yes	No	Yes	Yes	No	-0.2767	No
Panthera uncia	snow leopard	NA	No	Yes	No	No	No	No	NA	NA	NA
Pantholops hodgsonii	chiru	NT	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Papio anubis	olive baboon	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.35	No
Paradoxurus hermaphroditus	asian palm civet	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0918	No
Pedetes capensis	south african springhare	LC	No	Yes	No	No	No	No	NA	NA	NA
Peponocephala electra	melon-headed whale	LC	No	Yes	No	No	No	No	NA	NA	NA
Perognathus longimembris pacificus	pacific pocket mouse	EN	Yes	Yes	Yes	No	Yes	Yes	NA	0.0951	No
Peromyscus californicus insignis	california leaf-nosed bat	LC	No	Yes	No	No	No	No	NA	-0.0101	NA
Peromyscus crinitus	canyon deermouse	LC	No	Yes	No	No	No	No	NA	0.135	NA
Peromyscus eremicus	cactus deermouse	LC	No	Yes	No	No	No	No	NA	0.0177	NA
Peromyscus leucopus	white-footed deermouse	LC	No	Yes	No	No	No	No	NA	0.0432	NA
Peromyscus maniculatus bairdii	prairie deer mouse	LC	Yes	Yes	Yes	No	Yes	Yes	NA	0.0142	No
Peromyscus nasutus	northern rock deermouse	LC	No	Yes	No	No	No	No	NA	NA	NA
Peromyscus polionotus subgriseus	oldfield deermouse	LC	No	Yes	No	No	No	No	NA	0.0398	NA
Petromus typicus	dassie rat	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Phascolarctos cinereus	koala	VU	No	No	No	No	No	No	NA	NA	NA
Phataginus tricuspis	tree pangolin	EN	No	Yes	No	No	No	No	NA	NA	NA
Philantomba maxwellii	maxwell's duiker	LC	No	Yes	No	No	No	No	NA	0.1953	NA
Phoca largha	spotted seal	NA	No	Yes	No	No	No	No	NA	NA	NA
Phoca vitulina	harbor seal	LC	No	Yes	No	No	No	No	NA	0.1079	NA
Phocoena phocoena	harbor porpoise	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.634	NA
Phocoena sinus	vaquita	CR	No	Yes	No	No	No	No	NA	NA	NA
Phyllostomus discolor	pale spear-nosed bat	LC	No	Yes	No	No	No	Previous	NA	0.1418	NA
Physeter catodon	sperm whale	NA	No	Yes	No	No	No	No	NA	-0.507	NA
Piliocolobus tephrosceles	ugandan red colobus	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.1674	No
Pipistrellus kuhlii	kuhl's pipistrelle	LC	No	Yes	No	No	No	Previous	NA	-0.1194	NA
Pipistrellus pipistrellus	common pipistrelle	LC	Yes	Yes	Yes	No	No	Yes	Yes	NA	Yes
Pithecia pithecia	white-faced saki	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.3742	No
Platanista minor	indus river dolphin	EN	Yes	Yes	Yes	No	Yes	Yes	NA	NA	NA
Plecturocebus donacophilus	white-eared titi	LC	Yes	Yes	Yes	Yes	Yes	Yes	NA	NA	No
Pongo abelii	sumatran orangutan	CR	Yes	Yes	Yes	Yes	Yes	Yes	No	0.2519	No
Pontoporia blainvillei	la plata dolphin	VU	No	No	No	No	No	Yes	NA	NA	NA
Potos flavus	kinkajou	LC	No	Yes	No	No	No	No	NA	0.2666	NA
Prionailurus bengalensis euptilurus	leopard cat	LC	No	Yes	No	No	No	No	NA	NA	NA
Procapra przewalskii	przewalski's gazelle	EN	No	Yes	No	No	No	No	NA	NA	NA
Procavia capensis	rock hyrax	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.025	NA
Procyon lotor	raccoon	LC	No	Yes	No	No	No	No	NA	0.1008	NA
Prolemur simus	greater bamboo lemur	CR	No	Yes	No	No	No	No	NA	0.1943	NA
Propithecus coquereli	coquerel's sifaka	CR	Yes	Yes	Yes	Yes	Yes	Yes	No	0.0612	No
Przewalskium albirostris	white-lipped deer	NA	No	Yes	No	No	No	No	NA	NA	NA

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Psammomys obesus	fat sand rat	LC	Yes	Yes	Yes	No	No	Yes	NA	-0.0057	No
Pseudomys desertor	brown desert mouse	NA	No	Yes	No	No	No	No	NA	NA	NA
Pteronotus parnellii	parnell's mustached bat	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0247	Yes
Pteronura brasiliensis	giant otter	EN	Yes	Yes	Yes	No	Yes	Yes	No	-0.0927	No
Pteropus alecto	black flying fox	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.0266	Yes
Pteropus giganteus	indian flying fox	LC	No	Yes	No	No	No	No	NA	0.0801	NA
Pteropus pselaphon	bonin flying fox	EN	No	Yes	No	No	No	No	NA	NA	NA
Pteropus rufus	malagasy flying fox	VU	No	Yes	No	No	No	No	NA	0.2228	NA
Pteropus vampyrus	large flying fox	NT	Yes	Yes	Yes	No	Yes	Yes	No	-0.0298	Yes
Puma concolor	puma	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.1044	No
Puma yagouaroundi	jaguarundi	NA	No	Yes	No	No	No	No	NA	0.3252	NA
Pygathrix nemaeus	red-shanked douc langur	CR	Yes	Yes	Yes	Yes	Yes	Yes	No	0.2401	No
Rangifer tarandus tarandus	siberian reindeer	VU	Yes	Yes	(Yes)	No	(Yes)	(Yes)	No	0.0623	No
Rangifer tarandus granti	siberian reindeer	VU	No	Yes	(Yes)	No	(Yes)	(Yes)	NA	NA	NA
Raphicerus campestris	steenbok	LC	No	Yes	No	No	No	No	NA	-0.0364	NA
Rattus norvegicus	brown rat	LC	Yes	Yes	Yes	No	Yes	Previous	No	0.2479	No
Rattus rattus	roof rat	LC	No	Yes	No	No	No	No	NA	-0.226	NA
Redunca redunca	common reedbuck	LC	No	Yes	No	No	No	No	NA	-0.1577	NA
Rhinoceros unicornis	indian rhinoceros	VU	No	Yes	No	No	No	No	NA	-0.6088	NA
Rhinolophus ferrumequinum	greater horseshoe bat	LC	No	Yes	No	No	No	Previous	NA	-0.2243	NA
Rhinolophus sinicus	chinese rufous horseshoe bat	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	NA	Yes
Rhinopithecus bieti	black snub-nosed monkey	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	NA	No
Rhinopithecus roxellana	golden snub-nosed monkey	EN	Yes	Yes	Yes	Yes	Yes	Yes	No	0.2043	No
Rhizomys pruinosus	hoary bamboo rat	LC	No	Yes	No	No	No	Yes	NA	NA	NA
Rhombomys opimus	great gerbil	LC	No	Yes	No	No	No	No	NA	NA	NA
Rousettus aegyptiacus	egyptian fruit bat	LC	Yes	Yes	Yes	No	Yes	Previous	No	0.0297	Yes
Rousettus leschenaultii	Leschenault's rousette	NA	No	Yes	No	No	No	No	NA	NA	NA
Rousettus madagascariensis	malagasy rousette	VU	No	Yes	No	No	No	No	NA	NA	NA
Saguinus imperator	emperor tamarin	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	NA	No
Saiga tatarica	steppe saiga	CR	Yes	Yes	Yes	No	No	Yes	No	NA	No
Saimiri boliviensis boliviensis	bolivian squirrel monkey	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.485	No
Sapajus apella	tufted capuchin	LC	No	Yes	No	No	No	No	NA	NA	NA
Sarcophilus harrisii	tasmanian devil	EN	No	No	No	No	No	No	NA	NA	NA
Scalopus aquaticus	eastern mole	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.232	No
Sciurus carolinensis	eastern gray squirrel	LC	No	Yes	No	No	No	No	NA	0.1035	NA
Sciurus vulgaris	red squirrel	LC	No	Yes	No	No	No	No	NA	0.1467	NA
Semnopithecus entellus	northern plains gray langur	LC	Yes	Yes	Yes	Yes	Yes	Yes	No	0.4229	No
Sigmodon hispidus	hispid cotton rat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.2748	No
Solenodon paradoxus	hispaniolan solenodon	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.2864	No
Sorex araneus	common shrew	LC	Yes	Yes	Yes	No	Yes	Yes	No	-0.0903	No
Sousa chinensis	indo-pacific humpbacked dolphin	VU	No	Yes	No	No	No	No	NA	NA	NA
Spermophilus dauricus	daurian ground squirrel	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	-0.0631	No
Spilogale gracilis	western spotted skunk	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1257	No
Steno bredanensis	rough-toothed dolphin	NA	No	Yes	No	No	No	No	NA	NA	NA
Sturnira hondurensis	highland yellow-shouldered bat	LC	No	Yes	No	No	No	No	NA	0.0299	NA
Suricata suricatta	meerkat	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1275	No
Sus scrofa	pig	LC	Yes	Yes	Yes	No	Yes	Previous	No	-0.2139	No
Sylvicapra grimmia	bush duiker	LC	No	Yes	No	No	No	No	NA	0.0716	NA
Sylvilagus bachmani	brush rabbit	LC	No	Yes	No	No	No	No	NA	-0.0614	NA
Syncerus caffer	african buffalo	NT	No	Yes	No	No	No	No	NA	-0.2962	NA
Tadarida brasiliensis	mexican free-tailed bat	LC	Yes	Yes	Yes	No	Yes	Yes	Yes	NA	Yes

Species	Common Name	IUCN Status	Bio-diversity Analysis	TOGA	Cactus	Primate phastCons	Phylo geny	Repeat Annotated	Hiber nation	Brain Size Residual	Vocal Learning
Talpa occidentalis	spanish mole	LC	No	Yes	No	No	No	No	NA	NA	NA
Tamandua tetradactyla	southern tamandua	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.0304	NA
Tapirus indicus	malayan tapir	EN	Yes	Yes	Yes	No	Yes	Yes	NA	-0.2942	No
Tapirus terrestris	south american tapir	VU	Yes	Yes	Yes	No	Yes	Yes	NA	-0.3867	No
Taxidea taxus jeffersonii	american badger	LC	No	Yes	No	No	No	No	NA	0.0462	NA
Theropithecus gelada	gelada	LC	No	Yes	No	No	No	No	NA	0.4586	NA
Thryonomys swinderianus	greater cane rat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	-0.025	No
Tolypeutes matacus	southern three-banded armadillo	NT	Yes	Yes	Yes	No	Yes	Yes	NA	-0.102	NA
Tonatia saurophila	stripe-headed round-eared bat	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	NA
Trachypithecus francoisi	françois' langur	EN	No	Yes	No	No	No	No	NA	0.2656	NA
Tragelaphus imberbis	lesser kudu	NT	No	Yes	No	No	No	No	NA	-0.2706	NA
Tragelaphus scriptus	bushbuck	LC	No	Yes	No	No	No	No	NA	-0.0068	NA
Tragelaphus strepsiceros	greater kudu	LC	No	Yes	No	No	No	No	NA	-0.2679	NA
Tragulus javanicus	java lesser chevrotain	DD	Yes	Yes	Yes	No	Yes	Yes	No	-0.0231	No
Tragulus kanchil	lesser mouse-deer	LC	No	Yes	No	No	No	No	NA	NA	NA
Trichechus manatus latirostris	florida manatee	EN	Yes	Yes	Yes	No	Yes	Yes	No	-0.5879	NA
Tupaia belangeri	northern tree shrew	LC	No	Yes	No	No	No	No	NA	NA	NA
Tupaia chinensis	northern tree shrew	LC	Yes	Yes	Yes	No	Yes	Yes	No	NA	No
Tupaia tana	large treeshrew	LC	Yes	No	Yes	No	Yes	Yes	NA	NA	No
Tursiops aduncus	indo-pacific bottlenose dolphin	NT	No	Yes	No	No	No	No	NA	0.5732	NA
Tursiops truncatus	bottlenose dolphin	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.5329	Yes
Urocitellus parryii	daurian ground squirrel	LC	No	Yes	No	No	No	No	NA	NA	NA
Uromys caudimaculatus	giant uromys	NA	No	Yes	No	No	No	No	NA	NA	NA
Uropsilus gracilis	gracile shrew mole	LC	Yes	Yes	Yes	No	Yes	Yes	NA	NA	No
Ursus americanus	american black bear	LC	No	Yes	No	No	No	No	NA	0.3433	NA
Ursus arctos horribilis	brown bear	LC	No	Yes	No	No	No	No	NA	-0.0518	NA
Ursus maritimus	polar bear	VU	Yes	Yes	Yes	No	Yes	Yes	Yes	-0.1608	No
Ursus thibetanus thibetanus	asian black bear	VU	No	Yes	No	No	No	No	NA	0.0565	NA
Vicugna pacos	alpaca	LC	Yes	Yes	No	No	No	No	No	NA	NA
Vicugna pacos huacaya	alpaca	NA	No	Yes	Yes	No	Yes	Yes	NA	NA	NA
Vicugna vicugna mensalis	vicugna	NA	No	Yes	No	No	No	No	NA	NA	NA
Vombatus ursinus	common wombat	LC	No	No	No	No	No	No	NA	NA	NA
Vulpes lagopus	arctic fox	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1601	No
Vulpes vulpes	red fox	LC	No	Yes	No	No	No	No	NA	0.2776	NA
Xerus inauris	cape ground squirrel	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1921	No
Zalophus californianus	california sea lion	LC	Yes	Yes	Yes	No	Yes	Yes	No	0.1481	Yes
Zapus hudsonius	meadow jumping mouse	LC	Yes	Yes	Yes	No	No	Yes	Yes	0.0686	No
Ziphius cavirostris	cuvier's beaked whale	LC	Yes	Yes	Yes	No	No	Yes	No	-0.2056	NA

Table S1.
Species' names, species' taxonomy, analyses in which species were included, and species' phenotype information. "NA" means data was not available or not used in analysis. Annotations in parentheses indicate that neither the phenotype annotation information nor the assembly in the Cactus alignment had an indication of subspecies, but the TOGA alignments have two subspecies. "Brain size residual" is the residual of the brain size when regressing out body size. IUCN status is abbreviated as DD = Data Deficient; LC = Least Concern; VU = Vulnerable; NT = Near Threatened; EN = Endangered; CR = Critically Endangered; EW = Extinct in the Wild.

