PERSPECTIVES

Genomics expands the mammalverse

Diverse mammal genomes open a new portal to hidden aspects of evolutionary history

By **Nathan S. Upham**[1] and **Michael J. Landis**[2]

Mammal genomics has progressed at an uneven pace—half sloth, half cheetah—owing to various technical obstacles, including the complexity of eukaryotic genomes (*1*), difficulties obtaining high-quality DNA from wild animals (*2*), and conflicting evolutionary signatures (*3*). The 2022 completion of the telomere-to-telomere (T2T) human genome assembly was fueled by ultralong-read sequencing techniques only dreamt of two decades ago when the initial draft was published. Generating high-quality genomes across diverse mammal species is now possible, enabling the exploration of tightly packed, regulatory, and repetitive DNA regions. The mammalverse comprises ~6500 living species and >180 million years of genome evolution, ripe for investigation (*4*). On pages 366, 364, 371, 372, 363, and 365 of this issue, Christmas *et al.* (*5*), Kaplow *et al.* (*6*), Osmanski *et al.* (*7*), Wilder *et al.* (*8*), Moon *et al.* (*9*), and Foley *et al.* (*10*), respectively, explore this phylogenomic frontier, using the Zoonomia Consortium's new dataset of 240 species' genomes to investigate molecular-, population-, and species-level changes among placental mammals.

Introduced in Christmas *et al.*, the Zoonomia alignment does not rely on mapping to any single reference genome such as *Homo* or *Mus* and so provides flexibility for estimating evolutionary constraint versus lability across multiple types of structural rearrangements (such as inversions and translocations). To identify constrained genomic regions that have remained unchanged for millions of years, Christmas *et al.* investigated how protein-coding orthologs evolve relative to noncoding regions. Their multispecies analysis found that 3.6 million sites in the human genome are perfectly conserved relative to those of other placentals, far beyond the 191 sites predicted under neutral population-genetic theory assumptions, implicating the pervasive effects of purifying selection in removing damaging mutations. The team estimates that >10.7% of the human genome is evolutionarily constrained, exceeding previous estimates of 3 to 12%. Zoonomia expands the set of ultraconserved elements (here called zooUCEs) sevenfold over those previously available, creating a valuable resource for future evolutionary studies over various time scales.

Notably, nearly half of the conserved sites identified in the Zoonomia dataset fall within regions that are not annotated in the Encyclopedia of DNA Elements (ENCODE) database, meaning that their functions are unknown. To address this gap, Kaplow *et al.* introduced a machine learning method called Tissue-Aware Conservation Inference Tookit (TACIT) to predict when tissue-specific enhancer expression is associated with organismal phenotypes. Enhancers are often found in open chromatin regions of genomes where transcription factors bind and regulate gene expression. Kaplow *et al.* exploit this property to use the open chromatin regions, binding motifs, and known enhancers within the tissues of model species to train models to find similar associations in unannotated genomes. They found enhancer-to-phenotype correlations with brain size and behavior across placentals, including open chromatin regions that are nearby genes associated with human brain–size disorders, implying a possible general mechanism for brain-size evolution. More broadly, TACIT carries promise for uncovering enhancer-phenotype functions across the abundance of newly generated mammal genomes. However, this study also highlights the need for better planning to pair genome and transcriptome sampling with phenotypic data. With this approach, the long-standing goal of untangling the gene regulatory networks that underlie convergently evolved traits (*11*)—for example, the constrained sequences that regulate traits for mammal echolocation and subterranean living—grows closer to realization.

Further exploring the uncharacterized regions within mammal genomes, Osmanski *et al.* studied how transposable elements (TEs) evolve and accumulate over time. TEs are mobile genetic units that are increasingly studied as generators of variation, templates for refunctionalization, and historical records of past evolutionary dynamics. Osmanski *et al.* found that TEs make up 28 to 66% of typical mammalian genome content, with abundance and composition of TE copies varying idiosyncratically among mammal orders and families, but less so within families. Viewing each genome as an "ecosystem" populated by distinct TE types, they found that TE turnover tends to occur successively rather than in all-at-once sweeps, suggesting that TE types dominate briefly before a newer type arises. Notably, Osmanski *et al.* also found that carnivorous diets increased genomic susceptibility to DNA-based TEs, possibly through horizontal transfer from ingested prey or their viruses. Evidence that ecological traits can directly shape genome architecture is a fascinating demonstration of eco-evolutionary feedback.

