

Title: Evolutionary constraint and innovation across hundreds of placental mammals

Authors:

Matthew J. Christmas^{1†}, Irene M. Kaplow^{2,3†}, Diane P. Genereux⁴, Michael X. Dong¹, Graham M. Hughes⁵, Xue Li^{4,6,7}, Patrick F. Sullivan^{8,9}, Allyson G. Hindle¹⁰, Gregory Andrews⁷, Joel C. Armstrong¹¹, Matteo Bianchi¹, Ana M. Breit¹², Mark Diekhans¹¹, Cornelia Fanter¹⁰, Nicole M. Foley¹³, Daniel B. Goodman, Linda Goodman¹⁴, Kathleen C. Keough^{14,15,16}, Bogdan Kirilenko^{17,18,19}, Amanda Kowalczyk^{2,3}, Colleen Lawless⁵, Abigail L. Lind^{15,16}, Jennifer R. S. Meadows¹, Lucas R. Moreira^{4,7}, Ruby W. Redlich²⁰, Louise Ryan⁵, Ross Swofford⁴, Alejandro Valenzuela²¹, Franziska Wagner²², Ola Wallerman¹, Ashley R. Brown^{2,3}, Joana Damas²³, Kaili Fan⁷, John Gatesy²⁴, Jenna Grimshaw²⁵, Jeremy Johnson⁴, Sergey V. Kozyrev¹, Alyssa J. Lawler^{3,4,20}, Voichita D. Marinescu¹, Kathleen M. Morrill^{4,6,7}, Austin Osmani²⁵, Nicole S. Paulat²⁵, BaDoi N. Phan^{2,3,26}, Steven K. Reilly²⁷, Daniel E. Schäffer², Cynthia Steiner²⁸, Megan A. Supple²⁹, Aryn P. Wilder²⁸, Morgan E. Wirthlin^{2,3,30}, James R. Xue^{4,31}, Zoonomia Consortium§, Bruce W. Birren⁴, Steven Gazal³², Robert M. Hubble³³, Klaus-Peter Koepfli^{34,35,36}, Tomas Marques-Bonet^{37,38,39,40}, Wynn K. Meyer⁴¹, Martin Nweeia^{42,43,44,45}, Pardis C. Sabeti^{4,31,46}, Beth Shapiro^{29,47}, Arian F. A. Smit³³, Mark Springer⁴⁸, Emma Teeling⁵, Zhiping Weng⁷, Michael Hiller^{17,18,19}, Danielle L. Levesque¹², Harris A. Lewin^{23,49,50}, William J. Murphy¹³, Arcadi Navarro^{37,39,51,52}, Benedict Paten¹¹, Katherine S. Pollard^{15,16,53}, David A. Ray²⁵, Irina Ruf⁵⁴, Oliver A. Ryder^{28,55}, Andreas R. Pfenning^{2,3}, Kerstin Lindblad-Toh^{1,4†*}, Elinor K. Karlsson^{4,7,56†*}

Affiliations:

¹Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University; Uppsala, 751 32, Sweden.

²Department of Computational Biology, School of Computer Science, Carnegie Mellon University; Pittsburgh, PA 15213, USA.

³Neuroscience Institute, Carnegie Mellon University; Pittsburgh, PA 15213, USA.

⁴Broad Institute of MIT and Harvard; Cambridge, MA 02139, USA.

⁵School of Biology and Environmental Science, University College Dublin; Belfield, Dublin 4, Ireland.

⁶Morningside Graduate School of Biomedical Sciences, UMass Chan Medical School; Worcester, MA 01605, USA.

⁷Program in Bioinformatics and Integrative Biology, UMass Chan Medical School; Worcester, MA 01605, USA.

⁸Department of Genetics, University of North Carolina Medical School; Chapel Hill, NC 27599, USA.

⁹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet; Stockholm, Sweden.

¹⁰School of Life Sciences, University of Nevada Las Vegas; Las Vegas, NV 89154, USA.

¹¹Genomics Institute, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

¹²School of Biology and Ecology, University of Maine; Orono, ME 04469, USA.

¹³Veterinary Integrative Biosciences, Texas A&M University; College Station, TX 77843, USA.

¹⁴Fauna Bio Incorporated; Emeryville, CA 94608, USA.

¹⁵Department of Epidemiology & Biostatistics, University of California San Francisco; San Francisco, CA 94158, USA.

¹⁶Gladstone Institutes; San Francisco, CA 94158, USA.

- ¹⁷Faculty of Biosciences, Goethe-University; 60438 Frankfurt, Germany.
- ¹⁸LOEWE Centre for Translational Biodiversity Genomics; 60325 Frankfurt, Germany.
- ¹⁹Senckenberg Research Institute; 60325 Frankfurt, Germany.
- ²⁰Department of Biological Sciences, Mellon College of Science, Carnegie Mellon University; Pittsburgh, PA 15213, USA.
- ²¹Department of Experimental and Health Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra; Barcelona, 08003, Spain.
- ²²Museum of Zoology, Senckenberg Natural History Collections Dresden; 01109 Dresden, Germany.
- ²³The Genome Center, University of California Davis; Davis, CA 95616, USA.
- ²⁴Division of Vertebrate Zoology, American Museum of Natural History; New York, NY 10024, USA.
- ²⁵Department of Biological Sciences, Texas Tech University; Lubbock, TX 79409, USA.
- ²⁶Medical Scientist Training Program, University of Pittsburgh School of Medicine; Pittsburgh, PA 15261, USA.
- ²⁷Department of Genetics, Yale School of Medicine; New Haven, CT 06510, USA.
- ²⁸Conservation Genetics, San Diego Zoo Wildlife Alliance; Escondido, CA 92027, USA.
- ²⁹Department of Ecology and Evolutionary Biology, University of California Santa Cruz; Santa Cruz, CA 95064, USA.
- ³⁰Allen Institute for Brain Science; Seattle, WA 98109, USA.
- ³¹Department of Organismic and Evolutionary Biology, Harvard University; Cambridge, MA 02138, USA.
- ³²Keck School of Medicine, University of Southern California; Los Angeles, CA 90033, USA.
- ³³Institute for Systems Biology; Seattle, WA 98109, USA.
- ³⁴Center for Species Survival, Smithsonian's National Zoo and Conservation Biology Institute; Washington, DC 20008, USA.
- ³⁵Computer Technologies Laboratory, ITMO University; St. Petersburg 197101, Russia.
- ³⁶Smithsonian-Mason School of Conservation, George Mason University; Front Royal, VA 22630, USA.
- ³⁷Catalan Institution of Research and Advanced Studies (ICREA); Barcelona, 08010, Spain.
- ³⁸CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST); Barcelona, 08036, Spain.
- ³⁹Department of Medicine and Life Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra; Barcelona, 08003, Spain.
- ⁴⁰Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona; 08193, Cerdanyola del Vallès, Barcelona, Spain.
- ⁴¹Department of Biological Sciences, Lehigh University; Bethlehem, PA 18015, USA.
- ⁴²Department of Comprehensive Care, School of Dental Medicine, Case Western Reserve University; Cleveland, OH 44106, USA.
- ⁴³Department of Vertebrate Zoology, Canadian Museum of Nature; Ottawa, Ontario K2P 2R1, Canada.
- ⁴⁴Department of Vertebrate Zoology, Smithsonian Institution; Washington, DC 20002, USA.
- ⁴⁵Narwhal Genome Initiative, Department of Restorative Dentistry and Biomaterials Sciences, Harvard School of Dental Medicine; Boston, MA 02115, USA.
- ⁴⁶Howard Hughes Medical Institute; Chevy Chase, MD, USA.
- ⁴⁷Howard Hughes Medical Institute, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

⁴⁸Department of Evolution, Ecology and Organismal Biology, University of California Riverside; Riverside, CA 92521, USA.

⁴⁹Department of Evolution and Ecology, University of California Davis; Davis, CA 95616, USA.

⁵⁰John Muir Institute for the Environment, University of California Davis; Davis, CA 95616, USA.

⁵¹BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation; Barcelona, 08005, Spain.

⁵²CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST); Barcelona, 08003, Spain.

⁵³Chan Zuckerberg Biohub; San Francisco, CA 94158, USA.

⁵⁴Division of Messel Research and Mammalogy, Senckenberg Research Institute and Natural History Museum Frankfurt; 60325 Frankfurt am Main, Germany.

⁵⁵Department of Evolution, Behavior and Ecology, School of Biological Sciences, University of California San Diego; La Jolla, CA 92039, USA.

⁵⁶Program in Molecular Medicine, UMass Chan Medical School; Worcester, MA 01605, USA

†contributed equally, shared first author

‡contributed equally, shared last and corresponding author

§see list at end of manuscript

*Corresponding author. Email: kersli@broadinstitute.org (KLT) and elinor.karlsson@umassmed.edu (EKK)

Abstract: Evolutionary constraint and acceleration are powerful, cell-type agnostic measures of functional importance. Previous studies in mammals were limited by species number and reliance on human-referenced alignments. We explore the evolution of placental mammals, including humans, through reference-free whole-genome alignment of 240 species and protein-coding alignments for 428 species. We estimate 10.7% of the human genome is evolutionarily constrained. We resolve constraint to single nucleotides, pinpointing functional positions, and refine and expand by over seven-fold the catalog of ultraconserved elements. Overall, 48.5% of constrained bases are as yet unannotated, suggesting yet-to-be-discovered functional importance. Using species-level phenotypes and an updated phylogeny, we associate coding and regulatory variation with olfaction and hibernation. Focusing on biodiversity conservation, we identify genomic metrics that predict species at risk of extinction.

Main text:

Placental mammals, the evolutionary lineage that includes humans, are exceptionally diverse. With more than 5,000 extant species, from the 2 gram bumblebee bat to the 150,000 kilogram blue whale, mammals have adapted to almost every habitat on earth^{1,2}. How rapidly placental mammals diversified and whether this preceded or followed the major extinction event at the end of the Cretaceous period(**Fig. 1A**), have been difficult to determine due to the dearth of fossil and molecular data^{3,4}.

By comparing genomes, we can find positions whose unusual patterns of variation suggest functional importance⁵. Lack of variation (conservation) suggests purifying selection. Higher than expected variation in specific lineages (acceleration) suggests positive selection.

Exceptionally high conservation points to evolutionary constraint. Genomes also evolve through larger structural changes, including introgression and translocation of transposable elements⁶, deletions, and duplications⁷.

Associating evolutionary signals with species-level phenotypes can identify genetic variants underlying specific adaptations⁸, such as diet type⁹, echolocation¹⁰, and subterranean habitation¹¹. Coding regions tend to change more slowly than regulatory regions, which frequently arise and disappear within placental lineages, and at which nucleotide-level conservation is sometimes weak¹². The sensitivity and specificity of methods for detecting evolutionary signals and associating them with phenotypes depends on the number and diversity of species analyzed^{13,14}.

