RESEARCH ARTICLE

PRIMATE GENOMES

A global catalog of whole-genome diversity from 233 primate species

Lukas F. K. Kuderna[1,2]*†, Hong Gao[2], Mareike C. Janiak[3], Martin Kuhlwilm[1,4,5], Joseph D. Orkin[1,6], Thomas Bataillon[7], Shivakumara Manu[8,9], Alejandro Valenzuela[1], Juraj Bergman[7,10], Marjolaine Rousselle[7], Felipe Ennes Silva[11,12], Lidia Agueda[13], Julie Blanc[13], Marta Gut[13], Dorien de Vries[3], Ian Goodhead[3], R. Alan Harris[14], Muthuswamy Raveendran[14], Axel Jensen[15], Idrissa S. Chuma[16], Julie E. Horvath[17,18,19,20,21], Christina Hvilsom[22], David Juan[1], Peter Frandsen[22], Joshua G. Schraiber[2], Fabiano R. de Melo[23], Fabrício Bertuol[24], Hazel Byrne[25], Iracilda Sampaio[26], Izeni Farias[24], João Valsecchi[27,28,29], Malu Messias[30], Maria N. F. da Silva[31], Mihir Trivedi[9], Rogerio Rossi[32], Tomas Hrbek[24,33], Nicole Andriaholinirina[34], Clément J. Rabarivola[34], Alphonse Zaramody[34], Clifford J. Jolly[35], Jane Phillips-Conroy[36], Gregory Wilkerson[37], Christian Abee[37], Joe H. Simmons[37], Eduardo Fernandez-Duque[38], Sree Kanthaswamy[39], Fekadu Shiferaw[40], Dongdong Wu[41], Long Zhou[42], Yong Shao[41], Guojie Zhang[43,42,44,45,46], Julius D. Keyyu[47], Sascha Knauf[48], Minh D. Le[49], Esther Lizano[1,50], Stefan Merker[51], Arcadi Navarro[1,52,53,54], Tilo Nadler[55], Chiea Chuen Khor[56], Jessica Lee[57], Patrick Tan[58,59,56], Weng Khong Lim[59,58,60], Andrew C. Kitchener[61], Dietmar Zinner[62,63,64], Ivo Gut[13], Amanda D. Melin[65,66,67], Katerina Guschanski[68,15], Mikkel Heide Schierup[7], Robin M. D. Beck[3], Govindhaswamy Umapathy[8,9], Christian Roos[69], Jean P. Boubli[3], Jeffrey Rogers[14]*, Kyle Kai-How Farh[2]*, Tomas Marques Bonet[1,50,13,52]*

The rich diversity of morphology and behavior displayed across primate species provides an informative context in which to study the impact of genomic diversity on fundamental biological processes. Analysis of that diversity provides insight into long-standing questions in evolutionary and conservation biology and is urgent given severe threats these species are facing. Here, we present high-coverage whole-genome data from 233 primate species representing 86% of genera and all 16 families. This dataset was used, together with fossil calibration, to create a nuclear DNA phylogeny and to reassess evolutionary divergence times among primate clades. We found within-species genetic diversity across families and geographic regions to be associated with climate and sociality, but not with extinction risk. Furthermore, mutation rates differ across species, potentially influenced by effective population sizes. Lastly, we identified extensive recurrence of missense mutations previously thought to be human specific. This study will open a wide range of research avenues for future primate genomic research.

The order Primates includes over 500 recognized species that display an array of morphological, physiological, and behavioral adaptations (*1*). Spanning a broad range of social systems, locomotory styles, dietary specializations, and habitat preferences, these species rightly attract attention from scientists with equally diverse research interests. Because humans are members of the order Primates, we also find many important and informative biological parallels between ourselves and other primates. The analysis of nonhuman primate genomes has long been motivated by a desire to understand human evolutionary origins, human health, and disease. However, past comparative genomic analyses have mainly focused on a relatively small number of species (*2*, *3*), thus providing a limited understanding of genome variability in only a few key lineages, such as members of the great apes (*4*–*10*) or macaques (*11*–*13*). Furthermore, low numbers of wild-born individuals in these studies potentially result in assessments of diversity that may not reflect natural populations (*3*). To gain a more complete picture of how evolution has shaped genomic variation across primates, large-scale sequencing of many species and individuals is necessary, especially within previously neglected lineages such as strepsirrhines (lemurs, lorises, galagos, and relatives) and platyrrhines (monkeys of the Americas). The need for a more complete understanding of primate genetic diversity in the wild, and its determinants, is urgent given the current extinction crisis driven by climate change, habitat loss, and illegal trading and hunting (*14*). At present, 60% of the world's primate species are threatened with extinction, and current trends are likely to exacerbate the rates of biodiversity loss in the near future (*14*, *15*). The analysis of whole-genome sequences allows estimation of genetic diversity and evaluation of its association with ecological traits, degrees of inbreeding, and phylogenetic relationships, all metrics relevant to primate conservation genomics.

High-coverage genome sequences of 233 primate species

We sequenced the genomes of 703 individuals from 211 primate species on the Illumina NovaSeq 6000 platform (*16*). For 78% of individuals, the available amount of DNA permitted us to generate polymerase chain reaction–free libraries. We sequenced paired-end reads of 151 base pairs (bp) to an average production target of at least 100 gigabases (Gb), resulting in an average mapped coverage of 32.4× per individual (15.3 to 77.6×) (*16*). We expanded our dataset by including 106 individuals representing 29 species from previously published studies to maximize phylogenetic diversity (*8*, *17*–*24*). Altogether, we compiled data from 809 individuals from 233 primate species, amounting to 47% of the 521 currently recognized species (*14*). Our sampling covers 86% of primate genera (69), and all 16 families. More than 72% of individuals in this study are wild-born. Furthermore, 58% of species in our dataset are classified as threatened with extinction by the International Union for Conservation of Nature (IUCN) [i.e., classified in the categories vulnerable (VU), endangered (EN), and critically endangered (CR)], and 30 species are critically endangered. It is worth noting that among the species we sampled are some of the world's most endangered primates, which face an extremely high risk of extinction in the wild. Examples include the Western black crested gibbon (*Nomascus concolor*), with an estimated 1500 individuals left in the wild and scattered across an array of discontinuous habitats, and the northern sportive lemur (*Lepilemur septentrionalis*), with roughly 40 individuals estimated to remain in the wild, inhabiting an area potentially as small as 12 km^2 (*25*, *26*).

For 100 species, we generated sequencing data from more than one individual, and for 36 species from five or more individuals, 29 of which belong to newly sequenced species. We thus gathered broad primate taxonomic coverage by compiling species from all major geographic regions currently inhabited by primates, including the Americas, mainland Africa, Madagascar, and Asia (Fig. 1A). The data presented here provide the foundation for several additional studies in this issue, informing important and diverse topics including hybrid speciation and reticulation among primates (*27*) and predicting the landscape of tolerated mutations in the human genome (*28*).

Owing to technical challenges inherent to short-read assembly, we aligned our data to a backbone of 32 reference genomes for further analyses, most of which are derived from long-read sequencing technologies (*16*). These references are well distributed across the primate phylogeny and result in a median pairwise distance between the focal and reference species of 6.6×10^{-3} substitutions per site (0 to 4.1×10^{-2}), which is within the range of previous projects using a similar approach (*8*). To ensure our estimates of genetic diversity over these phylogenetic distances are minimally

biased, we compared pairs of diversity estimates in which reads from one species were mapped to its own reference as well as mapped to another species reference. Across 19 species pairs that fully cover the phylogenetic distances between focal species and reference in our data, we find heterozygosity estimates to be highly correlated (Pearson's $r = 0.97$, $p = 6.8 \times 10^{-12}$). Overall, we find a median value of 2.4 Gb per individual to be callable across all references, thus enabling genome-wide comparisons.