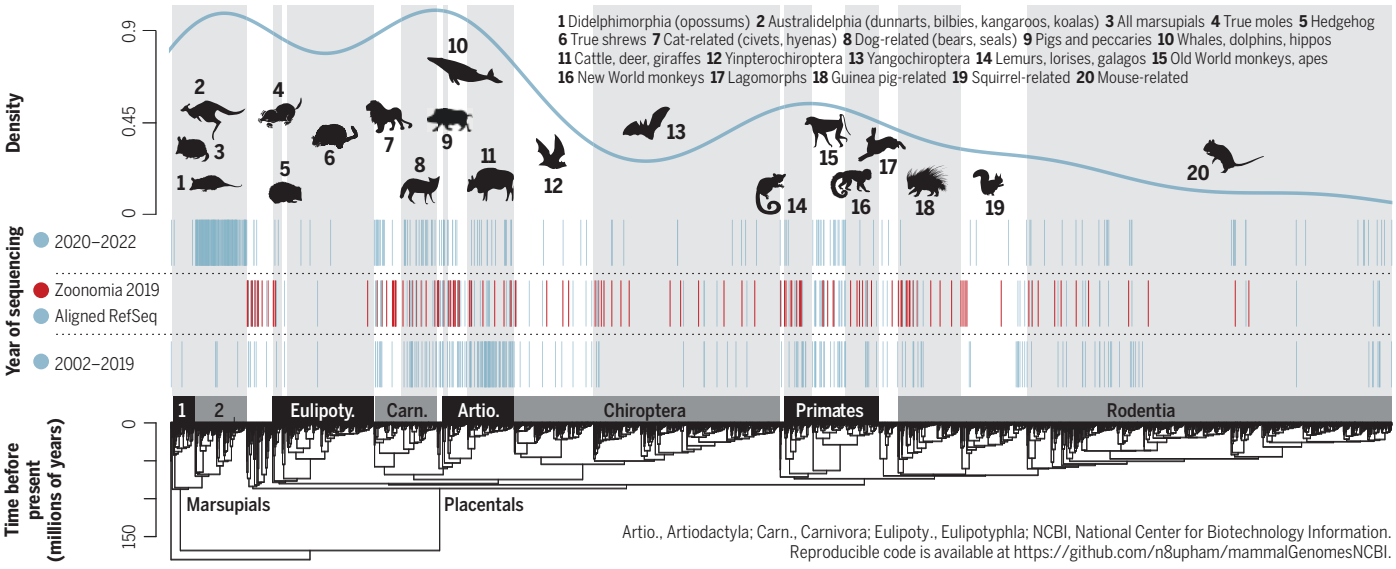
Life-history traits such as generation time are often closely related to effective population size (N_e), a genetic quantity that can contain information about past selection pressures. All else being equal, new mutations experience stronger selection and weaker genetic drift in larger populations, whereas drift outpaces selection in smaller populations, allowing TE insertions and other mutations to accumulate in eukaryotic genomes (*12*). Hence, the genetic variation within a single genome records the historical balance between selection and drift in relation to species life-history traits. Advancing this approach, Wilder *et al.* compared genome-wide estimates of N_e with modern-day census population size (N_c) across sequenced placental species. As predicted, they found that larger N_e/N_c ratios (shrinking populations) positively correlate with more-urgent conservation threat statuses today. These findings echo a recent study of the vaquita porpoise (*Phocoena sinus*) (*13*) regarding the value of genome-informed predictions of extinction risk, including identifying populations that have been historically small versus those recently reduced in size. In a related analysis, Moon *et al.* queried the genome of a famous sled dog from 1920s Alaska named Balto. Sequencing underbelly tissue from the taxidermied titan, they found that Balto had genetic variants for improved starch digestion, thicker fur, and overall higher diversity relative to modern Siberian huskies. Jointly, these studies highlight the irreplaceable value of museum specimens as historical baselines for measuring changes in genetic diversity (*14*).

Plunging deeper into the past, Foley *et al.* (*10*) analyzed how genomic patterns of genetic inheritance shifted in placental mammals after the dinosaur-annihilating meteor impact ~66 million years ago [the Cretaceous-Paleogene boundary (K-Pg)].

[1]School of Life Sciences, Arizona State University, Tempe, AZ, USA. [2]Department of Biology, Washington University, St. Louis, MO, USA. Email: nathan.upham@asu.edu; michael.landis@wustl.edu

Genomes relative to year and phylogenetic relationships

Genomes for 675 mammal species relative to the Mammalia phylogenetic tree of 5911 living species shows the disproportionate representation of large-bodied and high-latitude species. Shown is the consensus timescaled phylogeny from Upham *et al*. (*4*) and genome data downloaded from NCBI on 9 February 2023.



Artio., Artiodactyla; Carn., Carnivora; Eulipoty., Eulipotyphla; NCBI, National Center for Biotechnology Information. Reproducible code is available at https://github.com/n8upham/mammalGenomesNCBI.

