Genomic signatures of high-altitude adaptation and chromosomal polymorphism in geladas

Kenneth L. Chiou[1,2,3,4], Mareike C. Janiak[5,6], India A. Schneider-Crease[1,2,3], Sharmi Sen[7], Ferehiwot Ayele[8], Idrissa S. Chuma[9], Sascha Knauf[10,11], Alemayehu Lemma[8], Anthony V. Signore[12], Anthony M. D'Ippolito[13,14], Belayneh Abebe[15], Abebaw Azanaw Haile[16], Fanuel Kebede[16], Peter J. Fashing[17,18], Nga Nguyen[17,18], Colleen McCann[19,20], Marlys L. Houck[21], Jeffrey D. Wall[22], Andrew S. Burrell[23], Christina M. Bergey[24], Jeffrey Rogers[25], Jane E. Phillips-Conroy[26,27], Clifford J. Jolly[20,23], Amanda D. Melin[5,28,29], Jay F. Storz[12], Amy Lu[30], Jacinta C. Beehner[7,31], Thore J. Bergman[31,32], Noah Snyder-Mackler[1,2,3,4,33]

1. Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA.
2. School of Life Sciences, Arizona State University, Tempe, AZ, USA.
3. Department of Psychology, University of Washington, Seattle, WA, USA.
4. Nathan Shock Center of Excellence in the Basic Biology of Aging, University of Washington, Seattle, WA, USA.
5. Department of Anthropology and Archaeology, University of Calgary, Calgary, AB, Canada.
6. School of Science, Engineering, & Environment, University of Salford, Salford, UK.
7. Department of Anthropology, University of Michigan, Ann Arbor, MI, USA.
8. College of Veterinary Medicine and Agriculture, Addis Ababa University, Debre Zeit, Ethiopia.
9. Veterinary Unit, Conservation Science Department, Tanzania National Parks (TANAPA), Arusha, Tanzania.
10. Work Group Neglected Tropical Diseases, Infection Biology Unit, German Primate Center, Leibniz Institute for Primate Research, Göttingen, Germany.
11. Institute for International Animal Health/One Health, Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Greifswald, Island Riems, Germany.
12. School of Biological Sciences, University of Nebraska, Lincoln, NE, USA.
13. University Program in Genetics and Genomics, Duke University, Durham, NC, USA.
14. Center for Genomic and Computational Biology, Duke University, Durham, NC, USA.
15. African Wildlife Foundation, Debark, Ethiopia.
16. Ethiopian Wildlife Conservation Authority, Addis Ababa, Ethiopia.
17. Department of Anthropology and Environmental Studies Program, California State University Fullerton, Fullerton, CA, USA
18. Centre for Ecological and Evolutionary Synthesis, Department of Biosciences, University of Oslo, Oslo, Norway.
19. Department of Mammals, Bronx Zoo, Wildlife Conservation Society, New York, NY, USA.
20. New York Consortium in Evolutionary Primatology, New York, NY, USA.
21. Beckman Center for Conservation Research, San Diego Zoo Wildlife Alliance, Escondido, CA, USA.
22. Institute for Human Genetics, University of California San Francisco, San Francisco, CA, USA.
23. Department of Anthropology, New York University, New York, NY, USA.
24. Department of Genetics, Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ, USA.
25. Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.
26. Department of Neuroscience, Washington University School of Medicine, St. Louis, MO, USA.
27. Department of Anthropology, Washington University, St. Louis, MO, USA.
28. Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada.
29. Department of Medical Genetics, University of Calgary, Calgary, AB, Canada.
30. Department of Anthropology, Stony Brook University, Stony Brook, NY, USA.
31. Department of Psychology, University of Michigan, Ann Arbor, MI, USA.
32. Department of Ecology & Evolution, University of Michigan, Ann Arbor, MI, USA.
33. Center for Studies in Demography & Ecology, University of Washington, Seattle, WA, USA.
Corresponding authors: Kenneth L. Chiou (chiou@asu.edu) and Noah Snyder-Mackler (nsnyderm@asu.edu)

Abstract

Primates have adapted to numerous environments and lifestyles, but very few species are native to high elevations. Here, we investigated high-altitude adaptations in the gelada (*Theropithecus gelada*), a monkey endemic to the Ethiopian Plateau. We examined genome-wide variation in conjunction with measurements of hematological and morphological traits. Our new gelada reference genome is highly intact and assembled at chromosome-length levels. Unexpectedly, we identified a chromosomal polymorphism in geladas that could potentially contribute to reproductive barriers between populations. Compared to baboons at low altitude, we found that high-altitude geladas exhibit significantly expanded chest circumferences, potentially allowing for greater lung surface area for increased oxygen diffusion. We identified gelada-specific amino acid substitutions in the alpha-chain subunit of adult hemoglobin but found that gelada hemoglobin does not exhibit markedly altered oxygenation properties compared to lowland primates. We also found that geladas at high altitude do not exhibit elevated blood hemoglobin concentrations, in contrast to the normal acclimatization response to hypoxia in lowland primates. The absence of altitude-related polycythemia suggests that geladas are able to sustain adequate tissue-oxygen delivery despite environmental hypoxia. Finally, we identified numerous genes and genomic regions exhibiting accelerated rates of evolution, as well as gene families exhibiting expansions in the gelada lineage, potentially reflecting altitude-related selection. Our findings lend insight into putative mechanisms of high-altitude adaptation while suggesting promising avenues for functional hypoxia research.

Main text

Life at high altitude is associated with myriad environmental challenges including cold temperatures and reduced oxygen availability due to low barometric pressure. Consequently, organisms at high altitude have encountered strong evolutionary pressure to adapt to these challenges. Human populations living at high altitude, for example, have evolved physiological adaptations to hypoxia [1,2], providing compelling examples of strong directional selection operating over short evolutionary time frames.

Human populations began living at high altitude quite recently, from as little as 150 years to as long as 47,000 years ago [3]. This time frame pales in comparison to that of many nonhuman animals that have lived at high-altitude for millions of years. Such lineages would be expected to exhibit a greater number of fixed genetic and phenotypic differences relative to their closest lowland counterparts and provide a valuable comparative opportunity for understanding mechanisms underlying the evolution of high-altitude adaptations in humans and other animals. Comparative perspectives are particularly valuable for identifying both the shared and divergent routes that natural selection has taken at the nucleotide, protein, and pathway levels to facilitate adaptations to high-altitude life [4,5].

The gelada (*Theropithecus gelada*) is a cercopithecoid monkey—closely related to baboons (*Papio* spp.) and *Lophocebus*/*Rungwecebus* mangabeys [6,7]—endemic to Ethiopia (Fig. 1a–b). It is the only surviving member of the genus *Theropithecus*, which was found from South Africa to as far as Spain, Italy, and India up to 1 million years ago [8,9]. Geladas likely avoided the fate of their extinct congenerics by exploiting an extreme environment over the past several million years: the grassy plateaus of the Ethiopian highlands [10]. Consequently, geladas have adopted primarily grass-eating diets and are found mainly at elevations from 2,350 to 4,550 meters above sea level (Fig. 1c) [11], representing one of the highest altitudinal ranges of any extant primate species, matched only by some *Rhinopithecus* monkeys [12].

Perspectives on gelada high-altitude adaptations are particularly important given their close evolutionary affinity and shared biology with humans and may lend insights into treatments for diseases and disorders

associated with high altitude [13], including acute mountain sickness, high-altitude cerebral edema, and high-altitude pulmonary edema in high-altitude travelers, as well as chronic mountain sickness and preeclampsia in high-altitude residents. Furthermore, given the roles of hypoxia and ischemia in low-altitude diseases [14], a deeper understanding of mechanisms conferring resilience to hypoxia has the potential to inform treatment of diseases at all altitudes [15].

We sequenced and assembled the first gelada reference genome and combined it with detailed physiological, demographic, and morphological data collected from wild geladas to identify adaptations to their high-altitude environment. Through comparison with other mammals, we identified unique genomic adaptations to high altitude. We also analyzed population resequencing data from 70 wild and captive geladas to infer gelada population structure and demographic history. Curiously, we identified a chromosomal fission event that is polymorphic and possibly fixed between gelada populations. While sufficient breeding data are limited in captivity and precluded by lack of contact in the wild, given that heterozygous karyotypes tend to be associated with reduced reproductive success [16], this observation suggests a possible case of incipient speciation in primates with important conservation implications.

Sequencing, synteny, and annotation

We sequenced and assembled the genome of a wild adult female gelada from the Simien Mountains, Ethiopia, using a combination of two technologies: the linked-read 10x Genomics Chromium system [17] and Hi-C [18,19]. Initial assembly of the 10x linked-read data (55.7-fold coverage) yielded a highly intact assembly (contig N50: 134.4 Kb; scaffold N50: 57.3 Mb), which was substantially improved by incorporating Hi-C intrachromosomal contact data to produce a reference assembly with chromosome-length scaffolds and comparable contiguity and coverage to other recent nonhuman primate genomes (contig N50: 310.1 Kb; scaffold N50: 130.2 Mb; Extended Data Fig. 1a). A BUSCO analysis of genome completeness identified 12,267 genes—of which 12,098 are one-to-one orthologs—comprising 91.7% and 89.0% of expected genes present and complete in mammals and primates, respectively [20] (Extended Data Fig. 1c). In total, our assembly includes 20,683 protein-coding genes annotated by NCBI [21]. The gelada genome is highly syntenic with closely related genomes, showing strong collinearity with the anubis baboon (*Papio anubis*) genome (Panu 3.0 [22]) (Extended Data Fig. 1b).

Novel centric fission and putative incipient speciation

In assembling the genome of our reference individual, we identified an unexpected karyotype, 2n=44, that was not present in any other species in the papionin clade (macaques, drills/mandrills, mangabeys, baboons), which dates back to approximately 12 mya [23] and otherwise exhibits a conserved count of 21 chromosome pairs [24] (Fig. 2a). Our reference individual was homozygous for a centric fission [25] of chromosome 7, resulting in two new acrocentric chromosomes that we refer to as 7a and 7b (Fig. 2b,c). A single case of an apparently identical variant was previously reported in a captive gelada individual who was heterozygous for this variant (2n=43) but was interpreted as a rare structural anomaly in papionins [26]. We confirmed the homozygous 2n=44 karyotype via G-banding of fibroblasts in our reference individual and 3 additional unrelated individuals from the northernmost gelada population (Fig. 2c), demonstrating instead that this fissioned chromosome is a stable, possibly fixed variant in the Northern population. This variant appears to be unique to Northern geladas, with 2/2 wild geladas from central Ethiopia and 7/9 captive geladas from zoos—mainly of Central origin (Extended Data Fig. 2)—exhibiting the ancestral karyotype of 2n=42 (Fig. 2c and Extended Data Fig. 3). Two zoo geladas were heterozygous (2n=43), indicating a recent ancestor with a homozygous 2n=44 karyotype. These two heterozygous individuals had the most Northern ancestry (>10%) and the only Northern mitochondrial haplotype of all captive samples, suggesting that Northern ancestry can be traced through their maternal line (Fig. 2d and

Extended Data Fig. 2). Together, the evidence is consistent with our hypothesis that the fissioned chromosome is a uniquely Northern trait. The two individuals with heterozygous karyotypes had estimated proportions of Northern ancestry (0.241 and 0.115) consistent with them being F_2 (expected 0.250) and F_3 hybrids (expected 0.125), respectively, indicating that at least some heterokaryotypes are viable. Despite having opportunity, however, neither individual successfully reproduced in captivity, suggesting that the heterozygous karyotype could be associated with reproductive incompatibilities in hybrids, as is the case with other balanced chromosomal polymorphisms [27]. This conclusion remains speculative given the scarcity of detailed breeding data compounded by the low representation of Northern ancestry—and only three documented instances to date of the fissioned chromosome 7 variant—in captive geladas.