Zoonomia is the largest mammalian comparative genomics resource produced to date, with whole genomes aligned for 240 species (2.3-fold more families and ~4-fold more species than the mammals included in the earlier 100 Vertebrates alignment¹⁵) and protein-coding sequences aligned for 428 species. Our first publication described how species were selected, sequenced, and aligned using a reference-free approach that captures sequences missing from the human genome¹⁶. In a companion paper, we explore how evolutionary constraint is a powerful tool for mapping functional variants involved in human diseases¹⁷. Here, we address long-standing questions about the evolution of placental mammals, annotate regions of exceptional constraint and acceleration in the human genome, and investigate the origins of intriguing mammalian traits.

New resources for comparative genomics

Zoonomia includes two large alignments: the 241-way reference-free Cactus alignment, with 240 species (domestic dog has two representatives)¹⁶, and the 428-way human-referenced TOGA (Tool to infer Orthologs from Genome Alignments) alignment of protein-coding sequences (19,464 genes)¹⁸. We measured sequence conservation using phyloP for versions of the Cactus alignment projected onto the human, chimpanzee, mouse, dog, and little brown bat reference genomes (**Table S1**)⁵. For the primate lineage (43 species), we also measured conservation using phastCons, which models multibase elements and achieves more power with fewer species¹⁹. In **Table S2**, we list all species included with their common and binomial names, associated analyses, and phenotypes.

Phylogeny supports the “Long Fuse” Model

We used the Cactus alignment to infer a new phylogeny of placental mammals from near-neutrally evolving variable sites (N=466,232) (**Fig. 1B**) (detailed in²⁰). Divergence time analyses supported the “Long Fuse” model of mammalian diversification, where interordinal diversification took place in the Cretaceous and the majority of interordinal diversification took place after the Cretaceous-Paleogene (K-Pg) mass extinction event (Fig. 1A)²¹. In contrast, our analyses did not support the fossil record-derived “Explosive” model, which places all inter- and intraordinal diversification after the K-Pg event, and alternative scenarios posited by numerous analyses of evolutionarily constrained sequences^{22,23}.

The timetree and phylogenomic analysis, displaying a relative lack of phylogenomic discordance at interordinal nodes, suggests interordinal diversification occurred via allopatry during a period of continental fragmentation and sea level change prior to the K-Pg boundary (**Fig. 1A**)^{24,25}. Conversely, some post-K-Pg intraordinal diversification events show high phylogenomic discordance, consistent with a history of incomplete lineage sorting and gene flow. Notably, we highlight a pulse of predominantly intraordinal diversification immediately following the K-Pg boundary (**Fig. 1A**), a molecular result consistent with the fossil record²⁶.

Reconstructing the mammalian ancestor

We reconstructed a karyotype for the ancestor of all mammals (placental, marsupial, and monotreme) using high-contiguity genome assemblies for 32 species, representing all but four orders (detailed in ²⁷). Our sequence-based approach assigns 87% of the human genome to 20 ancestral chromosomes. We detected 365 rearrangements (145 new) and dated each on the mammalian phylogeny. On average ~2 breakpoints occur every million years (My), but this varies widely. There are forty-fold more on the branch to the therian ancestor (3.9/My) than between the placental and boreoeutherian ancestor (0.1/My). Nearly half the human genome is in synteny blocks >1Mb and shared across all mammals, with the longest (22.1 Mb) encompassing the *HOXD* gene cluster and other genes involved in embryonic morphogenesis(**Table S3**).

Transposable element activity

We cataloged transposable elements (TEs) in the genome assemblies for 243 species, producing by far the largest resource for investigating how these powerful agents of genomic change vary in placental mammals (detailed in ²⁸)(**Fig. S1**). TEs are repetitive, mobile DNA sequences that tend to insert as units of 100 to 10,000bp and accumulate (>1 million copies per genome). We found that TEs comprise 49.0%±7.5% of each genome on average with little variability across species, consistent with a model where TE accumulation is counterbalanced by DNA loss²⁹. Recent accumulation is positively correlated with genome size, suggesting insufficient time to purge TE insertions after a surge of activity. It is negatively correlated with TE diversity, suggesting genomic control mechanisms may limit the repertoire of active TEs³⁰. For specific TE classes, hotspots of accumulation dot the placental tree. For example, long interspersed nuclear elements (LINEs) are common in aardvarks and long terminal repeat (LTR) retrotransposons have accumulated in some rodents. Bats are a hotspot for horizontal transfer of DNA transposons (detailed in ³¹), having experienced over 200 such events, whereas only 11 transposons horizontally transferred into other lineages(**Table S4**).

Evolutionary constraint in humans

Zoonomia was designed to achieve single-base resolution of constraint in placental mammals by maximizing branch length¹⁶. With the 240 species selected, only 191 positions in the human genome (0.000006%) are expected to be perfectly conserved by chance. In the Cactus alignment, we observe 19,000-fold more than our expectation (3.6 million perfectly conserved positions).

We estimate that at least 10.7% (332 Mb) of the human genome is under constraint via purifying selection(**Fig. 2A**)⁵. While higher than estimated with the less well-powered 29 mammals dataset (5.5%)³², this is consistent with other estimates^{33–35}(**Table S1**). We referenced the alignment on four other species (chimpanzee, mouse, dog and little brown bat) and measured constraint in each at different false discovery rates.

We used the chimpanzee-referenced alignment to measure constraint at 10,032 human-specific deletions that had been functionally assessed using massively parallel reporter assays, including 717 associated with differences in enhancer activity (detailed in ³⁶). Because these positions are deleted in humans, the human-referenced alignment does not score them. Filtering based on phyloP scores from the Cactus alignment substantially increased concordance between predicted transcription factor binding differences and measured species-specific regulatory activity.

Alignment depth

At any position in the 241-way Cactus alignment, the number of species aligned varies widely. Eleven percent of the human genome is aligned to $\geq 95\%$ (≥ 228) of the species. Most of the genome (91%) aligns to at least five species (there are five other hominids in the alignment (**Fig. S2**). Highly conserved elements have deep alignments, and those evolving quickly align only to closely related species(**Fig. 2B**) (detailed in ³⁷). In regulatory sequence, the depth of the alignment varies with element type. ENCODE³⁸ promoter-like and enhancer-like candidate *cis*-regulatory elements (cCREs) align well to most species, but CTCF-only cCREs tend to change more quickly(**Fig. 2C**).

Single-base resolution of constraint

Single bases confidently (FDR<0.05) designated as evolutionarily constrained comprise 3.53% (101 Mb) of the human genome (**Fig. 2A**; **Table S1**). In the human genome, most (80%) are within 5bp of another constrained base, and 30% are in blocks ≥ 5 bp. Sequence data for ~140,000 individuals in TOPMed³⁹ shows variants at constrained bases (N=20,718,868) tend to have lower minor allele frequencies than unconstrained bases (0.0026 vs. 0.004; Wilcoxon rank sum exact test $p = 9.5 \times 10^{-13}$), consistent with purifying selection. Variants predicted to impact function have extremely low frequencies, regardless of phyloP score(**Fig. S3**).

Patterns of constraint in protein-coding sequence highlight the single-base resolution offered by the 241-way alignment. Most non-degenerate sites are constrained; most four-fold degenerate sites are not (74.1% vs 18.5%). Start codons, stop codons and splice sites are enriched for constraint (24x, 19x, and 25x greater; χ^2 test, $p < 2.2 \times 10^{-16}$). For coding bases, median phyloP is 4.9 (IQR=5.8) in the first position (non-degenerate for 17/20 amino acids), rises to 6.0 (IQR=4.0) in the second (non-degenerate in 19/20), and drops to 0.68 (IQR=2.7) in the third (non-degenerate for 2/20)(**Fig. S4**). Stop codons and non-degenerate codons (tryptophan and methionine) are constrained at all three positions(**Fig. 2D**). PhyloP scores are negatively correlated with the number of nucleotide options (Spearman rho = -0.51, $p < 2.2 \times 10^{-16}$)(**Fig. 2E**).

When codons encode amino acids at critical positions in the peptide, the variation in single-base constraint deviates from the pattern predicted from degeneracy alone. The codon for methionine is more conserved when functioning as a start codon than elsewhere(**Fig. 2F**). Similarly, cysteines in intra-peptide disulfide bridges (could cause misfolding when mutated) are more conserved than other cysteines(**Fig. 2F**). Four-fold degenerate sites overlapping ENCODE3 transcription factor binding sites ($N = 2,647,541$)⁴⁰ are more constrained than other four-fold degenerate sites ($N = 2,420,610$; χ^2 test, $p < 2.2 \times 10^{-16}$, **Fig. 2G**, **Fig. S5**).

Constraint is strongly enriched in coding sequence and moderately enriched in regulatory elements(**Fig. 2H**). The most constrained genes are cell cycle and developmental, including embryology, neurogenesis, and morphogenesis of most organ systems. The least constrained genes shape an animal's interaction with its environment: immunity (particularly response to pathogens), skin, smell, and taste(detailed in ¹⁷). Gene deserts (the longest 5% of intergenic regions; $N=873$) are more constrained when neighboring developmental transcription factors ($N=224$; $p_{\text{Wilcoxon}}=2.16 \times 10^{-15}$). cCREs in these gene deserts ($N=38,065$) are unusually conserved($p_{\text{Wilcoxon}} < 2.2 \times 10^{-16}$).

We used convolutional neural networks and publicly available ChIP-seq data to identify transcription factor binding sites (TFBS) genome-wide for 368 transcription factors(detailed in ³⁷). Most TFBS are unconstrained, but for each transcription factor, a distinct minority are constrained (**Fig. 3A**) and appear active in other species (H3K4me3 and H3K27ac modified; **Fig. S6**). Motif scores for bases in constrained TFBS correlate with the magnitude of constraint(**Fig. 3BC**). Motif logos calculated from constrained TFBS for CTCF, a highly conserved and ubiquitously expressed transcription factor, are nearly identical across species(**Fig. 3D**). Most human CTCF constrained TFBS sites are found in macaque, rat, mouse and dog, and their function, as measured by CTCF ChIP-seq, is retained(**Fig. 3EF**). Unconstrained TFBS are often shared with macaque but not with more distantly related species.

Regions of constraint

We defined constrained regions genome-wide using three approaches of varying stringency.

Zoonomia UltraConserved Elements (zooUCEs) are regions ≥ 20 bp where every position is identical in at least 235 (98%) of species. This refines and expands the original set of 481 UCEs: elements > 200 bp with identical sequence between human, mouse, and rat⁴¹. We discover 3,799 new UCEs, and 753 that overlap 318 original UCEs, for a total of 4,552 high-resolution zooUCEs averaging 28.9 ± 13.0 bp, with $27 \geq 100$ bp (**Fig. S7A**). ZooUCEs tend to be near genes in transcription-related and developmental biological processes(**Data S1**). The longest two (190bp and 161bp) are both in *POLAI*, which encodes the catalytic subunit of DNA polymerase-alpha, and are separated by a single base.