Although phylogenetic trees depict somewhat orderly relationships among species, the phylogenetic trees for individual genes of those same species often follow far more discordant histories. Much of this gene-tree discordance emerges from the random inheritance and sorting of gene variants among newly formed species, a process called incomplete lineage sorting (ILS). ILS is particularly common when ancestral species had large population sizes before diverging multiple times in rapid succession. Before the K-Pg event, placental ancestors are hypothesized to have been relatively long-lived with small population sizes, likely of similar size and ecology as modern treeshrews (order Scandentia; ~200 g), which could reduce ILS, whereas the ecological and demographic expansion of placentals after the K-Pg should promote rampant ILS. Confirming these predictions, Foley *et al*. found lower levels of ILS between older, pre–K-Pg relationships—for example, between all rodents and primates—and higher ILS between younger post–K-Pg relationships—for example, within bats, rodents, or primates. This work demonstrates how ILS, which was once considered "noise" in comparative datasets, can help reveal the histories of major ecological transitions.

Zooming out, mammal genomics is in a rapid expansion phase (see the figure). The number of distinct species with publicly available genomes rose by 180% since 2019 to now 675 mammals, led by Zoonomia (121 new) and a recent bolus of Australian marsupials (161 new). It is critical, however, to recognize that these genomes are disproportionately represented by large-bodied and high-latitude species. This bias relates to the sourcing of tissues for genome sequencing from zoo animals in the Global North, which often lack the known population origins and preserved specimens (such as skin, skull, and archived tissues) needed for later study (*15*). As a result, members of Carnivora, Artiodactyla (including whales), and Primates (~1100 species) have 285 species with genomes, whereas members of Chiroptera and Rodentia (~4000 species) have only 164. Variable genome quality further compounds these sampling biases, with only 76 mammal species assembled to the chromosome level and two-thirds of other genomes assembled too incompletely to identify typical repeat lengths (most assembled chunks are <1 Mb long). Thus, despite recent advances, the emerging field of T2T phylogenomics will need to remedy historical sampling gaps and improve legacy data to fully explore the mammalverse.

Of course, those missing mammal genomes present opportunities for new discoveries and insights. Future work should strive to evenly sample species relative to geographical realm, latitude, and elevation; island versus continental occurrence; body size, longevity, and other life-history traits; conservation status; and phylogenetic distinctiveness. Greater genus- and species-level sampling will help resolve ascertainment biases that may otherwise limit the generalizability of evolutionary inferences. For example, large-bodied organisms tend to evolve differently than small ones (smaller N_e in the former, leading to weaker selection), which is currently tipping the balance of generalizations about genome evolution toward rhinoceroses, elephants, and blue whales to the detriment of shrews, bats, and squirrels. Small-bodied mammals are expected to evolve more rapidly because of large N_e and short generation times, but both dynamics can be flipped when small species are range-restricted (for example, on mountains or islands) or long-lived (such as *Myotis* bats), which underscores their value for comparative genomic study. Sampling a greater diversity of mammals will also fill out phylogenetic representation below family-level lineages and refine the understanding of how genomes evolve over micro- and macroevolutionary time scales. The Zoonomia project, and others preceding it, have opened myriad new portals for exploring genome architecture, population structure, and global diversification in mammals, with findings that promise to astound in coming decades. ■

REFERENCES AND NOTES

1. J. Armstrong *et al*., *Nature* **587**, 246 (2020).
2. M. P. K. Blom, *Mol. Ecol.* **30**, 5935 (2021).
3. C. Scornavacca, F. Delsuc, N. Galtier, *Phylogenetics in the Genomic Era* (No commercial publisher, 2020); https://hal.archives-ouvertes.fr/hal-02535070v3.
4. N. S. Upham, J. A. Esselstyn, W. Jetz, *PLOS Biol*. **17**, e3000494 (2019).
5. M. Christmas *et al*., *Science* **380**, eabn3943 (2023).
6. I. Kaplow *et al*., *Science* **380**, eabm7993 (2023).
7. A. Osmanski *et al*., *Science* **380**, eabn1430 (2023).
8. A. Wilder *et al*., *Science* **380**, eabn5856 (2023).
9. K. Moon *et al*., *Science* **380**, eabn5887 (2023).
10. N. Foley *et al*., *Science* **380**, eabl8189 (2023).
11. S. Lamichhaney *et al*., *Philos. Trans. R. Soc. Lond. B Biol. Sci*. **374**, 20180248 (2019).
12. M. Lynch, *The Origins of Genome Architecture* (Sinauer Associates, 2007).
13. J. A. Robinson *et al*., *Science* **376**, 635 (2022).
14. D. C. Card *et al*., *Annu. Rev. Genet*. **55**, 633 (2021).
15. J. C. Buckner, R. C. Sanders, B. C. Faircloth, P. Chakrabarty, *eLife* **10**, e68264 (2021).

ACKNOWLEDGMENTS

The authors were supported by the National Science Foundation (DEB-2040347 to M.J.L.) and the National Institutes of Health (1R21AI164268-01 to N.S.U.).

10.1126/science.add2209

Genomics expands the mammalverse

Nathan S. Upham and Michael J. Landis

Science, **380** (6643), .
DOI: 10.1126/science.add2209