	Human			Chimp			Mouse			Dog			Little brown bat		
	Homo sapiens			*Pan troglodytes*			*Mus musculus*			*Canis lupus familiaris*			*Myotis lucifugus*		
genome (Gb)			3.09			3.1			2.73			2.33			2.03
	phyloP cutoff	% of genome	bases (Mb)	phyloP cutoff	% of genome	bases (Mb)	phyloP cutoff	% of genome	bases (Mb)	phyloP cutoff	% of genome	bases (Mb)	phyloP cutoff	% of genome	bases (Mb)
% constrained	n/a	10.70%	331.7	n/a	11.76%	358.8	n/a	8.99%	238.5	n/a	10.12%	244.8	n/a	18.04%	367.0
lower CI		10.70%	331.6		11.75%	358.3		8.96%	237.9		10.13%	244.2		18.04%	367.0
upper CI		10.74%	332.8		11.83%	360.7		9.02%	239.5		10.18%	245.5		18.04%	367.0
< 1% FDR	3.27	1.87%	58	3.26	1.80%	57	3.63	1.70%	47	3.52	2.11%	49.2	3.48	2.45%	50
< 3% FDR	2.59	2.67%	82	2.59	2.60%	81	3.02	2.10%	58	2.88	2.84%	66	2.85	3.27%	67
< 5% FDR *	2.27	3.26%	101	2.27	3.20%	99	2.7	2.60%	71	2.56	3.43%	79.8	2.54	3.85%	78
< 10% FDR	1.78	4.69%	145	1.78	4.60%	142	2.27	3.20%	88	2.1	4.69%	109.2	2.09	5.16%	105
< 20% FDR	1.25	7.56%	233	1.25	7.40%	229	1.79	4.40%	120	1.59	7.02%	163.4	1.61	7.38%	150

Table S2.
Genome-wide constraint and proportions of the genome under constraint at different false discovery rates for human, chimpanzee, mouse, dog, and little brown bat. Acronyms: Confidence Intervals (CI), False Discovery Rate (FDR), Megabases (Mb) and Gigabases (Gb).

species	Assembly	Mb constrained		% constrained		branch length to 9 closest species	Genome size (Mb)	scaffold N50 (Mb)	contig N50 (Mb)
		total	*FDR <0.05*	*total*	*FDR <0.05*				
dog	CanFam3.1	244.8	79.8	10.12%	3.30%	0.24	2411	45.9	0.3
human	GRCh38.p14	331.7	100.7	10.70%	3.20%	0.13	3100	67.8	57.9
mouse	GRCm39	238.5	70.8	8.99%	2.70%	0.67	2655	106.2	59.5
bat	Myoluc2.0	367.0	78.4	18.04%	3.90%	0.39	2035	4.3	0.1
chimp	Clint_PTRv2	358.8	98.7	11.76%	3.20%	0.13	3050	53.1	12.3
ρ (all constraint)		Mb constrained			cor	-0.36	-0.20	-0.70	-0.70
					p	0.55	0.78	0.23	0.23
		% constrainted			cor	-0.36	-0.20	-0.70	-0.70
					p	0.55	0.78	0.23	0.23
ρ (bases constrained at FDR<0.05)		Mb constrained			cor	**-0.97**	0.70	0.00	0.00
					p	**0.0048**	0.23	1.00	1.00
		% constrainted			cor	-0.05	-0.60	-0.90	-0.90
					p	0.93	0.35	0.08	0.08

Table S3. Constraint and genome quality information for mammals with Zoonomia phyloP scores. Bold indicates results that are significant ($p < 0.05$) after FDR correction.

direction	gene set	genes in set	overlap	expected overlap	ratio	p	FDR
conserved	GO:0006397: mRNA processing	487	81	27.4	2.96	0	0
conserved	GO:0198738: cell-cell signaling by wnt	460	79	25.9	3.05	0	0
conserved	GO:0007389: pattern specification process	433	76	24.4	3.12	0	0
conserved	GO:0008380: RNA splicing	417	74	23.5	3.15	0	0
conserved	GO:1903311: regulation of mRNA metabolic process	266	65	15.0	4.34	0	0
conserved	GO:0048706: embryonic skeletal system development	125	42	7.0	5.97	0	0
accelerated	GO:0098542: defense response to other organism	473	49	12.3	4.00	0	0
accelerated	GO:0043588: skin development	409	47	10.6	4.43	0	0
conserved	GO:0030900: forebrain development	375	64	21.1	3.03	8.9E-16	1.1E-13
accelerated	GO:0008544: epidermis development	455	46	11.8	3.90	1.9E-15	5.3E-13
conserved	GO:0045165: cell fate commitment	249	49	14.0	3.50	8.5E-15	9.1E-13
accelerated	GO:0001906: cell killing	157	27	4.1	6.63	4.7E-15	9.9E-13
conserved	GO:0048736: appendage development	177	40	10.0	4.02	2.2E-14	2.1E-12
conserved	GO:0010608: posttranscriptional regulation of gene expression	486	72	27.3	2.63	2.8E-14	2.2E-12
accelerated	GO:0006959: humoral immune response	242	28	6.3	4.46	3.3E-11	5.7E-09
accelerated	GO:0045088: regulation of innate immune response	369	31	9.6	3.24	8.8E-09	1.2E-06
accelerated	GO:0031349: positive regulation of defense response	427	31	11.1	2.80	2.4E-07	3.0E-05
accelerated	GO:0035821: modification of morphology or physiology of other organism	155	17	4.0	4.23	5.7E-07	6.0E-05
accelerated	GO:0007606: sensory perception of chemical stimulus	473	32	12.3	2.61	7.5E-07	7.1E-05
accelerated	GO:0030101: natural killer cell activation	85	12	2.2	5.44	1.9E-06	1.6E-04

Table S4.
GO Biological Process gene sets enriched for conserved or accelerated genes. We tested the top 5% most accelerated (low scoring) and most conserved (high scoring) genes as measured by mean phyloP score of coding sequence (Data S1) against a non-redundant representative set of Gene Ontology biological processes using WebGestalt.

	Gene set	Binomial P	Binomial Bonferroni P	Observed Regions	Observed Genes	Total Genes	Gene Set Cov	Term Cov
BP	GO:0000122: negative regulation of transcription from RNA polymerase II promoter	8.31E-291	1.07E-286	1327	197	796	0.1014	0.2475
BP	GO:0006366: transcription from RNA polymerase II promoter	3.65E-284	4.72E-280	1233	210	824	0.1081	0.2549
BP	GO:0003002: regionalization	4.86E-170	6.28E-166	684	110	312	0.0566	0.3526
BP	GO:0007417: central nervous system development	3.13E-164	4.04E-160	1271	199	875	0.1024	0.2274
BP	GO:0007389: pattern specification process	6.40E-155	8.26E-151	751	132	405	0.0679	0.3259
BP	GO:0045165: cell fate commitment	8.83E-149	1.14E-144	617	81	227	0.0417	0.3568
BP	GO:0021953: central nervous system neuron differentiation	1.63E-142	2.11E-138	551	67	163	0.0345	0.4110
BP	GO:0021516: dorsal spinal cord development	2.08E-132	2.69E-128	172	13	21	0.0067	0.6190
BP	GO:0043576: regulation of respiratory gaseous exchange	1.39E-120	1.80E-116	170	9	23	0.0046	0.3913
BP	GO:0009790: embryo development	1.02E-118	1.32E-114	1087	220	905	0.1132	0.2431
BP	GO:0030900: forebrain development	1.60E-114	2.07E-110	721	107	365	0.0551	0.2932
BP	GO:0060322: head development	6.14E-112	7.94E-108	991	165	713	0.0849	0.2314
BP	GO:0007420: brain development	1.07E-111	1.39E-107	965	156	675	0.0803	0.2311
BP	GO:0009792: embryo development ending in birth or egg hatching	3.17E-104	4.09E-100	749	132	560	0.0679	0.2357
BP	GO:0043009: chordate embryonic development	2.97E-103	3.84E-99	744	131	555	0.0674	0.2360
BP	GO:0009952: anterior/posterior pattern specification	1.38E-96	1.78E-92	394	63	196	0.0324	0.3214
BP	GO:0002087: regulation of respiratory gaseous exchange by neurological system process	4.02E-95	5.19E-91	128	6	13	0.0031	0.4615
BP	GO:0001708: cell fate specification	1.06E-94	1.36E-90	295	33	72	0.0170	0.4583
BP	GO:0048598: embryonic morphogenesis	4.43E-94	5.72E-90	770	143	550	0.0736	0.2600
BP	GO:0030182: neuron differentiation	5.27E-92	6.81E-88	1181	211	921	0.1086	0.2291
MF	GO:0000981: RNA polymerase II transcription factor activity, sequence-specific DNA binding	0.00E+00	0.00E+00	1644	227	766	0.1168	0.2963
MF	GO:0001067: regulatory region nucleic acid binding	0.00E+00	0.00E+00	1450	248	849	0.1276	0.2921
MF	GO:0000975: regulatory region DNA binding	0.00E+00	0.00E+00	1445	247	848	0.1271	0.2913
MF	GO:0044212: transcription regulatory region DNA binding	0.00E+00	0.00E+00	1445	247	846	0.1271	0.2920
MF	GO:0003690: double-stranded DNA binding	0.00E+00	0.00E+00	1393	224	809	0.1153	0.2769
MF	GO:1990837: sequence-specific double-stranded DNA binding	0.00E+00	0.00E+00	1377	216	731	0.1112	0.2955
MF	GO:0000976: transcription regulatory region sequence-specific DNA binding	0.00E+00	0.00E+00	1359	209	692	0.1076	0.3020
MF	GO:0001012: RNA polymerase II regulatory region DNA binding	0.00E+00	0.00E+00	1223	194	613	0.0998	0.3165
MF	GO:0000977: RNA polymerase II regulatory region sequence-specific DNA binding	0.00E+00	0.00E+00	1222	193	611	0.0993	0.3159
MF	GO:0000982: transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding	3.77E-292	1.57E-288	1027	136	400	0.0700	0.3400
MF	GO:0001159: core promoter proximal region DNA binding	3.20E-239	1.34E-235	867	127	379	0.0654	0.3351
MF	GO:0000987: core promoter proximal region sequence-specific DNA binding	1.31E-238	5.47E-235	863	125	376	0.0643	0.3324
MF	GO:0001227: transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific binding	4.34E-220	1.81E-216	661	59	226	0.0304	0.2611

MF	GO:0001228: transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	2.00E-202	8.35E-199	847	138	410	0.0710	0.3366
MF	GO:0000978: RNA polymerase II core promoter proximal region sequence-specific DNA binding	5.36E-201	2.24E-197	781	119	361	0.0612	0.3296
MF	GO:0001078: transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding	7.87E-151	3.29E-147	436	39	133	0.0201	0.2932
MF	GO:0001077: transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	8.21E-133	3.43E-129	603	102	274	0.0525	0.3723
MF	GO:0001205: transcriptional activator activity, RNA polymerase II distal enhancer sequence-specific binding	5.36E-123	2.24E-119	184	10	30	0.0051	0.3333
MF	GO:0003705: transcription factor activity, RNA polymerase II distal enhancer sequence-specific binding	5.14E-110	2.14E-106	312	26	100	0.0134	0.2600
MF	GO:0008134: transcription factor binding	1.42E-102	5.93E-99	678	128	531	0.0659	0.2411
CC	GO:0005667: transcription factor complex	1.74E-112	2.90E-109	517	75	315	0.0386	0.2381
CC	GO:0000785: chromatin	1.79E-77	2.98E-74	476	87	486	0.0448	0.1790
CC	GO:0000813: ESCRT I complex	1.55E-73	2.59E-70	70	1	10	0.0005	0.1000
CC	GO:1990429: peroxisomal importomer complex	1.91E-65	3.19E-62	43	1	2	0.0005	0.5000
CC	GO:0005730: nucleolus	1.11E-64	1.85E-61	674	132	930	0.0679	0.1419
CC	GO:0044427: chromosomal part	3.03E-64	5.05E-61	637	131	852	0.0674	0.1538
CC	GO:0005694: chromosome	2.41E-61	4.02E-58	694	155	975	0.0798	0.1590
CC	GO:0000790: nuclear chromatin	6.81E-56	1.13E-52	333	55	316	0.0283	0.1741
CC	GO:0044454: nuclear chromosome part	4.02E-49	6.70E-46	414	72	514	0.0371	0.1401
CC	GO:0036452: ESCRT complex	8.64E-49	1.44E-45	70	1	28	0.0005	0.0357
CC	GO:0000228: nuclear chromosome	6.90E-48	1.15E-44	434	80	550	0.0412	0.1455
CC	GO:0032993: protein-DNA complex	1.10E-39	1.84E-36	178	22	202	0.0113	0.1089
CC	GO:0016607: nuclear speck	3.14E-36	5.23E-33	322	77	373	0.0396	0.2064
CC	GO:0034451: centriolar satellite	8.44E-35	1.41E-31	58	3	26	0.0015	0.1154
CC	GO:0016604: nuclear body	1.09E-34	1.82E-31	498	123	721	0.0633	0.1706
CC	GO:0016363: nuclear matrix	2.69E-31	4.48E-28	154	20	104	0.0103	0.1923
CC	GO:0034399: nuclear periphery	2.05E-28	3.41E-25	164	24	129	0.0124	0.1860
CC	GO:0008023: transcription elongation factor complex	8.31E-27	1.38E-23	79	12	56	0.0062	0.2143
CC	GO:0043657: host cell	2.61E-26	4.35E-23	81	8	68	0.0041	0.1176
CC	GO:0016593: Cdc73/Paf1 complex	1.99E-25	3.32E-22	36	3	8	0.0015	0.3750

Table S5.
Gene set enrichment in ultraconserved elements Enrichment tested using GREAT version 4.0.4 with human genome version hg38. The association rule was Basal+extension: 5000 bp upstream, 1000 bp downstream, 1000000 bp max extension, curated regulatory domains included. BP = Biological Process; MF = Molecular Function; CC = Cellular Component.