Regions of contiguous constraint (RoCC) are defined as twenty or more constrained bases in a row(**Fig. S7B**). This criteria, less stringent than for zooUCEs, finds 273 RoCCs ≥ 500 bp and six ≥ 1 kb. The longest RoCC (1.36kb; chr2:172071926-172073285) encompasses four distal

enhancer-like ENCODE cCREs in an intron of the gene *METAP1D*. *METAP1D* encodes an essential mitochondrial protein conserved at least back to zebrafish⁴².

Finally, we divided the genome into 100kb bins (N=28,218) and identified 53 bins with significant constraint ($q < 0.05$; average 17.8% constrained bases vs 3.5% for genome; **Table S5**). These bins are enriched for transcription-related biological processes (**Data S1**) and overlap all four *HOX* gene clusters (**Fig. 2I**). Five are in gene deserts, and two neighbor conserved developmental transcription factors (*LMO4* and *BCL11A*).

Unannotated constraint

Almost half of all constrained bases (48.5%) are in regions with no annotations in the thousands of cell types, tissues, or conditions assayed by ENCODE3 (**Table S6**)³⁸. We grouped constrained bases <5bp apart in unannotated regions (excluding repeats, centromeres and telomeres) to define 424,179 UNannotated Intergenic Constrained RegioNs (UNICORNs) (mean size 38 ± 53 bp; Q95=131bp; 0.5% of genome; **Figs. 4A, S7C**). Most (77.0%) UNICORNs are within 500kb of the transcription start site for a protein-coding gene. UNICORNs have less variation than expected (lower minor allele frequencies; $p_{\text{Wilcoxon}} < 2.2 \times 10^{-16}$) and are less likely to overlap TOPMed³⁹ variants ($p_{\text{Wilcoxon}} < 2.2 \times 10^{-16}$) (**Fig. 4B**).

Many UNICORNs are likely functional under conditions not assayed in ENCODE3. Genomic elements that are active (annotated as open chromatin) in developing brain tissues⁴³, adult motor cortical neuron cell types⁴⁴, and narrowly defined regions of young adult brain⁴⁵ overlap 8.8%, 7.1%, and 8.6% of UNICORNs, respectively (17% collectively). This suggests UNICORNs demarcate elements important in tissue types or at time points that are difficult to access in functional analyses.

Regulatory evolution through transposable elements

Overall, about 11% of constrained bases are in repeats in the human alignment. Constraint is enriched in simple repeats and DNA transposons and depleted in short interspersed nuclear elements, LTRs, and satellite repeats (**Fig. S8A**). This likely reflects differences in functional importance. Simple repeats closer to genes, and thus more likely to influence gene expression⁴⁶, are more constrained (Spearman rho = -0.13, $p < 2.2 \times 10^{-16}$; **Fig. S8B**). DNA transposons are an ancient class of repeats known to acquire functional roles⁴⁷, such as the transcription factor *ZBED5* (70% constrained)⁴⁸. In contrast, the repeat classes depleted in constraint have been active more recently during primate evolution and may not yet have acquired function⁴⁹.

Most (87%) primate-specific TFBS overlap transposable elements (compared to 30% of non-primate TFBS), indicative of the important role of transposable elements in recent regulatory evolution⁵⁰ (**Fig. 3G**) (detailed in ³⁷). TFBS in transposable elements, and especially younger elements, tend to be less conserved and change more quickly (**Fig. S9A,B**)

Accelerated regions and topological associating domain boundaries

We developed an automated pipeline for identifying “accelerated” regions that are highly constrained across placental mammals but exceptionally variable in particular lineages, suggesting adaptive evolution (detailed in ⁵¹). Applying this to the human and chimp alignments, we found 312 human accelerated regions (zooHARs) and 141 in chimpanzees (zooCHARs). Accelerated regions in both species are mainly noncoding, have signatures of positive selection (82% and 86%, respectively), and reside near developmental and neurological genes⁵², like previous sets of HARs^{53–55}. zooHARs and zooCHARs overlap topological associating domains containing human-specific and chimp-specific structural variants, respectively⁵⁶. 3D genome structure is altered by human-specific structural variants in zooHARs, suggesting a mechanism for species-specific evolution through enhancer hijacking⁵¹.

Annotation and alignment of protein-coding genes

Annotating genes through alignment

Pairing the Cactus alignment with gene annotations from three species with exceptionally high-quality, high-contiguity genomes (human, mouse, and goat), we annotated genes in 238 placental species, including 129 without an NCBI or Ensembl annotation (**Table S2**)⁵⁷. Predicted orthologs have contiguous open reading frames spanning 90–110% of human protein length. Starting with an input set of 22,799 protein-coding genes, we annotated between 3,871 and 20,168 (averaging $12,905 \pm 2,925$) genes per genome. As expected, the number of genes annotated decreases as assembly contiguity drops (contig N50: Pearson $r=0.28$, $p=1 \times 10^{-5}$; scaffold N50: Pearson $r=0.37$, $p=3 \times 10^{-9}$) and as phylogenetic distance from human (the species with the highest quality gene annotations) increases (Pearson $r=-0.41$, $p=7 \times 10^{-11}$) (**Table S7**). These annotations, although incomplete, offer the first gene annotations for over a hundred species.

New 428-way protein-coding alignment in mammals

We constructed the largest-ever mammalian protein-coding alignment (428 species) using a new method integrating gene annotation, ortholog detection, and classification of genes as intact or lost (detailed in ¹⁸). TOGA (Tool to infer Orthologs from Genome Alignments) uses a machine learning model trained on human and mouse to annotate orthologs between human and each of the other species, and it can distinguish orthologs from paralogs or processed pseudogenes. It joins non-contiguous genes in fragmented assemblies, making it particularly useful for species where obtaining high-quality DNA is difficult, including rare or endangered species. We annotated, on average, 19,144 orthologous genes for ape assemblies and 17,779 genes for non-ape assemblies. Using a benchmarking set of mammalian single copy orthologs⁵⁸, TOGA improves annotation completeness for 108 of the 118 assemblies with an existing NCBI annotation.

Repeated loss of CMAH gene is associated with immunity evolution

For one gene known to be lost in humans, we used non-human referenced alignments to explore its persistence across placental mammals. Loss of *CMAH* was initially thought to be specific to the *Homo* lineage, prompting speculation it might explain distinctive human characteristics^{59–62}. *CMAH* hydroxylates the sialic acid Neu5Ac to Neu5Gc, and its loss restricts infection by pathogens dependent on Neu5Gc (e.g. *Plasmodium reichenowi*⁶³) but increases susceptibility to viruses that bind Neu5Ac (e.g. MERS coronavirus⁶⁴) and may drive speciation⁶⁵. We found three independent losses of *CMAH* in rodents and three in bats (two already described⁶²)(**Fig. S10**). *CMAH* accumulated multiple stop codons in the Cairo spiny mouse. Loss in mountain beaver and Brazilian guinea pig reflects diminished gene content and pseudogenization. Additional experiments are needed to confirm non-functionality.

Among the species with *CMAH* loss is the American mink (*Neovison vison*), the only non-human animal to suffer well-documented high mortality from SARS-CoV2. SARS-CoV2 likely lacks Neu5Gc⁶⁶ and thus may face fewer challenges in moving among mammalian hosts lacking *CMAH*^{67,68}. *CMAH* loss has been described in white-tailed deer⁶², another species susceptible to SARS-CoV-2 infection (without high mortality)⁶⁹, but we found fully intact *CMAH* homologs in more recent assemblies (GCA_002102435.1 and GCA_014726795.1). The connection between *CMAH* loss and zoonotic potential warrants further investigation, and monitoring species that have lost *CMAH* may help anticipate future zoonotic transfer events⁷⁰.

Connecting genotype to phenotype

Olfaction

We performed *de novo* olfactory receptor gene detection in 249 species and confirmed that counts vary widely (mean count = 1,217±683; N=249)(**Table S8**)⁷¹. Our reference-free approach is designed to avoid underestimating olfactory receptor gene counts in distantly related taxa and is robust to variation in genome assembly quality (count ~ contig N50: rho = 0.006, p-value= 0.93; count ~ scaffold N50: rho = 0.013; p-value= 0.84; N=236). African elephant has the largest repertoire (4,199 genes; 1,765 functional and 2,434 non-functional); killer whale has the smallest (154 genes: 24 functional and 130 non-functional)(**Fig. S11**).

Similar olfactory receptor gene counts in distantly related species may reflect shared selection pressures. Bumblebee bat and North Pacific right whale have among the smallest repertoires (332 and 392 genes respectively). Both species rely on echolocation⁷², although North Pacific right whale is also aquatic. Overall, whales (17 species) have fewer than any other clade, averaging 225±75 genes, consistent with previous work⁷³. Closely related species can have divergent counts. Anteaters, tree sloths, and armadillos (superorder Xenarthra) generally have large repertoires (2,784±1,089 genes, N=8). Brown-throated three-toed sloth, though, has only 545 genes (115 functional, 430 non-functional), far fewer than the two-toed sloth (2,314 genes: 929 functional, 1,385 non-functional) and Hoffman's two-toed sloth (3,839 genes, 1,268 functional, 2,571 non-functional). Three-toed sloths (family Bradypodidae) are less active and

have better vision than two-toed sloths (family Choloepodidae), which rely on olfaction to explore new locations⁷⁴. The olfactory variation between the two sloth families supports the hypothesis that they independently evolved suspensory behavior^{75–77}.

Olfactory receptor gene counts are strongly correlated with olfactory turbinal counts (Spearman $\rho=0.71$; phyloilm coefficient 0.014, phylogenetic permutations $p = 0.0013$; $N=63$) (**Fig. S12**)^{78,79}. Turbinals are an anatomic feature of the nasal cavity associated with olfaction capacity^{80–82}. With 65 phenotyped species, we discerned the connection between these two traits through genomic analysis, highlighting the potential of this approach for investigating multifactorial evolutionary responses to a selective pressure.

Hibernation

The evolutionary origins of hibernation and its core physiological state of metabolic depression (torpor) are unclear. Hibernation exists in every deep mammalian lineage, suggesting that it is either an ancestral trait repeatedly lost or a derived trait repeatedly gained^{83,84}. Understanding the genomic basis of its constituent physiology, including cellular recovery from repeated cooling and rewarming without apparent long-term harm, could inform therapeutics, critical care, and long-distance spaceflight.

If hibernation is ancestral, hibernator orthologs should more closely match the mammalian ancestor, revealing putative essential genes underlying this phenotype. We used Generalized Least Squares Forward Genomics¹³ to identify 100bp protein-coding regions where 22 deep hibernators (species capable of core temperature depression below 18°C for >24h) had higher sequence identity than 154 strict homeotherms to a reconstructed ancestral mammal genome. We found 14 regions in 11 genes more conserved ($p_{\text{FD}} < 0.01$) in deep hibernators than in strict homeotherms. Two genes, *MFN2* and *PINK1*, overlap four gene ontology (GO) Biological Process gene sets ($p_{\text{FDR}} < 0.05$) related to depolarization and degradation of damaged mitochondria, an organelle essential for metabolic depression⁸⁵. A third, *TXNIP*, also regulates mitophagy⁸⁶, with torpor-responsive gene expression occurring in several lineages of hibernators^{87–89} (**Table S9**).