Genetic diversity across primates

Heterozygosity in primates spans over an order of magnitude, with values ranging from 0.41×10^{-3} heterozygotes per base pair (het × bp^{-1}) to 7.14×10^{-3} het × bp^{-1} (Fig. 1C). We observe the lowest levels of diversity in the golden snub-nosed monkey (*Rhinopithecus roxellana*) at about one heterozygous position every 2400 bp. Only 15 species have a lower median genetic diversity than humans, the primate with by far the largest census size. Among these are several Asian colobines, but also the aye-aye, the western hoolock gibbon, and the Guinea baboon. There are marked differences in genetic diversity across genera, families, and geographic regions, with high-diversity species found among cercopithecines from mainland Africa and lemurs in Madagascar (Fig. 1B). Among cercopithecines, guenons of the genus *Cercopithecus* are almost exclusively responsible for high diversity with a median value of 4.54×10^{-3} het × bp^{-1}, more than double the primate-wide median. Some members of this tribe also show large historical effective population sizes, and there are several known instances of past and present interspecific hybridization (*29*–*32*). We further observe high diversity across several genera of lemurs, which are among the most endangered primates, primarily owing to rapid habitat loss and severe population decline. Examples include members of the true lemurs (*Eulemur* spp.), bamboo lemurs (*Hapalemur* spp.), and sifakas (*Propithecus* spp.).

We investigated whether genetic diversity estimates are correlated with extinction risk in primates, a subject of previous debate (*17*, *33*, *34*). Despite our broad sampling, we find no global relationship between numerically coded IUCN extinction risk categories and estimated heterozygosity [$p > 0.05$, phylogenetic generalized least squares (PGLS)] (Fig. 2A) (*16*). Because genetic diversity is strongly determined by long-term demographic history, rapid recent population declines such as those currently experienced by many primate species are unlikely to be detected in a cross-species comparison. Instead, temporal datasets within the same species are better suited to quantify recent changes in genetic diversity (*35*). Nevertheless, comparing genetic diversity for non-threatened [least concern (LC), near-threatened (NT)] and threatened (VU, EN, CR) species within the same family consistently uncovers lower diversity among species in the threatened categories for all families with more than one species in both categories, although not all comparisons reach statistical significance ($p < 0.05$, Mann–Whitney U test) (Fig. 2B). The only exception is Lorisidae, which showed no difference in genetic diversity between non-threatened and threatened species.

To further assess the potential impact of recent population decline, we analyzed runs of homozygosity (RoH) across species. We focused on tracts with a minimum length of one megabase (Mb), which in humans indicate recent inbreeding (*8*). The order-wide median fraction of the genome in RoH is 5.1%, and individual values vary substantially, reaching over 50%. We find critically endangered species, such as the white-headed langur (*Trachypithecus leucocephalus*), the eastern gorilla (*Gorilla beringei*), and mongoose lemur (*Eulemur mongoz*), among the species with the highest proportion of RoHs (Fig. 2C). However, some species not currently classified as threatened, such as Azara's owl monkey (*Aotus azarae*) and the northern greater galago (*Otolemur garnettii*), also have a high fraction of the genome in RoHs. Although the overall conservation status of these two species might not be worrisome, some individuals may belong to smaller local populations, which can exacerbate inbreeding. We find 13 critically endangered species with lower than the primate-wide average fractions of their genomes in RoHs, among them the three douc langur species (*Pygathrix*

[1]IBE, Institute of Evolutionary Biology (UPF-CSIC), Department of Medicine and Life Sciences, Universitat Pompeu Fabra. PRBB, C. Doctor Aiguader N88, 08003 Barcelona, Spain. [2]Illumina Artificial Intelligence Laboratory, Illumina Inc.; Foster City, CA 94404, USA. [3]School of Science, Engineering & Environment, University of Salford, Salford M5 4WT, UK. [4]Department of Evolutionary Anthropology, University of Vienna, Djerassiplatz 1, 1030 Vienna, Austria. [5]Human Evolution and Archaeological Sciences (HEAS), University of Vienna, Austria. [6]Département d'anthropologie, Université de Montréal, 3150 Jean-Brillant, Montréal, QC H3T 1N8, Canada. [7]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [8]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India. [9]Laboratory for the Conservation of Endangered Species, CSIR-Centre for Cellular and Molecular Biology, Hyderabad 500007, India. [10]Section for Ecoinformatics and Biodiversity, Department of Biology, Aarhus University, Aarhus, Denmark. [11]Research Group on Primate Biology and Conservation, Mamirauá Institute for Sustainable Development, Estrada da Bexiga 2584, CEP 69553-225, Tefé, Amazonas, Brazil. [12]Evolutionary Biology and Ecology (EBE), Département de Biologie des Organismes, Université libre de Bruxelles (ULB), Av. Franklin D. Roosevelt 50, CP 160/12, B-1050 Brussels Belgium. [13]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri I Reixac 4, 08028 Barcelona, Spain. [14]Human Genome Sequencing Center and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA. [15]Department of Ecology and Genetics, Animal Ecology, Uppsala University, SE-75236 Uppsala, Sweden. [16]Tanzania National Parks, Arusha, Tanzania. [17]North Carolina Museum of Natural Sciences, Raleigh, NC 27601, USA. [18]Department of Biological and Biomedical Sciences, North Carolina Central University, Durham, NC 27707, USA. [19]Department of Biological Sciences, North Carolina State University, Raleigh, NC 27695, USA. [20]Department of Evolutionary Anthropology, Duke University, Durham, NC 27708, USA. [21]Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA. [22]Copenhagen Zoo, 2000 Frederiksberg, Denmark. [23]Universidade Federal de Viçosa, Viçosa, Brazil. [24]Universidade Federal do Amazonas, Departamento de Genética, Laboratório de Evolução e Genética Animal (LEGAL), Manaus, Amazonas 69080-900, Brazil. [25]Department of Anthropology, University of Utah, Salt Lake City. UT 84102, USA. [26]Universidade Federal do Para, Bragança, Para, Brazil. [27]Research Group on Terrestrial Vertebrate Ecology, Mamirauá Institute for Sustainable Development, Tefé, Amazonas, Brazil. [28]Rede de Pesquisa para Estudos sobre Diversidade, Conservação e Uso da Fauna na Amazônia – RedeFauna, Manaus, Amazonas, Brazil [29]Comunidad de Manejo de Fauna Silvestre en la Amazonía y en Latinoamérica – ComFauna, Iquitos, Loreto, Peru. [30]Universidade Federal de Rondônia, Porto Velho, Rondônia, Brazil. [31]Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brazil. [32]Instituto de Biociências, Universidade Federal do Mato Grosso, Cuiabá, MT, Brazil. [33]Department of Biology, Trinity University, San Antonio, TX 78212, USA. [34]Life Sciences and Environment, Technology and Environment of Mahajanga, University of Mahajanga, Mahajanga, Madagascar. [35]Department of Anthropology, New York University, New York, NY 10003, USA. [36]Department of Neuroscience, Washington University School of Medicine in St. Louis, St. Louis, MO 63110, USA. [37]Keeling Center for Comparative Medicine and Research, MD Anderson Cancer Center, Bastrop TX 78602, USA. [38]Department of Anthropology, Yale University, New Haven, CT 06511, USA. [39]School of Mathematical and Natural Sciences, Arizona State University, Phoenix, AZ 85004, USA. [40]Guinea Worm Eradication Program, The Carter Center Ethiopia, Addis Ababa, Ethiopia. [41]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China. [42]Center for Evolutionary and Organismal Biology, Zhejiang University School of Medicine, Hangzhou 310058, China. [43]Villum Centre for Biodiversity Genomics, Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark. [44]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China. [45]Liangzhu Laboratory, Zhejiang University Medical Center, 1369 West Wenyi Road, Hangzhou 311121, China. [46]Women's Hospital, School of Medicine, Zhejiang University, 1 Xueshi Road, Shangcheng District, Hangzhou 310006, China. [47]Tanzania Wildlife Research Institute (TAWIRI), Head Office, P.O. Box 661, Arusha, Tanzania. [48]Institute of International Animal Health/One Health, Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, 17493 Greifswald–Insel Riems, Germany. [49]Department of Environmental Ecology, Faculty of Environmental Sciences, University of Science and Central Institute for Natural Resources and Environmental Studies, Vietnam National University, Hanoi, Vietnam. [50]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain. [51]Department of Zoology, State Museum of Natural History Stuttgart, Stuttgart, Germany. [52]Institució Catalana de Recerca i Estudis Avançats (ICREA) and Universitat Pompeu Fabra. Pg. Luís Companys 23, 08010 Barcelona, Spain. [53]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Av. Doctor Aiguader, N88, 08003 Barcelona, Spain. [54]BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, C. Wellington 30, 08005 Barcelona, Spain. [55]Cuc Phuong Commune, Nho Quan District, Ninh Binh Province, Vietnam. [56]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore. [57]Mandai Nature, 80 Mandai Lake Road, Singapore. [58]SingHealth Duke-NUS Institute of Precision Medicine (PRISM), Singapore. [59]Cancer and Stem Cell Biology Program, Duke-NUS Medical School, Singapore [60]SingHealth Duke-NUS Genomic Medicine Centre, Singapore. [61]Department of Natural Sciences, National Museums Scotland, Chambers Street, Edinburgh EH1 1JF, UK, and School of Geosciences, Drummond Street, Edinburgh EH8 9XP, UK. [62]Cognitive Ethology Laboratory, Germany Primate Center, Leibniz Institute for Primate Research, 37077 Göttingen, Germany. [63]Department of Primate Cognition, Georg-August-Universität Göttingen, 37077 Göttingen, Germany. [64]Leibniz ScienceCampus Primate Cognition, 37077 Göttingen, Germany. [65]Department of Anthropology and Archaeology, University of Calgary, 2500 University Dr NW, Calgary, AB T2N 1N4, Canada. [66]Department of Medical Genetics, University of Calgary, 3330 Hospital Drive NW, HMRB 202, Calgary, AB T2N 4N1, Canada. [67]Alberta Children's Hospital Research Institute, University of Calgary, 3330 Hospital Drive NW, HMRB 202, Calgary, AB T2N 4N1, Canada. [68]Institute of Ecology and Evolution, School of Biological Sciences, University of Edinburgh, Edinburgh, UK. [69]Gene Bank of Primates and Primate Genetics Laboratory, German Primate Center, Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Göttingen, Germany.
***Corresponding author. Email: lkuderna@illumina.com (L.F.K.K.); jr13@bcm.edu (J.R.); kfarh@illumina.com (K.K.-H.F.); tomas.marques@upf.edu (T.M.B.)**
†Present address: Illumina Artificial Intelligence Laboratory, Illumina Inc., Foster City, CA 94404, USA.