chr	start	end	length	qval	.std.resid	fraction constrained at FDR<0.05	fraction coding	gene(s)
chr2	178500000	178600000	100000	0.02534	4.06	0.376	0.52226	PLEKHA3, TTN
chr7	27100000	27200000	100000	0.000025	6.02	0.333	0.09744	HOXA
chr2	176100000	176200000	100000	0.000754	5.15	0.263	0.07629	HOXD
chr2	144400000	144600000	200000	0.000022	6.15, 5.94	0.232	0.00811	ZEB2
chr17	48500000	48600000	100000	0.005084	4.62	0.222	0.05957	HOXB
chr12	53900000	54100000	200000	0.00014	4.31, 5.58	0.215	0.03836	HOXC
chr1	27500000	27600000	100000	0.00216	4.84	0.206	0.04808	AHDC1
chr2	60400000	60600000	200000	0.000066	4.19, 5.74	0.192	0.01562	BCL11A
chr7	114600000	114700000	100000	0.010962	4.42	0.186	0.01984	FOXP2
chr1	87300000	87400000	100000	0.00278	4.77	0.180	0.00491	LMO4
chr18	25100000	25200000	100000	0.001778	4.92	0.174	0.00105	ZNF521
chr10	113100000	113200000	100000	0.010962	4.42	0.173	0.01511	TCF7L2
chr1	10700000	10800000	100000	0.000442	5.26	0.172	0.00000	CASZ1
chr15	36900000	37100000	200000	0.001091	4.35, 5.04	0.169	0.00547	MEIS2
chr15	96300000	96400000	100000	0.010969	4.4	0.168	0.01281	NR2F2
chr3	114400000	114500000	100000	0.00216	4.86	0.167	0.00000	ZBTB20
chr10	129800000	130000000	200000	0.001091	4.79, 5.03	0.167	0.00942	EBF3
chr2	172000000	172100000	100000	0.013449	4.32	0.160	0.01728	DLX1, DLX2, METAP1D
chr3	62400000	62500000	100000	0.023289	4.09	0.160	0.01293	CADPS
chr10	61900000	62100000	200000	0.017361	4.18, 3.96	0.155	0.01793	ARID5B
chr2	58900000	59000000	100000	0.000992	5.08	0.154	0.00000	LINC01122
chr1	44200000	44300000	100000	0.047713	3.88	0.151	0.02023	DMAP1, ERI3
chr5	141800000	141900000	100000	0.019259	4.14	0.151	0.03913	PCDH1
chr3	71000000	71100000	100000	0.018405	4.15	0.151	0.00795	FOXP1
chr14	33600000	33700000	100000	0.014583	4.25	0.150	0.00297	NPAS3
chr5	158900000	159100000	200000	0.013874	4.29, 4.27	0.150	0.00274	EBF1
chr16	73000000	73100000	100000	0.00216	4.85	0.150	0.00000	ZFHX3
chr5	139600000	139700000	100000	0.016727	4.2	0.147	0.01376	CXXC5, UBE2D2
chr8	76600000	76700000	100000	0.000159	5.52	0.145	0.00000	ZFHX4
chr2	66400000	66700000	300000	0.000387	4.26, 5.33, 4.4	0.145	0.00548	MEIS1
chr10	21500000	21600000	100000	0.016727	4.19	0.141	0.03657	MLLT10, SKIDA1
chr2	143300000	143400000	100000	0.008396	4.5	0.140	0.00000	ARHGAP15
chr15	36600000	36700000	100000	0.012204	4.36	0.135	0.00469	CDIN1
chr2	143600000	143700000	100000	0.012204	4.36	0.135	0.00134	ARHGAP15
chr10	75700000	75800000	100000	0.012856	4.33	0.134	0.00046	LRMDA
chr2	63000000	63100000	100000	0.02831	4.02	0.134	0.01547	EHBP1, OTX1
chr9	125700000	125900000	200000	0.013763	4.3, 4.21	0.133	0.00153	MAPKAP1, PBX3
chr9	123700000	123800000	100000	0.025537	4.05	0.132	0.00211	DENND1A
chr3	169200000	169500000	300000	0.006499	4.56, 4.03, 4.29	0.132	0.00112	MECOM

chr2	163700000	163800000	100000	0.016727	4.19	0.132	0.00024	FIGN
chr3	181700000	181800000	100000	0.047781	3.86	0.131	0.00950	SOX2
chr1	62900000	63000000	100000	0.025537	4.05	0.128	0.00000	LINC01739
chr10	101400000	101500000	100000	0.003673	4.7	0.127	0.00273	BTRC
chr2	59900000	60000000	100000	0.047713	3.87	0.126	0.00000	
chr10	100600000	100800000	200000	0.03856	3.93, 3.85	0.125	0.00367	PAX2
chr13	72500000	72600000	100000	0.047781	3.87	0.117	0.00000	
chr9	106800000	106900000	100000	0.010962	4.42	0.116	0.00000	ZNF462
chr5	88600000	88700000	100000	0.018096	4.16	0.111	0.00000	
chr1	87500000	87600000	100000	0.021706	4.11	0.110	0.00000	
chr15	84400000	84500000	100000	0.014583	4.25	0.098	0.00000	
chr1	90800000	90900000	100000	0.031255	3.99	0.092	0.00000	
chr17	20700000	20800000	100000	0.014583	4.25	0.091	0.00000	
chrX	74100000	74200000	100000	0.047713	3.87	0.053	0.00000	
chr19	8900000	9000000	100000	0.000395	-5.31	0.008	0.38947	MUC16

Table S6.
Constrained or accelerated regions (≥100 Kb) in the human genome. Regions with significantly high (positive standardized residual) or low (negative standardized residual, only one window) constraint ($q < 0.05$ from a linear model).

			Proportion of constraint		phyloP	
Gene desert	**Element**	**N**	**Mean**	**SD**	**Mean**	**SD**
Bordering developmental transcription factor	Deserts	224	0.04	0.02	0.12	0.18
	cCREs within deserts	38065	0.12	0.2	0.53	1.16
Not bordering developmental transcription factor	Deserts	649	0.02	0.01	0.06	0.13
	cCREs within deserts	90087	0.08	0.15	0.32	0.85

Table S7.
The proportion of bases under constraint is higher in gene deserts neighboring developmental transcription factors. Gene deserts are defined as the longest 5% of intergenic regions.

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
HIF1A	bHLH	16257	90830	1.00	0.035	0.45	0.703
ZBTB17	C2H2 ZF	3579	7145	0.95	3.9E-08	0.65	8.8E-03
NFIA	SMAD	15111	48067	0.95	4.1E-08	0.53	0.041
KLF14	C2H2 ZF	691	803	0.93	3.2E-10	0.67	5.9E-04
ZNF148	C2H2 ZF	7102	18365	0.93	4.2E-06	0.72	5.4E-03
ETV2	Ets	4520	7386	0.92	6.1E-06	0.76	2.8E-03
ESR1	Nuclear receptor	76409	572776	0.92	8.7E-03	0.63	0.178
NFIC	SMAD	23072	52406	0.92	1.0E-06	0.48	0.067
EBF3	EBF1	11491	46846	0.91	5.6E-04	0.89	1.4E-03
CEBPZ	Unknown	1211	1022	0.91	5.7E-06	0.36	0.209
NFIB	SMAD	6226	17244	0.91	2.4E-06	0.49	0.061
SP1	C2H2 ZF	27114	83180	0.91	1.1E-04	0.70	0.017
SP3	C2H2 ZF	5825	5329	0.91	1.1E-04	0.66	0.026
ZNF281	C2H2 ZF	1441	1380	0.90	3.7E-06	0.50	0.060
KLF10	C2H2 ZF	5306	5138	0.90	5.4E-08	0.64	2.2E-03
GRHL1	Grainyhead	3470	10449	0.90	8.5E-08	0.39	0.090
SOX10	HMG/Sox	1906	2955	0.90	2.6E-06	0.10	0.716
EBF1	EBF1	12967	68205	0.90	1.9E-04	0.75	7.4E-03
ZNF768	C2H2 ZF	628	11704	0.89	1.9E-06	0.48	0.052
NFYB	Unknown	4116	4325	0.89	4.4E-06	0.28	0.291
ZBTB7A	C2H2 ZF	5558	16061	0.89	5.1E-05	0.84	3.3E-04
NFYA	CBF/NF-Y	2938	3182	0.88	7.6E-04	0.27	0.458
HSF2	HSF	70	281	0.88	1.5E-07	0.75	1.0E-04
ETV4	Ets	10161	17049	0.88	8.1E-04	0.47	0.169
THAP11	THAP finger	1781	1232	0.87	5.3E-06	0.56	0.019
SP4	C2H2 ZF	12632	14916	0.87	9.0E-08	0.33	0.119
CTCF	C2H2 ZF	75291	433692	0.86	6.5E-05	0.50	0.068
ETV1	Ets	15214	26155	0.85	8.2E-03	0.63	0.094
NFE2L1	bZIP	4040	12268	0.84	1.3E-03	0.41	0.209
NFATC1	Rel	492	1245	0.83	7.3E-04	0.67	0.018
ZNF652	C2H2 ZF	3935	18705	0.83	8.3E-05	0.43	0.097
ETS1	Ets	37449	118771	0.82	0.012	0.73	0.041
MAZ	C2H2 ZF	13194	22812	0.82	0.013	0.77	0.026
KLF15	C2H2 ZF	7208	8718	0.82	2.0E-03	0.46	0.157
KLF9	C2H2 ZF	15912	33500	0.82	2.1E-03	0.42	0.195
ZNF467	C2H2 ZF	4557	19604	0.82	6.3E-08	0.42	0.025
ZBTB33	C2H2 ZF	1047	1679	0.82	3.9E-03	-0.49	0.150
ERG	Ets	53960	205856	0.81	0.014	0.66	0.073
MYF5	bHLH	7529	11479	0.81	0.014	0.35	0.397
SRF	MADS box	8256	39910	0.81	1.4E-04	0.32	0.227
EGR2	C2H2 ZF	17045	68197	0.81	4.5E-03	0.54	0.108
NFKB2	Rel	5629	27339	0.81	2.5E-03	0.61	0.045
ELK4	Ets	5571	6787	0.81	0.015	0.21	0.616
LYL1	bHLH	11059	50209	0.80	0.030	0.46	0.299

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
GLIS3	C2H2 ZF	5638	24841	0.80	1.7E-03	0.68	0.015
ZNF436	C2H2 ZF	491	3830	0.80	1.1E-08	0.85	1.3E-10
EGR1	C2H2 ZF	18848	51720	0.80	5.2E-03	0.60	0.068
KLF1	C2H2 ZF	9243	15629	0.80	3.0E-03	0.66	0.029
ZNF320	C2H2 ZF	727	5951	0.80	2.2E-05	0.66	1.5E-03
GRHL2	Grainyhead	20573	100773	0.80	2.4E-05	0.25	0.294
TP73	p53	3901	20260	0.79	8.2E-05	-0.16	0.517
CTCFL	C2H2 ZF	28837	36093	0.79	1.2E-03	0.42	0.153
ERF	Ets	12748	33505	0.79	0.020	0.63	0.091
NFATC2	Rel	2069	7497	0.79	0.011	0.62	0.073
ZFP42	C2H2 ZF	970	1296	0.79	2.5E-03	-0.35	0.260
KLF3	C2H2 ZF	3088	4744	0.78	0.012	0.38	0.310
NFYC	Unknown	7890	16400	0.78	7.3E-03	-0.16	0.660
GABPA	Ets	22241	71990	0.78	0.023	0.52	0.185
POU2F2	Homeodomain; POU	5572	20522	0.78	0.024	0.36	0.383
TFE3	bHLH	8549	23408	0.77	0.024	0.36	0.377
ELF2	Ets	2183	1981	0.77	0.026	0.53	0.179
EGR3	C2H2 ZF	19579	75145	0.77	0.026	0.64	0.086
HAND1	bHLH	300	830	0.77	1.4E-03	0.39	0.164
ETV6	Ets	8588	17890	0.77	0.016	0.70	0.036
KLF5	C2H2 ZF	17727	53041	0.77	0.027	0.64	0.090
ELK1	Ets	5762	6398	0.76	0.027	0.38	0.351
TFAP2C	AP-2	36011	179035	0.76	0.017	0.72	0.027
SPDEF	Ets	6626	16827	0.76	0.017	0.17	0.666
SOX9	HMG/Sox	1456	2913	0.76	2.5E-04	0.18	0.486
ZKSCAN1	C2H2 ZF	6049	25613	0.76	4.2E-03	0.64	0.025
TP53	p53	18979	77748	0.76	6.6E-04	-0.07	0.792
RFX2	RFX	5058	6595	0.75	3.1E-03	0.49	0.090
ZFX	C2H2 ZF	32058	242180	0.75	0.013	0.66	0.038
THRA	Nuclear receptor	124	393	0.75	1.3E-03	0.37	0.169
YY2	C2H2 ZF	3412	2289	0.74	9.0E-03	0.01	0.980
ZNF143	C2H2 ZF	12119	41451	0.74	2.5E-03	0.20	0.483
ELK3	Ets	13021	39903	0.74	0.036	0.45	0.267
JUN	bZIP	50006	273381	0.74	0.037	0.09	0.830
NFE2L2	bZIP	9914	28541	0.74	0.015	0.42	0.230
ZNF554	C2H2 ZF	5985	17731	0.73	0.024	0.45	0.222
RUNX2	Runt	22329	108143	0.73	0.061	-0.05	0.923
SIX1	Homeodomain	1189	1661	0.73	4.5E-03	0.53	0.065
ZBTB7B	C2H2 ZF	2962	3734	0.73	3.0E-03	0.53	0.050
ELF4	Ets	4979	6745	0.73	0.040	0.74	0.036
NFKB1	Rel	9655	43398	0.73	0.011	0.59	0.056
ARNTL	bHLH	7045	11767	0.73	2.7E-04	0.21	0.372
STAT2	STAT	636	1462	0.72	5.1E-03	0.58	0.039
KLF4	C2H2 ZF	27644	135494	0.72	0.043	0.24	0.569
ETV7	Ets	3478	5305	0.72	0.044	0.41	0.316