If heterozygous karyotypes are associated with reduced reproductive success [16], however, the novel chromosomal variant could represent a barrier to hybridization underlying speciation between Northern and Central geladas. These groups are typically considered subspecies—*T. gelada gelada* (Northern) and *T. gelada obscurus* (Central)—but show evidence of being distinct evolutionary units that would qualify as species under the phylogenetic species concept [28]. If the heterokaryotype is associated with reduced fitness, these populations would further qualify as species under the biological species concept, cementing the case for taxonomic revision and reconsideration of conservation priorities. Furthermore, this would provide a possible case study of chromosomal rearrangements underlying speciation, the mechanisms of which remain poorly understood [29,30].

Conservation and population genomics

To better understand the demographic history of geladas, including historical population sizes and population divergence, we sequenced the whole genomes of 70 captive and wild geladas from multiple parts of Ethiopia (n=3 wild Central geladas; n=50 wild Northern geladas; n=17 captive geladas of Central origin) as well as 20 hamadryas baboons from Filoha, Ethiopia [31] (median coverage = 11.5x; Supplementary Table 1). Our sample did not include any individuals from the Southern gelada population, which is difficult to access but represents another distinct evolutionary unit [28] (Fig. 1a).

The geladas in our sample fell into two distinct populations corresponding to previously described subspecies: the Northern population, which encompasses all wild individuals from the Simien Mountains, and the Central population, which encompasses wild individuals from Guassa Community Conservation Area as well as the majority of individuals from zoos (Fig. 2d; based on unsupervised clustering [32]). A small number (n = 3) of zoo individuals showed elevated (>9%) fractions of Northern ancestry, including the two zoo animals found to have 2n=43 heterozygous karyotypes. These cases are likely explained by captive breeding of parents from different populations. We found no evidence of interbreeding between wild gelada populations and our demographic analysis indicated that the effective population sizes of the two gelada populations (Northern and Central) began to diverge about 500 thousand years ago (Fig. 3a). It is therefore most likely that the chromosomal fission arose in Northern geladas following this population divergence.

We found higher genetic diversity in the Northern gelada population than the Central gelada population (Fig. 3b). Central geladas had lower median heterozygosity than Northern geladas (Wilcoxon, P = 8.7e-07) and also significantly longer runs of homozygosity, specifically for runs <1 Mb (Wilcoxon, P = 8.5e-08) and 1–3 Mb (P = 0.008). Both gelada populations were significantly less genetically diverse than hamadryas baboons (median heterozygosity: P = 2.6e-08 [Northern], P = 7.4e-06 [Central]), perhaps reflecting their limited geographic distribution and habitat discontinuity compared to baboons.

Physiological adaptations to high-altitude hypoxia

Hemoglobin-oxygen affinity

Many animals that have adapted to high altitude have evolved an increased affinity of oxygen to hemoglobin, which can minimize the decline in arterial oxygen saturation in spite of environmental hypoxia [33]. We therefore examined the genes encoding the alpha- and beta-chain subunits of adult hemoglobin in geladas. We found two amino acid substitutions in hemoglobin-alpha, at sites 12 and 23, that are unique to geladas relative to other primates (Fig. 4a). These substitutions, along with all coding sequences for this protein, were invariant across all 70 geladas that were sequenced in this study. To test whether these substitutions alter functional properties of the protein, we measured hemoglobin-oxygen binding affinity from purified adult hemoglobin of geladas, humans, and three species of baboons (Supplementary Table 8). In the presence of allosteric cofactors—the experimental condition most relevant to *in vivo* conditions—we found no differences in P_{50} (the partial pressure at which hemoglobin is 50% saturated) of gelada hemoglobin compared to that of humans ($P = 0.053$) or baboons ($P = 0.950$) (Fig. 4b). Thus, the amino acid substitutions found in gelada hemoglobins do not appear to be associated with increased hemoglobin-oxygen affinity, in contrast with the pattern generally observed in high-altitude birds [33] and some high-altitude mammals [4,34,35] but mirroring a similar lack of increased oxygen affinity in snow leopard hemoglobin [36].

Hemoglobin concentration and erythrocytosis

In lowland mammals, a typical response to chronic hypoxia is an increase in red blood cell production (erythrocytosis), which in humans is associated with chronic mountain sickness in high-altitude residents. Erythrocytosis is particularly prevalent in Andean highlanders and is altitude-dependent in its severity [37,38], while hemoglobin concentrations are not markedly elevated in highland Tibetans [37–39]. To test for erythrocytosis in geladas, we compared hemoglobin concentrations from 92 wild geladas (44 of which were not sequenced and thus not included in our genomic dataset) sampled at high altitude (3250–3600 m) to values reported from captive geladas [40] and baboons [41] at low altitude. We found that hemoglobin concentrations in geladas at high altitude were not elevated and were in fact significantly lower than hemoglobin concentrations in either captive geladas ($P = 0.005$) or baboons ($P < 0.001$). The absence of elevated hemoglobin concentrations in wild geladas living at high altitude is consistent with patterns documented in other hypoxia-adapted alpine mammals [42,43] and, among humans, most closely resembles the Tibetan phenotype (Fig. 4c). Since red blood cell production is induced by reduced oxygenation of renal tissue, the absence of an elevated hemoglobin concentration in wild geladas living at >3000 m suggests that the animals are able to sustain adequate tissue-oxygen delivery in spite of the reduced availability of oxygen. Such physiological compensation may be attributable to evolved and/or plastic changes in any number of cardiorespiratory or circulatory traits that govern oxygen transport.

Pulmonary adaptations at high altitude

We also tested the hypothesis that geladas might compensate for hypoxia by expanding their lung volumes, which is a known high-altitude developmental adaptation spurred by rapid lung growth in early life [44]. Expanded lungs in high-altitude animals maximize the pulmonary diffusing capacity for oxygen by proliferating alveolar units and increasing surface area for gas exchange [45,46]. To test for correlates of expanded lung volumes in geladas, we compared chest circumferences in wild geladas (n=78) to an extensive database of baboon morphometric measurements (n=482) [47–49]. We found that, controlling for sex and baboon species, geladas had significantly larger chest circumferences compared to baboons when also controlling for body mass ($P = 1.63e-42$), waist circumference ($P = 3.23e-31$), or both ($P = 8.58e-46$) (Fig. 4d). These results indicate that geladas have significantly expanded relative chest circumferences compared to baboons, which parallels the larger chest dimensions exhibited by native Andean highlanders [50]. This finding is consistent with the possibility of expanded lung volumes, which we

did not directly measure in this study. It is currently unknown whether these differences are plastic responses to high-altitude hypoxia or a genetically controlled component of the adaptive toolkit in geladas. Comparison to chest dimensions in captive geladas born and reared at low altitude, which are currently unavailable, would help distinguish between these two possibilities.

Genomic evidence of adaptation to high altitude

To identify signatures of adaptations to high altitude in the gelada genome, we first focused on two forms of genetic change that could underlie changes in phenotype: coding mutations that alter protein function and gene duplications that alter gene dosage and/or division of labor among protein isoforms.

We tested for evidence of positive selection in gelada coding sequences using two complementary d_N/d_S tests—the site-based model implemented in PAML [51] and the gene-based model implemented in BUSTED [52]—using a consensus chronogram [53,54]. We assigned proteins from 40 taxa (Extended Data Fig. 4 and Supplementary Table 2) to single-copy orthogroups [55] and, after filtering (Methods), included 6,105 protein-coding genes in our analysis. We identified 103 genes exhibiting significant signatures of positive selection (FDR-adjusted $P < 0.05$) using both d_N/d_S approaches.

To test for gene duplication resulting in significant expansions of gene families, we assigned proteins from 40 taxa (Extended Data Fig. 4 and Supplementary Table 2) to gene families in the TreeFam9 database [56,57]. We then tested for gene family size changes using birth-death models implemented in CAFE [58]. We identified 108 gene families exhibiting significant expansions in gene family size (FDR-adjusted $P < 0.05$).

Positive selection on protein-coding sequences

We found several compelling candidate genes for high-altitude adaptation among 103 total genes with significant signatures of positive selection (FDR-adjusted $P < 0.05$; Supplementary Table 3). These included four genes involved in the hypoxia-inducible factor (HIF) pathway (*ITGA2*, *NOTCH4, FERMT1*, *MLPH*). We also identified several that have been identified as candidate genes in human hypoxia-adapted populations, including *FRAS1*, which is involved in renal agenesis and exhibits signatures of positive selection in Tibetans [59] and Ethiopians [60], *HMBS*, which is involved in heme biosynthesis and exhibits a signature of positive selection in Nepalese Sherpa [61], and *TNRC18*, a largely unknown gene that is linked to selection in Bajau breath-hold divers [62]. Other notable candidate genes include *AQP1*, which plays an important role in fluid clearance and edema formation following acute lung injury [63], *COX15*, which is involved in heme α biosynthesis and cytochrome *c* oxidase assembly [64] and exhibits signatures of positive selection in high-altitude rhesus macaques [65], *DHCR24*, which is involved in the induction of heme oxygenase 1 (HO-1) [66] and exhibits a signature of positive selection in alpine sheep [67], and *CYGB*, which is part of the globin family and encodes an oxygen-binding respiratory protein [68].

At the pathway level, we found that signatures of positive selection were enriched (FDR-adjusted $P < 0.1$; Supplementary Table 4) for processes related to classical functions associated with high-altitude adaptation, including oxygen sensing (response to hypoxia; angiogenesis; cellular response to hypoxia) [69], response to oxidative (response to hydrogen peroxide) and other stress (response to glucocorticoid; MAPK cascade) [70,71], and female reproduction (in utero embryonic development; response to estradiol; ovulation cycle process; female pregnancy) [72–74]. In addition, we identified several enriched processes related to neural function (axon guidance; positive regulation of neuron projection development; chemical synaptic transmission; brain development), cell growth and proliferation (response to insulin; negative regulation of cell population proliferation; negative regulation of canonical Wnt signaling pathway), and cardiac function (cell-cell signaling involved in cardiac conduction).

While we found a high degree of overlap between putatively selected pathways in geladas and human populations living at high altitudes, aside from notable examples listed above, few candidate genes identified by our analysis were shared with candidate genes identified by studies of high-altitude human populations [2,15] or other high-altitude primates [12,65]. This may reflect differences between studies designed to identify adaptive changes between conspecific populations (e.g., F_{ST}-based studies) versus those designed to detect adaptive divergence between species (e.g., d_N/d_S-based studies). It may also be limited by the relative paucity of comparable genomic studies of high-altitude primates. Nevertheless, at present our results suggest that gelada adaptations to shared physiological challenges at high altitude may largely involve different suites of genes—in line with similar findings among birds [75]—underscoring their putative utility as a novel model for understanding adaptations to hypoxia.

Gene family expansion

We identified 108 gene families with significant expansion in the gelada lineage compared to 43 gene families with significant contractions (FDR-adjusted $P < 0.05$; Supplementary Table 5). Significant expansions included the genes *CENPF* and *SART1*, which each exist as a single copy in most primates, including the common ancestor of geladas and baboons, but have expanded to include five copies and four copies respectively in the gelada lineage (Extended Data Fig. 5b,c). *CENPF* encodes a kinetochore protein (CENPF) that regulates chromosome alignment and separation during mitosis and also protects centromeric cohesion [76]. Interestingly, CENPF is a marker for cell proliferation in human malignancies [77] and is strongly upregulated in response to hypoxia in bone marrow mesenchymal cell cultures [78], suggesting that it may also play a role in the response to high-altitude hypoxia. The protein encoded by *SART1* suppresses activation of HIF-1 by promoting the ubiquitination of HIF-1α. Expansion of the *SART1* family may therefore be a possible adaptation for suppressing constitutive HIF-1 activation under conditions of chronic environmental hypoxia.