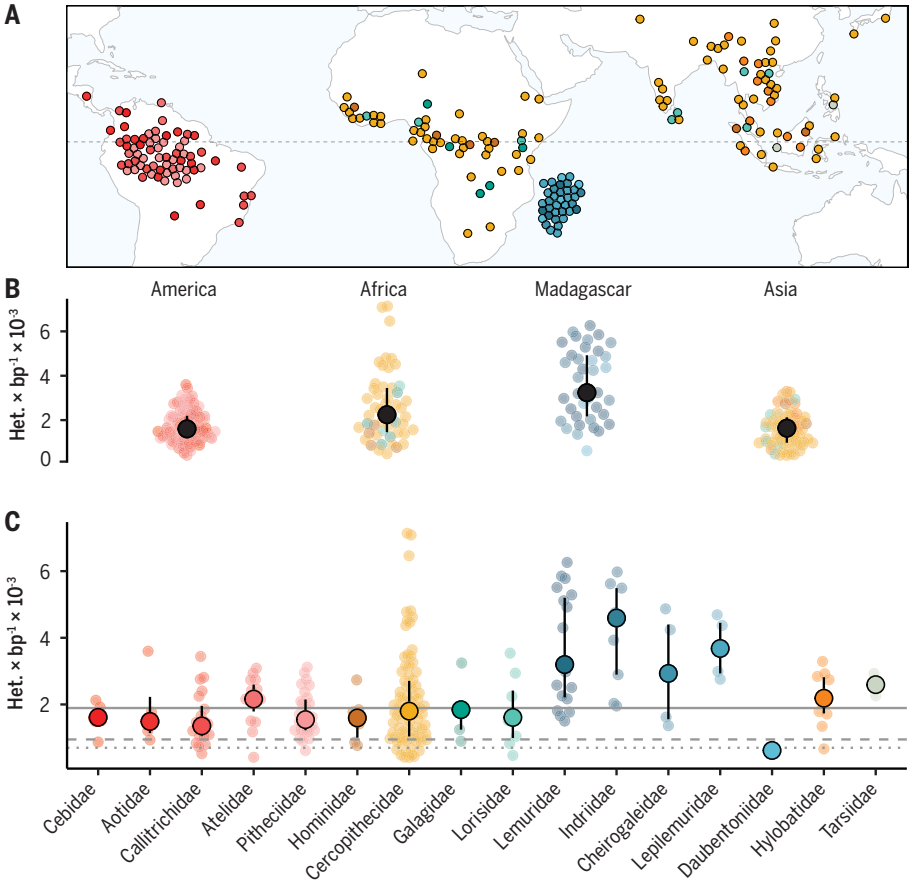

Fig. 1. Genetic diversity in primates across geographic regions and families. (**A**) Sampling range of species analyzed in this project. Each point represents the approximate species range centroid of all sampled species with available ranges. Points are repelled to avoid overplotting. (**B**) Heterozygosity stratified by geographic region. Solid black circles and whiskers represent median values and interquartile range. (**C**) Median species heterozygosity by family. Solid circles and whiskers represent median and interquartile range. Solid gray line denotes primate-wide median heterozygosity; dashed and dotted lines denote human heterozygosity for African and bottlenecked out-of-Africa populations, respectively. Points are colored according to the family a species belongs to, as denoted on the *x* axis of (C).

cinerea, *P. nemaeus*, *P. nigripes*), red-tailed sportive lemur (*Lepilemur ruficaudatus*), and Verreaux's sifaka (*Propithecus verreauxi*). We find no overall relationship between extinction risk and degree of inbreeding deduced from the total fraction of the genome in RoHs (Pearson's $r = 0.03$, $p = 0.71$). This implies that RoHs are not a good predictor of extinction risk in primates and suggests that many critically endangered species are threatened by nongenetic factors, likely reflecting population declines that have been too fast to be detectable on the genomic level. Given the potential importance of functional variation to conservation efforts, we sought to quantify the proportion of loss of functional variation in each lineage (*34*, *36*). To this end, we quantified stop-gain and missense mutations and normalized them by the number of synonymous mutations to account for lineage-specific differences in evolutionary rates. We found inverse relationships between the missense/synonymous ratios (Pearson's $r = -0.35$, $p = 9.3 \times 10^{-8}$) and, to a lesser extent, stop-gain/synonymous ratios and heterozygosity across primates, suggesting effects of purifying selection on deleterious variation, although the latter does not reach statistical significance (Pearson's $r = -0.12$, $p = 0.082$). We do not find deleterious variations as measured by the stop-gain/synonymous ratio to be correlated with extinction risk (Pearson's $r < 0.01$, $p = 0.94$). Nevertheless, we caution that the varying quality of the references and their annotations, together with potential changes in gene structure between the references and analyzed species, might add noise to the comparisons across our references.

A time-calibrated nuclear phylogeny of primates

We generated a genome-wide nuclear phylogeny of ultraconserved elements (UCEs) and 500 bp of their flanking regions, a widely used marker that enables easy detection of sequence orthologs across species (*37*). To this end, we identified the location of ~3500 UCE probes across all primate genomes and generated individual gene trees for each locus using a maximum-likelihood approach (*38*–*40*). We used the resulting trees as input for a coalescent analysis to obtain the topology of the species tree, which has strong support at most nodes and recovers all currently recognized primate families, tribes, and genera as monophyletic (*41*–*44*). We used a newly established set of 27 well-justified fossil calibration points to constrain the timing of key phylogenetic divergences among different lineages (*45*). We estimate the split between Haplorhini and Strepsirrhini to have happened between 63.3 and 58.3 million years (Ma) ago, and thus the radiation of crown Primates is entirely within the Paleocene. We find the deepest divergence within tarsiers to be notably recent at 15.2 to 9.5 Ma, which, together with fossil evidence, implies considerable extinction along the long branch leading to extant tarsiers (*46*–*49*). All interfamilial relationships within our phylogeny receive strong support [posterior probability (PP) = 1], except for the position of Aotidae (owl monkeys), which is weakly supported as sister to Callitrichidae (marmosets and tamarins) rather than Cebidae (capuchin and squirrel monkeys) (PP = 0.56). We consider the precise relationship among these three families to remain uncertain. Lastly, we estimate the human–chimpanzee divergence between 9.0 and 6.9 Ma, and thus slightly older than other recent analyses, although these overlap our confidence intervals (*41*–*43*).