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
RXRA	Nuclear receptor	19649	85156	0.72	0.108	0.58	0.232
RARG	Nuclear receptor	1559	4572	0.72	4.0E-03	0.45	0.104
TP63	p53	28370	229466	0.71	2.8E-03	-0.04	0.900
VDR	Nuclear receptor	12138	57906	0.71	2.8E-03	0.54	0.037
ATOH8	bHLH	170	254	0.71	9.3E-03	0.03	0.936
EHF	Ets	10482	39662	0.71	0.031	0.60	0.090
ZNF354A	C2H2 ZF	18	445	0.71	1.2E-10	0.54	6.4E-06
MAFK	bZIP	12247	83268	0.71	3.2E-03	0.14	0.608
ZFP28	C2H2 ZF	247	2180	0.70	3.6E-04	0.61	3.3E-03
SPI1	Ets	63841	759763	0.70	0.011	0.48	0.110
RUNX1	Runt	36078	240995	0.70	0.081	0.04	0.931
ZSCAN22	C2H2 ZF	4272	9796	0.70	0.017	0.69	0.020
RARA	Nuclear receptor	18693	70200	0.70	0.082	0.46	0.303
SREBF1	bHLH	1648	2428	0.70	3.2E-04	0.03	0.902
KLF12	C2H2 ZF	3346	3139	0.70	0.017	0.34	0.307
KLF13	C2H2 ZF	2950	3025	0.69	0.018	0.33	0.322
OSR2	C2H2 ZF	4058	14834	0.69	0.084	0.61	0.143
ZNF547	C2H2 ZF	171	1045	0.69	1.0E-03	0.36	0.132
FOXE1	Forkhead	786	1208	0.69	0.013	0.27	0.394
ZIC2	C2H2 ZF	8467	27662	0.69	6.8E-03	0.44	0.112
ETV5	Ets	15682	56169	0.68	0.133	0.55	0.260
MAX	bHLH	40106	207403	0.68	0.134	0.70	0.122
MZF1	C2H2 ZF	1657	6380	0.68	0.063	0.53	0.177
SIX2	Homeodomain	15479	55040	0.68	0.011	0.45	0.125
ELF3	Ets; AT hook	25119	150098	0.68	0.045	0.68	0.042
ELF5	Ets	338	312	0.68	0.016	0.59	0.045
TFEB	bHLH	2017	4003	0.67	0.066	0.39	0.343
USF2	bHLH	13609	55954	0.67	0.068	0.34	0.409
IKZF1	C2H2 ZF	21210	155326	0.67	0.068	0.72	0.043
RFX5	RFX	6430	12550	0.67	0.047	0.57	0.112
NR1H3	Nuclear receptor	4252	11810	0.67	0.035	0.34	0.339
ZNF146	C2H2 ZF	1176	31923	0.66	2.2E-09	0.42	5.1E-04
SPIB	Ets	7052	40304	0.66	0.019	0.21	0.514
NFAT5	Rel	3753	13227	0.65	0.022	0.58	0.049
MAFB	bZIP	466	903	0.64	0.024	0.14	0.660
MAFF	bZIP	10670	61227	0.64	0.014	0.00	0.987
HSF1	HSF	4868	39984	0.64	0.019	0.75	3.4E-03
SP2	C2H2 ZF	9624	12933	0.64	0.124	0.02	0.962
ZNF8	C2H2 ZF	445	7995	0.64	2.6E-03	0.31	0.185
FLI1	Ets	53494	261615	0.64	0.125	0.51	0.243
BCL11A	C2H2 ZF	5422	14147	0.63	0.027	0.27	0.391
PBX3	Homeodomain	3186	8747	0.63	0.049	0.46	0.182
ZNF140	C2H2 ZF	279	3591	0.63	5.5E-04	0.45	0.021
ELF1	Ets	26633	80769	0.63	0.129	0.54	0.207
FOXM1	Forkhead	8023	38048	0.63	0.052	0.19	0.601

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
FOS	bZIP	39298	160580	0.63	0.131	0.18	0.702
BCL6	C2H2 ZF	11656	69462	0.63	3.1E-03	0.36	0.122
JUNB	bZIP	19987	96718	0.63	0.071	0.43	0.251
NFE2	bZIP	9434	41631	0.63	0.039	0.32	0.344
MNT	bHLH	12192	77110	0.63	0.183	0.65	0.160
BACH1	bZIP	7290	20126	0.63	0.097	0.45	0.264
MYCN	bHLH	48496	328938	0.62	0.136	0.22	0.642
RELA	Rel	22663	128758	0.62	0.055	0.62	0.058
TFAP2A	AP-2	22331	99119	0.62	0.101	0.42	0.298
RELB	Rel	6755	30705	0.62	0.057	0.43	0.209
ZFP64	C2H2 ZF	8602	36066	0.61	0.027	0.77	2.1E-03
IRF1	IRF	6506	31647	0.61	0.012	0.60	0.014
RFX3	RFX	4335	3923	0.61	0.021	0.16	0.581
MAFG	bZIP	4275	23808	0.61	0.028	0.28	0.350
PRDM14	C2H2 ZF	5891	42493	0.60	0.113	0.78	0.023
ZNF76	C2H2 ZF	3022	3429	0.60	6.9E-07	0.65	3.7E-08
ESR2	Nuclear receptor	6724	24973	0.60	0.017	0.40	0.143
RUNX3	Runt	9871	52721	0.60	0.153	-0.05	0.913
FOSL2	bZIP	32906	120401	0.60	0.090	0.36	0.344
FOSB	bZIP	9444	21404	0.59	0.124	0.19	0.657
CDX2	Homeodomain	15509	116218	0.59	0.044	0.15	0.636
RFX1	RFX	14904	24710	0.58	0.028	0.05	0.861
FOXP1	Forkhead	18096	125879	0.58	0.099	0.08	0.828
E2F7	E2F	543	652	0.58	0.099	0.05	0.901
ZNF16	C2H2 ZF	41	570	0.57	2.2E-05	0.61	3.9E-06
SOX3	HMG/Sox	1691	2161	0.56	0.145	-0.30	0.463
BATF3	bZIP	1222	6853	0.56	0.028	-0.12	0.669
WT1	C2H2 ZF	2810	7524	0.56	0.187	0.66	0.108
NR2F2	Nuclear receptor	28590	107858	0.56	0.189	0.33	0.470
XBP1	bZIP	4496	17745	0.56	0.093	-0.73	0.016
ZIC5	C2H2 ZF	10420	28910	0.56	0.038	0.34	0.242
ZNF331	C2H2 ZF	196	2213	0.55	4.7E-03	0.65	4.3E-04
MYOG	bHLH	14888	45638	0.54	0.165	-0.42	0.297
NKX2-1	Homeodomain	19415	143548	0.54	0.267	0.58	0.230
HIC1	C2H2 ZF	3255	15543	0.54	0.211	0.36	0.433
ZSCAN31	C2H2 ZF	465	1300	0.54	0.090	0.49	0.126
LEF1	HMG/Sox	1106	2028	0.53	0.091	0.15	0.665
ZBTB48	C2H2 ZF	6346	50391	0.53	0.221	0.50	0.257
BACH2	bZIP	4257	16463	0.53	0.097	0.19	0.577
BCL6B	C2H2 ZF	269	817	0.52	0.120	0.31	0.379
MITF	bHLH	13891	77376	0.52	0.121	0.08	0.829
ESRRA	Nuclear receptor	11366	55434	0.52	0.229	0.29	0.522
ZNF260	C2H2 ZF	205	4695	0.52	1.0E-05	0.35	5.0E-03
NFIL3	bZIP	5253	34388	0.52	0.123	-0.10	0.788
ZNF324	C2H2 ZF	1109	4865	0.52	0.020	0.21	0.373

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
PKNOX1	Homeodomain	15147	66277	0.51	0.160	0.63	0.070
IRF2	IRF	5111	18367	0.51	0.063	0.44	0.118
ZNF341	C2H2 ZF	6780	24796	0.51	0.199	0.64	0.087
NRF1	Unknown	13959	30541	0.51	0.135	-0.14	0.695
ZNF563	C2H2 ZF	360	1879	0.50	0.024	0.64	2.2E-03
NR2F1	Nuclear receptor	17994	78770	0.49	0.260	0.21	0.650
MECOM	C2H2 ZF	4762	32212	0.49	0.324	-0.15	0.777
KLF6	C2H2 ZF	8667	15061	0.48	0.187	-0.07	0.849
ZNF528	C2H2 ZF	862	4763	0.48	0.069	0.45	0.095
STAT5B	STAT	7463	33201	0.48	0.159	0.60	0.067
ZFP82	C2H2 ZF	62	359	0.48	0.017	0.49	0.016
NEUROD1	bHLH	29746	77291	0.47	0.200	-0.44	0.236
CREM	bZIP	12782	30676	0.47	0.292	0.07	0.884
ZIM3	C2H2 ZF	1563	19679	0.47	0.051	0.56	0.016
BATF	bZIP	7372	36800	0.46	0.110	0.24	0.434
ZNF274	C2H2 ZF	294	1276	0.46	1.2E-04	0.30	0.015
NR1H4	Nuclear receptor	378	1503	0.46	0.117	0.31	0.307
USF1	bHLH	15270	73163	0.46	0.256	0.06	0.881
MLX	bHLH	4482	8323	0.45	0.190	-0.42	0.222
ATF7	bZIP	11604	81117	0.45	0.190	0.01	0.974
NR2F6	Nuclear receptor	9388	42822	0.45	0.122	0.15	0.633
HOXC9	Homeodomain	6972	17289	0.45	0.063	-0.14	0.581
ZNF449	C2H2 ZF	2220	8690	0.44	0.199	0.35	0.317
HNF4G	Nuclear receptor	5052	17988	0.44	0.135	0.13	0.681
ATF2	bZIP	12197	98370	0.44	0.206	0.01	0.984
HLF	bZIP	3784	27607	0.44	0.208	0.09	0.812
SNAI1	C2H2 ZF	1325	1826	0.43	0.340	-0.02	0.963
ZEB1	C2H2 ZF; Homeodomain	16008	48033	0.43	0.341	-0.12	0.797
HAND2	bHLH	17581	39994	0.42	0.294	-0.19	0.660
ZNF322	C2H2 ZF	470	3088	0.42	0.070	0.27	0.266
OLIG2	bHLH	13389	22128	0.42	0.298	0.07	0.861
TCF4	bHLH	33426	241897	0.41	0.270	-0.47	0.204
FOSL1	bZIP	19165	56083	0.41	0.315	0.13	0.751
DDIT3	bZIP	3402	13460	0.41	0.243	-0.07	0.843
RBPJ	CSL	18963	86519	0.41	0.318	0.31	0.448
FOXO3	Forkhead	5384	14449	0.40	0.281	0.03	0.938
TEAD1	TEA	25446	160936	0.40	0.321	0.17	0.693
SOX5	HMG/Sox	10525	39846	0.40	0.282	-0.43	0.249
PHOX2B	Homeodomain	22494	169693	0.40	0.221	-0.16	0.633
TBX21	T-box	8457	48464	0.40	0.285	-0.20	0.615
GFI1	C2H2 ZF	3169	27725	0.40	0.256	0.24	0.505
ZBTB18	C2H2 ZF	459	485	0.40	0.291	-0.02	0.956
FOXJ2	Forkhead	415	933	0.39	0.381	0.24	0.604
ZNF24	C2H2 ZF	3788	45806	0.39	0.111	0.34	0.172
IRF4	IRF	10101	48847	0.38	0.180	0.27	0.347

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
NR2C1	Nuclear receptor	2526	7620	0.37	0.464	0.12	0.828
TBX5	T-box	6603	31257	0.37	0.193	-0.01	0.977
ZBTB12	C2H2 ZF	543	2680	0.36	0.336	0.44	0.239
SNAI2	C2H2 ZF	25087	210209	0.36	0.425	-0.31	0.493
TFAP4	bHLH	31591	105658	0.36	0.378	-0.28	0.497
ZNF329	C2H2 ZF	236	848	0.36	0.062	0.78	9.0E-07
STAT5A	STAT	2700	10408	0.35	0.392	0.79	0.021
NEUROG2	bHLH	33888	117939	0.35	0.402	-0.38	0.353
ATOH1	bHLH	29035	69109	0.35	0.363	-0.51	0.158
TWIST1	bHLH	18438	31643	0.34	0.403	-0.38	0.358
ZNF317	C2H2 ZF	2401	21298	0.34	0.193	0.00	0.990
SMAD4	SMAD	1512	6806	0.34	0.338	0.22	0.543
PPARG	Nuclear receptor	27511	180147	0.34	0.184	0.19	0.476
ZNF18	C2H2 ZF	2013	11896	0.34	0.374	-0.04	0.917
ZNF667	C2H2 ZF	66	240	0.34	0.158	0.35	0.147
SOX2	HMG/Sox	31175	156883	0.33	0.463	-0.11	0.821
NR5A2	Nuclear receptor	1658	7595	0.33	0.383	0.20	0.612
OTX2	Homeodomain	30199	215466	0.33	0.429	-0.32	0.439
NR2C2	Nuclear receptor	6772	27155	0.33	0.528	0.22	0.670
GATA3	GATA	27498	232478	0.32	0.537	-0.15	0.770
ATF4	bZIP	6920	65938	0.32	0.371	-0.16	0.669
ZSCAN29	C2H2 ZF	1182	4739	0.32	0.315	0.09	0.792
TEAD4	TEA	24535	149998	0.31	0.455	0.12	0.784
HNF1B	Homeodomain	6449	24482	0.31	0.303	-0.36	0.224
ZNF41	C2H2 ZF	119	1400	0.31	0.243	0.45	0.081
MAF	bZIP	11639	44387	0.31	0.330	0.21	0.507
PBX2	Homeodomain	4572	16560	0.30	0.394	0.05	0.886
MEF2D	MADS box	1582	3007	0.30	0.365	-0.35	0.292
MYOD1	bHLH	64616	262043	0.30	0.466	-0.40	0.322
HNF1A	Homeodomain	4903	15851	0.30	0.317	-0.26	0.397
STAT4	STAT	4278	19007	0.30	0.371	0.55	0.077
REST	C2H2 ZF	20551	72127	0.30	0.215	0.19	0.425
PBX1	Homeodomain	5221	19771	0.30	0.405	0.06	0.869
ONECUT2	CUT; Homeodomain	22974	204370	0.29	0.442	-0.36	0.345
TEAD2	TEA	109	302	0.29	0.274	0.25	0.346
TCF12	bHLH	29375	79650	0.29	0.489	-0.29	0.481
STAT3	STAT	23691	117049	0.29	0.454	0.64	0.065
AHRR	bHLH	963	2306	0.29	0.493	-0.42	0.294
PAX5	Paired box	11208	53368	0.28	0.539	0.04	0.938
NR1H2	Nuclear receptor	2305	5336	0.28	0.355	0.05	0.877
ATF3	bZIP	1827	2077	0.28	0.434	-0.31	0.380
HNF4A	Nuclear receptor	27045	183783	0.28	0.357	0.09	0.770
ZNF136	C2H2 ZF	9	283	0.28	0.106	0.39	0.020
FEZF1	C2H2 ZF	5586	24348	0.28	0.412	0.46	0.153
ZNF770	C2H2 ZF	2661	54010	0.27	0.032	0.81	4.0E-15