We did a second, distinct analysis of hibernation by applying RERconverge⁹⁰ to our TOGA alignments. We detected 764 genes with disparate evolutionary rates in 21 deep hibernators versus 146 non-hibernators ($p_{\text{FDR}} < 0.05$; **Data S3**, **Fig. S13**). 511 genes score as slower-evolving in hibernators, and 253 as more quickly evolving. The duration and depth of hibernation are highly variable across mammals, suggesting evolutionary honing of the hibernation phenotype in different lineages. Consistent with this, both fast and slow evolving genes overlap 6,505 genes identified as torpor-modified in hibernator brain⁸⁷

Genes evolving more quickly in hibernators are enriched in GO pathways relating to synaptic transmission; neuroprotection in the cold is well-documented for hibernator brains, as is their ability to rapidly remodel synapses upon rewarming⁹¹. Genes evolving more slowly in hibernators are enriched in GO pathways related to DNA repair, suggesting that an effective

response to DNA damage may be essential to successful hibernation⁹². In both gene lists, metabolic processes are an over-represented pathway, consistent with the importance of metabolic regulation in hibernation^{85,93}.

Regulatory evolution

We developed a forward genomics toolkit for associating *cis*-regulatory elements with specific phenotypes (detailed in ⁹⁴). This Tissue-Aware Conservation Inference Toolkit (TACIT) uses open chromatin sequence data available for a few species to predict open chromatin (a proxy for *cis*-regulatory element activity) in many species at open chromatin region orthologs found by the Cactus alignment(**Fig. 5A**)⁹⁵. TACIT then associates open chromatin predictions with phenotype annotations. By using open chromatin predictions, TACIT does not require tissue-specific open chromatin data in each species, which is costly and logistically challenging to obtain.

We applied TACIT to neural traits with substantial species-level phenotype data: brain size (158 species) and vocal learning (181 species). We first predicted open chromatin using models trained on data from different parts of the brain and then tested for associations. For brain size, we found 34 motor cortex (**Fig. 5B**) and 13 parvalbumin neuron associations. Motor cortex open chromatin regions near microcephaly or macrocephaly genes had stronger associations (1-sided $p_{\text{Wilcoxon}}=0.0073$).

For vocal learning, we tested for both regulatory and coding changes using TACIT and RERconverge (detailed in ⁹⁷). Vocal learning is the ability to mimic non-innate sounds, and it likely evolved convergently in humans, bats, cetaceans, and pinnipeds⁹⁸. Using TACIT, we associated open chromatin regions near *GALC*⁹⁹(**Fig. 5C**) and other speech disorder-related genes with vocal learning. This included a region near *TSHZ3*, a transcription factor whose disruption results in absence or delay of human speech¹⁰⁰.

Biodiversity

We built a new tool to identify species in urgent need of conservation using genomic features correlated to elevated risk of extinction (detailed in ¹⁰¹). While no single genomic summary statistic is diagnostic of conservation classification by the International Union for Conservation of Nature (IUCN), threatened species have, on average, lower heterozygosity (phylogenetic regression $p=0.016$)¹⁶, smaller long-term effective population sizes (N_e)($p<3.30\times 10^{-5}$), larger recent declines in population size (ratio of historic N_e to contemporary population size; $p=0.012$), higher proportions of missense mutations ($p=3.63\times 10^{-4}$), and fewer mutations at extremely conserved sites, as measured by kurtosis of the phyloP distribution($p=0.004$; **Fig. 6**). The predictive value of N_e is especially striking given that historic N_e reflects population dynamics more than 10,000 years ago. Both long-term population history and contemporary declines appear to imperil species survival¹⁰¹.

Applying predictive models using genomic metrics to species lacking IUCN classifications, we assess the Upper Galilee Mountains blind mole rat as unlikely to be threatened (12-14% probability). The risk is higher for Java lesser chevrotain (27-63%) and for killer whale (53-69%), with several populations already listed as endangered¹⁰². Our findings suggest a single reference genome, an increasingly achievable resource¹⁰³, may be valuable for identifying species in need of further study and immediate intervention.

Conclusion

In 2011, the landmark 29 Mammals paper demonstrated that evolutionary constraint is a powerful indicator of function³² and estimated that sequencing a few hundred placental mammals would achieve single-nucleotide resolution of constraint across the human genome. Zoonomia realizes that vision. In combination with existing resources such as ENCODE³⁸ and TOPMed³⁹ and new technologies including regulatory genomic assays^{94,97}, chromatin conformation capture^{27,51}, and massively parallel reporter assays³⁶, Zoonomia's alignments of hundreds of placental species address fundamental questions in mammalian evolution.

Zoonomia addresses the central goal of medical genomics: to identify genetic variants that influence disease risk and understand their biological mechanisms¹⁰⁴. Constraint scores confirm strong conservation of protein-coding sequences and discern subtle correlation to other functions, like transcription factor binding. Unsurprisingly, *HOX* gene clusters are exceptionally constrained. More surprisingly, constraint is nearly as strong in gene deserts neighboring developmental genes. Constrained regions overall, including those from Zoonomia's updated list of ultraconserved elements and new list of UNICORNs, are strongly enriched for transcriptional regulation and developmental genes. Even in regions of only modest nucleotide-level conservation, our alignments help predict activity conservation at orthologs via prediction of tissue and cell type-specific *cis*-regulatory elements across species. Hundreds of other regions, mainly non-coding, are accelerated, especially in primates and humans⁵¹. They are enriched for topologically associated domains containing human-specific structural variants, perhaps reflecting adaptation via altered 3D genome structure.

Zoonomia marks a new era in non-human and comparative genomics while underscoring the critical need for high-quality mammalian phenotypes. Our alignments facilitate annotation of coding, regulatory, and functionally constrained regions in hundreds of species. Forward genomic approaches can find genes and regulatory elements underlying traits that, while medically relevant, are difficult to study in humans because our species is either not exceptional (e.g., hibernation and olfaction) or has minimal natural variation (e.g., brain size and vocal learning). Using just the 27% of Zoonomia species for which anatomic turbinal-count data are available, we discovered a strong correlation with olfactory receptor gene counts. Even so, achieving the richer data sets needed to study other phenotypes, evaluate pattern robustness, and address broader prospects — for example, the possibility of predicting genomic features from fossils⁹ — will require close collaborations between genomics researchers and scientists with

expertise in morphology, physiology, and behavior to prioritize species for paired sequencing and phenotyping.

Comparative genomics projects are classically motivated by the potential to advance human biomedicine, all while relying on biodiversity imperiled by human activity¹⁰⁵. We show that even a single reference genome per species, the core data type for comparative genomics, can support biodiversity conservation by guiding extinction risk assessment¹⁰¹. With genome assemblies increasingly accessible¹⁰³, conservation scientists will be better equipped to identify potentially threatened populations earlier, when management efforts are more effective.

Through close and enduring partnerships among researchers in genomics, morphology, behavior, and conservation, resources from Zoonomia and other comparative-genomics projects can simultaneously address questions in human health and trait mapping while guiding conservation of the biodiversity that is essential to these discoveries.

Acknowledgments:

We thank everyone who collected the samples and phenotype data essential for this project. We especially thank the San Diego Zoo Wildlife Alliance staff, E. Baitchman, R. Johnston, and Zoo New England, Broad Institute Genomics Platform, the SNP & SEQ Technology Platform (National Genomics Infrastructure Sweden and Science for Life Laboratory), and the Texas A&M High Performance Research Computing Center. Species silhouettes were sourced from PhyloPic.

Data and materials availability:

Zoonomia data resources are available through the UCSC genome browser, NCBI, and github. Data and materials for companion papers can be found within the specific papers.

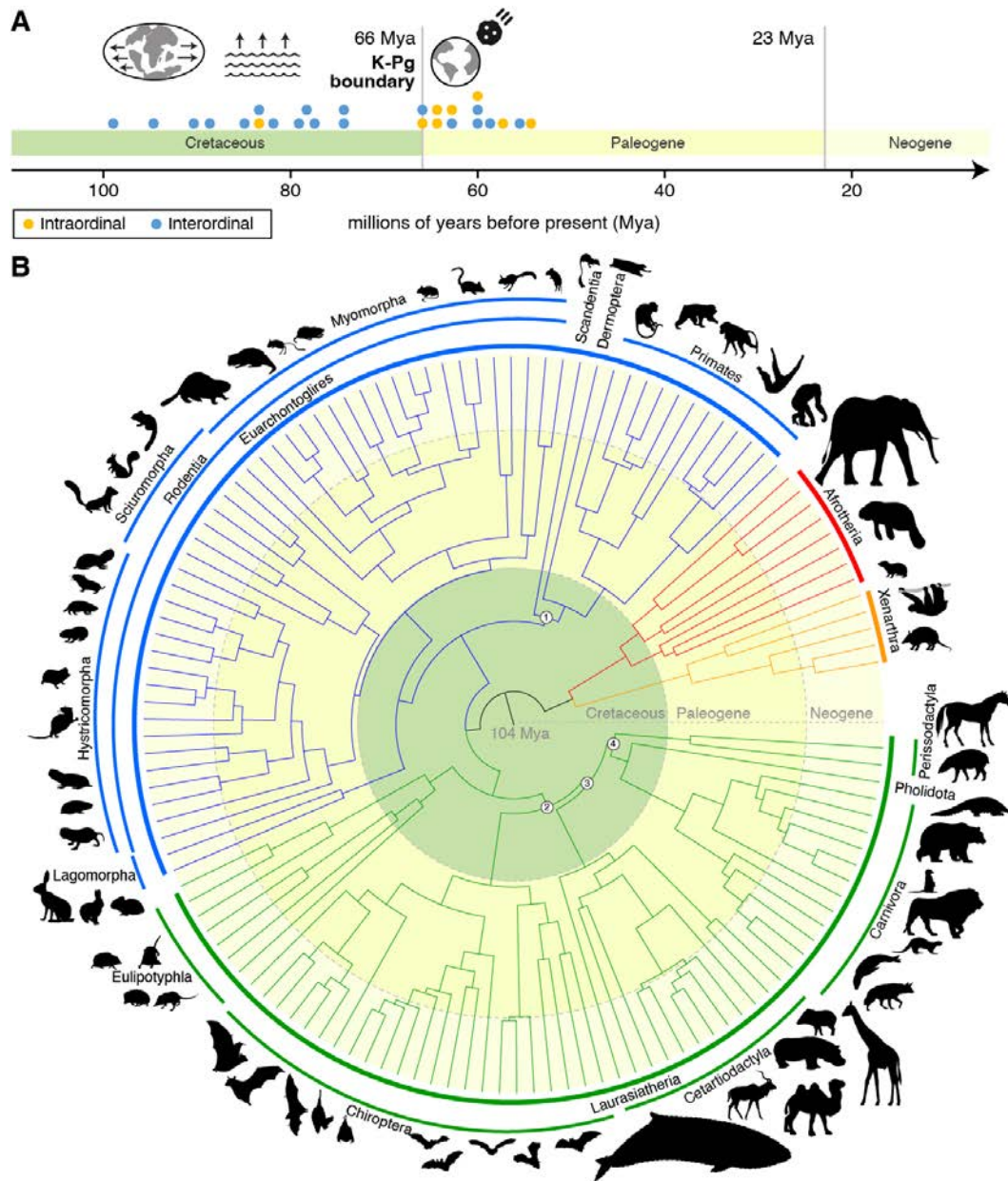


Figure 1: New placental mammal phylogeny supports Long Fuse model of diversification.