Taking advantage of our rich resequencing data, we generated a tree topology that includes two individuals per species for all species with more than one sequenced individual. We observe paraphyletic or polyphyletic placements of these individuals in 17 species, possibly calling several currently established species boundaries into question (Fig. 3). These cases could result from genetic structure interpreted as species delimitation, incomplete lineage sorting, or hybridization, and most are also observed at the mitochondrial level (*16*, *50*–*53*). Although some instances of hybridization have previously been described, such as among different species of langurs (*54*), we find most of the paraphyletic or polyphyletic placements among platyrrhines. These include 13 species, among them capuchins, squirrel monkeys, howler monkeys, uakaris, sakis, and titis, and point to the need for more taxonomic studies using genomic data in this group (*55*). Finally, we retrieve previously unknown phylogenetic relationships for species that were sequenced for the first time in this study, such as different species of howler monkeys (e.g., *Alouatta puruensis*, or *A. juara*).

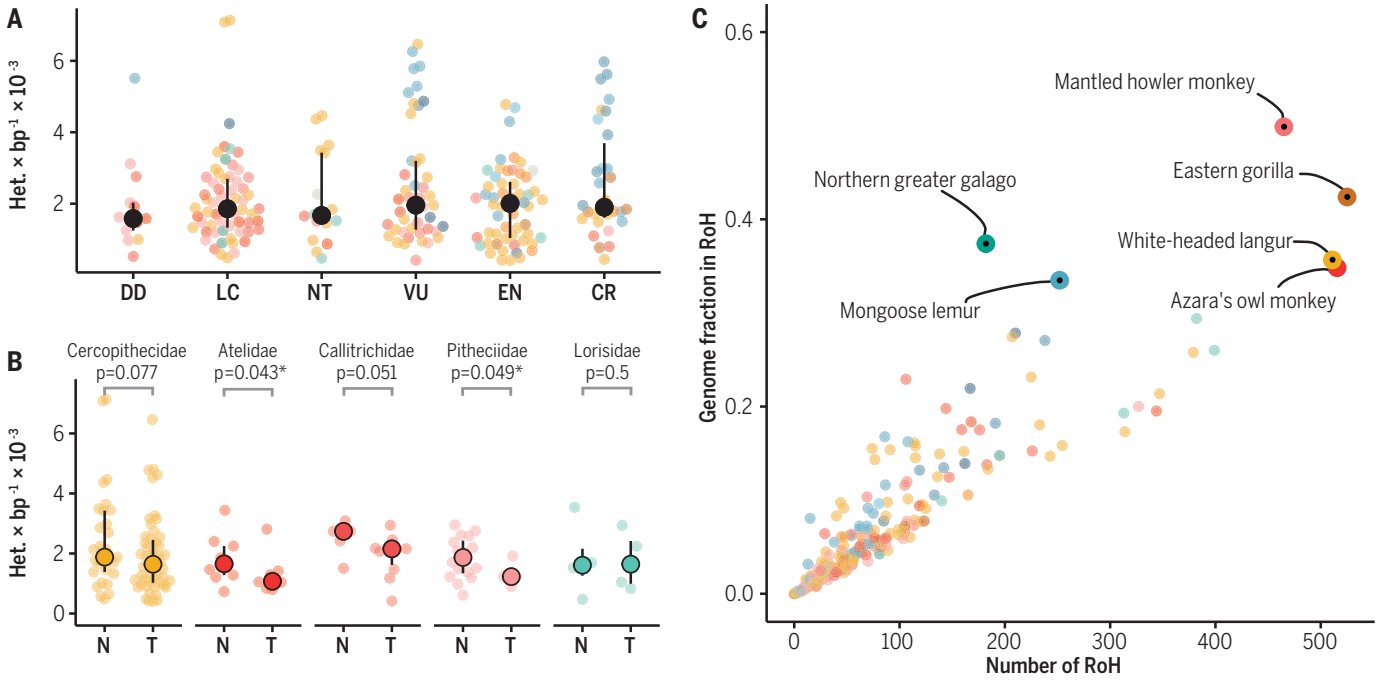

Fig. 2. Runs of homozygosity and impact of extinction risk on diversity (**A**) Relationship between IUCN extinction risk categories and heterozygosity. Solid black circles and bars denote median and IQR. (**B**) Partition into threatened (T: VU, EN, CR) and nonthreatened (N: LC, NT) categories for all families with more than one species in either partition. Significant differences ($p < 0.05$, one-sided rank-sum test) are marked with an asterisk. (**C**) Median number of tracts of homozygosity versus median proportion of the genome in runs of homozygosity per species. Species with a fraction over 1/3 are highlighted. Solid black dots within highlights denote threatened species (VU, EN, CR).

Determinants of diversity and mutation rate

We used the topology of the species tree and 614 UCE alignments, for which we had full species coverage, to estimate branch lengths as the number of substitutions per site. We combined this with our dated phylogeny and published estimates of generation times to estimate mutation rates per generation for all primate species from their substitution rates (*16*). Although we caution that we cannot rule out potential biases in these estimates, such as the effects of selection or uncertainties in fossil calibration, they agree well with published estimates for overlapping species on the basis of trio sequencing (Spearman's $r = 0.85$, $p = 0.02$; Fig. 4C). Our estimated mutation rates (μ) per generation vary between 0.25×10^{-8} and 1.62×10^{-8} (Fig. 4A), showing a considerably larger range than previously reported (*56*). We observe the lowest estimate per generation in Lemuridae and find highly variable estimates across some families such as Cebidae and Lorisidae, which also have variable generation times (8 to 17 and 4.6 to 9 years per generation, respectively). The highest estimates of μ are in great apes. We find a significant and positive correlation between μ per generation and the generation time (Spearman's $r = 0.36$, $p = 1.89 \times 10^{-8}$), which partly counteracts a generation-time effect on the yearly mutation rate. The latter is therefore larger in species with a shorter generation time (Fig. 4E). Together, variation in effective population size (N_e) and generation time explain roughly half of the observed variation in mutation rates among extant species.

We used our estimates of μ and estimates of genetic diversity π based on median heterozygosity to get an estimate of the effective population sizes $N_e = \pi/(4 \times \mu)$. We find multiple species belonging to different families of lemurs, as well as several species of guenons within the Cercopithecidae, with the largest N_e estimates, often exceeding 2×10^5 (Fig. 4B). For several critically endangered lemur species, e.g., the northern sportive lemur (*Lepilemur septentrionalis*), the red-tailed sportive lemur (*Lepilemur ruficaudatus*), or the Alaotra reed lemur (*Hapalemur alaotrensis*), these likely surpass census sizes by a considerable margin. We find multiple members of the genera *Cercopithecus* and *Eulemur* exhibiting high N_e values, which may be driven by interspecific hybridization observed in these species. Conversely, we observe comparatively low N_e estimates in great apes, lorises, and platyrrhines (Fig. 4B) (*16*).

The drift-barrier hypothesis (*57*, *58*) predicts that μ per generation should decrease with N_e, because new mutations affecting fitness are predominantly deleterious, and the ability to select for lower mutation rate increases with the population size. We tested for a relationship between μ and N_e, while controlling for the relationship between μ and generation time in a PGLS model, and observed a significantly lower mutation rate for species with higher N_e. We find around 45% of the variation in μ to be explained by N_e, thus lending apparent support to the drift-barrier hypothesis (*59*). However, we caution that although this pattern is consistent with the drift-barrier hypothesis, N_e is estimated by the division of π by μ, which at least partially explains the negative relationship. Additionally, our estimates of μ assume homogeneous levels of evolutionary constraint on the UCEs and flanking regions used to estimate divergence time and substitution rate. Should there be a strong covariation between substitution rates in these regions and effective size in branches, underlying variation in N_e along the branches of the phylogeny can act as a confounder of apparent variation in mutation rates and thus further complicate a formal test of the drift-barrier hypothesis.