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
CREB3L1	bZIP	3407	6797	0.27	0.425	-0.44	0.181
GATA6	GATA	26993	185173	0.27	0.521	-0.21	0.622
CREB5	bZIP	4563	18328	0.27	0.456	-0.01	0.985
STAT6	STAT	595	2246	0.27	0.404	0.64	0.024
ZNF250	C2H2 ZF	16	1034	0.25	0.042	0.45	2.1E-04
SOX4	HMG/Sox	8560	44302	0.25	0.582	-0.75	0.053
ZNF257	C2H2 ZF	240	957	0.25	0.456	0.18	0.597
ZNF680	C2H2 ZF	708	7562	0.25	0.351	0.53	0.034
YY1	C2H2 ZF	36744	179897	0.25	0.462	-0.31	0.360
HMBOX1	Homeodomain	1750	18848	0.25	0.593	-0.03	0.955
FOXK2	Forkhead	4176	22064	0.25	0.637	-0.26	0.613
ZNF549	C2H2 ZF	750	4241	0.24	0.372	0.25	0.348
GATA1	GATA	19082	168673	0.23	0.662	-0.09	0.869
TBX3	T-box	919	3286	0.23	0.587	-0.38	0.349
PDX1	Homeodomain	14538	41390	0.22	0.542	-0.09	0.806
MXI1	bHLH	6111	11278	0.22	0.723	-0.44	0.453
MYC	bHLH	33217	193926	0.22	0.574	-0.35	0.358
ZNF134	C2H2 ZF	590	3823	0.21	0.442	0.07	0.806
FOXA3	Forkhead	9126	48328	0.21	0.622	-0.03	0.945
E2F4	E2F	7435	12549	0.20	0.631	-0.02	0.960
ZNF816	C2H2 ZF	186	1667	0.19	0.474	-0.04	0.873
GATA2	GATA	32383	251835	0.18	0.727	-0.25	0.638
CLOCK	bHLH	3100	7089	0.18	0.694	-0.36	0.433
BHLHE40	bHLH	12105	46740	0.18	0.697	-0.21	0.658
NR3C1	Nuclear receptor	29283	190610	0.18	0.520	-0.22	0.439
JUND	bZIP	30938	123695	0.18	0.701	-0.07	0.886
ATF1	bZIP	8689	21316	0.18	0.651	-0.54	0.131
ZNF350	C2H2 ZF	435	2063	0.17	0.521	0.09	0.734
SOX17	HMG/Sox	6556	34539	0.17	0.686	-0.60	0.119
NR4A1	Nuclear receptor	5277	20650	0.17	0.689	0.49	0.217
TAL1	bHLH	18339	118897	0.16	0.724	-0.23	0.620
HOXA9	Homeodomain	690	2354	0.16	0.652	-0.12	0.749
PGR	Nuclear receptor	28299	219287	0.15	0.586	-0.18	0.513
STAT1	STAT	6853	23854	0.15	0.695	0.63	0.069
GATA4	GATA	29197	272128	0.15	0.745	-0.20	0.670
ZBTB6	C2H2 ZF	3496	14395	0.14	0.668	0.25	0.438
CEBPD	bZIP	5597	26061	0.12	0.733	-0.35	0.315
ZNF121	C2H2 ZF	1015	37591	0.12	0.334	0.55	3.8E-06
SCRT2	C2H2 ZF	5367	28488	0.12	0.738	-0.51	0.136
SCRT1	C2H2 ZF	9604	77265	0.11	0.786	-0.38	0.319
MEIS1	Homeodomain	3974	13343	0.10	0.792	-0.05	0.895
FOXA2	Forkhead	77920	730071	0.10	0.811	0.02	0.970
TBP	TBP	5429	36543	0.10	0.819	0.17	0.681
HOXB13	Homeodomain	34736	275861	0.10	0.804	0.06	0.888
FOXA1	Forkhead	64001	718656	0.09	0.797	0.10	0.785

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
ZBTB26	C2H2 ZF	4058	5989	0.08	0.861	-0.17	0.713
E2F6	E2F	13829	37067	0.07	0.871	0.10	0.822
CREB1	bZIP	38466	134004	0.07	0.915	-0.49	0.408
TCF7L2	HMG/Sox	12865	54211	0.06	0.876	-0.05	0.904
EOMES	T-box	4537	21633	0.02	0.951	-0.68	0.046
PITX3	Homeodomain	7130	26201	0.02	0.960	-0.20	0.666
AHR	bHLH	9004	30993	0.02	0.964	-0.40	0.372
ZNF490	C2H2 ZF	922	12335	0.02	0.881	0.34	0.012
MEF2C	MADS box	1636	4062	0.02	0.957	-0.40	0.250
PAX6	Homeodomain; Paired box	1495	2934	0.02	0.944	-0.33	0.201
AR	Nuclear receptor	45728	453586	0.01	0.985	-0.47	0.352
NKX3-1	Homeodomain	2540	19613	0.00	0.995	-0.06	0.878
CEBPG	bZIP	11235	110717	0.00	0.994	-0.36	0.309
DUX4	Homeodomain	20636	258936	0.00	0.993	-0.05	0.884
ZNF708	C2H2 ZF	193	634	0.00	0.991	0.24	0.475
NKX2-5	Homeodomain	3000	12640	-0.01	0.974	-0.09	0.691
TCF7L1	HMG/Sox	6624	24551	-0.01	0.983	0.07	0.865
FOXP2	Forkhead	2310	4143	-0.02	0.972	-0.51	0.239
ZNF418	C2H2 ZF	57	318	-0.02	0.915	0.01	0.947
MEF2B	MADS box	10844	34212	-0.02	0.951	-0.33	0.350
FOXO1	Forkhead	8254	33056	-0.03	0.955	0.32	0.487
POU5F1	Homeodomain; POU	14372	74816	-0.06	0.839	-0.55	0.028
NANOG	Homeodomain	19172	78100	-0.07	0.807	-0.42	0.104
KLF16	C2H2 ZF	1670	1842	-0.08	0.887	-0.31	0.550
MYNN	C2H2 ZF	2437	10041	-0.08	0.768	-0.24	0.377
CEBPB	bZIP	58732	788738	-0.09	0.818	-0.23	0.548
ARNT	bHLH	8573	27846	-0.11	0.855	-0.35	0.562
MEF2A	MADS box	5083	14186	-0.13	0.729	-0.40	0.249
TFDP1	E2F	5692	8865	-0.14	0.771	-0.19	0.687
MYB	Myb/SANT	16240	111913	-0.15	0.751	-0.04	0.935
ASCL1	bHLH	25692	69797	-0.18	0.693	-0.52	0.232
EPAS1	bHLH	5213	11401	-0.19	0.763	-0.67	0.219
PRDM1	C2H2 ZF	3342	30879	-0.22	0.549	-0.28	0.426
TCF3	bHLH	20554	73536	-0.22	0.674	-0.32	0.538
CEBPA	bZIP	27674	270376	-0.25	0.492	-0.29	0.413
ZNF263	C2H2 ZF	9023	64556	-0.25	0.582	-0.10	0.838
ZNF382	C2H2 ZF	4655	49686	-0.26	0.322	-0.35	0.168
FOXK1	Forkhead	1514	2030	-0.26	0.580	-0.23	0.618
MEIS2	Homeodomain	8024	36598	-0.27	0.666	-0.62	0.268
HOXA13	Homeodomain	923	3255	-0.33	0.389	-0.17	0.671
GFI1B	C2H2 ZF	2169	7668	-0.33	0.424	-0.11	0.801
ZNF85	C2H2 ZF	194	1282	-0.33	6.9E-03	0.05	0.712
ZNF586	C2H2 ZF	67	3886	-0.35	4.0E-03	0.48	6.2E-05
FOXH1	Forkhead	3686	29192	-0.37	0.372	-0.32	0.439
SOX13	HMG/Sox	8759	53503	-0.42	0.305	0.22	0.604

TF	TF Family	# Const-rained	# Unconst-rained	Correlation (Constrained)	p Correlation (Constrained)	Correlation (Unconstrained)	p Correlation (Unconstrained)
E2F1	E2F	16895	48151	-0.45	0.375	-0.44	0.378
PAX7	Homeodomain; Paired box	2172	6725	-0.50	0.167	-0.39	0.300
EN1	Homeodomain	5534	17148	-0.52	0.290	-0.67	0.144
TCF7	HMG/Sox	3449	11608	-0.63	0.127	-0.95	1.1E-03
PAX3	Homeodomain; Paired box	12830	70494	-0.67	0.048	-0.28	0.460
LHX2	Homeodomain	1089	1870	-0.72	0.067	-0.70	0.080
LHX9	Homeodomain	2986	6773	-0.74	0.093	-0.66	0.154
ISL1	Homeodomain	10625	26540	NA	NA	-0.03	0.964

Table S8.
Transcription factor binding sites for 367 transcription factors were categorized as constrained or unconstrained. We used a two-component Gaussian mixture model to classify sites as constrained or unconstrained using convolutional neural networks and publicly available ChIP-seq data for over 600 ENCODE3 chromatin immunoprecipitation experiments spanning hundreds of cell and tissue types.

Annotation type	Name	Date accessed	Details	Link
Coding regions	GENCODE v37	21.05.2021	UTRs and exons for all protein-coding genes	https://www.gencodegenes.org/human/release_37.html
			Promoters (TSS +/- 1kb)	Manually calculated from each TSS
Regulatory features	ENCODE3 cCREs	03.06.2021	candidate *cis*-regulatory elements, including promoter-like signatures (PLS), proximal enhancer-like signatures (pELS) and distal enhancer-like signatures (dELS)	https://screen.encodeproject.org/
	ENCODE3 DHS	03.06.2021	DNase hypersensitive sites in 243 cell lines	https://doi.org/10.1038/s41586-020-2528-x
	ENCODE3 ChIA-PET anchors	03.06.2021	Chromatin loop anchors identified in 24 cell types	https://doi.org/10.1038/s41586-020-2151-x
	UCSC Promoters	20.08.2021	Experimentally validated promoters generated by the Eukaryotic Promoter Database.	https://genome.ucsc.edu/cgi-bin/hgTrackUi?db=mm10&c=chrX&g=epdNew
	Human_promoter_Villar	20.08.2021	Promoters active in human liver	https://doi.org/10.1016/j.cell.2015.01.006
	Enhancer_Andersson	20.08.2021	Atlas of active enhancers across human cell types and tissues	https://doi.org/10.1038/nature12787
	Enhancer_Hoffman	20.08.2021	Enhancers identified from ENCODE data	https://doi.org/10.1093/nar/gks1284
	Human_Enhancer_Villar	20.08.2021	Enhancers active in human liver	https://doi.org/10.1016/j.cell.2015.01.006
	SuperEnhancer_Hnisz	20.08.2021	Catolog of super-enhancers in 86 human cell and tissue types	https://doi.org/10.1016/j.cell.2013.09.053

Table S9.
Annotation sources used for identifying genomic regions outside of annotations when defining UNICORNs.

Element	TE Family	Autonomy	Library Consensus Length (bp)	Species Involved	Full-Length Hits
HysCri-1.134	hAT	Non-autonomous	192	*Heterocephalus glaber*	303
				Hystrix cristata	257
nhAT2_ML	hAT	Non-autonomous	204	*Cheirogaleus medius*	91
				Nycticebus coucang	439
NycCou-1.114	hAT	Non-autonomous	215	*Nycticebus coucang*	123
PMER1	hAT	Non-autonomous	90	*Otolemur garnettii*	1857
npiggy1_Mm	PiggyBac	Non-autonomous	240	*Microcebus murinus*	890
				Mirza coquereli	518
ScaAqu-1.134	PiggyBac	Non-autonomous	520	*Scalopus aquaticus*	453
ScaAqu-1.148	PiggyBac	Non-autonomous	240	*Scalopus aquaticus*	143
ScaAqu-1.172	PiggyBac	Non-autonomous	517	*Scalopus aquaticus*	225
HipAmp-1.103	Tc-Mariner	Non-autonomous	966	*Hippopotamus amphibius*	368
OdoVir-5.847	Tc-Mariner	Autonomous	1283	*Balaenoptera acutorostrata*	985
				Balaenoptera bonaerensis	398
				Delphinapterus leucas	922
				Eschrichtius robustus	524
				Eubalaena japonica	511
				Hippopotamus amphibius	332
				Inia geoffrensis	483
				Kogia breviceps	477
				Lipotes vexillifer	728
				Mesoplodon bidens	490
				Monodon monoceros	516
				Neophocaena asiaeorientalis	820
				Orcinus orca	875
				Phocoena phocoena	521
				Platanista gangetica	484
				Pontoporia blainvillei	277
				Tursiops truncatus	688
				Ziphius cavirostris	464
ProCoq-1.279	Tc-Mariner	Non-autonomous	131	*Cheirogaleus medius*	170
				Daubentonia madagascariensis	413
				Eulemur flavifrons	168
				Eulemur fulvus	192
				Indri indri	220
				Lemur catta	168
				Microcebus murinus	117
				Mirza coquereli	127
				Propithecus coquereli	164

Table S10.
Putative horizontally transferred transposable elements involving non-chiropterans.
Outside of bats, we found 11 instances of potential TE introduction through horizontal transfer (transfer from one species to another in the absence of reproduction). In contrast, in bats, we identified 222 putative horizontal transfers, including Tc-Mariner, hAT, and piggyBac elements.