We also identified biological processes that were associated with signatures of gene family expansion (FDR-adjusted $P < 0.1$; Supplementary Table 6). We found that signatures of gene expansion were significantly associated with processes related to the hypoxia response (regulation of transcription from RNA polymerase II promoter in response to hypoxia) as well as the DNA damage response (e.g., DNA repair; nucleotide-excision repair; DNA incision, 5'-to lesion; DNA duplex unwinding), which may reflect degrees of DNA damage due to elevated levels of ultraviolet radiation at high altitude [79,80]. Other enriched processes included those related to immune function (e.g., NIK/NF-kappaB signaling; stimulatory C-type lectin receptor signaling pathway; viral transcription; IL-1-mediated signaling pathway; T cell receptor signaling pathway; TNF-mediated signaling pathway), cell proliferation (e.g., Wnt signaling pathway; planar cell polarity pathway; MAPK cascade), oxidative phosphorylation (mitochondrial respiratory chain complex I assembly; mitochondrial electron transport, NADH to ubiquinone; oxidation-reduction process), and hematopoiesis (regulation of hematopoietic stem cell differentiation).

Accelerated evolution in the gelada lineage

To investigate the emergence of gelada-specific features and to expand our analysis to non-coding regions of the genome including regulatory elements [81], we identified and characterized genomic regions that are highly conserved through evolution but exhibit a greater number of changes in the gelada lineage. A similar approach has been used to identify “human accelerated regions” (HARs) that are possible hallmarks of human evolution [82,83], tend to be developmental gene regulatory elements or in non-coding RNA regions [84], and are putatively linked to uniquely human social behavior and cognition [85].

We used an approach modeled on that of Pollard *et al*. [83] to define uniquely accelerated regions in the gelada lineage, which we refer to as “gelada accelerated regions”, or GARs. We analyzed 60,345 conserved alignment blocks across a total of 57 mammalian taxa (Methods), including geladas, and

identified a total of 29 GARs (FDR-adjusted $P < 0.2$; Extended Data Fig. 6 and Supplementary Table 9). We identified fewer GARs than reported counts of HARs, which range from approximately 200–3000 at similar thresholds [83,86], likely due to differences in filtering, thresholding, and other aspects of methodology.

We found that biological processes including response to hypoxia, regulation of angiogenesis, cellular response to oxidative stress, lung development, and respiratory gaseous exchange by respiratory system were significantly enriched for signatures of acceleration (FDR-adjusted $P < 0.01$; Supplementary Table 10), indicating that classical functions related to oxygen metabolism at high altitude have undergone accelerated evolution in the gelada lineage. Other enriched processes included those related to temperature regulation (e.g., positive regulation of cold-induced thermogenesis, temperature homeostasis) and brain development (e.g., neurogenesis, brain morphogenesis)

Of the 29 GARs, 13 (44.9%) were located in intergenic regions, ten (34.5%) in introns, one (3.4%) in a 5' UTR region, and the remaining five (17.2%) in coding sequences. Many of these GARs were nominally regulatory: 13 GARs (44.9%) were associated with regulatory hallmarks of enhancer activity in at least one primary tissue or cell type in humans (Methods). Of these putative enhancers, 11 are associated with hallmarks of enhancer activity in human fetal tissues or were nearest to genes that are involved in developmental processes including in utero embryonic development and post-embryonic development. These results indicate that a large fraction of GARs may function as developmental enhancers, similar to HARs [86]. Two additional GARs (GAR5 and GAR8) are located in regions showing strong evidence of being transcriptional start sites across many human tissues (> 60 cell/tissue types each). While accelerated evolution within these regions may indicate shifts in the activity of regulatory elements that perform similar functions in geladas, this is not necessarily the case. In fact, accelerated evolution in some regions could plausibly be related to the evolution of new regulatory functions in the gelada lineage. Disentangling these competing possibilities will require high-quality profiling of putative regulatory elements in geladas as well as functional validation of these regions.

Strikingly, in two cases, multiple GARs were found near the same genes. These genes were *RBFOX1*, which was the closest gene to GAR28 and GAR29, and *ZNF536*, which was the closest gene to GAR26 and GAR27. In both cases, GARs were at least 500 kb apart from one another and in low linkage disequilibrium (mean r^2: 0.07–0.17). Both *RBFOX1* and *ZNF536* are linked to brain expression and function. The protein encoded by *RBFOX1* is an important regulator of neuronal excitation [87] while the protein encoded by *ZNF536* is a negative regulator of neuronal differentiation [88]. None of the associated regions showed hallmarks of transcription factor binding or chromatin accessibility in human tissues and cells.

Other identified GARs also were located nearest to genes involved in brain function. These genes included *RTN4RL1* (GAR21), which is involved in postnatal brain development and regulating regeneration of axons, *GIGYF2* (GAR16), which encodes a regulator of vesicular transport and IGF-1 signaling in the central nervous system [89], *CNTN4* (GAR3), which has been linked to neuropsychiatric disorders and fear conditioning [90], and *NFASC* (GAR1), which is linked to neurite outgrowth and adhesion [91]. The accelerated evolution of GARs near multiple genes related to neural function in geladas may reflect the sensitivity of the brain to the metabolic pressures of high-altitude hypoxia [92,93].

Several GARs were located nearest to genes that are involved in the response to hypoxia or oxidative stress, suggesting that they might be adaptations to high-altitude environments [71]. Two GARs were located nearest to genes—*HTATIP2* (GAR17) and a novel *LDHB*-like gene in geladas (ENSTGEG00000009621; GAR7)—predicted to be involved in oxidation reduction. One GAR was located in the intron of *RCAN1* (GAR4), which is involved in the response to oxidative stress and regulation of

angiogenesis [94]. Another GAR was located in the 5' UTR region of *FBN1* (GAR8), which is hypoxia responsive [95] and more highly expressed at higher elevations among yaks [96].

Intriguingly, we found that one GAR, GAR18, was nearest to the gene *SOX6*, which encodes a protein that plays an essential role in erythroid cell differentiation and is necessary for basal and stress erythropoiesis [97,98]. GAR18 was found 2,651 bp upstream of *SOX6* and was associated with regulatory hallmarks of enhancer activity in five primary cell types (Supplementary Table 9). Given its position and putative function as an enhancer in humans, GAR18 could suppress hypoxia-induced erythropoiesis by decreasing or disabling enhancer activity, providing a direct link to the lack of altitude-associated erythropoiesis that we observed in wild geladas.

Conclusion

The first assembled gelada genome provides novel insights into the unique adaptations of this charismatic Ethiopian primate. We identified a novel and stable karyotype that appears to be at extremely high frequency and possibly fixed in the Northern population of geladas. Given that chromosomal rearrangements tend to be associated with infertility in heterozygous karyotypes, our findings suggest that geladas may encompass at least two distinct biological species. This finding is important for at least two reasons. First, a taxonomic revision would roughly halve the populations of each gelada species and, consequently, alter their conservation status and ultimately increase resources to protect them. Second, the centric fission of chromosome 7 is an extraordinarily recent example of a stable chromosomal variant in a long-lived primate. It therefore provides a unique opportunity to study karyotypic evolution, the birth of new centromeres, and the role of chromosomal rearrangements in speciation in a primate closely related to humans. We note, however, that these conclusions are currently speculative given that data on reproductive outcomes in heterokaryotypic geladas are limited and likely to remain so due to a lack of contact between populations in the wild and the low representation of Northern ancestry in captivity.

By combining morphometric, hematological, and genomic data, we identified a suite of gelada-specific traits that may confer adaptation to their high-altitude environment, including evidence for increased lung capacity and positive selection in a number of hypoxia-related genes, and gelada-lineage-specific accelerated regulatory regions. Interestingly, while we found gelada-specific amino acid substitutions in hemoglobin, these changes did not alter oxygen-binding affinity, which highlights the need for functional assays to validate purely sequence-based findings. With this in mind, our genome assembly and gelada-specific genetic changes provide multiple avenues for future research on the function of the protein-coding and regulatory changes unique to geladas. This research thus builds upon our current understanding of the mechanisms of adaptation to extreme environments and provides an avenue for research that may have a transformative impact on the study and treatment of hypoxia-related conditions.

Methods

Animal procedures

Capture and release

Samples and data collected for this study were obtained from wild geladas in the Simien Mountains National Park (~3,000–4,550 meters above sea level) as part of continuous long-term research conducted by the Simien Mountains Gelada Research Project (SMGRP). Beginning in 2017, the SMGRP has carried out annual capture-and-release campaigns during which animals were temporarily immobilized through remote-distance injection. Briefly, a mixture of ketamine (7.5 mg/kg) and medetomidine (0.04 mg/kg) was injected using darts delivered by a blowpipe (Telinject USA, Inc). Following data and sample collection under the supervision of licensed veterinarians and veterinary technicians, immobilization was reversed with atipamezole (0.2 mg/kg). Animals were monitored by project staff throughout their recovery until they were visibly unimpaired and had returned to their social units. All research was conducted with permission of the Ethiopian Wildlife and Conservation Authority (EWCA) following all laws and guidelines in Ethiopia. Animal procedures were conducted with approval by the Institutional Animal Care and Use Committees (IACUCs) of the University of Washington (protocol 4416-01) and Arizona State University (20-1754R). This research conformed to the American Society of Primatologists/International Primatological Society Code of Best Practices for Field Primatology.

Morphometrics

While animals were sedated, we collected morphometric measurements including body mass, chest circumference, and waist circumference. Body mass was measured by a hanging digital scale to 0.05 kg precision. Chest circumference and waist circumference were measured using flexible tape to 0.1 cm precision. Chest circumference was defined as the maximum circumference of the trunk, taken at the maximum anterior projection of the thoracic cage. Waist circumference was defined as the minimum circumference between the pelvis and the thoracic cage. We report measurements from a total of 78 wild geladas sampled from 2017–2020.

Biological sample collection

Whole blood was obtained from all chemically immobilized individuals by femoral venipuncture and collected into K3 EDTA S-Monovette collection tubes (Sarstedt). 1 ml of whole blood was cryopreserved in liquid nitrogen, ~50 µl was used for hematology, and the remainder was fractionated by centrifugation using a ficoll gradient. Fibroblasts were cultured from small biopsy punches of ear tissue that were stored in RPMI supplemented with 20% FBS and 10% DMSO. To maximize viability, these samples were frozen in steps by first storing in styrofoam at -20°C overnight, then transferring to liquid nitrogen.

We also measured hemoglobin concentrations using an AimStrip 78200 digital hemoglobin meter. We loaded 10 µl of venous blood into provided test strips and recorded hemoglobin concentrations (g/dl) using the digital meter. We report measurements from a total of 92 wild geladas sampled from 2017–2020.

Other DNA sources

Apart from primary DNA samples collected for this project (n=50), we obtained additional DNA samples from other sources, including DNA extracts from 20 wild hamadryas baboons from Filoha, Ethiopia (contributed by C. Jolly and J. Phillips-Conroy), which were previously determined to have unadmixed ancestry [31], DNA extracts from 17 zoo geladas (n=1 contributed by the Wildlife Conservation Society/Bronx Zoo, n=16 contributed by the San Diego Zoo Wildlife Alliance), and muscle tissue from 3

wild geladas (contributed by N. Nguyen and P. Fashing). A full list of DNA samples used for this research is provided in Supplementary Table 1.