To further disentangle what factors might contribute to the levels of genetic diversity and mutation rates, we compiled a list of 32 traits that can be summarized by grouping them into the broader categories of body mass, life history, activity budget, ranging patterns, climatic niche, social organization, sexual selection, diet composition, social systems, mating systems, and natal

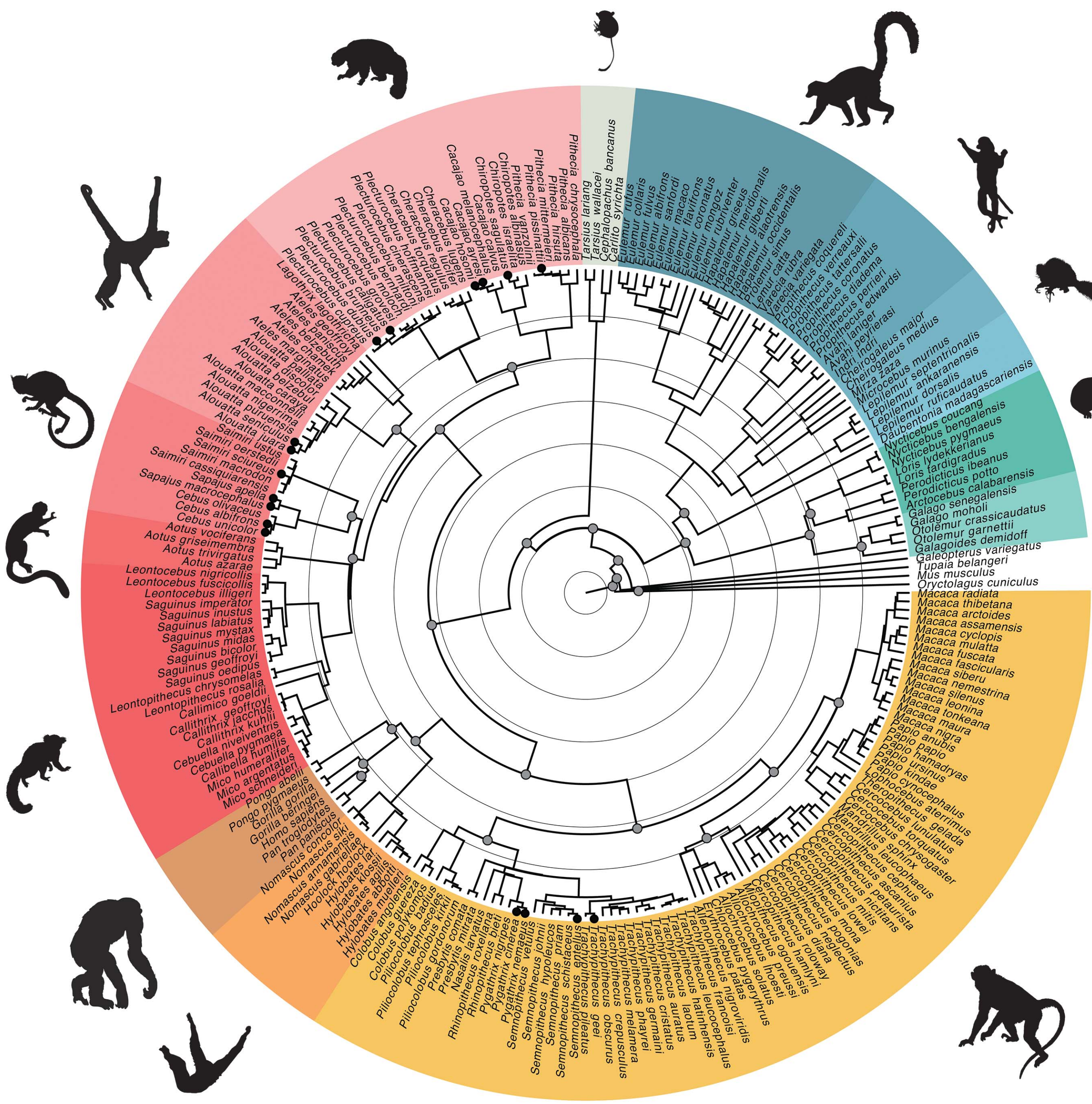

Fig. 3. Fossil-calibrated nuclear time tree. Concentric background circles mark 10-million-year intervals; solid gray circles in internal nodes show fossil calibration points (*36*); species marked with solid circles at tips show paraphyly or polyphyly when including additional individuals to estimate the topology.

dispersal mode (*60–62*). To account for potential phylogenetic inertia in trait evolution, we generated PGLS models using either genetic diversity or mutation rate as the response variable and individual traits as the predictors. We find traits within mating systems, activity budget, climatic niche, ranging patterns, and life history to be significant predictors of diversity ($p < 0.05$), and traits within the former three categories remaining so after accounting for multiple testing (Benjamini-Hochberg correction, false discovery rate = 0.05). Species organized in single-male polygynous mating systems show lower diversity than the background ($r^2_{pred} = 0.11$, $p_{corr} = 1.53 \times 10^{-2}$), consistent with expectations of reduced contribution of allelic diversity from males (*63*). Within the climatic niche, we observe a gradient of diversity declining from south to north ($r^2_{pred} = 0.28$, $p_{corr} = 1.45 \times 10^{-5}$), which is driven by highly diverse lemur species in the Southern Hemisphere. We also find a significant correlation with mean temperature and amount of precipitation ($r^2_{pred} = 0.33$, $p_{corr} = 1.97 \times 10^{-4}$). It is worth noting that these measurements are not highly correlated with each other (Pearson's r −0.27 to 0.17), and the relationships are thus at least partly independent. Lastly, within the activity budget, we find the amount of time spent socializing to be correlated with diversity ($r^2_{pred} = 0.11$, $p_{corr} = 5.56 \times 10^{-3}$). However, we caution that the measurement of activity budget is difficult to standardize across species, and interpreting this relationship is thus challenging. We find no significant impact of life-history traits such as body mass or longevity on genetic diversity