Human Phenotype	Rank	Hyper raw P-value	Hyper FDR Q-value	Hyper fold enrich-ment	Hyper fore-ground regions	Hyper total regions	Hyper region set coverage	Hyper fore-ground genes hit	Total genes annotated
Moderate visual impairment	1	1.51E-12	1.01E-08	5.606	26	171	0.00193	2	2
Ambiguous genitalia, male	2	1.67E-11	5.56E-08	3.478	40	424	0.00297	5	7
Vitelliform-like macular lesions	3	2.52E-11	5.60E-08	4.786	27	208	0.00200	3	3
Hyperpigmentation of the fundus	5	7.48E-11	9.98E-08	4.567	27	218	0.00200	2	3
Failure to thrive secondary to recurrent infections	6	8.70E-11	9.67E-08	5.086	24	174	0.00178	2	3
Calcinosis	7	1.54E-10	1.47E-07	4.425	27	225	0.00200	2	4
Methylmalonic aciduria	8	1.57E-10	1.31E-07	2.567	58	833	0.00430	6	14
Calf muscle hypertrophy	9	1.69E-10	1.25E-07	2.789	50	661	0.00371	6	13
Recurrent opportunistic infections	10	1.78E-10	1.19E-07	4.916	24	180	0.00178	2	4
Forehead hyperpigmentation	11	2.71E-10	1.65E-07	7.043	17	89	0.00126	1	1
Posterior wedging of vertebral bodies	11	2.71E-10	1.65E-07	7.043	17	89	0.00126	1	1
High iliac wings	11	2.71E-10	1.65E-07	7.043	17	89	0.00126	1	1
Spontaneous abortion	14	3.26E-10	1.55E-07	3.541	34	354	0.00252	3	4
Muscle hypertrophy of the lower extremities	15	4.44E-10	1.97E-07	2.711	50	680	0.00371	6	14
Impaired social interactions	16	1.04E-09	4.34E-07	2.160	75	1280	0.00556	10	28
Reticular pigmentary degeneration	17	1.10E-09	4.31E-07	5.696	19	123	0.00141	1	1
Diaphyseal cortical sclerosis	18	2.72E-09	1.01E-06	7.091	15	78	0.00111	1	1
Quadriceps muscle atrophy	18	2.72E-09	1.01E-06	7.091	15	78	0.00111	1	1
Exaggerated startle response	20	2.81E-09	9.37E-07	2.796	43	567	0.00319	7	11
Short stepped shuffling gait	21	4.17E-09	1.32E-06	4.214	24	210	0.00178	2	3
Methylmalonic acidemia	22	4.53E-09	1.37E-06	2.790	42	555	0.00312	5	10
Abnormal social behavior	23	1.13E-08	3.29E-06	2.041	75	1355	0.00556	10	29
Short diaphyses	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	1
Severe hydrops fetalis	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	1
Anterior rib punctate calcifications	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	1
Sternal punctate calcifications	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	1
Punctate vertebral calcifications	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	2
Abnormality of the vertebral spinous processes	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	1
Short 3rd metacarpal	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	2
Patchy variation in bone mineral density	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	1
Calcinosis cutis	24	1.76E-08	4.90E-06	5.784	16	102	0.00119	1	1
Hamartoma of tongue	33	2.24E-08	4.54E-06	4.766	19	147	0.00141	3	4
Long clavicles	34	2.87E-08	5.64E-06	2.518	45	659	0.00334	6	8
Radial club hand	35	4.11E-08	7.83E-06	8.630	11	47	0.00082	1	2
Poor eye contact	36	5.90E-08	1.09E-05	2.128	61	1057	0.00452	8	22
Stippled chondral calcification	37	6.39E-08	1.15E-05	4.975	17	126	0.00126	2	2
Intermittent hyperpnea at rest	38	6.85E-08	1.20E-05	5.267	16	112	0.00119	1	1
Renal aminoaciduria	38	6.85E-08	1.20E-05	5.267	16	112	0.00119	1	1
Absent toenail	40	8.72E-08	1.45E-05	2.942	32	401	0.00237	3	5
Hypoplastic areola	41	1.01E-07	1.64E-05	3.465	25	266	0.00185	1	1
Abnormal rib ossification	43	1.34E-07	2.07E-05	2.946	31	388	0.00230	3	5
Biliary cirrhosis	44	1.62E-07	2.45E-05	3.872	21	200	0.00156	2	5

Human Phenotype	Rank	Hyper raw P-value	Hyper FDR Q-value	Hyper fold enrich-ment	Hyper fore-ground regions	Hyper total regions	Hyper region set coverage	Hyper fore-ground genes hit	Total genes annotated
Brushfield spots	45	1.77E-07	2.63E-05	4.643	17	135	0.00126	2	4
Intrahepatic biliary dysgenesis	45	1.77E-07	2.63E-05	4.643	17	135	0.00126	2	3
Unilateral narrow palpebral fissure	47	1.81E-07	2.58E-05	2.700	35	478	0.00260	2	2
Decreased methylcobalamin	48	2.09E-07	2.91E-05	2.778	33	438	0.00245	3	6
Decreased methionine synthase activity	48	2.09E-07	2.91E-05	2.778	33	438	0.00245	3	6
Calcific stippling	50	2.34E-07	3.12E-05	3.944	20	187	0.00148	3	4
Enuresis nocturna	51	2.68E-07	3.51E-05	3.292	25	280	0.00185	1	2
Scleroderma	52	2.90E-07	3.72E-05	3.605	22	225	0.00163	3	4

Table S11.
Top 50 gene sets enriched for genes near primate-specific, transposable element-derived CTCF TFBS. Enrichment analysis of cis-regulatory regions with GREAT.

Order	*Species*	olfactory receptor genes			# Olfactory Turbinals	assembly	BUSCO % complete	contig N50	scaffold N50
		functional	non-functional	total					
Afrosoricida	*Echinops telfairi*	822	556	1378	-	GCF_000313985.1	90%	20428	45764842
Afrosoricida	*Chrysochloris asiatica*	879	756	1635	-	GCA_004027935.1	96%	6444	6492
Afrosoricida	*Microgale talazaci*	886	770	1656	-	GCA_004026705.1	64%	56946	65143
Hyracoidea	*Procavia capensis*	658	724	1382	14	GCA_004026925.1	94%	37294	39432
Hyracoidea	*Heterohyrax brucei*	661	806	1467	-	GCA_004026845.1	77%	62841	67904
Macroscelidae	*Elephantulus edwardii*	670	484	1154	-	GCA_004027355.1	95%	16488	17713
Proboscidae	*Loxodonta africana*	1765	2434	4199	52	GCF_000001905.1	93%	46401353	69023
Sirenia	*Trichechus manatus latirostris*	470	755	1225	10	GCF_000243295.1	94%	37750	14442683
Tubulidentata	*Orycteropus afer afer*	1033	1228	2261	-	GCA_004365145.1	94%	20354	22965
Cetartiodactyla	*Camelus ferus*	468	499	967	36	GCF_000311805.1	96%	2005940	90263
Cetartiodactyla	*Camelus dromedarius*	499	494	993	36	GCF_000767585.1	96%	4188677	69131
Cetartiodactyla	*Camelus bactrianus*	491	528	1019	-	GCF_000767855.1	94%	139019	8812066
Cetartiodactyla	*Saiga tatarica*	428	610	1038	-	GCA_004024985.1	29%	6406	6453
Cetartiodactyla	*Vicugna pacos huacaya*	527	555	1082	-	GCA_000767525.1	95%	213649	5303709
Cetartiodactyla	*Giraffa tippelskirchi*	570	668	1238	-	GCA_001651235.1	86%	47894	212164
Cetartiodactyla	*Beatragus hunteri*	607	764	1371	-	GCA_004027495.1	76%	57444	69303
Cetartiodactyla	*Okapia johnstoni*	586	787	1373	-	GCA_001660835.1	90%	39571	111538
Cetartiodactyla	*Antilocapra americana*	613	785	1398	-	GCA_004027515.1	96%	59271	70316
Cetartiodactyla	*Ovis canadensis canadensis*	639	815	1454	-	GCA_004026945.1	77%	55973	69397
Cetartiodactyla	*Hemitragus hylocrius*	663	792	1455	-	GCA_004026825.1	80%	66552	85340
Cetartiodactyla	*Ovis aries*	714	742	1456	36	GCF_000298735.2	96%	100009711	150472
Cetartiodactyla	*Pantholops hodgsonii*	609	859	1468	-	GCF_000400835.1	91%	18674	2772860
Cetartiodactyla	*Bos indicus*	735	741	1476	-	GCA_000247795.2	94%	28375	106310653
Cetartiodactyla	*Catagonus wagneri*	696	826	1522	-	GCA_004024745.1	96%	66027	91723
Cetartiodactyla	*Capra hircus*	780	793	1573	34	GCF_001704415.1	96%	26244591	87277232
Cetartiodactyla	*Ammotragus lervia*	740	868	1608	-	GCA_002201775.1	95%	52017	1301762
Cetartiodactyla	*Moschus moschiferus*	676	946	1622	-	GCA_004024705.1	94%	34798	47482
Cetartiodactyla	*Capra aegagrus*	764	895	1659	-	GCA_000978405.1	95%	91317560	19347
Cetartiodactyla	*Odocoileus virginianus texanus*	763	924	1687	-	GCA_002102435.1	93%	850721	122019
Cetartiodactyla	*Tragulus javanicus*	885	875	1760	24	GCA_004024965.1	96%	78495	86670
Cetartiodactyla	*Bubalus bubalis*	751	1030	1781	-	GCF_000471725.1	96%	1412388	21938
Cetartiodactyla	*Sus scrofa*	997	900	1897	54	GCF_000003025.5	96%	576008	69480
Cetartiodactyla	*Rangifer tarandus tarandus*	765	1134	1899	42	GCA_004026565.1	82%	77671	89062
Cetartiodactyla	*Bos taurus*	914	1123	2037	46	GCF_000003205.7	96%	6806220	276285
Cetartiodactyla	*Bos mutus*	783	1270	2053	-	GCF_000298355.1	94%	1407960	22822
Cetartiodactyla	*Elaphurus davidianus*	953	1142	2095	-	GCA_002443075.1	95%	59950	2844142
Cetartiodactyla	*Bison bison bison*	836	1333	2169	-	GCF_000754665.1	90%	19971	7192658
Cetartiodactyla	*Hippopotamus amphibius*	994	1196	2190	-	GCA_004027065.1	96%	74418	86527
Carnivora	*Leptonychotes weddellii*	269	418	687	-	GCF_000349705.1	80%	23664	904031
Carnivora	*Mirounga angustirostris*	255	438	693	-	GCA_004023865.1	94%	52833	64640
Carnivora	*Neomonachus schauinslandi*	269	432	701	-	GCF_002201575.1	95%	112698	29518589
Carnivora	*Odobenus rosmarus divergens*	389	457	846	-	GCF_000321225.1	96%	89951	2616778
Carnivora	*Zalophus californianus*	323	537	860	-	GCA_004024565.1	96%	90567	132377
Carnivora	*Lycaon pictus*	537	425	962	-	GCA_001887905.1	85%	5555	63240551
Carnivora	*Puma concolor*	583	445	1028	-	GCF_003327715.1	84%	27187	100532876
Carnivora	*Felis nigripes*	517	514	1031	-	GCA_004023925.1	41%	16007	18631
Carnivora	*Panthera tigris altaica*	636	443	1079	-	GCF_000464555.1	86%	30032	8860407
Carnivora	*Vulpes lagopus*	571	513	1084	-	GCA_004023825.1	83%	73976	102053
Carnivora	*Acinonyx jubatus*	580	514	1094	18	GCA_004027535.1	96%	42476	65411
Carnivora	*Canis lupus familiaris*	581	517	1098	27	GCF_000002285.3	96%	45876610	267478
Carnivora	*Pteronura brasiliensis*	516	639	1155	-	GCA_004024605.1	84%	90879	119023
Carnivora	*Ailurus fulgens*	686	476	1162	-	GCA_002007465.1	95%	99577	2983736
Carnivora	*Enhydra lutris kenyoni*	518	647	1165	-	GCF_002288905.1	91%	244529	38751465
Carnivora	*Hyaena hyaena*	562	604	1166	-	GCA_004023945.1	96%	51677	66490
Carnivora	*Panthera onca*	599	576	1175	-	GCA_004023805.1	80%	62836	116574
Carnivora	*Felis catus*	683	503	1186	18	GCF_000181335.2	96%	18072971	45189
Carnivora	*Panthera pardus*	691	527	1218	-	GCA_001857705.1	96%	21701857	20993
Carnivora	*Cryptoprocta ferox*	638	598	1236	-	GCA_004023885.1	95%	128639	173473
Carnivora	*Mustela putorius furo*	707	625	1332	22	GCF_000239315.1	89%	9567452	66755
Carnivora	*Mellivora capensis*	663	732	1395	-	GCA_004024625.1	81%	50213	59143
Carnivora	*Spilogale gracilis*	700	751	1451	-	GCA_004023965.1	78%	63054	85970
Carnivora	*Suricata suricatta*	707	834	1541	-	GCA_004023905.1	87%	148487	186735
Carnivora	*Helogale parvula*	875	676	1551	-	GCA_004023845.1	84%	113567	179119
Carnivora	*Paradoxurus hermaphroditus*	831	735	1566	22	GCA_004024585.1	74%	62870	71823
Carnivora	*Ailuropoda melanoleuca*	655	958	1613	-	GCF_002007445.1	95%	127363	129245720
Carnivora	*Ursus maritimus*	869	757	1626	-	GCF_000687225.1	88%	46506	15940661
Carnivora	*Mungos mungo*	860	784	1644	20	GCA_004023785.1	96%	180702	236501
Cetartiodactyla	*Orcinus orca*	24	130	154	0	GCF_000331955.2	95%	12735091	70300
Cetartiodactyla	*Phocoena phocoena*	25	131	156	0	GCA_004363495.1	83%	89111	115969
Cetartiodactyla	*Monodon monoceros*	28	135	163	0	GCA_004026685.1	95%	67024	86766
Cetartiodactyla	*Neophocaena asiaeorientalis asiaeorientalis*	27	138	165	0	GCA_003031525.1	94%	6341296	86003
Cetartiodactyla	*Pontoporia blainvillei*	18	149	167	-	GCA_004363935.1	30%	2541	2541