(A) The majority of interordinal diversification (blue circles) occurred in the Cretaceous, coincident with continental fragmentation and sea level changes. A second pulse of intra-ordinal diversification (yellow circles) occurred after the mass extinction event at the K-Pg boundary. (B) A phylogeny based on divergence times estimated using ~460kb of near-neutrally evolving sequence for 240 species resolves recalcitrant relationships in the placental mammal phylogeny (black numbers in white circles) including (1) Euarchonta (primates, colugos, and treeshrews); (2) Scrotifera (bats, cetartiodactyls, perissodactyls, carnivorans, and pangolins); (3) Feruungulata (all scrotiferans excluding bats); and (4) Ferae (carnivorans and pangolins) + Perissodactyla.

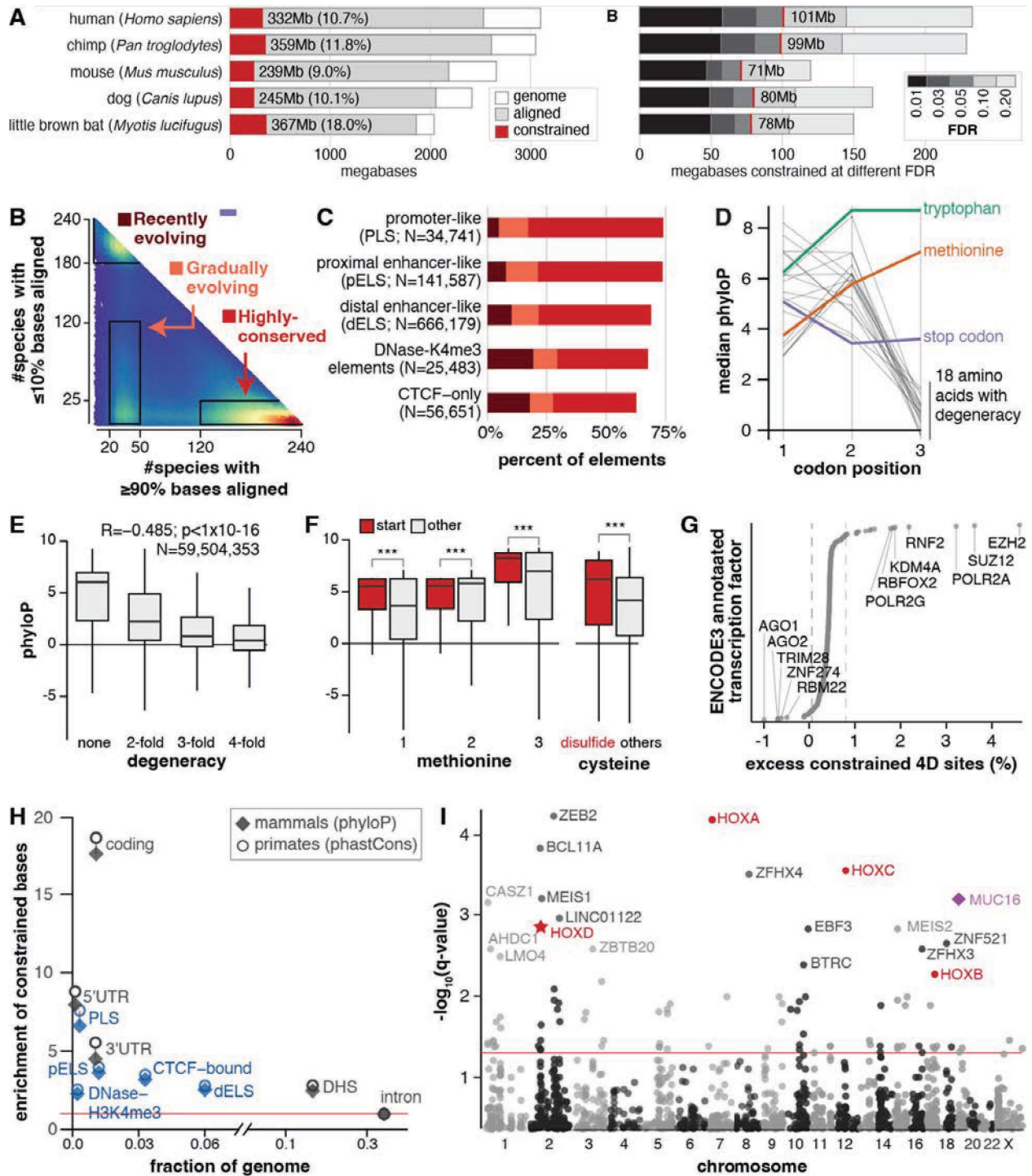


Figure 2: Comparing 240 species resolves mammalian constraint to single bases.

(A) We estimated a lower-bound on the proportion of the genome under purifying selection (constraint) by comparing genome-wide phyloP to putatively neutrally evolving ancestral repeats and identified single bases as significantly constrained at varying FDR thresholds. (B) At 924,641 human cCREs³⁸, we compared the number of species with poor alignments (y-axis) to the number with good alignments (x-axis) and found three clusters that are (C) non-randomly distributed across cCRE types. (D) Non-degenerate codons have higher phyloP scores. (E)

Non-degenerate coding positions are most constrained; phyloP scores fall with increasing degeneracy. **(F)** Coding positions are more constrained if they function as start sites or form disulfide bridges (***) $p_{\text{Wilcoxon}} < 1 \times 10^{-16}$. In E and F, the box represents 25%/75% quartiles, the horizontal line the median, and the vertical line the 5-95% quartiles. **(G)** Coding positions that are TFBS for some transcription factors are enriched for constraint. Dotted grey lines=5th and 95th percentiles. **(H)** Functional regions are enriched for constraint. cCREs are in blue. Red line: no enrichment. **(I)** The most constrained 100kb bins include all four HOX clusters (red text), with *HOXD* (star) overlapping the longest synteny block shared across mammals⁵¹. One bin (purple diamond) significantly lacks constraint and contains *MUC16*, a mucin secreted in innate immune responses. Red line: $q=0.05$. Bins with $q < 0.006$ are labelled.

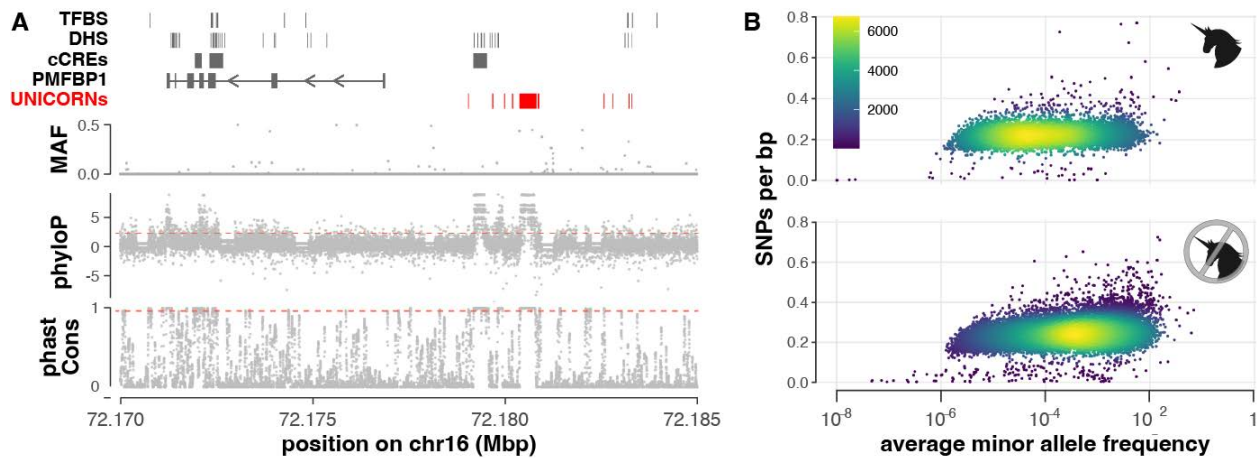


Figure 4: Constraint highlights unannotated regions that are likely functional.

(A) Example UNICORNs (red) on human chromosome 16, the largest of which is 418 bp and located 3.5kb upstream of *PMFBP1*'s transcription start site. Bottom two tracks: mammal phyloP scores and primate phastCons scores. Grey dots represent single bases. Red dotted lines represent 5% FDR threshold for phyloP ($\text{phyloP} > 2.270$) and threshold for phastCons capturing equivalent genome proportion ($\text{phastCons base score} \geq 0.961$). UNICORNs lack coding or regulatory annotations (top track) and have low diversity in human populations (second track and B). (B) Number of SNPs per base pair (0-1) against mean minor allele frequency for human SNPs within UNICORNs (top) or within a random set of unannotated sequences (bottom). Allele frequencies were \log_{10} transformed. UNICORNs contain fewer variants, and those have lower allele frequencies than in the random set (Wilcoxon rank sum test, $p < 2.2 \times 10^{-16}$). Human variants and allele frequencies obtained from TOPMed data freeze 8³⁹.