Sample collection, sequencing, and assembly

10x Genomics Chromium library generation and sequencing

High molecular weight DNA was extracted from cryopreserved whole blood of an adult eight year-old female gelada (DIX) from the Simien Mountains using the Gentra Puregene Blood Kit (Qiagen) following manufacturer instructions and quality-checked using pulsed-field gel electrophoresis. Linked-read libraries were then prepared using the Chromium Genome Reagent Kit v2 (10x Genomics) following manufacturer instructions. Finished libraries were sequenced to 55.7x coverage on two lanes of the Illumina HiSeq X platform using 2×150 bp sequencing.

Hi-C library generation and sequencing

Approximately seven million peripheral blood mononuclear cells (PBMCs) were isolated by ficoll gradient, washed, counted, fixed in formalin, and cryopreserved. Hi-C libraries were later prepared from cryopreserved formalin-fixed PBMCs following Rao *et al*. [18] and sequenced on the Illumina NextSeq 500 platform using 2×81 bp sequencing.

Genome assembly

Chromium-derived reads were assembled using Supernova v2.0.1 [17] with default parameters. Resulting scaffolds were then further assembled incorporating Hi-C data through the 3D *de novo* assembly (3D-DNA) pipeline v170123 [19]. Hi-C contact maps and the draft assembly with chromosome-length scaffolds were edited using Juicebox Assembly Tools [99] to correct visually apparent misjoins. Finally, gaps were closed using GapCloser v1.12 [100] to produce the final assembly (Tgel 1.0).

Transcriptome sequencing and genome annotation

RNA was obtained from cultured fibroblast cell lines derived from a biopsy ear punch taken from the same adult female gelada (DIX) and an unrelated male gelada (DRT_2017_018) from the Simien Mountains National Park, Ethiopia. Total RNA was extracted using the Quick RNA Miniprep Plus kit (Zymo Research). RNA-seq libraries were prepared using the NEBNext Ultra II RNA Library Prep kit (New England Biolabs) following instructions for a 200 bp insert. Finished libraries were sequenced on the Illumina NextSeq 500 platform using 2×81 bp sequencing.

The final assembly (Tgel 1.0) was deposited in GenBank (accession GCA_003255815.1) and annotated *de novo* by the National Center for Biotechnology Information (NCBI) using the Eukaryotic Genome Annotation Pipeline [21]. Fibroblast RNA-seq short reads were submitted to the Sequence Read Archive (accessions SRX4071585 and SRX4100999) and included in the annotation pipeline.

BUSCO assessment

To assess the completeness of the Tgel 1.0 genome assembly, we used BUSCO v4.0.6 [101] and compared our assembly against common orthologs in both the mammalian (dataset mammalia_odb10, creation date 2019-11-20) and primate (dataset primates_odb10, 2019-11-20) lineages (Extended Data Fig. 1c). BUSCO was run with default settings, using the following versions of third party components: python v3.7.4, NCBI BLAST v2.2.31, Augustus v3.2.3, HMMER v3.1b2, SEPP v4.3.10, and Prodigal v2.6.3.

Synteny analysis

We assessed the synteny between chromosomal scaffolds of Tgel 1.0 and the anubis baboon reference genome, Panu 3.0 [22], using two approaches. In the first approach, we computed pairwise alignments between genomes using nucmer from MUMmer v3.23 [102], using a cluster size of 400, a minimum match length of 10 bp, and a maximum of 500 bp between clusters. We then used the delta-filter utility program from MUMmer to retain only alignments with a minimum identity of 40% and a minimum overlap of 1% between query and reference alignments. We then plotted links between assemblies (Extended Data Fig. 1b). In the second approach, we used the reference-free method implemented in Smash v1.0 [103] to identify syntenic blocks and to visualize chromosome rearrangements (Extended Data Fig. 7). We ran Smash using default settings.

Karyotype assessment

Our Hi-C data revealed a distinct lack of contacts between scaffolds corresponding to nonoverlapping segments of chromosome 7 on the baboon reference genome (Fig. 2b). To test for a possible chromosomal fission in our reference individual, we performed G-banded karyotyping on fibroblasts cultured from the same individual, which confirmed that our reference had a homozygous karyotype with a centromeric fission in chromosome 7 (Fig 2c and Extended Data Fig. 3), resulting in two new acrocentric chromosomes that we refer to as 7a and 7b and a full karyotype of 2n=44. We tested for the presence of this centric fission in additional captive and wild geladas from zoos, northern Ethiopia (Simien Mountains), and central Ethiopia (Guassa). We counted chromosomes for zoo and northern Ethiopian geladas using karyotyping (with or without chromosome banding), taking advantage of the availability of live cells either through the Frozen Zoo® (San Diego Zoo Wildlife Alliance) or through samples collected by our project.

For wild central Ethiopian geladas, for which live cells were not available, we tested for the presence of a chromosomal fission by generating and analyzing Hi-C data from ethanol-fixed tissue samples. We generated Hi-C libraries using the Proximo Hi-C animal kit (Phase Genomics) following manufacturer instructions and sequenced them on the Illumina iSeq platform. We then ran the 3D-DNA pipeline v170123 [19] separately from each Hi-C library using our reference gelada chromosomal scaffolds as input. Resulting contact maps were then assessed for the presence/absence of contacts between 7a and 7b, both visually (Extended Data Fig. 3c) and by permutation. For permutations, we simulated the null distribution of interchromosomal contacts (i.e., contacts between distinct chromosomes excluding chromosome 7) by dividing the reference genome into 10 million base-pair windows, then randomly sampling windows without replacement until the combined sizes added up to the lengths of 7a and 7b, respectively. We then determined the frequency of Hi-C contacts between windows assigned to these simulated “chromosomes”. In all cases, we determined significant overrepresentation of contacts between arms of chromosome 7 relative to our simulated null distributions, thus rejecting the hypothesis that chromosome 7 is exclusively fissioned (i.e., 2n=44). Because the relative proportion of contacts between 7a and 7b surpassed estimates generated from baboon Hi-C data, we also rejected the possibility of a heterozygous karyotype (i.e., 2n=43), as baboons—along with all non-gelada papionins—are not known to exhibit a fissioned chromosome 7.

Hemoglobin-oxygen ($Hb\text{-}O_2$) affinity

To identify unique amino acid substitutions in geladas, we used amino acid sequences for the hemoglobin alpha (HBA) and beta (HBB) subunits from our reference assembly and aligned them to corresponding amino sequences obtained from the UniProtKB database (Fig. 4a and Supplementary Table 7). We

aligned amino sequences using Clustal Omega v1.2.4 [104] and visualized the resulting alignment using Mesquite [105] (Fig. 4a).

After discovering unique substitutions in the gelada HBA, we confirmed that these substitutions were fixed in geladas by extracting coding sequence regions for the gelada HBA protein (RefSeq ID XP_025230451.1) from a population VCF file including 70 gelada individuals (see population resequencing methods below). We found no variants across all coding sequences, indicating that proteins sampled from any individual are representative of gelada HBA.

We quantified Hb-O_2 affinity using total hemoglobin purified from hemolysates of one individual each of gelada, hamadryas baboon, Guinea baboon, anubis baboon, and human (Supplementary Table 8) using previously described methods [106]. Briefly, proteins were purified by anion-exchange FPLC, removing endogenous organic phosphates and yielding stripped samples. Using purified Hb solutions (0.4 mM heme), we measured O_2 equilibrium curves in the absence (stripped) or presence of allosteric effectors 0.1 M KCl and 2,3-diphosphoglycerate (DPG; at 2-fold molar excess). Reactions were run at 37°C in 0.1 M HEPES buffer with 0.5 mM EDTA. P_{50} values were measured at three different pH levels: ~7.2, ~7.4, and ~7.7. A linear least-squares regression comparing pH and log P_{50} was computed and the resulting equations were used to correct P_{50} values to pH 7.4 for each of the gelada, baboon (three species combined), and human datasets.

We predicted that gelada hemoglobins would display increased oxygen affinity compared to those of baboons and humans. To test these predictions, we used two approaches. First, we plotted predicted P_{50} values at pH 7.4 for each taxon with error bars ± the standard errors of the estimates (SEEs) from the respective regression models (after exponentiation to reverse the log scales for each). We then assessed the resulting error bars for overlap, with overlap indicating no statistical difference in predicted P_{50} (Fig. 4b). Second, we calculated the differences in gelada vs. baboon and gelada vs. human predicted log P_{50} at pH 7.4 as well as the standard errors of the differences, then performed one-sided *t*-tests using the equations specified by Rees & Henry [107] (case assuming homogeneity of variance; equations 3–6) to test the alternative hypotheses that gelada log P_{50} < baboon log P_{50} and gelada log P_{50} < human log P_{50}. We performed these comparisons for both the stripped condition and the condition in the presence of allosteric effectors.

Analysis of blood hemoglobin concentrations

We compared venous blood hemoglobin concentrations from this study (n=92, mean elevation ≈ 3,250 m) to corresponding measurements in zoo geladas (n=42, mean elevation ≈ 100 m) [40] and captive hamadryas baboons (n=1023, mean elevation ≈ 50 m) [41]. As the reference zoo gelada data do not differentiate sexes, we performed all comparisons with both sexes grouped together. We tested for differences of means between (1) wild geladas vs. zoo geladas and (2) wild geladas vs. captive hamadryas baboons using Welch's *t*-test with the means, standard deviations, and sample sizes of each population as input.

To validate our decision to analyze both sexes together, we replicated all findings when (1) comparing each sex from the wild to geladas of unknown sex from zoos and (2) comparing each sex from the wild to the corresponding sex in captive hamadryas baboons (Extended Data Fig. 8). As the direction of our findings did not change with any of these analyses, it is unlikely that our findings could be affected by sex-differences in hemoglobin concentrations.

For visual comparison, we plotted the means and standard deviations for hemoglobin concentration values from the Simien Mountains and zoo gelada reference values together with data from a metaanalysis of human populations across altitudes [108]. To facilitate comparison, we excluded hemoglobin concentration values from infants and juveniles and combined values between adult males and females of each human population at each reported altitude. As hemoglobin values were provided across 1 km ranges, we assigned a single elevation value as the midpoint of each 1 km range (we assigned 5,500 m for values in the category >5,000 m). We highlight differences in altitude-based hemoglobin concentrations among populations by fitting separate regression lines for (1) Tibetans and Sherpa and (2) all other human populations (Fig. 4c).

Analysis of chest circumference

We analyzed relative chest circumference in geladas by controlling for either body mass, waist circumference, or both (Fig. 4d). We combined gelada measurement data with decades of baboon measurement data collected by two authors (J. Phillip-Conroy and C. Jolly). We restricted our analysis to adults over six years of age, estimated either from dentition [47] or calculated from known birth dates, and removed baboons of mixed species ancestry. After filtering, our comparison consisted of n=78 geladas and n=482 baboons. We tested for significantly larger lung volumes in geladas by running linear models with chest circumference as the dependent variable and sex, genus (*Papio* vs. *Theropithecus*), and an interaction term between genus and species as covariates to take into account the nested nature of species within genera. Additionally, in each of three linear models, we included (1) body mass, (2) waist circumference, or (3) both body mass and waist circumference as additional covariates to control for aspects of body size. For all models, we adjusted *P* values for one-sided hypothesis testing using the *t* distribution.

Ortholog determination

We compared protein and gene sequences from the gelada genome to a dataset of 39 additional (40 total) mammalian genomes (Extended Data Fig. 4 and Supplementary Table 2) obtained from NCBI and Ensembl. Homologous relationships were determined using two approaches. In the first approach, we used a *de novo* orthology inference approach implemented in OrthoFinder [55] to assign proteins to orthogroups, which we then used to identify single-copy orthologous sequences and coding sequences under positive selection. In the second approach, we used the hidden Markov model approach implemented in HMMER3 [109] to assign proteins to curated phylogenetically based gene families in the TreeFam9 database [56,57], which we then used to identify expansions and contractions of gene families.