Order	*Species*	olfactory receptor genes			# Olfactory Turbinals	assembly	BUSCO % complete	contig N50	scaffold N50
		functional	non-functional	total					
Cetartiodactyla	*Inia geoffrensis*	28	142	170	0	GCA_004363515.1	68%	24570	26707
Cetartiodactyla	*Tursiops truncatus*	19	151	170	0	GCA_001922835.1	76%	44299	26555543
Cetartiodactyla	*Kogia breviceps*	18	166	184	0	GCA_004363705.1	65%	26201	28812
Cetartiodactyla	*Delphinapterus leucas*	32	155	187	0	GCF_002288925.1	95%	19885328	159142
Cetartiodactyla	*Platanista minor*	29	185	214	-	GCA_004363435.1	50%	20879	23933
Cetartiodactyla	*Lipotes vexillifer*	20	203	223	0	GCF_000442215.1	95%	31902	2419148
Cetartiodactyla	*Mesoplodon bidens*	32	192	224	0	GCA_004027085.1	69%	28959	33532
Cetartiodactyla	*Ziphius cavirostris*	42	255	297	0	GCA_004364475.1	41%	3606	3608
Cetartiodactyla	*Balaenoptera acutorostrata scammoni*	93	221	314	11	GCF_000493695.1	94%	22690	12843668
Cetartiodactyla	*Balaenoptera bonaerensis*	71	245	316	-	GCA_000978805.1	53%	8410	20082
Cetartiodactyla	*Eschrichtius robustus*	91	243	334	-	GCA_004363415.1	80%	68559	94414
Cetartiodactyla	*Eubalaena japonica*	109	283	392	-	GCA_004363455.1	72%	34866	39813
Chiroptera	*Craseonycteris thonglongyai*	139	193	332	-	GCA_004027555.1	67%	23218	25762
Chiroptera	*Miniopterus natalensis*	287	221	508	-	GCF_001595765.1	94%	29777	4315193
Chiroptera	*Rhinolophus sinicus*	236	279	515	-	GCA_001888835.1	96%	37803	3754400
Chiroptera	*Pteronotus parnellii*	239	277	516	-	GCA_000465405.1	47%	9502	22675
Chiroptera	*Rhinolophus ferrumequinum*	245	274	519	-	GCA_007922735.1	96%	127195	156956
Chiroptera	*Hipposideros armiger*	288	250	538	-	GCA_001890085.1	90%	39863	2328177
Chiroptera	*Miniopterus schreibersii*	265	304	569	-	GCA_004026525.1	85%	82539	108707
Chiroptera	*Mormoops blainvillei*	252	337	589	-	GCA_004026545.1	86%	142682	156292
Chiroptera	*Noctilio leporinus*	342	260	602	-	GCA_004026585.1	87%	135651	191494
Chiroptera	*Eptesicus fuscus*	361	245	606	-	GCF_000308155.1	94%	21392	13454942
Chiroptera	*Megaderma lyra*	264	349	613	28	GCA_004026885.1	88%	71674	96489
Chiroptera	*Macrotus californicus*	323	301	624	-	GCA_007922815.1	70%	16386	16925
Chiroptera	*Hipposideros galeritus*	256	402	658	-	GCA_004027415.1	72%	33704	37985
Chiroptera	*Eidolon helvum*	278	387	665	-	GCA_000465285.1	72%	27684	12668
Chiroptera	*Myotis davidii*	319	359	678	-	GCF_000327345.1	87%	15182	3454484
Chiroptera	*Myotis brandtii*	411	274	685	-	GCF_000412655.1	89%	23289	3225832
Chiroptera	*Nycticeius humeralis*	374	334	708	-	GCA_007922795.1	54%	14267	15095
Chiroptera	*Desmodus rotundus*	363	354	717	-	GCA_002940915.2	93%	26869735	80250
Chiroptera	*Myotis lucifugus*	463	269	732	10	GCF_000147115.1	87%	64330	4293315
Chiroptera	*Pipistrellus pipistrellus*	391	351	742	-	GCA_004026625.1	64%	26707	33992
Chiroptera	*Anoura caudifer*	383	367	750	-	GCA_004027475.1	88%	143417	185021
Chiroptera	*Myotis myotis*	427	323	750	-	GCA_004026985.1	96%	19322	21824
Chiroptera	*Tadarida brasiliensis*	338	418	756	-	GCA_004025005.1	67%	22617	24311
Chiroptera	*Lasiurus borealis*	412	355	767	-	GCA_004026805.1	68%	35165	38543
Chiroptera	*Antrozous pallidus*	416	362	778	-	GCA_007922775.1	50%	13442	13958
Chiroptera	*Murina aurata feae*	442	423	865	-	GCA_004026665.1	64%	23400	26051
Chiroptera	*Pteropus alecto*	373	499	872	-	GCF_000325575.1	95%	31841	15954802
Chiroptera	*Artibeus jamaicensis*	459	449	908	12	GCA_004027435.1	70%	32365	35347
Chiroptera	*Macroglossus sobrinus*	391	526	917	-	GCA_004027375.1	91%	338389	453401
Chiroptera	*Micronycteris hirsuta*	479	460	939	-	GCA_004026765.1	77%	62961	68868
Chiroptera	*Tonatia saurophila*	473	527	1000	-	GCA_004024845.1	86%	141649	165561
Chiroptera	*Rousettus aegyptiacus*	488	531	1019	-	GCA_004024865.1	96%	100260	125332
Chiroptera	*Pteropus vampyrus*	476	589	1065	-	GCF_000151845.1	90%	21866	5954017
Chiroptera	*Carollia perspicillata*	490	847	1337	-	GCA_004027735.1	48%	10341	10739
Dermoptera	*Galeopterus variegatus*	831	982	1813	-	GCA_004027255.1	95%	33956	37578
Lagomorpha	*Ochotona princeps*	477	373	850	12	GCF_000292845.1	92%	42347	26863993
Lagomorpha	*Lepus americanus*	527	440	967	20	GCA_004026855.1	53%	15057	16725
Lagomorpha	*Oryctolagus cuniculus cuniculus*	639	405	1044	16	GCF_000003625.3	89%	35972871	64648
Eulipotyphla	*Condylura cristata*	570	390	960	-	GCF_000260355.1	89%	46163	55520359
Eulipotyphla	*Erinaceus europaeus*	814	453	1267	14	GCF_000296755.1	94%	21359	3264618
Eulipotyphla	*Crocidura indochinensis*	774	685	1459	-	GCA_004027635.1	21%	4927	4929
Eulipotyphla	*Uropsilus gracilis*	809	691	1500	-	GCA_004024945.1	63%	43618	55035
Eulipotyphla	*Sorex araneus*	1048	613	1661	14	GCF_000181275.1	92%	22623	22794405
Eulipotyphla	*Scalopus aquaticus*	978	867	1845	-	GCA_004024925.1	83%	72421	94879
Eulipotyphla	*Solenodon paradoxus*	821	1209	2030	-	GCA_004363575.1	91%	236847	407682
Perissodactyla	*Ceratotherium simum cottoni*	746	986	1732	-	GCA_004027795.1	55%	21199	23005
Perissodactyla	*Ceratotherium simum simum*	908	1013	1921	-	GCF_000283155.1	96%	92960	26277727
Perissodactyla	*Equus przewalskii*	761	1307	2068	-	GCF_000696695.1	87%	57610	513800
Perissodactyla	*Tapirus terrestris*	880	1271	2151	-	GCA_004025025.1	84%	164193	186384
Perissodactyla	*Equus asinus asinus*	898	1281	2179	-	GCF_001305755.1	95%	66737	3776412
Perissodactyla	*Tapirus indicus*	936	1338	2274	-	GCA_004024905.1	87%	225792	308930
Perissodactyla	*Diceros bicornis*	992	1337	2329	-	GCA_004027315.1	85%	87375	115504
Perissodactyla	*Equus caballus*	931	1594	2525	74	GCF_000002305.2	96%	112381	46749900
Perissodactyla	*Dicerorhinus sumatrensis sumatrensis*	1180	1858	3038	-	GCA_002844835.1	97%	89455	614498
Pholidota	*Manis pentadactyla*	571	1129	1700	-	GCA_000738955.1	71%	28718	117920
Pholidota	*Manis javanica*	478	1416	1894	-	GCF_001685135.1	85%	38922	204728
Pholidota	*Manis tricuspis*	703	1386	2089	-	GCA_004765945.1	67%	23755	25241
Primates	*Colobus angolensis palliatus*	177	275	452	-	GCF_000951035.1	93%	38363	7840981
Primates	*Piliocolobus tephrosceles*	202	254	456	-	GCA_002776525.1	94%	98446	131283114
Primates	*Saimiri boliviensis boliviensis*	265	218	483	-	GCF_000235385.1	94%	38823	18744880
Primates	*Pygathrix nemaeus*	191	312	503	-	GCA_004024825.1	75%	51434	68569

Order	*Species*	olfactory receptor genes			# Olfactory Turbinals	assembly	BUSCO % complete	contig N50	scaffold N50
		functional	non-functional	total					
Primates	*Rhinopithecus roxellana*	202	306	508	-	GCF_000769185.1	96%	1549224	77151
Primates	*Semnopithecus entellus*	184	332	516	4	GCA_004025065.1	58%	25064	29955
Primates	*Nasalis larvatus*	197	339	536	-	GCA_004027105.1	59%	34532	44943
Primates	*Rhinopithecus bieti*	187	355	542	-	GCF_001698545.1	92%	36417	2225337
Primates	*Nomascus leucogenys*	217	328	545	-	GCF_000146795.2	96%	52956880	35148
Primates	*Cercopithecus neglectus*	262	309	571	-	GCA_004027615.1	50%	9789	10270
Primates	*Alouatta palliata*	281	320	601	-	GCA_004027835.1	74%	51304	72427
Primates	*Ateles geoffroyi*	299	311	610	-	GCA_004024785.1	74%	59033	73111
Primates	*Aotus nancymaae*	340	282	622	-	GCA_000952055.2	96%	8268663	126456
Primates	*Plecturocebus donacophilus*	250	377	627	-	GCA_004027715.1	61%	41149	46445
Primates	*Chlorocebus sabaeus*	267	366	633	-	GCF_000409795.2	95%	90449	81825804
Primates	*Callithrix jacchus*	325	308	633	6	GCA_002754865.1	96%	155284	129239660
Primates	*Pithecia pithecia*	259	380	639	-	GCA_004026645.1	78%	63292	83104
Primates	*Saguinus imperator*	318	325	643	-	GCA_004024885.1	80%	53787	65636
Primates	*Macaca mulatta*	316	339	655	-	GCF_000772875.2	96%	4193270	107172
Primates	*Macaca fascicularis*	326	332	658	4	GCF_000364345.1	95%	88649475	86040
Primates	*Erythrocebus patas*	294	373	667	-	GCA_004027335.1	65%	28523	34535
Primates	*Papio anubis*	308	363	671	8	GCA_000264685.2	94%	585721	149817
Primates	*Cebus capucinus imitator*	348	324	672	-	GCF_001604975.1	95%	41196	5274112
Primates	*Cercocebus atys*	326	352	678	-	GCF_000955945.1	95%	112942	12849131
Primates	*Macaca nemestrina*	312	372	684	-	GCF_000956065.1	95%	106897	15219753
Primates	*Pongo abelii*	289	396	685	-	GCA_002880775.3	96%	98475126	11074009
Primates	*Cebus albifrons*	338	352	690	-	GCA_004027755.1	59%	25652	31156
Primates	*Mandrillus leucophaeus*	351	347	698	-	GCF_000951045.1	92%	31346	3186748
Primates	*Gorilla gorilla gorilla*	303	437	740	-	GCA_900006655.3	93%	20634945	9406846
Primates	*Pan paniscus*	318	436	754	-	GCF_000258655.2	95%	8197324	66676
Primates	*Pan troglodytes*	360	400	760	-	GCA_002880755.3	96%	53103722	12421315
Primates	*Homo sapiens*	372	448	820	8	GCA_000001405.27	100%	59364414	56413054
Primates	*Propithecus coquereli*	519	381	900	12	GCF_000956105.1	91%	28129	5604909
Primates	*Indri indri*	458	468	926	-	GCA_004363605.1	67%	26447	28620
Primates	*Eulemur flavifrons*	513	413	926	-	GCA_001262665.1	92%	27331	413352
Primates	*Lemur catta*	532	451	983	20	GCA_004024665.1	88%	158439	215715
Primates	*Nycticebus coucang*	519	532	1051	-	GCA_004027815.1	65%	17693	19724
Primates	*Otolemur garnettii*	672	387	1059	12	GCF_000181295.1	96%	27100	13852661
Primates	*Microcebus murinus*	670	468	1138	14	GCA_000165445.3	96%	108171978	210702
Primates	*Mirza coquereli*	619	540	1159	-	GCA_004024645.1	71%	60596	79947
Primates	*Cheirogaleus medius*	621	572	1193	12	GCA_004024725.1	88%	90923	118572
Primates	*Eulemur fulvus*	618	581	1199	-	GCA_004027275.1	60%	21310	22561
Primates	*Daubentonia madagascariensis*	700	769	1469	27	GCA_004027145.1	92%	298834	379919
Carnivora	*Cricetulus griseus*	851	517	1368	-	GCA_900186095.1	97%	62039716	97133
Rodentia	*Ctenodactylus gundi*	306	368	674	-	GCA_004027205.1	92%	218543	354548
Rodentia	*Jaculus jaculus*	421	338	759	-	GCF_000280705.1	93%	15675	22080993
Rodentia	*Allactaga bullata*	462	408	870	-	GCA_004027895.1	61%	30651	36308
Rodentia	*Thryonomys swinderianus*	523	421	944	-	GCA_004025085.1	60%	18955	21523
Rodentia	*Psammomys obesus*	507	465	972	-	GCA_002215935.1	96%	76398	
Rodentia	*Petromus typicus*	517	487	1004	-	GCA_004026965.1	68%	29916	35766
Rodentia	*Ellobius talpinus*	417	590	1007	-	GCA_001685095.1	54%	8259	15246
Rodentia	*Ellobius lutescens*	426	607	1033	-	GCA_001685075.1	84%	11648	242123
Rodentia	*Meriones unguiculatus*	552	487	1039	-	GCA_004026785.1	91%	69839	100883
Rodentia	*Muscardinus avellanarius*	598	456	1054	-	GCA_004027005.1	73%	44294	59013
Rodentia	*Ctenomys sociabilis*	527	530	1057	-	GCA_004027165.1	75%	37562	49073
Rodentia	*Chinchilla lanigera*	625	465	1090	-	GCF_000276665.1	95%	61105	21893125
Rodentia	*Myocastor coypus*	580	516	1096	10	GCA_004027025.1	66%	28446	35982
Rodentia	*Octodon degus*	613	490	1103	-	GCF_000260255.1	95%	19847	12091372
Rodentia	*Pedetes capensis*	585	536	1121	-	GCA_007922755.1	80%	54009	74432
Rodentia	*Rhizomys pruinosus*	614	523	1137	-	GCA_004026225.1	94%	2627	2627
Rodentia	*Mesocricetus auratus*	647	546	1193	14	GCF_000349665.1	86%	22511	12753307
Rodentia	*Perognathus longimembris pacificus*	615	579	1194	-	GCA_004363475.1	66%	17686	24714
Rodentia	*Glis glis*	694	518	1212	-	GCA_004027185.1	56%	26062	30338
Rodentia	*Acomys cahirinus*	720	492	1212	-	GCA_004027535.1	73%	42476	65411
Rodentia	*Dipodomys ordii*	707	529	1236	-	GCF_000151885.1	90%	48087	11931245
Rodentia	*Aplodontia rufa*	620	625	1245	-	GCA_004027875.1	73%	33245	37811
Rodentia	*Onychomys torridus*	654	606	1260	-	GCA_004026725.1	95%	18287	20878
Rodentia	*Ondatra zibethicus*	628	637	1265	-	GCA_004026605.1	83%	73746	89093
Rodentia	*Dipodomys stephensi*	673	604	1277	-	GCA_004024685.1	68%	30710	36811
Rodentia	*Microtus ochrogaster*	779	500	1279	-	GCF_000317375.1	95%	21250	17270019
Rodentia	*Zapus hudsonius*	695	594	1289	-	GCA_004024765.1	65%	23350	26350
Rodentia	*Graphiurus murinus*	716	603	1319	-	GCA_004027655.1	58%	22138	28463
Rodentia	*Dolichotis patagonum*	509	848	1357	-	GCA_004027295.1	74%	27983	31026
Rodentia	*Peromyscus maniculatus bairdii*	865	507	1372	14	GCF_000500345.1	96%	3760915	36367
Rodentia	*Xerus inauris*	632	742	1374	-	GCA_004024805.1	78%	64054	83865
Rodentia	*Cavia aperea*	556	824	1380	-	GCA_000688575.1	63%	1039	27928671
Rodentia	*Hystrix cristata*	684	714	1398	22	GCA_004026905.1	80%	57102	64768
Rodentia	*Mus spretus*	874	556	1430	-	GCA_001624865.1	95%	17887	131945496