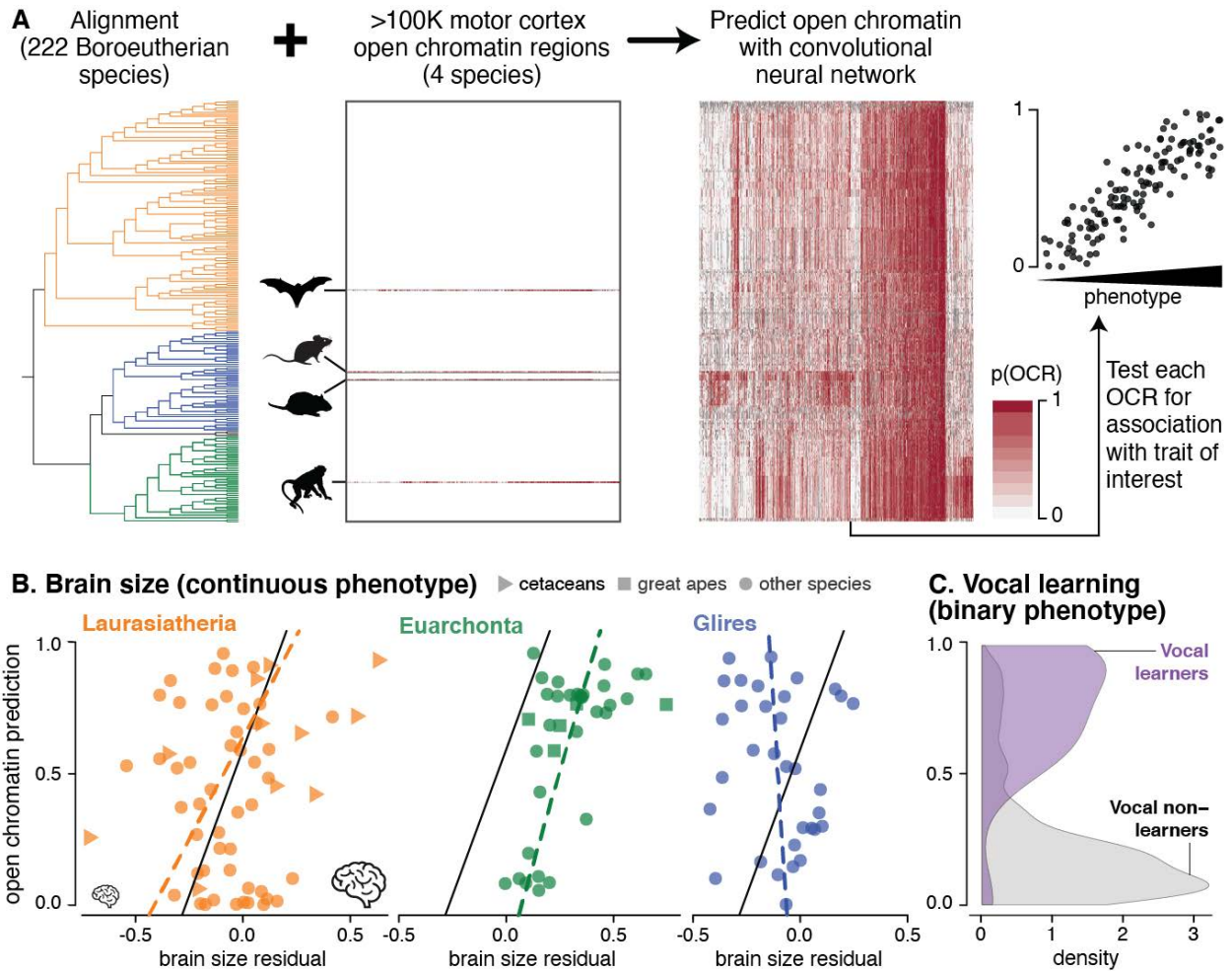


Figure 5: Forward genomics associates genomic differences with brain size and vocal learning.

(A) TACIT trains a convolutional neural network on open chromatin regions in a few species and then predicts open chromatin in specific tissues or cell types in many others and screens for phenotype associations. (B) TACIT associates motor cortex open chromatin with brain size (a quantitative trait), and here we show results for the rhesus macaque open chromatin region chr10:48660711-48661679 in the *MACROD2* locus, which is implicated in nervous system development¹⁰⁶. The solid line shows the phylolm line of best fit (coeff. = 0.451, permutations $p_{FDR} = 0.0375$). The dashed lines show phylolm lines of best fit for each clade as a visual aid. (C) We also applied TACIT to associate motor cortex open chromatin with vocal learning, a binary trait, and here we show results for the Egyptian fruit bat open chromatin region PVIL01002568.1:139004-139596 in the *GALC* locus (permutations $p_{FDR} = 0.0376$), which was also associated with open chromatin specifically active in the orofacial motor cortex of bat⁹⁷ and which has previously been associated with severe speech and motor disabilities in humans⁹⁹.

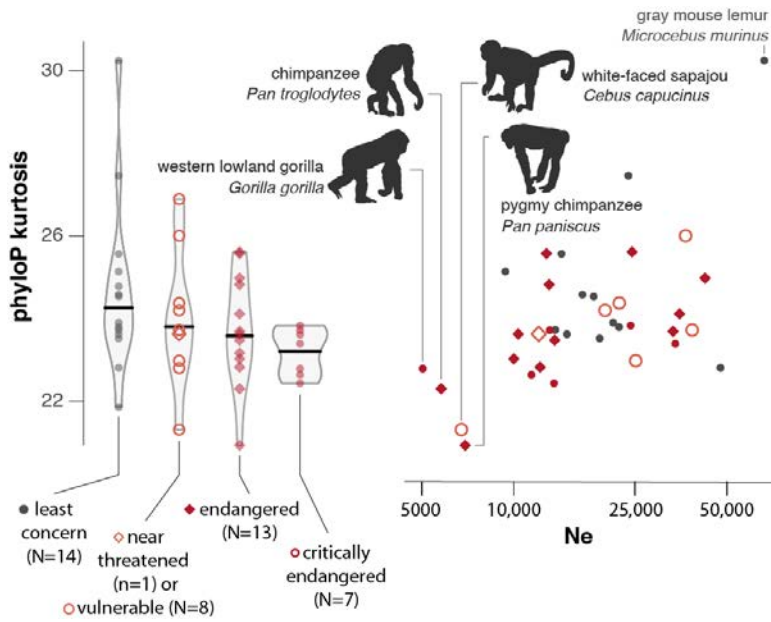


Figure 6. Genomic metrics distinguish at-risk species.

Primates categorized at increasing levels of extinction risk tend to have fewer mutations at sites with extreme phyloP scores, measured as phyloP kurtosis, and smaller effective population sizes (N_e). The four species (all at risk) with the lowest N_e and the species with the highest N_e are labeled. Screening several genome metrics in parallel provides initial assessment of extinction risk that can help prioritize species for deeper assessment¹⁰¹.

Consortium list:

Gregory Andrews¹, Joel C. Armstrong², Matteo Bianchi³, Bruce W. Birren⁴, Kevin R. Bredemeyer⁵, Ana M. Breit⁶, Matthew J. Christmas³, Hiram Clawson², Joana Damas⁷, Federica Di Palma^{8,9}, Mark Diekhans², Michael X. Dong³, Eduardo Eizirik¹⁰, Kaili Fan¹, Cornelia Fanter¹¹, Nicole M. Foley⁵, Karin Forsberg-Nilsson^{12,13}, Carlos J. Garcia¹⁴, John Gatesy¹⁵, Steven Gazal¹⁶, Diane P. Genereux⁴, Linda Goodman¹⁷, Jenna Grimshaw¹⁴, Michaela K. Halsey¹⁴, Andrew J. Harris⁵, Glenn Hickey¹⁸, Michael Hiller^{19,20,21}, Allyson G. Hindle¹¹, Robert M. Hubley²², Graham M. Hughes²³, Jeremy Johnson⁴, David Juan²⁴, Irene M. Kaplow^{25,26}, Elinor K. Karlsson^{1,4,27}, Kathleen C. Keough^{17,28,29}, Bogdan Kirilenko^{19,20,21}, Klaus-Peter Koepfli^{30,31,32}, Jennifer M. Korstian¹⁴, Amanda Kowalczyk^{25,26}, Sergey V. Kozyrev³, Alyssa J. Lawler^{4,26,33}, Colleen Lawless²³, Thomas Lehmann³⁴, Danielle L. Levesque⁶, Harris A. Lewin^{7,35,36}, Xue Li^{1,4,37}, Abigail Lind³⁸, Kerstin Lindblad-Toh^{3,4}, Ava Mackay-Smith³⁹, Voichita D. Marinescu³, Tomas Marques-Bonet^{40,41,42,43}, Victor C. Mason⁴⁴, Jennifer R. S. Meadows³, Wynn K. Meyer⁴⁵, Jill E. Moore¹, Lucas R. Moreira^{1,4}, Diana D. Moreno-Santillan¹⁴, Kathleen M. Morrill^{1,4,37}, Gerard Muntané²⁴, William J. Murphy⁵, Arcadi Navarro^{40,42,46,47}, Martin Nweeia^{48,49,50,51}, Sylvia Ortmann⁵², Austin Osmanski¹⁴, Benedict Paten², Nicole S. Paulat¹⁴, Andreas R. Pfenning^{25,26}, BaDoi N. Phan^{25,26,53}, Katherine S. Pollard^{28,29,54}, Henry E. Pratt¹, David A. Ray¹⁴, Steven K. Reilly³⁹, Jeb R. Rosen²², Irina Ruf⁵⁵, Louise Ryan²³, Oliver A. Ryder^{56,57}, Pardis C. Sabeti^{4,58,59}, Daniel E. Schäffer²⁵, Aitor Serres²⁴, Beth Shapiro^{60,61}, Arian F. A. Smit²², Mark Springer⁶², Chaitanya Srinivasan²⁵, Cynthia Steiner⁵⁶, Jessica M. Storer²², Kevin A. M. Sullivan¹⁴, Patrick F. Sullivan^{63,64}, Elisabeth Sundström³, Megan A. Supple⁶⁰, Ross Swofford⁴, Joy-El Talbot⁶⁵, Emma Teeling²³, Jason Turner-Maier⁴, Alejandro Valenzuela²⁴, Franziska Wagner⁶⁶, Ola Wallerman³, Chao Wang³, Juehan Wang¹⁶, Zhiping Weng¹, Aryn P. Wilder⁵⁶, Morgan E. Wirthlin^{25,26,67}, James R. Xue^{4,58}, Xiaomeng Zhang^{4,25,26}

Affiliations:

¹Program in Bioinformatics and Integrative Biology, UMass Chan Medical School; Worcester, MA 01605, USA.

²Genomics Institute, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

³Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University; Uppsala, 751 32, Sweden.

⁴Broad Institute of MIT and Harvard; Cambridge, MA 02139, USA.

⁵Veterinary Integrative Biosciences, Texas A&M University; College Station, TX 77843, USA.

⁶School of Biology and Ecology, University of Maine; Orono, ME 04469, USA.

⁷The Genome Center, University of California Davis; Davis, CA 95616, USA.

⁸Genome British Columbia; Vancouver, BC, Canada.

⁹School of Biological Sciences, University of East Anglia; Norwich, UK.

¹⁰School of Health and Life Sciences, Pontifical Catholic University of Rio Grande do Sul; Porto Alegre, 90619-900, Brazil.

¹¹School of Life Sciences, University of Nevada Las Vegas; Las Vegas, NV 89154, USA.

¹²Biodiscovery Institute, University of Nottingham; Nottingham, UK.

¹³Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University; Uppsala, 751 85, Sweden.

¹⁴Department of Biological Sciences, Texas Tech University; Lubbock, TX 79409, USA.