We used OrthoFinder to identify candidate single-copy orthologs, using the longest translation form of each gene as inputs. We defined single-copy orthogroups as orthogroups for which the number of assigned genes in any taxon was either 0 or 1. This definition takes into account the possibility that genes are missing in some of the analyzed genomes due to incomplete assembly or annotation, resulting in 0 copies.

While the longest translation forms of each gene are useful for homolog determination as they maximize the amount of sequence information, they are not necessarily optimal for alignment as they tend to introduce nonshared exons leading to a greater number of misaligned positions. To address this problem, we used the protein alignment optimizer heuristic implemented in the software PALO [110] to select optimal isoforms for the analysis. Rather than selecting the longest isoform, PALO selects isoforms that are most similar in length. Because the PALO algorithm is combinatorial by nature, the computational burden increases exponentially with the number of possible length combinations to test and the software imposes an internal limit of 100 million length combinations. For orthogroups surpassing this threshold, we thus

implemented a stepwise strategy in which we rank-ordered taxa according to their number of unique protein lengths, then ran PALO in the largest possible group of taxa for which the product of their protein counts did not exceed 100 million. After selecting one protein length per taxon for this group, we repeated the procedure as necessary until all species could be run. When multiple isoforms shared the optimum length selected by PALO for a given taxon, we selected one at random.

We used HMMER3 [109] to assign proteins to gene families in the TreeFam9 database. The longest translation forms of each gene were used as inputs and the gene family with the highest bit score was assigned to each gene.

Positive selection on protein-coding genes analysis

To identify proteins under positive selection, we generated alignments for all single-copy orthologs identified by OrthoFinder and using isoforms selected by PALO. We aligned amino acid sequences using Clustal Omega v1.2.4 [104], then generated codon alignments using the pal2nal.pl script from the PhaME toolkit [111]. We excluded all alignments for which (1) fewer than 36 taxa (< 90%) had sequences and (2) the total alignment was either less than 120 nucleotides (< 40 codons/amino acids) long or less than 25% the length of the full gelada protein. We then tested for positive selection using two d_N/d_S-based approaches: (1) branch-site models implemented in PAML v4.9 [51] and (2) gene-wide models implemented in HyPhy [112]. For our PAML analysis, we ran likelihood ratio tests on codon alignments using the “M2a” model of positive selection in the program codeml (model = 2, NSsites = 2). For our HyPhy analysis, we ran likelihood ratio tests for episodic positive selection using BUSTED [52]. For both analyses, we used a consensus chronogram downloaded from TimeTree [53,54] including all 40 taxa (Extended Data Fig. 4) as input into our models, with missing branches removed as necessary for each alignment. We corrected all *P* values using a Benjamini-Hochberg procedure [113] (Supplementary Table 3).

Gene family expansion analysis

We tested for significant gene family size changes using our previously described protein assignments to the TreeFam9 [56,57] database. Expansions and contractions were determined using CAFE 4.2 [58], which uses a probabilistic graphical model based on a random birth/death process to calculate the probability of transitions (λ) in gene family size from parent to child nodes in a phylogenetic tree. For this analysis, we allowed the program to estimate the most likely value of lambda (6.09e-4) and used a consensus chronogram from TimeTree [53,54] (Extended Data Fig. 4) as input. We used Viterbi *P* values from CAFE to assess branch-specific patterns of expansion or contraction. Because branch-specific Viterbi *P* value indicate rapid evolution but not necessarily expansion, we defined expanded gene families as those that were both larger in *T. gelada* relative to the most recent common ancestor (MRCA) and significant at a false discovery rate (FDR) threshold of 20%. Gene families that thus exhibited significant expansion were interpreted as putative targets of selection in geladas (Extended Data Fig. 5 and Supplementary Table 5).

Gene Ontology enrichment analyses

We performed Gene Ontology (GO) [114,115] enrichment analyses in order to identify biological processes that are differentially associated with signatures of positive selection and gene family expansion.

For our protein positive selection analysis, we downloaded GO annotations associated with all ENSEMBL genes in our analysis, obtained using biomaRt [116]. For each orthogroup, we then assigned the combined, non-redundant set of GO terms and filtered to include only terms in the biological process ontology. We then tested for enrichment of low *P* values using one-sided threshold-independent Kolmogorov–Smirnov (KS) tests implemented using topGO [117], which corrects for the correlated nature of the GO graph

network. We implemented tests in topGO using the “weight01” algorithm, excluding GO terms with fewer than 10 associated genes. We report enriched biological processes that passed a threshold (FDR-adjusted $P < 0.1$) using both our PAML and BUSTED *P* values analyzed separately (Supplementary Table 4).

For our gene family expansions analysis, we annotated gene families in the TreeFam9 database [57] based on provided mappings of gene families to accessioned proteins in the UniprotKB database [118]. We used the Uniprot accessions to assign proteins to ENSEMBL genes using biomaRt [116], then linked the combined, non-redundant set of GO terms associated with genes from the human (GRCh38) genome to each TreeFam9 family. We tested for enrichment of low *P* values using KS tests with identical settings and filters to those used in our protein positive selection analysis. We used branch-specific *P* values for the gelada branch from CAFE as input. Because *P* values from CAFE are nondirectional and ranged from 0 to 0.5, however, we first rank-ordered *P* values according to the strength of evidence for expansion by subtracting the *P* values from 1 whenever the gene family contracted in size in the gelada branch (i.e., fewer genes in the gelada branch compared to the ancestral gelada-baboon node). We considered all biological processes with an FDR-adjusted $P < 0.1$ to be significantly enriched (Supplementary Table 6).

Gelada accelerated region analysis

We used an accelerated region approach modeled on that of Pollard *et al.* [83] on genome-wide alignment blocks to identify regions encountering accelerated evolution in the gelada lineage, which we refer to as “gelada accelerated regions”, or GARs.

We first obtained whole-genome alignment blocks for the “57 mammals EPO” dataset from ENSEMBL [119] (release 101), which includes Tgel 1.0 and 56 additional mammalian genomes in multiple alignment format (MAF). We subsetted alignment blocks to include only terminal branches (i.e., excluding ancestral sequences), then preprocessed MAF files using mafTools [120] to remove duplicate species (mafDuplicateFilter), set human (*Homo sapiens*) as the reference species (mafRowOrderer), index all blocks to the positive strand on the reference (mafStrander), and to sort blocks by position (mafSorter). We subsetted blocks to include only gelada and the species trio of human/mouse/rat, then performed local realignment within MAF blocks using MAFFT [121,122] to correct misalignments. We next used MafFilter [123,124] to define and extract conserved alignment blocks that met the following criteria within the human/mouse/rat species trio: (1) a block length of ≥ 50 bp, (2) gaps in no more than 10% of positions within a 50 bp window, and (3) variable sites (including gaps) in no more than 10% of positions within a 50 bp window. We used these criteria because they were within the range of effective parameters evaluated by Pollard *et al.* [83], but maximized the number of genomic regions available for downstream analyses. We retained blocks encompassing the most inclusive set of coordinate positions that passed these criteria.

We filtered all alignment blocks by the criteria described above using AlnFilter and EntropyFilter from the MafFilter software package [124]. Notably, both of these algorithms are designed to identify and remove sites failing filters across sliding windows. Blocks are then normally split to remove sites failing filters within any window and trimmed blocks containing residual coordinates are returned as output. Because our pipeline instead required identifying and retaining sites passing filters, we modified and recompiled the source code of MafFilter and its Bio++ dependencies [125,126] to direct windows failing filters to the output and windows passing filters to the “trash”. In so doing, we took advantage of a feature of MafFilter by which windows failing filters are optionally redirected to a “trash” MAF file, with adjacent coordinate

sites merged into contiguous blocks for perusal. By directing coordinates passing filters to the “trash” and by using the resulting blocks as inputs for the remainder of the pipeline, we were able to induce the desired behavior from the software.

After identifying sites passing filters, we extracted their coordinates using MafFilter OutputCoordinates and used the resulting file to extract the corresponding positions from the MAF blocks containing all species using maf_parse from the PHAST software package [127]. We then repeated local alignment of MAF blocks using MAFFT and indexed blocks to the gelada reference genome using mafRowOrderer and mafStrander from mafTools to create the final MAF blocks. A total of 60,345 blocks passing filters were included in our analysis.

We tested for acceleration within blocks using the CONS model described by Pollard *et al.* [83] and the phylogenetic tree included with the ENSEMBL “57 mammals EPO ” dataset. The CONS model fits a general time-reversible model (REV) on aligned sequences using phyloFit from the PHAST v1.5 software package [127]. Acceleration is assessed by a likelihood ratio test (LRT), comparing a phylogenetic model in which branches are scaled across the tree in equal proportions and a model in which the foreground branch (gelada) is scaled separately from the remainder of the tree. The LRT statistic is the log ratio of the likelihood of the latter (alternate) model to the former (null) model multiplied by 2. We calculated significance from the LRT statistic using the chi-squared distribution.

To assess the distribution of acceleration scores within blocks, we also ran phyloP from the PHAST v1.5 package, which tests for acceleration or conservation at the nucleotide level. We ran phyloP for all 60,345 blocks in our analysis with the same phylogenetic model (REV) and phylogenetic tree, with the *P*-value reporting mode set to “CONACC” to distinguish between signals of conservation and acceleration. We then extracted genomic positions and CONACC *P* values from the phyloP output files.

We defined GARs as blocks for which FDR-adjusted $P < 0.2$ from the CONS model. A total of 29 blocks passed this threshold (Extended Data Fig. 6 and Supplementary Table 9), which we classified as either exonic, intronic, or intergenic based on overlap with annotated regions in the ENSEMBL GFF3 file. To identify biological processes associated with these blocks, we matched all 60,345 blocks passing filters to their nearest genes using GenomicRanges [128] in R, then downloaded all associated GO biological processes to genes using biomaRt [116]. To identify candidate regulatory elements, we also matched blocks with overlapping ChromHMM chromatin state annotations (15-state model, 127 epigenomes) obtained from the Roadmap Consortium [129]. We focused on 8 states that show putatively regulatory hallmarks (i.e., enrichment of ChIP-seq binding sites and enrichment of DNase peaks): active transcription start site (TSS), flanking active TSS, transcribed at genes 5' and 3', genic enhancers, enhancers, bivalent/poised TSS, flanking bivalent/poised TSS, and bivalent enhancer.

For two pairs of GARs (GAR26–GAR27 and GAR28–GAR29) that were nearest to the same genes, we estimated linkage disequilibrium between each GAR within each pair. To perform this analysis, we used whole-genome gelada variant data (described below) and calculated r^2 between sites using VCFtools v0.1.16 [130]. We limited our sample to geladas, excluded indels and non-biallelic SNVs, and filtered to only sites within the boundaries of each GAR ± 1000 bp. We calculated r^2 between all pairs of sites between GARs by setting a minimum distance of 500 kb between sites (arguments: --geno-r2 --ld-window-bp-min 500000), then calculated mean r^2 across all pairs of sites.

Because of the relatively low number of detected GARs (n = 29), we tested for enrichment of accelerated scores among biological pathways using threshold-independent Kolmogorov–Smirnov (KS) tests. To annotate alignment blocks, we assigned GO biological processes based on the nearest genes as previously described. To limit spurious matching of functions to blocks, we excluded all blocks with a

distance greater than 10,000 bp from their nearest gene, resulting in 41,910 blocks representing 4,145 genes retained for analysis. We then tested for enrichment through topGO [117] using the “weight01” algorithm. We excluded GO terms that were associated with fewer than 10 genes in our analysis and adjusted all *P* values using a Benjamini-Hochberg procedure [113] to correct for multiple hypothesis testing. We considered all biological processes with an FDR-adjusted $P < 0.01$ to be significantly enriched (Supplementary Table 10).