Order	*Species*	olfactory receptor genes			# Olfactory Turbinals	assembly	BUSCO % complete	contig N50	scaffold N50
		functional	non-functional	total					
Rodentia	*Mus pahari*	942	536	1478	-	GCA_900095145.2	96%	111406228	29465
Rodentia	*Mus musculus*	964	523	1487	12	GCF_000001635.26	97%	52589046	32273079
Rodentia	*Sigmodon hispidus*	841	653	1494	14	GCA_004025045.1	86%	67983	101373
Rodentia	*Mus caroli*	993	524	1517	-	GCA_900094665.2	96%	122627250	30917
Rodentia	*Dinomys branickii*	734	862	1596	-	GCA_004027595.1	81%	64805	77918
Rodentia	*Ictidomys tridecemlineatus*	859	833	1692	-	GCF_000236235.1	94%	44137	8192786
Rodentia	*Cricetomys gambianus*	829	867	1696	-	GCA_004027575.1	83%	80761	110049
Rodentia	*Marmota marmota marmota*	873	826	1699	10	GCF_001458135.1	95%	66492	31340621
Rodentia	*Rattus norvegicus*	1073	679	1752	12	GCF_000001895.5	95%	100461	14986627
Rodentia	*Cuniculus paca*	752	1025	1777	18	GCA_004365215.1	94%	3890	3892
Rodentia	*Spermophilus dauricus*	864	1009	1873	-	GCA_002406435.1	91%	34849	1761345
Rodentia	*Cavia tschudii*	666	1215	1881	-	GCA_004027695.1	85%	64807	91436
Rodentia	*Capromys pilorides*	952	951	1903	-	GCA_004027915.1	34%	4080	4081
Rodentia	*Castor canadensis*	729	1301	2030	-	GCA_004027675.1	91%	49369	55723
Rodentia	*Cavia porcellus*	703	1335	2038	18	GCF_000151735.1	93%	80583	27942054
Rodentia	*Fukomys damarensis*	960	1113	2073	-	GCF_000743615.1	91%	5314287	44830
Rodentia	*Hydrochoerus hydrochaeris*	685	1566	2251	12	GCA_004027455.1	88%	148451	202224
Rodentia	*Nannospalax galili*	994	1273	2267	-	GCF_000622305.1	95%	30353	3618479
Rodentia	*Heterocephalus glaber*	820	1700	2520	-	GCF_000247695.1	95%	47778	20532749
Rodentia	*Dasyprocta punctata*	1429	1804	3233	-	GCA_004363535.1	71%	38650	43703
Scandenta	*Tupaia tana*	874	981	1855	-	GCA_004365275.1	27%	2971	2973
Scandenta	*Tupaia chinensis*	1053	1168	2221	8	GCF_000334495.1	94%	25938	3670124
Cingulata	*Chaetophractus vellerosus*	1114	1694	2808	-	GCA_004027955.1	15%	1606	1606
Cingulata	*Tolypeutes matacus*	1034	2095	3129	-	GCA_004025125.1	28%	9441	10217
Cingulata	*Dasypus novemcinctus*	1390	2646	4036	-	GCF_000208655.1	88%	26277	1687935
Pilosa	*Bradypus variegatus*	115	430	545	-	GCA_004027775.1	21%	1900	1900
Pilosa	*Choloepus didactylus*	929	1385	2314	32	GCA_004027855.1	41%	7167	7289
Pilosa	*Tamandua tetradactyla*	960	1476	2436	26	GCA_004025105.1	62%	18654	19789
Pilosa	*Myrmecophaga tridactyla*	1229	1940	3169	26	GCA_004026745.1	67%	36434	41255
Pilosa	*Choloepus hoffmanni*	1268	2571	3839	-	GCA_000164785.2	95%	366442	64490

Table S12.
Olfactory receptor gene and turbinal counts for 249 placental mammal species.

Gene	region	Forward or Reverse	# Trait Loss Species	# Trait Preserving Species	Perfect Match Margin	GLS p	p adjusted	total	significant
RPE65	chr1:68444529-68444685	r	154	22	-0.064	1.04E-08	0.00045	17	4
DDOST	chr1:20654221-20654374	r	153	22	-0.058	1.11E-08	0.00045	12	2
UBE2Q1	chr1:154552088-154552185	r	148	22	-0.092	2.75E-08	0.00088	12	1
MEF2D	chr1:156477006-156477219	r	140	22	-0.093	5.75E-08	0.00154	18	1
DHRS3	chr1:12580521-12580677	r	153	22	-0.076	1.44E-07	0.00314	12	2
FUBP1	chr1:77965066-77965238	r	153	20	-0.069	1.78E-07	0.00314	26	1
RPE65	chr1:68429768-68429941	r	153	22	-0.121	1.98E-07	0.00314	17	4
PTP4A2	chr1:31908411-31908972	r	146	19	-0.126	2.06E-07	0.00314	20	1
LARP4B	chr10:830862-830987	r	153	22	-0.048	2.34E-07	0.00314	23	1
TXNIP	chr1:145995139-145995305	r	150	22	-0.096	2.54E-07	0.00314	12	2
PINK1	chr1:20639884-20639997	f	151	22	-0.123	2.80E-07	0.00321	10	2
TXNIP	chr1:145994538-145994802	r	151	22	-0.117	5.93E-07	0.00635	12	2
DHRS3	chr1:12568339-12568426	r	154	22	-0.136	7.36E-07	0.00739	12	2
MFN2	chr1:12009590-12009731	f	152	22	-0.085	1.01E-06	0.00954	24	2
PINK1	chr1:20648990-20649237	f	151	21	-0.109	1.68E-06	0.01420	10	2
PINK1_AS	chr1:20648990-20649237	r	151	21	-0.109	1.68E-06	0.01420	5	1
RPE65	chr1:68431278-68431404	r	150	22	-0.118	1.77E-06	0.01422	17	4
RPE65	chr1:68440847-68441003	r	153	22	-0.070	2.50E-06	0.01912	17	4
NCDN	chr1:35560304-35561301	f	148	22	-0.048	6.28E-06	0.04459	13	1
MFN2	chr1:12004823-12004934	f	153	22	-0.143	6.96E-06	0.04459	24	2
CALML3	chr10:5525085-5525535	f	102	21	-0.060	7.20E-06	0.04459	2	1
RP11_116G8.5	chr10:5525085-5525535	r	102	21	-0.060	7.20E-06	0.04459	1	1
CACNA1E	chr1:181785710-181785826	f	153	21	-0.085	7.36E-06	0.04459	57	1
DDOST	chr1:20653619-20653777	r	153	22	-0.101	7.72E-06	0.04459	12	2
OVGP1	chr1:111421272-111421463	r	145	19	-0.292	7.84E-06	0.04459	7	1
QSOX1	chr1:180194211-180194405	f	148	22	-0.200	8.05E-06	0.04459	14	1
TMEM9	chr1:201151756-201151864	r	151	22	-0.211	8.52E-06	0.04498	6	1
SHC1	chr1:154965537-154965783	r	151	22	-0.105	8.68E-06	0.04498	16	1

Table S13.
Regions associated with hibernation in GLS forward genomics analysis.

gene set	description	genes in set	overlap	expected overlap	ratio	p	FDR	genes
GO:1904923	regulation of autophagy of mitochondrion in response to mitochondrial depolarization	13	2	0.01	> 100	7.52E-05	0.39	PINK1; MFN2
GO:0098779	positive regulation of mitophagy in response to mitochondrial depolarization	8	2	0.01	> 100	3.23E-05	0.51	PINK1; MFN2
GO:1904925	positive regulation of autophagy of mitochondrion in response to mitochondrial depolarization	12	2	0.01	> 100	6.52E-05	0.51	PINK1; MFN2
GO:0098780	response to mitochondrial depolarisation	18	2	0.02	> 100	1.36E-04	0.53	PINK1; MFN2
GO:1904707	positive regulation of vascular associated smooth muscle cell proliferation	33	2	0.03	69.32	4.22E-04	1.00	MEF2D; MFN2
GO:0016242	negative regulation of macroautophagy	35	2	0.03	65.36	4.71E-04	1.00	PINK1; QSOX1
GO:0042572	retinol metabolic process	52	2	0.05	43.99	1.00E-03	1.00	RPE65; DHRS3
GO:1904705	regulation of vascular associated smooth muscle cell proliferation	55	2	0.05	41.59	1.12E-03	1.00	MEF2D; MFN2
GO:0008286	insulin receptor signaling pathway	60	2	0.05	38.13	1.32E-03	1.00	RPE65; SHC1
GO:0072655	establishment of protein localization to mitochondrion	74	2	0.06	30.91	1.98E-03	1.00	PINK1; MFN2

Table S14.
Top ten GO Biological Process gene sets enriched for genes associated with hibernation in GLS forward genomics analysis. Gene sets overlapping < 2 genes are excluded.

direction in hibernators	gene set	genes in set	overlap	expected overlap	ratio	p	FDR	genes
faster evolving	GO:0006970: response to osmotic stress	76	2	0.0462	43.30	9.14E-04	0.583	SCN2A; SLC12A5
faster evolving	GO:0045785: positive regulation of cell adhesion	392	3	0.2382	12.59	0.0014	0.583	DENND6A; HSPD1; CRK
faster evolving	GO:0009266: response to temperature stimulus	201	2	0.1222	16.37	0.0062	0.983	SLC12A5; HSPD1
faster evolving	GO:0016358: dendrite development	219	2	0.1331	15.03	0.0073	0.983	SLC12A5; CRK
faster evolving	GO:0002449: lymphocyte mediated immunity	238	2	0.1446	13.83	0.0086	0.983	HSPD1; CRK
faster evolving	GO:0002347: response to tumor cell	20	1	0.0122	82.27	0.0121	0.983	HSPD1
faster evolving	GO:0051131: chaperone-mediated protein complex assembly	20	1	0.0122	82.27	0.0121	0.983	HSPD1
faster evolving	GO:0051271: negative regulation of cellular component movement	301	2	0.1829	10.93	0.0135	0.983	ADAMTS9; CRK
faster evolving	GO:0040013: negative regulation of locomotion	314	2	0.1908	10.48	0.0146	0.983	ADAMTS9; CRK
faster evolving	GO:0046677: response to antibiotic	316	2	0.1920	10.41	0.0148	0.983	HSPD1; CRK
slower evolving	GO:0044782: cilium organization	409	3	0.3038	9.87	0.0029	1.000	CENPJ; CFAP410; ARMC9
slower evolving	GO:0009112: nucleobase metabolic process	36	1	0.0267	37.40	0.0264	1.000	ALDH6A1
slower evolving	GO:0042769: DNA damage response, detection of DNA damage	39	1	0.0290	34.52	0.0286	1.000	CFAP410
slower evolving	GO:1901987: regulation of cell cycle phase transition	386	2	0.2867	6.98	0.0319	1.000	CENPJ; RIOK2
slower evolving	GO:1904951: positive regulation of establishment of protein localization	437	2	0.3246	6.16	0.0401	1.000	CENPJ; RIOK2
slower evolving	GO:0033627: cell adhesion mediated by integrin	67	1	0.0498	20.09	0.0487	1.000	ITGBL1
slower evolving	GO:0044772: mitotic cell cycle phase transition	487	2	0.3617	5.53	0.0488	1.000	CENPJ; RIOK2
slower evolving	GO:0030101: natural killer cell activation	85	1	0.0631	15.84	0.0614	1.000	ELF4
slower evolving	GO:0072527: pyrimidine-containing compound metabolic process	99	1	0.0735	13.60	0.0712	1.000	ALDH6A1
slower evolving	GO:0007229: integrin-mediated signaling pathway	101	1	0.0750	13.33	0.0725	1.000	ITGBL1

Table S15.
GO Biological Process gene sets enriched for genes evolving faster or slower in hibernators in RERconverge analysis. We tested all significant genes against a non-redundant representative set of Gene Ontology Biological Processes using WebGestalt.

Supplementary data files

Data S1. (separate file)
Average phyloP scores for protein-coding genes.

Data S2. (separate file)
Input data for analysis of constraint in 100kb bins. chr: Chromosome window is located on; start: start position of window; end: end position of window; zoonomiaBases: Number of bases within window with phyloP scores; zoonomiaConsBases: Number of bases within window with phyloP > 2.27 (5% FDR constraint positions); zoonomiaNspeciesLow: Number of bases within window with a low (≤24) number of species aligning; zooN, A, C, G, T: Number of each base within window; zooFrac: Fraction of bases in window under constraint (phyloP > 2.27); k24gt90.sum: Number of bases with k24 > 0.9 (K24 UCSC track, a measure of mappability); cds.distinct.sum: Number of bases in CDS; cCRE.sum: Number of bases in cCREs; dhs.sum: Number of bases overlapping DHS.

Data S3. (separate file)
Hibernation RERConverge results. Results include p-values, phylogenetic permulations p-values, and Bayes factor values to indicate whether a gene's relative evolutionary rate is associated with hibernation.