¹⁵Division of Vertebrate Zoology, American Museum of Natural History; New York, NY 10024, USA.

¹⁶Keck School of Medicine, University of Southern California; Los Angeles, CA 90033, USA.

¹⁷Fauna Bio Incorporated; Emeryville, CA 94608, USA.

¹⁸Baskin School of Engineering, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

¹⁹Faculty of Biosciences, Goethe-University; 60438 Frankfurt, Germany.

²⁰LOEWE Centre for Translational Biodiversity Genomics; 60325 Frankfurt, Germany.

²¹Senckenberg Research Institute; 60325 Frankfurt, Germany.

²²Institute for Systems Biology; Seattle, WA 98109, USA.

²³School of Biology and Environmental Science, University College Dublin; Belfield, Dublin 4, Ireland.

²⁴Department of Experimental and Health Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra; Barcelona, 08003, Spain.

²⁵Department of Computational Biology, School of Computer Science, Carnegie Mellon University; Pittsburgh, PA 15213, USA.

²⁶Neuroscience Institute, Carnegie Mellon University; Pittsburgh, PA 15213, USA.

²⁷Program in Molecular Medicine, UMass Chan Medical School; Worcester, MA 01605, USA.

²⁸Department of Epidemiology & Biostatistics, University of California San Francisco; San Francisco, CA 94158, USA.

²⁹Gladstone Institutes; San Francisco, CA 94158, USA.

³⁰Center for Species Survival, Smithsonian's National Zoo and Conservation Biology Institute; Washington, DC 20008, USA.

³¹Computer Technologies Laboratory, ITMO University; St. Petersburg 197101, Russia.

³²Smithsonian-Mason School of Conservation, George Mason University; Front Royal, VA 22630, USA.

³³Department of Biological Sciences, Mellon College of Science, Carnegie Mellon University; Pittsburgh, PA 15213, USA.

³⁴Senckenberg Research Institute and Natural History Museum Frankfurt; 60325 Frankfurt am Main, Germany.

³⁵Department of Evolution and Ecology, University of California Davis; Davis, CA 95616, USA.

³⁶John Muir Institute for the Environment, University of California Davis; Davis, CA 95616, USA.

³⁷Morningside Graduate School of Biomedical Sciences, UMass Chan Medical School; Worcester, MA 01605, USA.

³⁸NA.

³⁹Department of Genetics, Yale School of Medicine; New Haven, CT 06510, USA.

⁴⁰Catalan Institution of Research and Advanced Studies (ICREA); Barcelona, 08010, Spain.

⁴¹CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST); Barcelona, 08036, Spain.

⁴²Department of Medicine and Life Sciences, Institute of Evolutionary Biology (UPF-CSIC), Universitat Pompeu Fabra; Barcelona, 08003, Spain.

⁴³Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona; 08193, Cerdanyola del Vallès, Barcelona, Spain.

⁴⁴Institute of Cell Biology, University of Bern; 3012, Bern, Switzerland.

⁴⁵Department of Biological Sciences, Lehigh University; Bethlehem, PA 18015, USA.

⁴⁶BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation; Barcelona, 08005, Spain.

⁴⁷CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology (BIST); Barcelona, 08003, Spain.

⁴⁸Department of Comprehensive Care, School of Dental Medicine, Case Western Reserve University; Cleveland, OH 44106, USA.

⁴⁹Department of Vertebrate Zoology, Canadian Museum of Nature; Ottawa, Ontario K2P 2R1, Canada.

⁵⁰Department of Vertebrate Zoology, Smithsonian Institution; Washington, DC 20002, USA.

⁵¹Narwhal Genome Initiative, Department of Restorative Dentistry and Biomaterials Sciences, Harvard School of Dental Medicine; Boston, MA 02115, USA.

⁵²Department of Evolutionary Ecology, Leibniz Institute for Zoo and Wildlife Research; 10315 Berlin, Germany.

⁵³Medical Scientist Training Program, University of Pittsburgh School of Medicine; Pittsburgh, PA 15261, USA.

⁵⁴Chan Zuckerberg Biohub; San Francisco, CA 94158, USA.

⁵⁵Division of Messel Research and Mammalogy, Senckenberg Research Institute and Natural History Museum Frankfurt; 60325 Frankfurt am Main, Germany.

⁵⁶Conservation Genetics, San Diego Zoo Wildlife Alliance; Escondido, CA 92027, USA.

⁵⁷Department of Evolution, Behavior and Ecology, School of Biological Sciences, University of California San Diego; La Jolla, CA 92039, USA.

⁵⁸Department of Organismic and Evolutionary Biology, Harvard University; Cambridge, MA 02138, USA.

⁵⁹Howard Hughes Medical Institute; Chevy Chase, MD, USA.

⁶⁰Department of Ecology and Evolutionary Biology, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

⁶¹Howard Hughes Medical Institute, University of California Santa Cruz; Santa Cruz, CA 95064, USA.

⁶²Department of Evolution, Ecology and Organismal Biology, University of California Riverside; Riverside, CA 92521, USA.

⁶³Department of Genetics, University of North Carolina Medical School; Chapel Hill, NC 27599, USA.

⁶⁴Department of Medical Epidemiology and Biostatistics, Karolinska Institutet; Stockholm, Sweden.

⁶⁵Iris Data Solutions, LLC; Orono, ME 04473, USA.

⁶⁶Museum of Zoology, Senckenberg Natural History Collections Dresden; 01109 Dresden, Germany.

⁶⁷Allen Institute for Brain Science; Seattle, WA 98109, USA

References

1. Jones, K. E. & Safi, K. Ecology and evolution of mammalian biodiversity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2451–2461 (2011).
2. Jones, K. E. *et al.* PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals: Ecological Archives E090-184. *Ecology* **90**, 2648–2648 (2009).
3. Davies, T. W., Bell, M. A., Goswami, A. & Halliday, T. J. D. Completeness of the eutherian mammal fossil record and implications for reconstructing mammal evolution through the Cretaceous/Paleogene mass extinction. *Paleobiology* **43**, 521–536 (2017).
4. Springer, M. S., Foley, N. M., Brady, P. L., Gatesy, J. & Murphy, W. J. Evolutionary Models for the Diversification of Placental Mammals Across the KPg Boundary. *Front. Genet.* **10**, 1241 (2019).
5. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
6. Kazazian, H. H., Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
7. Feulner, P. G. D. & De-Kayne, R. Genome evolution, structural rearrangements and speciation. *J. Evol. Biol.* **30**, 1488–1490 (2017).
8. Hiller, M. *et al.* A ‘forward genomics’ approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep.* **2**, 817–823 (2012).
9. Wagner, F. *et al.* Reconstruction of evolutionary changes in fat and toxin consumption reveals associations with gene losses in mammals: a case study for the lipase inhibitor PNLIPRP1 and the xenobiotic receptor NR1I3. *Journal of Evolutionary Biology*.
10. Marcovitz, A. *et al.* A functional enrichment test for molecular convergent evolution finds a

- clear protein-coding signal in echolocating bats and whales. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 21094–21103 (2019).
11. Partha, R. *et al.* Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *Elife* **6**, (2017).
12. Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015).
13. Prudent, X., Parra, G., Schwede, P., Roscito, J. G. & Hiller, M. Controlling for Phylogenetic Relatedness and Evolutionary Rates Improves the Discovery of Associations Between Species' Phenotypic and Genomic Differences. *Mol. Biol. Evol.* **33**, 2135–2150 (2016).
14. Armstrong, J. *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
15. Conservation Track Settings. <http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=cons100way>.
16. Zoonomia Consortium. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587**, 240–245 (2020).
17. Meadows, J. *et al.* Leveraging Base Pair Mammalian Constraint to Understand Genetic Variation and Human Disease. *Science*.
18. Kirilenko, B. M. *et al.* TOGA integrates gene annotation with orthology inference at scale. *Science*.
19. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
20. Foley, N. M. *et al.* A genomic timescale for placental mammal evolution. *Science*.
21. Meredith, R. W. *et al.* Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011).
22. O'Leary, M. A. *et al.* The Placental Mammal Ancestor and the Post-K-Pg Radiation of

- Placentals. *Science* **339**, 662–667 (2013).
23. Esselstyn, J. A., Oliveros, C. H., Swanson, M. T. & Faircloth, B. C. Investigating Difficult Nodes in the Placental Mammal Tree with Expanded Taxon Sampling and Thousands of Ultraconserved Elements. *Genome Biol. Evol.* **9**, 2308–2321 (2017).
24. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**, 226–229 (1996).
25. Eizirik, E., Murphy, W. J. & O’Brien, S. J. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* **92**, 212–219 (2001).
26. Archibald, J. D. & Deutschman, D. H. *J. Mamm. Evol.* **8**, 107–124 (2001).
27. Damas, J. & Lewin, H. High-resolution reconstruction and evolution of the mammalian ancestor karyotype. *PNAS?*
28. Osmanski, A. B. *et al.* Insights into mammalian TE diversity via the curation of 200+ mammalian genome assemblies. *Science*.
29. Kapusta, A., Suh, A. & Feschotte, C. Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1460–E1469 (2017).
30. Yang, L., Scott, L. & Wichman, H. A. Tracing the history of LINE and SINE extinction in sigmodontine rodents. *Mob. DNA* **10**, 22 (2019).
31. Paulat, N. S. *et al.* Chiropterans are a hotspot for horizontal transfer of DNA transposons in Mammalia. *Science*.
32. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
33. Ponting, C. P. & Hardison, R. C. What fraction of the human genome is functional? *Genome Res.* **21**, 1769–1776 (2011).
34. Rands, C. M., Meader, S., Ponting, C. P. & Lunter, G. 8.2% of the Human genome is

- constrained: variation in rates of turnover across functional element classes in the human lineage. *PLoS Genet.* **10**, e1004525 (2014).
35. Graur, D. An Upper Limit on the Functional Fraction of the Human Genome. *Genome Biol. Evol.* **9**, 1880–1885 (2017).
36. Xue, J. R. *et al.* The functional and evolutionary impacts of human-specific deletions in conserved elements. *Science*.
37. Andrews, G., Fan, K., Pratt, H. & Weng, Z. Landscapes of Human cis-regulatory Elements and Transcription Factor Binding Sites Evolutionarily Constrained in the Mammalian Lineage. *Science*.
38. ENCODE Project Consortium *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
39. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
40. Wang, J. *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812 (2012).
41. Bejerano, G. *et al.* Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004).
42. Ma, A. C. H. *et al.* Methionine aminopeptidase 2 is required for HSC initiation and proliferation. *Blood* **118**, 5448–5457 (2011).
43. Markenscoff-Papadimitriou, E. *et al.* A Chromatin Accessibility Atlas of the Developing Human Telencephalon. *Cell* **182**, 754–769.e18 (2020).
44. Bakken, T. E. *et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111–119 (2021).
45. Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome*

- Res.* **28**, 1243–1252 (2018).
46. Fotsing, S. F. *et al.* The impact of short tandem repeat variation on gene expression. *Nat. Genet.* **51**, 1652–1659 (2019).
47. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368 (2007).
48. Hayward, A., Ghazal, A., Andersson, G., Andersson, L. & Jern, P. ZBED evolution: repeated utilization of DNA transposons as regulators of diverse host functions. *PLoS One* **8**, e59940 (2013).
49. Cechova, M. *et al.* High satellite repeat turnover in great apes studied with short- and long-read technologies. Preprint at <https://doi.org/10.1101/470054>.
50. Trizzino, M. *et al.* Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* **27**, 1623–1633 (2017).
51. Keough, K. C. *et al.* Human-specific evolution at multiple scales and dimensions. *Science*.
52. McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
53. Hubisz, M. J. & Pollard, K. S. Exploring the genesis and functions of Human Accelerated Regions sheds light on their role in human evolution. *Curr. Opin. Genet. Dev.* **29**, 15–21 (2014).
54. Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L. R. & Pollard, K. S. Many human accelerated regions are developmental enhancers. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130025 (2013).
55. Kostka, D., Hubisz, M. J., Siepel, A. & Pollard, K. S. The role of GC-biased gene conversion in shaping the fastest evolving regions of the human genome. *Mol. Biol. Evol.* **29**, 1047–1057 (2012).