Whole-genome population resequencing and analysis

Library generation and sequencing

DNA was extracted from whole-blood samples or muscle samples using the DNeasy Blood & Tissue Kit (QIAGEN), following manufacturer recommendations for maximizing yield and quality. Concentration was assessed by Qubit 3 (Invitrogen) and 50 ng of DNA were used as input for whole-genome sequencing (WGS). Libraries were prepared using the Nextera DNA Library Prep protocol (Illumina catalog #FC-121-1030). Briefly, DNA was added to a 10 µl reaction containing 5 µl of TD buffer and 1 µl of tagment DNA enzyme (TDE1), then incubated at 55°C for 5 minutes. Tagmentation reactions were cleaned using 2x concentration of Ampure XP beads (Beckman Coulter), then 10 µl of cleaned DNA were added to a 24 µl PCR reaction including 1x NEBNext Q5 master mix (New England Biolabs) and 0.42 µM each of indexed P5/P7 primers. Libraries were amplified using six cycles of PCR and cleaned using 0.65x concentration of Ampure XP beads (Beckman Coulter). Libraries were pooled equimolarly and sequenced on either the Illumina HiSeq X or NovaSeq 6000 platforms (2×151 bp sequencing) to a median coverage of 11.54x.

Mapping and genotyping

We mapped reads to either the gelada reference genome (Tgel 1.0) or the anubis baboon reference genome (Panubis 1.0) [131] using the speedseq align v0.1.2 pipeline [132], which includes reference mapping with BWA-MEM [133], duplicate marking and discordant-read/split-read extraction with SAMBLASTER [134], and position sorting and BAM file indexing with SAMBAMBA [135]. To mitigate potential mapping biases, we used the dataset mapped to Tgel 1.0 for analyses including only geladas and the dataset mapped to Panubis 1.0 for analyses including both geladas and hamadryas baboons.

We genotyped reads using a pipeline implemented in GATK v4.1.2.0. We genotyped on a per-sample basis using GATK HaplotypeCaller to generate GVCF files. We then performed joint genotyping across samples using GATK GenotypeGVCFs, after first creating a GenomicsDB workspace using GATK GenomicsDBImport. We filtered variants using GATK VariantFiltration with the filters "QD < 2.0, MQ < 40.0, FS > 60.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, and SOR > 3.0", then extracted sites passing filters using BCFtools [136,137].

We used the resulting genotypes to recalibrate base quality scores using the GATK BaseRecalibrator and ApplyBQSR workflows. We then repeated per-sample variant calling, joint genotyping, and variant filtration to sequentially improve our genotype qualities. We performed a total of two rounds of base quality score recalibration bootstrapping in this manner, then repeated our genotyping pipeline a final time to generate final genotypes in VCF format. Our final VCFs included 48,744,921 variants mapped to Tgel 1.0 (chromosomes 1–21 and X) and 35,615,864 variants mapped to Panubis 1.0 (chromosomes 1–20, X, and Y).

Population structure analysis

Separately from our GATK genotyping pipeline, we implemented the genotyping uncertainty models in ANGSD [138] and PCAngsd [32] to analyze population structure. We used BAM files mapped to the gelada genome (Tgel 1.0) as input. We then used ANGSD v0.921 to generate genotype likelihoods in beagle

format (arguments: -GL 1 -doGlf 2 -doMajorMinor 1 -doMaf 2 -minMaf 0.05 -SNP_pval 1e-6 -minQ 20 -minMapQ 30 -skipTriallelic 1 -minInd 22 -doDepth 1 -doCounts 1). We then used PCAngsd v0.95 to estimate admixture proportions (arguments: -admix -admix_alpha 5000).

Determination of geographic provenience

To determine the provenience of the 17 zoo individuals in our study, we assembled mitochondrial DNA sequences from WGS reads for all individuals in our dataset and aligned them to the mitochondrial DNA dataset of Zinner *et al.* [139], which consists of cytochrome *b* + hypervariable region I (HVI) D-loop sequences from wild geladas across their natural distribution. We assembled complete mitochondrial sequences using GetOrganelle v1.7.5 [140], which uses Bowtie2 [141], BLAST [142], and SPAdes [143] to assemble circular genomes *de novo* from WGS data. To reduce computational burden, we limited our input to 2 million read pairs (~600 Mb) per individual randomly sampled using seqtk v1.3 [144] and incorporated a reference mitochondrial genome assembly (GenBank accession FJ785426.1) [145] as an input seed sequence. We then extracted the cytochrome *b* + HVI region for each sample by aligning to Zinner *et al.* [139] sequences using EMBOSS water v6.6.0 [146].

We combined all new gelada cytochrome *b* + HVI sequences (Supplementary Table 1) with 61 gelada haplotypes from Zinner *et al.* [139], one sequence from a hamadryas baboon in our dataset (FIL001), and one sequence from a rhesus macaque (GenBank accession NC_005943.1) [147]. We then aligned all nonredundant haplotypes using Clustal Omega v1.2.4 [104]. To infer a phylogenetic tree, we ran IQ-TREE v2.1.2 [148] with two partitions, protein-coding (1–1134) and non-coding (1135–1737) sequences, and set the rhesus macaque sequence as the outgroup. We used the ModelFinder option [149] within IQ-TREE to select the best nucleotide substitution models according to the the Bayesian Information Criterion (HKY+F+G4 and HKY+F+I+G4 for protein-coding and non-coding partitions, respectively) and ran 10,000 ultrafast bootstrap replicates (Extended Data Fig. 2).

Demographic history

We estimated demographic histories using MSMC2 v.2.1.2 [150,151] (Fig. 3a). We used per-sample VCFs, described earlier, and generated a mask file to exclude sites with excessively low or high coverage. We used Mosdepth [152] to calculate mean sample coverage and to generate a BED file per-sample marking sites with sequencing depth above a minimum of 50% mean coverage and below a maximum of 250% mean coverage. We then merged VCF files and mask files using the generate_multihetsep.py script and ran MSMC2 using the resulting file as input. We set 11.67 years as the average generation time, which we derived by calculating the average maternal age at birth from the SMGRP longitudinal life history database, and 0.5×10^{-8} as the mutation rate (μ), which is derived from estimates in anubis baboons [153]. For plotting, we excluded samples with < 10x mean coverage.

Analysis of heterozygosity and runs of homozygosity

We calculated average heterozygosity and identified runs of homozygosity for 90 individuals from the Central and Northern gelada populations, a captive gelada group, and a population of hamadryas baboons (*Papio hamadryas*) from Filoha, Ethiopia (Fig. 3b). We used variants called from data mapped to the anubis baboon reference genome (Panubis 1.0) [131]. Heterozygosity was calculated as per-site average in 100 kb windows with a 10 kb slide across all autosomes in the Panubis 1.0 reference for each individual. Windows were generated with BEDtools makewindows v2.29.2 [154] and number and percent callable sites within each window were identified with BEDtools intersect v1.10.2 [154]. A window was considered part of a run of homozygosity if its average heterozygosity was below 0.0002. We identified runs of adjacent windows with Ho < 0.0002 with the rle function in R v3.6.0 [155] and calculated the number of callable bases contained within runs of homozygosity <1 Mb, 1–3 Mb, and >3 Mb in length.

Data availability

All genomic data, including the Tgel 1.0 assembly (GenBank accession number GCA_003255815.1) and short-read sequencing data, are available through National Center for Biotechnology Information (NCBI) repositories and are linked to BioProject accession number PRJNA470999. Gelada hematological and morphological data are available on Dryad (https://doi.org/10.5061/dryad.fbg79cnvq).

All requests for biological material from the Simien Mountains used for this manuscript will be considered and granted depending on availability. For other biological materials, requests should be made to the contributors of those materials, which are specified in the manuscript.

Code availability

All code written for this project is available on GitHub (https://github.com/smacklab/gelada-genome).

Acknowledgments

We are grateful, first and foremost, to those who made this research possible, particularly the research staff (Esheti Jejaw, Ambaye Fenta, Setey Girmay, Dereje Bewket, and Atirsaw Adwana), logistical support staff (Tariku W/Aregay and Shiferaw Asrat), and assistants and students of the Simien Mountains Gelada Research Project—especially Julie Jarvey and Megan Gomery—as well as the Ethiopian Wildlife Conservation Authority (EWCA) for permission and support for working in the Simien Mountains National Park. We are also grateful to EWCA, the Amhara Regional Government, and Mehal Meda Woreda for permission to conduct research at Guassa Community Conservation Area; and to Badiloo Muluyee, Ngadaso Subsebey, Bantilka Tessema, Tasso Wudimagegn, and many field assistants for important logistical research support there. We thank David McDonald and the Cellular Imaging Core at the Fred Hutchinson Cancer Research Center for assistance with karyotyping. We are additionally grateful to Sierra Sams and Sarah Ford for assistance with lab work, and to Michael Montague, Kelley Harris, Abigail Bigham, Graham Scott, Ivan Liachko, Zev Kronenberg, Olga Dudchenko, Noah Simons, Nelson Ting, and Julien Dutheil for feedback through various stages of this research.

Support for this research was provided by the National Science Foundation (BCS 2010309 to NS-M, BCS 1848900 to NS-M, BCS 2013888 to NS-M, BCS 1723237 to NSM, BCS 1723228 to ALu], BCS 0715179 to TJB, OIA 1736249 to JFS, IOS 2114465 to JFS, IOS 1255974 to JCB, and IOS 1854359 to JCB, the National Institutes of Health (NIA R00AG051764 to NS-M and NHLBI R01HL087216 to JFS), the University of Washington Royalty Research Fund, the San Diego Zoo, and the German Research Foundation (DFG KN1097/3-1 to SK). KLC was supported by a National Institutes of Health fellowship (NIA T32AG000057). MCJ was supported by the Natural Environment Research Council (NE/T000341/1) and the Natural Sciences and Engineering Research Council Discovery Accelerator Grant. ISC (Schneider-Crease) is supported by the ASU Center for Evolution and Medicine.

Author Contributions Statement

NS-M, KLC, and MCJ conceived the research. KLC, MCJ, IAS-C, ADM, ALu, JCB, TJB, and NS-M designed the study. KLC, IAS-C, SS, FA, ISC, SK, ALe, BA, JCB, TJB, and NS-M collected field gelada samples and data, facilitated by AAH and FK. PJF, NN, CM, MLH, JDW, ASB, CMB, JR, JEP-C, and CJJ contributed samples and/or data. AVS and JFS designed, performed, and analyzed Hb-O_2 affinity experiments. KLC, AMD, and NS-M generated genomic data. KLC, MCJ, and NS-M performed genomic analyses. KLC, MCJ, and NS-M wrote the paper. All authors revised and approved the final manuscript.

Competing interests statement

The authors declare no competing interests.

Figure legends

Figure 1. The gelada at high altitude. (A) Geladas form three main populations that are each geographically restricted to highland areas of Ethiopia. Presence points are shown from the sample of Zinner *et al.* [139]. (B) An adult male gelada in the Simien Mountains (photo © India Schneider-Crease). (C) Geladas are found almost exclusively from 2,350 to 4,550 meters above sea level, constituting one of the highest altitudinal ranges of any primate species.