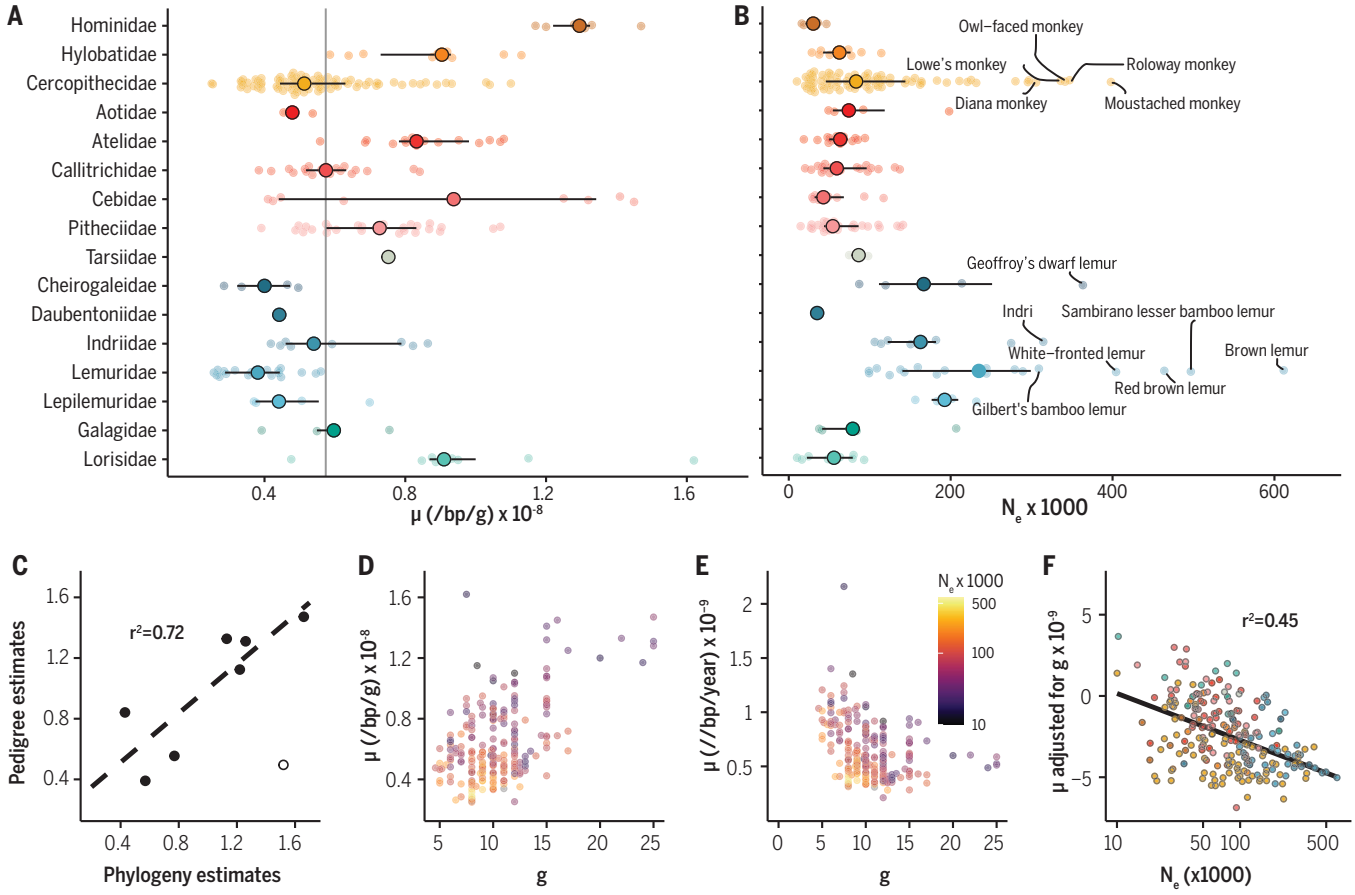

Fig. 4. Estimates of mutation rates and effective population size. (**A**) Distribution of estimates of the per-generation mutation rate across primate families (μ). Large solid circles denote median, and horizontal bars denote the interquartile range. The gray line denotes the primate-wide median. (**B**) Distribution of N_e estimates across primate families. Species with effective population size above 3×10^5 are highlighted. (**C**) Comparison of pedigree-based estimates of μ for great apes (*79*, *80*), olive baboon (*81*), rhesus macaque (*82*), and common marmoset (*83*) show a high correlation between the two estimates (Spearman's $r = 0.85$, $p = 0.02$). The open circle denotes the estimate for the mouse lemur (*84*), which was excluded from the comparison as an outlier (*16*). Data for trio estimates were derived from (*85*). (**D**) Positive correlation between estimates of per-generation mutation rates and generation times (g) (Pearson's $r = 0.53$, $p = 2.1 \times 10^{-17}$). (**E**) Inverse relationship between yearly mutation rate and generation time. Circles in (D) and (E) are colored by the effective population size N_e (Pearson's $r = -0.34$, $p = 3.1 \times 10^{-7}$). (**F**) Relationship between per-generation mutation rate, adjusted by first regressing the effects of generation time, and effective population size. The relationship is highly significant after phylogenetic correction ($r^2 = 0.45$, $p < 0.001$).

within primates, although body mass is significant before accounting for multiple testing. These relationships have been previously described, albeit for broader evolutionary distances, including a wider range of genetic diversity and body mass (*64*, *65*). We additionally calculated the relationship of the traits above to our mutation rate estimates. After correcting for multiple testing, we did not find any significant predictors of μ.

Variants specific to the human lineage

Finally, we revisited a previously published catalog of 647 high-frequency human-specific missense changes, i.e., amino acid–altering variants that putatively emerged specifically in the human lineage and quickly rose to high frequency or fixation (*66*). This catalog was mainly defined by looking at derived sites segregating at high frequency in anatomically modern humans, at which archaic hominins (Neanderthals and Denisovans) carry the ancestral allele. Although insufficient to explain the whole spectrum of human uniqueness, such a catalog should contain prime candidates for some of its molecular underpinnings. We sought to determine how often the putatively human-specific derived allele occurs at orthologous positions across the genomes of other primate species analyzed in this study. We find 63% (406) of high-frequency human-specific missense changes to occur in at least one other primate species and 55% in more than two, segregating at high frequency (>0.9) within the sampled individuals of a species (Fig. 5). This suggests that mutational recurrence generally might be widespread across primates. We find mutation pairs in recurrent high-frequency human-specific missense changes enriched in T-C and A-G mutations, and to a lesser extent in C-T and G-A compared with nonrecurrent ones.

We leveraged our data to generate a more stringent picture of the mutations that arose specifically in the human lineage and have not emerged elsewhere in primates. We identified alleles present in anatomically modern humans at a frequency of at least 99.9% that differ in state from a set of four high-coverage archaic hominins genomes (*67*–*70*). We ensured that the human allele represents the derived state by requiring the ancestral allele to be present at a frequency of >99% in a genetic diversity panel of 139 previously published great ape genomes (*8*, *9*, *71*, *72*). The resulting 24,374 candidates include a conservative set of 124 missense coding mutations affecting 107 different genes, among which are 17 previously undescribed changes affecting 12 genes (*66*).

We further sought to detect which genes have not shown frequent allele recurrence in other primate species. To this end, we removed variants that we found to reoccur in >1% of

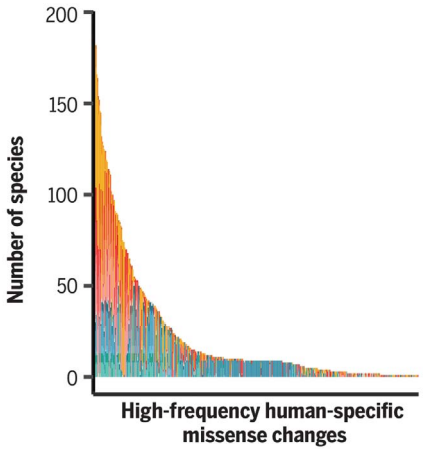

Fig. 5. Recurrent putative high-frequency human-specific missense changes. Each bar on the *x* axis represents a high-frequency human-specific missense change with the same allele found in a different species. Color schemes are the same as presented in Figs. 1 and 2.

species at a frequency of >0.1%. In this set, we find 89 missense changes, affecting 80 distinct genes. We observe no enrichment for functional categories or association to diseases among them. Within our catalog, we also find the two amino acid differences with demonstrated functional differences between humans and Neanderthals: The ancestral allele in *NOVA1* (neuro-oncological ventral antigen 1) leads to a slower development of cortical organoids and modifies synaptic protein interactions (*73*); the human-derived allele of the adenylosuccinate lyase gene (*ADSL*) leads to a reduced de novo synthesis of purines in the brain (*74*). Furthermore, changes in mitotic spindle-associated genes previously reported to be under positive selection (*SPAG5, KIF18A*) maintain their status as distinctively human (*75*). This may have had an impact on neurogenesis during development (*76*), although this hypothesis has not been experimentally validated. We find a specifically human change in *TMPRSS2*, a main factor in the response to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection with known functional variants that have possibly been under selection in some human populations (*77*).

Analogous to the above, we additionally generated a catalog of sites that are fixed across great apes but differ from rhesus macaque (*Macaca mulatta*). Among these 11.2 million variants, we find 1 million without observed recurrences beyond apes, corresponding to mutations specific to the great ape lineage. These contain 3792 missense variants affecting 2970 different genes that are significantly enriched for multiple cilia-related functional categories, such as axoneme assembly, motile cilium assembly, nonmotile cilium assembly, cilium-dependent cell motility, and epithelial cilium movement involved in extracellular fluid movement, suggesting that the evolution of ape-specific features of cilia have been important in shaping the lineage leading to our own species. The disruption of normally functioning cilia can lead to an array of heterogeneous pathologies in humans, collectively known as ciliopathies. Among 187 genes with established links to different ciliopathies, we find 30% to be affected by ape-specific missense changes (*78*) ($p < 0.01$, Fisher's exact test). More generally, we also find an overall significant enrichment of genes with nonrecurrent ape-specific missense changes among genes with disease association in OMIM (Online Mendelian Inheritance in Man) ($p < 0.01$, Fisher's exact test), suggesting that—to some degree—variants that give rise to the ape-specific phenotype, and thus ultimately also to the human one, affect a greater proportion of the genes that make us susceptible to diseases than would be expected by chance.