56. Kronenberg, Z. N. *et al.* High-resolution comparative analysis of great ape genomes. *Science* **360**, (2018).
57. Howe, K. L. *et al.* Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).
58. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
59. Chou, H.-H. *et al.* Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11736–11741 (2002).
60. Hayakawa, T., Aki, I., Varki, A., Satta, Y. & Takahata, N. Fixation of the human-specific CMP-N-acetylneuraminic acid hydroxylase pseudogene and implications of haplotype diversity for human evolution. *Genetics* **172**, 1139–1146 (2006).
61. Irie, A., Koyama, S., Kozutsumi, Y., Kawasaki, T. & Suzuki, A. The molecular basis for the absence of N-glycolylneuraminic acid in humans. *J. Biol. Chem.* **273**, 15866–15871 (1998).
62. Peri, S., Kulkarni, A., Feyertag, F., Berninsone, P. M. & Alvarez-Ponce, D. Phylogenetic Distribution of CMP-Neu5Ac Hydroxylase (CMAH), the Enzyme Synthetizing the Proinflammatory Human Xenoantigen Neu5Gc. *Genome Biol. Evol.* **10**, 207–219 (2018).
63. Dankwa, S. *et al.* Ancient human sialic acid variant restricts an emerging zoonotic malaria parasite. *Nat. Commun.* **7**, 11187 (2016).
64. Li, W. *et al.* Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E8508–E8517 (2017).
65. Ghaderi, D. *et al.* Sexual selection by female immunity against paternal antigens can fix loss of function alleles. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17743–17748 (2011).

66. Breiman, A., Ruvén-Clouet, N. & Le Pendu, J. Harnessing the natural anti-glycan immune response to limit the transmission of enveloped viruses such as SARS-CoV-2. *PLoS Pathog.* **16**, e1008556 (2020).
67. Varki, N. M. & Varki, A. Diversity in cell surface sialic acid presentations: implications for biology and disease. *Lab. Invest.* **87**, 851–857 (2007).
68. Galili, U. Evolution in primates by ‘Catastrophic-selection’ interplay between enveloped virus epidemics, mutated genes of enzymes synthesizing carbohydrate antigens, and natural anti-carbohydrate antibodies. *Am. J. Phys. Anthropol.* **168**, 352–363 (2019).
69. Kuchipudi, S. V. *et al.* Multiple spillovers and onward transmission of SARS-Cov-2 in free-living and captive White-tailed deer (*Odocoileus virginianus*). *bioRxiv* 2021.10.31.466677 (2021) doi:10.1101/2021.10.31.466677.
70. Carlson, C. J. *et al.* The future of zoonotic risk prediction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200358 (2021).
71. Hughes, G. M. *et al.* The Birth and Death of Olfactory Receptor Gene Families in Mammalian Niche Adaptation. *Mol. Biol. Evol.* **35**, 1390–1406 (2018).
72. Puechmaille, S. J. *et al.* The evolution of sensory divergence in the context of limited gene flow in the bumblebee bat. *Nat. Commun.* **2**, 573 (2011).
73. Liu, A. *et al.* Convergent degeneration of olfactory receptor gene repertoires in marine mammals. *BMC Genomics* **20**, 1–14 (2019).
74. Nyakatura, J. A. The convergent evolution of suspensory posture and locomotion in tree sloths. *J. Mamm. Evol.* **19**, 225–234 (2012).
75. Nyakatura, J. A., Petrovitch, A. & Fischer, M. S. Limb kinematics during locomotion in the two-toed sloth (*Choloepus didactylus*, Xenarthra) and its implications for the evolution of the sloth locomotor apparatus. *Zoology* vol. 113 221–234 Preprint at

<https://doi.org/10.1016/j.zool.2009.11.003> (2010).

76. Billet, G. *et al.* High morphological variation of vestibular system accompanies slow and infrequent locomotion in three-toed sloths. *Proceedings of the Royal Society B: Biological Sciences* vol. 279 3932–3939 Preprint at <https://doi.org/10.1098/rspb.2012.1212> (2012).
77. Presslee, S. *et al.* Palaeoproteomics resolves sloth relationships. *Nat Ecol Evol* **3**, 1121–1130 (2019).
78. Ho, L. & Ané, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397–408 (2014).
79. Saputra, E., Kowalczyk, A., Cusick, L., Clark, N. & Chikina, M. Phylogenetic Permutations: A Statistically Rigorous Approach to Measure Confidence in Associations in a Phylogenetic Context. *Mol. Biol. Evol.* **38**, 3004–3021 (2021).
80. Bird, D. J., Amirkhanian, A., Pang, B. & Van Valkenburgh, B. Quantifying the cribriform plate: influences of allometry, function, and phylogeny in Carnivora. *Anat. Rec.* **297**, 2080–2092 (2014).
81. Martinez, Q. *et al.* Convergent evolution of olfactory and thermoregulatory capacities in small amphibious mammals. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8958–8965 (2020).
82. Larochelle, R. & Baron, G. Comparative morphology and morphometry of the nasal fossae of four species of North American shrews (Soricinae). *Am. J. Anat.* **186**, 306–314 (1989).
83. Lovegrove, B. G. The evolution of endothermy in Cenozoic mammals: a plesiomorphic-apomorphic continuum. *Biol. Rev. Camb. Philos. Soc.* **87**, 128–162 (2012).
84. Lovegrove, B. G., Lobban, K. D. & Levesque, D. L. Mammal survival at the Cretaceous–Palaeogene boundary: metabolic homeostasis in prolonged tropical hibernation in tenrecs. *Proceedings of the Royal Society B: Biological Sciences* **281**, 20141304 (2014).
85. Staples, J. F. Metabolic suppression in mammalian hibernation: the role of mitochondria. *J.*

- Exp. Biol.* **217**, 2032–2036 (2014).
86. Huang, C. *et al.* Thioredoxin interacting protein (TXNIP) regulates tubular autophagy and mitophagy in diabetic nephropathy through the mTOR signaling pathway. *Sci. Rep.* **6**, 29196 (2016).
87. Fu, R. *et al.* Dynamic RNA Regulation in the Brain Underlies Physiological Plasticity in a Hibernating Mammal. *Front. Physiol.* **11**, 624677 (2020).
88. Giroud, S. *et al.* The Torpid State: Recent Advances in Metabolic Adaptations and Protective Mechanisms†. *Front. Physiol.* **11**, 623665 (2020).
89. Sun, H., Wang, J., Xing, Y., Pan, Y.-H. & Mao, X. Gut transcriptomic changes during hibernation in the greater horseshoe bat (*Rhinolophus ferrumequinum*). *Front. Zool.* **17**, 21 (2020).
90. Kowalczyk, A. *et al.* RERconverge: an R package for associating evolutionary rates with convergent traits. *Bioinformatics* **35**, 4815–4817 (2019).
91. Ohe, C. G. von der, von der Ohe, C. G., Garner, C. C., Darian-Smith, C. & Heller, H. C. Synaptic Protein Dynamics in Hibernation. *Journal of Neuroscience* vol. 27 84–92 Preprint at <https://doi.org/10.1523/jneurosci.4385-06.2007> (2007).
92. Cerri, M. *et al.* Hibernation for space travel: Impact on radioprotection. *Life Sci. Space Res.* **11**, 1–9 (2016).
93. Quinones, Q. J., Ma, Q., Zhang, Z., Barnes, B. M. & Podgoreanu, M. V. Organ protective mechanisms common to extremes of physiology: a window through hibernation biology. *Integr. Comp. Biol.* **54**, 497–515 (2014).
94. Kaplow, I. M. *et al.* Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. *Science*.
95. Kaplow, I. M. *et al.* Inferring mammalian tissue-specific regulatory conservation by

- predicting tissue-specific differences in open chromatin. *bioRxiv* 2020.12.04.410795 (2021)
doi:10.1101/2020.12.04.410795.
96. Phan, B. N. *et al.* Fine-mapping candidate neurological disorder-associated regulatory variants using cell type-aware comparative genomic. *Science*.
97. Wirthlin, M. E. *et al.* Vocal learning-associated convergent evolution in mammalian regulatory elements and proteins. *Science*.
98. Janik, V. M. & Slater, P. J. B. Vocal learning in mammals. *Advances in the Study of Behaviour* **26**, 59–100 (1997).
99. Orsini, J. J., Escolar, M. L., Wasserstein, M. P. & Caggana, M. Krabbe Disease. in *GeneReviews*® (eds. Adam, M. P. *et al.*) (University of Washington, Seattle, 2000).
100. Caubit, X. *et al.* TSHZ3 deletion causes an autism syndrome and defects in cortical projection neurons. *Nat. Genet.* **48**, 1359–1369 (2016).
101. Wilder, A. P., Supple, M. A. & Zoonomia consortium. Identifying genomic features of extinction risk to conserve biodiversity using comparative genomics. *Science*.
102. Vélez-Espino, L. A. *et al.* Relative importance of chinook salmon abundance on resident killer whale population growth and viability. *Aquat. Conserv.* **25**, 756–780 (2015).
103. Rhie, A. *et al.* Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).
104. Lappalainen, T. & MacArthur, D. G. From variant to function in human disease genetics. *Science* **373**, 1464–1468 (2021).
105. Ceballos, G. *et al.* Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Science Advances* (2015) doi:10.1126/sciadv.1400253.
106. Ito, H. *et al.* Biochemical and Morphological Characterization of a Neurodevelopmental Disorder-Related Mono-ADP-Ribosylhydrolase, MACRO Domain Containing 2. *Dev.*

Neurosci. **40**, 278–287 (2018).