Figure 2. Unique karyotypic evolution in geladas. (A) Apart from geladas, the papionin clade exhibits an extremely conserved karyotype of 42 diploid chromosomes. Twenty species with known karyotypes sampled by Stanyon *et al.* [24] are shown with the consensus chronogram from TimeTree [53,54]. (B) Hi-C contact map reveals a distinct lack of contacts between the arms of chromosome 7. (C) G-banded karyotyping and analysis of genomic rearrangements reveal strong synteny between fissioned chromosomes and the intact arms of chromosome 7 in Central geladas and baboons, respectively. (D) STRUCTURE bar plot showing ancestry proportions from 70 resequenced gelada genomes reveals two main populations differentiating Northern (orange) and Central (green) geladas. Zoo animals are of mainly Central ancestry, but two individuals with the highest levels of Northern ancestry are also heterozygous for the centric fission characteristic of Northern geladas.

Figure 3. Historical demography and genomic diversity among gelada populations. (A) Multiple Sequentially Markovian Coalescent model reveals a historical divergence in effective population size between Northern and Central geladas, occurring roughly 500 k.y.a. (B) Analysis of genomic diversity reveals that geladas have lower heterozygosity and a higher portion of the genome in runs of homozygosity (ROHs) relative to hamadryas baboons, indicating less genetic diversity and a lower effective population size. Within geladas, the Northern population is more diverse than the Central population according to both metrics. Pairwise tests were run using two-sided Wilcoxon rank-sum tests, corrected for multiple comparisons using a Benjamini-Hochberg procedure. Comparisons with FDR-adjusted $P < 0.001$ (indicated by "***") or FDR-adjusted $P < 0.05$ (indicated by "*") are annotated. Exact FDR-adjusted *P* values for these comparisons are as follows. Heterozygosity: baboon vs. N. gelada (7.9e-8), baboon vs. C. gelada (7.4e-6), N. gelada vs. C. gelada (1.3e-6); Gigabases in ROHs < 1 Mb: baboon vs. N. gelada (3.1e-12), baboon vs. C. gelada (2.2e-7), N. gelada vs. C. gelada (2.5e-7); Gigabases in ROHs 1–3 Mb: N. gelada vs. C. gelada (1.8e-2).

Figure 4. Gelada blood and lung phenotypes at high altitude. (A) Protein alignment reveals two unique substitutions in the alpha subunit of hemoglobin (Hb) in gelada. (B) Hb–O_2 affinity assays, however, do not find evidence of increased oxygen binding affinity (i.e., lower P_{50}) of gelada Hb. Predicted P_{50} at pH 7.4 is shown along with error bars representing the standard error of the estimate. Measurements were collected from gelada (n=1), baboon (n=3), and human (n=1) samples at three pH values per sample (~7.2, ~7.4, and ~7.7) and corrected to pH 7.4 using linear regression. Measurements are shown for experiments run in either the presence or absence of allosteric cofactors 0.1 M KCl and 2,3-DPG. (C) Geladas at high altitude do not exhibit elevated Hb concentrations (erythrocytosis) at high altitude, in contrast to most humans with the notable exception of Tibetans and Sherpa. Values for human populations are plotted from the metaanalysis by Gassmann *et al.* [108]. The mean ± standard deviation is shown for zoo (n=42) and wild (Simiens) geladas (n=92). (D) Comparison of gelada chest circumferences to those of five baboon species reveals that geladas maintain larger chest circumferences relative to their body mass and waist circumference, respectively.

References

1. Beall, C. M. Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integr. Comp. Biol.* **46**, 18–24 (2006).

2. Bigham, A. W. Genetics of human origin and evolution: high-altitude adaptations. *Curr. Opin. Genet. Dev.* **41**, 8–13 (2016).

3. Ossendorf, G. *et al.* Middle Stone Age foragers resided in high elevations of the glaciated Bale Mountains, Ethiopia. *Science* **365**, 583–587 (2019).
4. Storz, J. F. & Cheviron, Z. A. Physiological genomics of adaptation to high-altitude hypoxia. *Annu Rev Anim Biosci* **9**, 149–171 (2021).
5. Storz, J. F. High-altitude adaptation: mechanistic insights from integrated genomics and physiology. *Mol. Biol. Evol.* **38**, 2677–2691 (2021).
6. Pozzi, L. *et al.* Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* **75**, 165–183 (2014).
7. Pugh, K. D. & Gilbert, C. C. Phylogenetic relationships of living and fossil African papionins: combined evidence from morphology and molecules. *J. Hum. Evol.* **123**, 35–51 (2018).
8. Jolly, C. J. The classification and natural history of *Theropithecus* (*Simopithecus*) (Andrews, 1916) baboons of the African Plio-Pleistocene. *Bull. Br. Mus. Nat. Hist. Bot.* **22**, 1–123 (1972).
9. Hughes, J. K., Elton, S. & O'Regan, H. J. *Theropithecus* and 'Out of Africa' dispersal in the Plio-Pleistocene. *J. Hum. Evol.* **54**, 43–77 (2008).
10. Jablonski, N. G. *Theropithecus: The Rise and Fall of a Primate Genus*. (Cambridge University Press, 1993).
11. Yalden, D. W., Largen, M. J. & Kock, D. Catalogue of the mammals of Ethiopia. 3. Primates. *Monit Zool Ital Suppl* **9**, 1–52 (1977).
12. Yu, L. *et al.* Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. *Nat. Genet.* **48**, 947–952 (2016).
13. West, J. B. The physiologic basis of high-altitude diseases. *Ann. Intern. Med.* **141**, 789–800 (2004).
14. Lee, J. W., Ko, J., Ju, C. & Eltzschig, H. K. Hypoxia signaling in human diseases and therapeutic targets. *Exp. Mol. Med.* **51**, 1–13 (2019).
15. Azad, P. *et al.* High-altitude adaptation in humans: from genomics to integrative physiology. *J. Mol. Med.* **95**, 1269–1282 (2017).
16. King, M. *Species Evolution: The Role of Chromosome Change*. (Cambridge University Press, 1995).
17. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).

18. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
19. Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
20. Waterhouse, R. M. *et al.* BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
21. Thibaud-Nissen, F., Souvorov, A., Murphy, T., DiCuccio, M. & Kitts, P. *Eukaryotic Genome Annotation Pipeline*. (National Center for Biotechnology Information, 2013).
22. Rogers, J. *et al.* The comparative genomics and complex population history of *Papio* baboons. *Science Advances* **5**, eaau6947 (2019).
23. Raaum, R. L., Sterner, K. N., Noviello, C. M., Stewart, C.-B. & Disotell, T. R. Catarrhine primate divergence dates estimated from complete mitochondrial genomes: concordance with fossil and nuclear DNA evidence. *J. Hum. Evol.* **48**, 237–257 (2005).
24. Stanyon, R. *et al.* Primate chromosome evolution: ancestral karyotypes, marker order and neocentromeres. *Chromosome Res.* **16**, 17–39 (2008).
25. Perry, J., Slater, H. R. & Choo, K. H. A. Centric fission — simple and complex mechanisms. *Chromosome Res.* **12**, 627–640 (2004).
26. Muleris, M., Dutrillaux, B. & Chauvier, G. Mise en évidence d'une fission centromérique hétérozygote chez un mâle *Theropithecus gelada* et comparaison chromosomique avec les autres Papioninae. *Génét. Sél. Evol.* **15**, 177–184 (1983).
27. Weber, A. F., Buoen, L. C., Terhaar, B. L., Ruth, G. R. & Momont, H. W. Low fertility related to 1/29 centric fusion anomaly in cattle. *J. Am. Vet. Med. Assoc.* **195**, 643–646 (1989).
28. Trede, F. *et al.* Geographic distribution of microsatellite alleles in geladas (Primates, Cercopithecidae): evidence for three evolutionary units. *Zool. Scr.* **49**, 659–667 (2020).
29. Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358 (2001).
30. Faria, R. & Navarro, A. Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol. Evol.* **25**, 660–669 (2010).

31. Bergey, C. M., Phillips-Conroy, J. E., Disotell R, T. & Jolly, C. J. Dopamine pathway is highly diverged in primate species that differ markedly in social behavior. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 6178–6181 (2016).
32. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
33. Storz, J. F. Hemoglobin–oxygen affinity in high-altitude vertebrates: is there evidence for an adaptive trend? *J. Exp. Biol.* **219**, 3190–3203 (2016).
34. Signore, A. V. *et al.* Adaptive Changes in Hemoglobin Function in High-Altitude Tibetan Canids Were Derived via Gene Conversion and Introgression. *Mol. Biol. Evol.* **36**, 2227–2237 (2019).
35. Signore, A. V. & Storz, J. F. Biochemical pedomorphosis and genetic assimilation in the hypoxia adaptation of Tibetan antelope. *Sci Adv* **6**, eabb5447 (2020).
36. Janecka, J. E. *et al.* Genetically based low oxygen affinities of felid hemoglobins: lack of biochemical adaptation to high-altitude hypoxia in the snow leopard. *J. Exp. Biol.* **218**, 2402–2409 (2015).
37. Beall, C. M., Brittenham, G. M., Macuaga, F. & Barragan, M. Variation in hemoglobin concentration among samples of high-altitude natives in the Andes and the Himalayas. *Am. J. Hum. Biol.* **2**, 639–651 (1990).
38. Beall, C. M. *et al.* Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara. *Am. J. Phys. Anthropol.* **106**, 385–400 (1998).
39. Beall, C. M. *et al.* Natural selection on *EPAS1* (*HIF2α*) associated with low hemoglobin concentration in Tibetan highlanders. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11459–11464 (2010).
40. International Species Information System. *Reference ranges for physiological values in captive wildlife*. (International Species Information System, 2002).
41. Harewood, W. J. *et al.* Biochemistry and haematology values for the baboon (*Papio hamadryas*): the effects of sex, growth, development and age. *J. Med. Primatol.* **28**, 19–31 (1999).
42. Storz, J. F., Scott, G. R. & Cheviron, Z. A. Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *J. Exp. Biol.* **213**, 4125–4136 (2010).
43. Storz, J. F. & Scott, G. R. Life ascending: mechanism and process in physiological adaptation to high-altitude hypoxia. *Annu. Rev. Ecol. Evol. Syst.* **50**, 503–526 (2019).

44. Frisancho, A. R. Developmental adaptation to high altitude hypoxia. *Int. J. Biometeorol.* **21**, 135–146 (1977).
45. Hsia, C. C. W., Carbayo, J. J. P., Yan, X. & Bellotto, D. J. Enhanced alveolar growth and remodeling in Guinea pigs raised at high altitude. *Respir. Physiol. Neurobiol.* **147**, 105–115 (2005).
46. Llapur, C. J. *et al.* Increased lung volume in infants and toddlers at high compared to low altitude. *Pediatr. Pulmonol.* **48**, 1224–1230 (2013).
47. Phillips-Conroy, J. E., Jolly, C. J. & Brett, F. L. Characteristics of hamadryas-like male baboons living in anubis baboon troops in the Awash hybrid zone, Ethiopia. *Am. J. Phys. Anthropol.* **86**, 353–368 (1991).
48. Jolly, C. J. & Phillips-Conroy, J. E. Testicular size, developmental trajectories, and male life history strategies in four baboon taxa. in *Reproduction and Fitness in Baboons: Behavioral, Ecological, and Life History Perspectives* (eds. Swedell, L. & Leigh, S. R.) 257–275 (Springer, 2006).
49. Bernstein, R. M., Drought, H., Phillips-Conroy, J. E. & Jolly, C. J. Hormonal correlates of divergent growth trajectories in wild male anubis (*Papio anubis*) and hamadryas (*P. hamadryas*) baboons in the Awash River Valley, Ethiopia. *Int. J. Primatol.* **34**, 732–752 (2013).
50. Beall, C. M. A comparison of chest morphology in high altitude Asian and Andean populations. *Hum. Biol.* **54**, 145–163 (1982).
51. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
52. Murrell, B. *et al.* Gene-wide identification of episodic selection. *Mol. Biol. Evol.* **32**, 1365–1371 (2015).
53. Hedges, S. B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32**, 835–845 (2015).
54. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
55. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
56. Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic*

Acids Res. **34(D1)**, D572–D580 (2006).