REFERENCES AND NOTES

1. A. B. Rylands, R. A. Mittermeier, in *Primate Behavioral Ecology, 6th edition*, K. B. Strier, Ed. (Routledge, New York, 2021), pp. 407–428.
2. L. F. Kuderna, P. Esteller-Cucala, T. Marques-Bonet, *Curr. Opin. Genet. Dev.* **62**, 65–71 (2020).
3. J. D. Orkin, L. F. K. Kuderna, T. Marques-Bonet, *Annu. Rev. Anim. Biosci.* **9**, 103–124 (2021).
4. Chimpanzee Sequencing and Analysis Consortium, *Nature* **437**, 69–87 (2005).
5. D. P. Locke *et al.*, *Nature* **469**, 529–533 (2011).
6. A. Scally *et al.*, *Nature* **483**, 169–175 (2012).
7. K. Prüfer *et al.*, *Nature* **486**, 527–531 (2012).
8. J. Prado-Martinez *et al.*, *Nature* **499**, 471–475 (2013).
9. A. Nater *et al.*, *Curr. Biol.* **27**, 3576–3577 (2017).
10. Z. N. Kronenberg *et al.*, *Science* **360**, eaar6343 (2018).
11. R. A. Gibbs *et al.*, *Science* **316**, 222–234 (2007).
12. W. C. Warren *et al.*, *Science* **370**, eabc6617 (2020).
13. C. Xue *et al.*, *Genome Res.* **26**, 1651–1662 (2016).
14. A. Estrada *et al.*, *Sci. Adv.* **3**, e1600946 (2017).
15. C. S. Mantyka-Pringle *et al.*, *Biol. Conserv.* **187**, 103–111 (2015).
16. See supplementary materials.
17. Zoonomia Consortium, *Nature* **587**, 240–245 (2020).
18. L. Wang *et al.*, *Gigascience* **8**, giz098 (2019).
19. Z. Liu *et al.*, *Mol. Biol. Evol.* **37**, 952–968 (2020).
20. B. J. Evans *et al.*, *R. Soc. Open Sci.* **4**, 170351 (2017).
21. Z. Fan *et al.*, *Mol. Phylogenet. Evol.* **127**, 376–386 (2018).
22. D. Vanderpool *et al.*, *PLOS Biol.* **18**, e3000954 (2020).
23. L. Yu *et al.*, *Nat. Genet.* **48**, 947–952 (2016).
24. N. Osada, K. Matsudaira, Y. Hamada, S. Malaivijitnond, *Genome Biol. Evol.* **13**, evaa209 (2021).
25. E. E. Louis *et al.*, Lepilemur septentrionalis. The IUCN Red List of Threatened Species 2020; https://dx.doi.org/10.2305/IUCN.UK.2020-2.RLTS.T11622A115567059.en.
26. C. Coudrat, B. Rawson, P. Phiaphalath, F. Pengfei, C. Roos, M. H. Nguyen, IUCN Red List of Threatened Species: Nomascus concolor. *IUCN Red List of Threatened Species* (2015); https://www.iucnredlist.org/species/39775/17968556.
27. E. F. Sørensen *et al.*, *Science* **380**, eabn8153 (2023).
28. H. Gao *et al.*, *Science* **380**, eabn8197 (2023).
29. T. van der Valk *et al.*, *Mol. Biol. Evol.* **37**, 183–194 (2020).
30. K. M. Detwiler, *Int. J. Primatol.* **40**, 28–52 (2019).
31. Y. A. de Jong, T. M. Butynski, *Primate Conserv.* **25**, 43–56 (2010).
32. H. Svardal *et al.*, *Nat. Genet.* **49**, 1705–1713 (2017).
33. D. Spielman, B. W. Brook, R. Frankham, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15261–15264 (2004).
34. J. C. Teixeira, C. D. Huber, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015096118 (2021).
35. T. van der Valk, D. Díez-Del-Molino, T. Marques-Bonet, K. Guschanski, L. Dalén, *Curr. Biol.* **29**, 165–170.e6 (2019).
36. J. A. Robinson *et al.*, *Sci. Adv.* **5**, eaau0757 (2019).
37. B. C. Faircloth *et al.*, *Syst. Biol.* **61**, 717–726 (2012).
38. S. Naser-Khdour, B. Q. Minh, W. Zhang, E. A. Stone, R. Lanfear, *Genome Biol. Evol.* **11**, 3341–3352 (2019).
39. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, *Mol. Biol. Evol.* **35**, 518–522 (2018).
40. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, *Nat. Methods* **14**, 587–589 (2017).
41. P. Perelman *et al.*, *PLOS Genet.* **7**, e1001342 (2011).
42. M. S. Springer *et al.*, *PLOS ONE* **7**, e49521 (2012).
43. M. D. Reis *et al.*, *Syst. Biol.* **67**, 594–615 (2018).
44. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, *BMC Bioinformatics* **19** (suppl 6), 153 (2018).
45. D. de Vries, R. M. D. Beck, *Palaeontol. Electron.* **26**, 1–52 (2023).
46. J. B. Rossie, X. Ni, K. C. Beard, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4381–4385 (2006).
47. J. S. Zijlstra, L. J. Flynn, W. Wessels, *J. Hum. Evol.* **65**, 544–550 (2013).
48. Y. Chaimanee, R. Lebrun, C. Yamee, J.-J. Jaeger, *Proc. Biol. Sci.* **278**, 1956–1963 (2011).
49. X. Ni, Q. Li, L. Li, K. C. Beard, *Science* **352**, 673–677 (2016).
50. J. Tung, L. B. Barreiro, *Curr. Opin. Genet. Dev.* **47**, 61–68 (2017).
51. C. Fontsere, M. de Manuel, T. Marques-Bonet, M. Kuhlwilm, *BioEssays* **41**, e1900123 (2019).
52. J. Sukumaran, L. L. Knowles, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 1607–1612 (2017).
53. M. C. Janiak *et al.*, *Mol. Ecol.* **31**, 3888–3902 (2022).
54. C. Roos *et al.*, *BMC Evol. Biol.* **11**, 77 (2011).
55. M. G. M. Lima *et al.*, *Mol. Phylogenet. Evol.* **124**, 137–150 (2018).
56. M. Chintalapati, P. Moorjani, *Curr. Opin. Genet. Dev.* **62**, 58–64 (2020).
57. T. Ohta, *Nature* **246**, 96–98 (1973).
58. W. Sung, M. S. Ackerman, S. F. Miller, T. G. Doak, M. Lynch, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18488–18492 (2012).
59. M. Lynch, *Trends Genet.* **26**, 345–352 (2010).
60. J. M. Kamilar, N. Cooper, *Philos. Trans. R. Soc. London B Biol. Sci.* **368**, 20120341 (2013).
61. A. R. DeCasien, S. A. Williams, J. P. Higham, *Nat. Ecol. Evol.* **1**, 112 (2017).
62. S. Shultz, C. Opie, Q. D. Atkinson, *Nature* **479**, 219–222 (2011).
63. B. Charlesworth, *Nat. Rev. Genet.* **10**, 195–205 (2009).
64. H. Ellegren, N. Galtier, *Nat. Rev. Genet.* **17**, 422–433 (2016).
65. J. Romiguier *et al.*, *Nature* **515**, 261–263 (2014).
66. M. Kuhlwilm, C. Boeckx, *Sci. Rep.* **9**, 8463 (2019).
67. K. Prüfer *et al.*, *Science* **358**, 655–658 (2017).
68. F. Mafessoni *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15132–15136 (2020).
69. R. E. Green *et al.*, *Science* **328**, 710–722 (2010).
70. D. Reich *et al.*, *Nature* **468**, 1053–1060 (2010).
71. Y. Xue *et al.*, *Science* **348**, 242–245 (2015).
72. M. de Manuel *et al.*, *Science* **354**, 477–481 (2016).
73. C. A. Trujillo *et al.*, *Science* **371**, eaax2537 (2021).
74. V. Stepanova *et al.*, *eLife* **10**, e58741 (2021).
75. S. Peyrégne, J. Kelso, B. M. Peter, S. Pääbo, *eLife* **11**, e75464 (2022).
76. S. Pääbo, *Cell* **157**, 216–226 (2014).
77. S. Jeon *et al.*, *Mol. Cells* **44**, 680–687 (2021).
78. J. F. Reiter, M. R. Leroux, *Nat. Rev. Mol. Cell Biol.* **18**, 533–547 (2017).
79. S. Besenbacher, C. Hvilsom, T. Marques-Bonet, T. Mailund, M. H. Schierup, *Nat. Ecol. Evol.* **3**, 286–292 (2019).
80. M. D. Kessler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2560–2569 (2020).
81. F. L. Wu *et al.*, *PLOS Biol.* **18**, e3000838 (2020).
82. L. A. Bergeron *et al.*, *Gigascience* **10**, giab029 (2021).
83. C. Yang *et al.*, *Nature* **594**, 227–233 (2021).
84. C. Ryan Campbell *et al.*, bioRxiv (2020), p. 724880.
85. L. A. Bergeron *et al.*, *eLife* **11**, e73577 (2022).