57. Schreiber, F., Patricio, M., Muffato, M., Pignatelli, M. & Bateman, A. TreeFam v9: a new website, more species and orthology-on-the-fly. *Nucleic Acids Res.* **42(D1)**, D922–D925 (2013).
58. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
59. Deng, L. *et al.* Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders. *Natl Sci Rev* **6**, 1201–1222 (2019).
60. Alkorta-Aranburu, G. *et al.* The genetic architecture of adaptations to high altitude in Ethiopia. *PLoS Genet.* **8**, e1003110 (2012).
61. Jeong, C. *et al.* Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat. Commun.* **5**, 3281 (2014).
62. Ilardo, M. A. *et al.* Physiological and genetic adaptations to diving in sea nomads. *Cell* **173**, 569–580.e15 (2018).
63. Tan, J. *et al.* Expression of aquaporin-1 and aquaporin-5 in a rat model of high-altitude pulmonary edema and the effect of hyperbaric oxygen exposure. *Dose Response* **18**, 1559325820970821 (2020).
64. Bareth, B. *et al.* The heme *a* synthase Cox15 associates with cytochrome *c* oxidase assembly intermediates during Cox1 maturation. *Mol. Cell. Biol.* **33**, 4128–4137 (2013).
65. Szpiech, Z. A., Novak, T. E., Bailey, N. P. & Stevison, L. S. Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evol. Lett.* **5**, 408–421 (2021).
66. Wu, B. J. *et al.* High-density lipoproteins inhibit vascular endothelial inflammation by increasing 3β-hydroxysteroid-Δ24 reductase expression and inducing heme oxygenase-1. *Circ. Res.* **112**, 278–288 (2013).
67. Zhu, S. *et al.* Genome-wide association study using individual single-nucleotide polymorphisms and haplotypes for erythrocyte traits in Alpine Merino sheep. *Front. Genet.* **11**, 848 (2020).
68. Pesce, A. *et al.* Neuroglobin and cytoglobin: fresh blood for the vertebrate globin family. *EMBO Rep.* **3**, 1146–1151 (2002).

69. Bigham, A. W. & Lee, F. S. Human high-altitude adaptation: forward genetics meets the HIF pathway. *Genes Dev.* **28**, 2189–2204 (2014).

70. McLean, C. J., Booth, C. W., Tattersall, T. & Few, J. D. The effect of high altitude on saliva aldosterone and glucocorticoid concentrations. *Eur. J. Appl. Physiol. Occup. Physiol.* **58**, 341–347 (1989).

71. Dosek, A., Ohno, H., Acs, Z., Taylor, A. W. & Radak, Z. High altitude and oxidative stress. *Respir. Physiol. Neurobiol.* **158**, 128–131 (2007).

72. Beall, C. M. Ages at menopause and menarche in a high-altitude Himalayan population. *Ann. Hum. Biol.* **10**, 365–370 (1983).

73. Moore, L. G. Maternal O_2 transport and fetal growth in Colorado, Peru, and Tibet high-altitude residents. *Am. J. Hum. Biol.* **2**, 627–637 (1990).

74. Keyes, L. E. *et al.* Intrauterine growth restriction, preeclampsia, and intrauterine mortality at high altitude in Bolivia. *Pediatr. Res.* **54**, 20–25 (2003).

75. Natarajan, C. *et al.* Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* **354**, 336–339 (2016).

76. Holt, S. V. *et al.* Silencing Cenp-F weakens centromeric cohesion, prevents chromosome alignment and activates the spindle checkpoint. *J. Cell Sci.* **118**, 4889–4900 (2005).

77. Landberg, G., Erlanson, M., Roos, G., Tan, E. M. & Casiano, C. A. Nuclear autoantigen p330d/CENP-F: a marker for cell proliferation in human malignancies. *Cytometry* **25**, 90–98 (1996).

78. Martin-Rendon, E. *et al.* Transcriptional profiling of human cord blood CD133+ and cultured bone marrow mesenchymal stem cells in response to hypoxia. *Stem Cells* **25**, 1003–1012 (2007).

79. Piazena, H. The effect of altitude upon the solar UV-B and UV-A irradiance in the tropical Chilean Andes. *Solar Energy* **57**, 133–140 (1996).

80. Wang, Q.-W., Hidema, J. & Hikosaka, K. Is UV-induced DNA damage greater at higher elevation? *Am. J. Bot.* **101**, 796–802 (2014).

81. King, M.-C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975).

82. Pollard, K. S. *et al.* An RNA gene expressed during cortical development evolved rapidly in humans.

Nature **443**, 167–172 (2006).

83. Pollard, K. S. *et al.* Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* **2**, e168 (2006).
84. 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).
85. Doan, R. N. *et al.* Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354.e12 (2016).
86. 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).
87. Gehman, L. T. *et al.* The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711 (2011).
88. Qin, Z. *et al.* ZNF536, a novel zinc finger protein specifically expressed in the brain, negatively regulates neuron differentiation by repressing retinoic acid-induced gene transcription. *Mol. Cell. Biol.* **29**, 3633–3643 (2009).
89. Ruiz-Martinez, J. *et al. GIGYF2* mutation in late-onset Parkinson's disease with cognitive impairment. *J. Hum. Genet.* **60**, 637–640 (2015).
90. Oguro-Ando, A. *et al. Cntn4*, a risk gene for neuropsychiatric disorders, modulates hippocampal synaptic plasticity and behavior. *Transl. Psychiatry* **11**, 106 (2021).
91. Koticha, D. *et al.* Cell adhesion and neurite outgrowth are promoted by neurofascin NF155 and inhibited by NF186. *Mol. Cell. Neurosci.* **30**, 137–148 (2005).
92. Hochachka, P. W. *et al.* The brain at high altitude: hypometabolism as a defense against chronic hypoxia? *J. Cereb. Blood Flow Metab.* **14**, 671–679 (1994).
93. Hornbein, T. F. The high-altitude brain. *J. Exp. Biol.* **204**, 3129–3132 (2001).
94. Wu, Y. & Song, W. Regulation of RCAN1 translation and its role in oxidative stress-induced apoptosis. *FASEB J.* **27**, 208–221 (2013).
95. Luo, S., Zou, R., Wu, J. & Landry, M. P. A probe for the detection of hypoxic cancer cells. *ACS Sens* **2**, 1139–1145 (2017).

96. Qi, X. *et al.* The transcriptomic landscape of yaks reveals molecular pathways for high altitude adaptation. *Genome Biol. Evol.* **11**, 72–85 (2019).
97. Dumitriu, B. *et al.* Sox6 is necessary for efficient erythropoiesis in adult mice under physiological and anemia-induced stress conditions. *PLoS One* **5**, e12088 (2010).
98. Cantù, C. *et al.* Sox6 enhances erythroid differentiation in human erythroid progenitors. *Blood* **117**, 3669–3679 (2011).
99. Dudchenko, O. *et al.* The Juicebox Assembly Tools module facilitates *de novo* assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *bioRxiv* 254797 (2018).
100. Li, R. *et al.* De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010).
101. 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).
102. Marçais, G. *et al.* MUMmer4: A fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).
103. Pratas, D., Silva, R. M., Pinho, A. J. & Ferreira, P. J. S. G. An alignment-free method to find and visualise rearrangements between pairs of DNA sequences. *Sci. Rep.* **5**, 10203 (2015).
104. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
105. Maddison, W. & Maddison, D. *Mesquite: a modular system for evolutionary analysis*. (2019).
106. Zhu, X. *et al.* Divergent and parallel routes of biochemical adaptation in high-altitude passerine birds from the Qinghai-Tibet Plateau. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 1865–1870 (2018).
107. Rees, D. G. & Henry, C. J. K. On comparing the predicted values from two simple linear regression lines. *Statistician* **37**, 299–306 (1988).
108. Gassmann, M. *et al.* The increase in hemoglobin concentration with altitude varies among human populations. *Ann. N. Y. Acad. Sci.* **1450**, 204–220 (2019).
109. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39(W2)**, W29–W37 (2011).

110. Villanueva-Cañas, J. L., Laurie, S. & Albà, M. M. Improving genome-wide scans of positive selection by using protein isoforms of similar length. *Genome Biol. Evol.* **5**, 457–467 (2013).

111. Shakya, M. *et al.* Standardized phylogenetic and molecular evolutionary analysis applied to species across the microbial tree of life. *Sci. Rep.* **10**, 1723 (2020).

112. Kosakovsky Pond, S. L., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).

113. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

114. Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

115. Gene Ontology Consortium. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43(D1)**, D1049–D1056 (2015).

116. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

117. Alexa, A. & Rahnenführer, J. *topGO: enrichment analysis for Gene Ontology*. (2019). doi:10.18129/B9.bioc.topGO.

118. Magrane, M. & UniProt Consortium. UniProt Knowledgebase: a hub of integrated protein data. *Database* **2011**, bar009 (2011).

119. Herrero, J. *et al.* Ensembl comparative genomics resources. *Database* **2016**, bav096 (2016).

120. Earl, D. *et al.* Alignathon: a competitive assessment of whole-genome alignment methods. *Genome Res.* **24**, 2077–2089 (2014).

121. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

122. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

123. Dutheil, J. Y., Gaillard, S. & Stukenbrock, E. H. MafFilter: a highly flexible and extensible multiple genome alignment files processor. *BMC Genomics* **15**, 53 (2014).

124. Dutheil, J. Y. Processing and analyzing multiple genomes alignments with MafFilter. in *Statistical*

Population Genomics (ed. Dutheil, J. Y.) 21–48 (Springer US, 2020).

125. Dutheil, J. *et al.* Bio++: a set of C++ libraries for sequence analysis, phylogenetics, molecular evolution and population genetics. *BMC Bioinformatics* **7**, 188 (2006).
126. Guéguen, L. *et al.* Bio++: efficient extensible libraries and tools for computational molecular evolution. *Mol. Biol. Evol.* **30**, 1745–1750 (2013).
127. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
128. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
129. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
130. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
131. Batra, S. S. *et al.* Accurate assembly of the olive baboon (*Papio anubis*) genome using long-read and Hi-C data. *Gigascience* **9**, giaa134 (2020).
132. Chiang, C. *et al.* SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat. Methods* **12**, 966–968 (2015).
133. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
134. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
135. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
136. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
137. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).
138. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

139. Zinner, D. *et al.* Phylogeography, mitochondrial DNA diversity, and demographic history of geladas (*Theropithecus gelada*). *PLoS One* **13**, e0202303 (2018).

140. Jin, J.-J. *et al.* GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **21**, 241 (2020).

141. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

142. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

143. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

144. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**, e0163962 (2016).

145. Hodgson, J. A. *et al.* Successive radiations, not stasis, in the South American primate fauna. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 5534–5539 (2009).

146. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).

147. Gokey, N. G. *et al.* Molecular analyses of mtDNA deletion mutations in microdissected skeletal muscle fibers from aged rhesus monkeys. *Aging Cell* **3**, 319–326 (2004).

148. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).

149. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

150. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).

151. Schiffels, S. & Wang, K. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. in *Statistical Population Genomics* (ed. Dutheil, J. Y.) 147–166 (Springer US, 2020).

152. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).

153. Wu, F. L. *et al.* A comparison of humans and baboons suggests germline mutation rates do not track cell divisions. *PLoS Biol.* **18**, e3000838 (2020).

154. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

155. R Core Team. R: a language and environment for statistical computing. (2013).