ACKNOWLEDGMENTS

The authors thank the Veterinary and Zoology staff at Wildlife Reserves Singapore for help in obtaining the tissue samples, as well

as the Lee Kong Chian Natural History Museum for storage and provision of the tissue samples. We thank H. Doddapaneni, D. M. Muzny, and M. C. Gingras for their support of sequencing at the Baylor College of Medicine Human Genome Sequencing Center. We appreciate the support of R. Gibbs, director of HGSC, for this project and thank Baylor College of Medicine for internal funding. We thank P. Karanth (IISc) and H. N. Kumara (SACON) for collecting and providing some of the samples from India. We acknowledge the support provided by the Council of Scientific and Industrial Research (CSIR), India, to G.U. for the sequencing at the Centre for Cellular and Molecular Biology (CCMB), India. Silhouettes in Fig. 3 were obtained from phylopic.org. The silhouette for *Propithecus* is credited to Terpsichores and has been published under CC BY-SA 3.0. All other silhouettes are under public domain. This is Duke Lemur Center publication #1559. E.F.D. thanks the Ministry of Production and the Environment of Formosa Province in Argentina for the research presented here. Samples from Amazônia, Brazil, were accessed under SisGen no. A8F3D55. **Funding:** L.F.K.K. was supported by an EMBO STF 8286. M.K. was supported by "la Caixa" Foundation (ID 100010434), fellowship code LCF/BQ/PR19/11700002, and by the Vienna Science and Technology Fund (WWTF) [10.47379/VRG20001]. J.D.O. was supported by "la Caixa" Foundation (ID 100010434) and the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 847648. The fellowship code is LCF/BQ/PI20/11760004. F.E.S. has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 801505. FES also received funds from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (Process nos.: 303286/2014-8, 303579/2014-5, 200502/2015-8, 302140/2020-4, 300365/2021-7, 301407/2021-5, 301925/2021-6), International Primatological Society (Conservation grant), The Rufford Foundation (14861-1, 23117-2, 38786-B), the Margot Marsh Biodiversity Foundation (SMA-CCO-G0023, SMA-CCOG0037), and Primate Conservation Inc. (no. 1713 and no. 1689). Fieldwork for samples collected in the Brazilian Amazon was funded by grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq/SISBIOTA Program 563348/2010-0), Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM/SISBIOTA 2317/2011), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES AUX 3261/2013) to IPF. Sampling of nonhuman primates in Tanzania was funded by the German Research Foundation (KN1097/3-1 to S.K. and RO3055/2-1 to C.R.) and by the US National Science Foundation (BNS83-03506 to J.P.-C.). No animals in Tanzania were sampled purposely for this study. Details of the original study on *Treponema pallidum* infection can be requested from S.K. Sampling of baboons in Zambia was funded by US NSF grant BCS-1029451 to J.P.-C., C.J.J., and J.R. The research reported in this manuscript was also funded by the Vietnamese Ministry of Science and Technology's Program 562 (grant no. ĐTĐL. CN-64/19) to M.D.L. A.N.C. is supported by PID2021-127792NB-I00 funded by MCIN/AEI/10.13039/501100011033 (FEDER Una manera de hacer Europa)" and by "Unidad de Excelencia María de Maeztu", funded by the AEI (CEX2018-000792-M) and Departament de Recerca i Universitats de la Generalitat de Catalunya (GRC 2021 SGR 0467). A.D.M. was supported by the National Sciences and Engineering Research Council of Canada and Canada Research Chairs program. T.M.B. is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 864203), PID2021-126004NB-100 (MICIIN/FEDER, UE) and Secretaria d'Universitats i Recerca and CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2021 SGR 00177). M.C.J, D.d.V. I.G., R.M.D.B., and J.P.B. were supported by a UKRI NERC standard grant (NE/T000341/1). S.M.A. was supported by a BINC fellowship from the Department of Biotechnology (DBT), India. *Aotus azarae* samples from Argentina where obtained with grant support to E.F.-D. from the Zoological Society of San Diego, Wenner-Gren Foundation, the L.S.B. Leakey Foundation, the National Geographic Society, the US National Science Foundation (NSF-BCS-0621020, 1232349, 1503753, 1848954; NSF-RAPID-1219368, NSF-FAIN-1952072; NSF-DDIG-1540255; NSF-REU 0837921, 0924352, 1026991), and the US National Institute on Aging (NIA- P30 AG012836-19, NICHD R24 HD-044964-11). J.H.S. was supported in part by the NIH under award number P40OD024628 - SPF Baboon Research Resource. K.G. was supported by the Swedish Research Council VR (2020-03398). This research is supported by the National Research Foundation Singapore under its National Precision Medicine Programme (NPM) Phase II Funding (MOH-000588) and administered by the Singapore Ministry of Health's National Medical Research Council. **Author contributions:** Conceptualization: T.M.B., K.K.-H.F., J.R. Methodology & analysis: L.F.K.K., H.G., M.C.J., M.K., J.D.O., S.M., A.V., J.B., M.R., R.M.D.B., T.M.B., K.K.-H.F., J.R., T.B., Y.S., L.Z., J.G.S., D.d.V., I.G., A.J., J.P.B., M.R., R.A.H. Fieldwork & sample acquisition: J.P.B., C.R., G.U., K.G., F.E.S., F.R.D.-M., F.B., H.B., I.S., I.F., J.V., M.M., M.N.F.d.S., M.T., R.R., T.H., A.M., D.Z., A.C.K., W.K.L., C.C.K., P.T., J.L., S.M., M.D.L., S.K., J.D.K., F.S., E.F.-D., J.H.S., C.A., G.W., J.P.-C., C.J.J., A.Z., C.J.R., N.A., C.H., P.F., I.S.C., J.H., J.R. Topic section leaders: L.F.K.K., J.P.B., M.H.S., R.M.D.B., K.G., C.R., G.U., A.M., T.M.B. Sequencing: L.A., M.G., J.E.H., J.B., G.U., E.L., R.A.H., M.R. Supervision: T.M.B., K.K.-H.F., J.R., M.H.S., R.M.D.B., G.Z., D.W., D.J., J.P.B. Writing – original draft: L.F.K.K., T.M.B. Writing – review and editing: All authors. **Competing interests:** L.F.K.K., H.G., J.G.S., and K.K.-H.F. are employees of Illumina Inc. as of the submission of this manuscript. **Data and materials availability:** All sequencing data have been deposited at the European Nucleotide Archive under the accession number PRJEB49549.

SUPPLEMENTARY MATERIALS

View/request a protocol for this paper from *Bio-protocol*.

Submitted 19 December 2021; accepted 6 February 2023
10.1126/science.abn7829