RESEARCH ARTICLE SUMMARY

RUMINANT GENOMICS

Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits

Lei Chen*, Qiang Qiu*, Yu Jiang*, Kun Wang*, Zeshan Lin*, Zhipeng Li*, Faysal Bibi, Yongzhi Yang, Jinhuan Wang, Wenhui Nie, Weiting Su, Guichun Liu, Qiye Li, Weiwei Fu, Xiangyu Pan, Chang Liu, Jie Yang, Chenzhou Zhang, Yuan Yin, Yu Wang, Yue Zhao, Chen Zhang, Zhongkai Wang, Yanli Qin, Wei Liu, Bao Wang, Yandong Ren, Ru Zhang, Yan Zeng, Rute R. da Fonseca, Bin Wei, Ran Li, Wenting Wan, Ruoping Zhao, Wenbo Zhu, Yutao Wang, Shengchang Duan, Yun Gao, Yong E. Zhang, Chunyan Chen, Christina Hvilsom, Clinton W. Epps, Leona G. Chemnick, Yang Dong, Siavash Mirarab, Hans Redlef Siegismund, Oliver A. Ryder, M. Thomas P. Gilbert, Harris A. Lewin, Guojie Zhang†, Rasmus Heller†, Wen Wang†

INTRODUCTION: The ruminants are one of the most successful mammalian lineages, exhibiting extensive morphological and ecological diversity and containing several key livestock species, such as cattle, buffalo, yak, sheep, and goat. Ruminants have evolved several distinct characteristics such as a multichambered stomach, cranial appendages (headgear), specialized dentition, a highly cursorial locomotion, and a wide range of body size variations. Despite their biological prominence and value to human societies, the evolutionary history of ruminants has not been fully resolved, and the molecular mechanisms underlying their particular characteristics remains largely unknown.

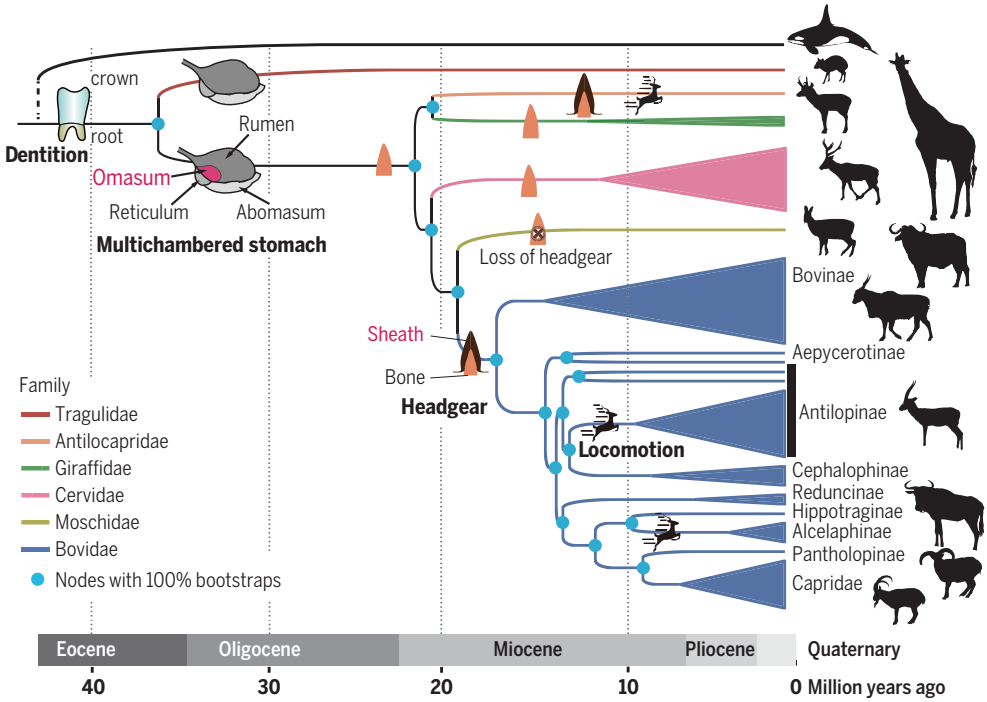
RATIONALE: We seek to resolve the controversies in the ruminant phylogeny and reveal the genetic basis underpinning the evolutionary innovations in ruminants. Here, we report the newly sequenced genomes of 44 ruminant species, covering about half the genera and all six extant Ruminantia families. We included seven published ruminant genomes (five bovids and two cervids) to reconstruct the phylogenetic tree by using improved time calibrations. We also reconstructed the Pleistocene demographic histories of these ruminant species using whole-genome heterozygosity information. Together with transcriptomic data of 516 samples from 68 tissues of four species, we conducted comparative genomic analyses to reveal candidate genes and regulatory elements that might have contributed to the evolution of the distinct ruminant characteristics.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aav6202

RESULTS: Using whole-genome orthologous sequences obtained from 51 ruminants, we have produced a new well-supported ruminant phylogenetic tree. The new tree resolves previous controversies over the deep branches of ruminant families, as well as the highly radiated Bovidae family. We estimated the emergence of crown Ruminantia to the late Oligocene (39.1 million to 32.3 million years ago) and that of Pecora to the Neocene (23.3 million to 20.8 million years ago). Investigations of demographic history revealed massive population decline events that occurred in most ruminant species, starting from ~100,000 to 50,000 years ago, which was temporally and spatially concurrent with the increased human activities on different continents during this period. We further identified many genomic changes that associate with important evolutionary innovations, such as the multichambered stomach, headgear, body size variation, cursorial locomotion, and dentition.

CONCLUSION: Our results demonstrate the power of using comparative phylogenomic approaches in resolving the deep branches of phylogeny that result from rapid radiations. The data and results presented in this study provide valuable resources and insights into the evolution of ruminant and mammalian biology. ■



Phylogeny and trait evolution of ruminants. The phylogenic tree of ruminants is presented with the species within same families and subfamilies collapsed. The ruminants have many textbook examples of distinct traits. The four-chambered stomach with omasum chamber is a key innovation evolved in pecoran ruminants. Headgear keratinous sheath only appear in Bovidae and Antilocapridae lineages. Many ruminants have evolved high-crowned or hypsodont teethes. The Antilocapridae and two bovid lineages are among the mammals with highest cursorial locomotion ability.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. E-mail: wwang@mail.kiz.ac.cn (W.W.); rheller@bio.ku.dk (R.H.); guojie.zhang@bio.ku.dk (G.Z.)
Cite this article as L. Chen *et al.*, *Science* **364**, eaav6202 (2019). DOI: 10.1126/science.aav6202

RESEARCH ARTICLE

RUMINANT GENOMICS

Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits

Lei Chen[1]*, Qiang Qiu[1]*, Yu Jiang[2]*, Kun Wang[1]*, Zeshan Lin[1]*, Zhipeng Li[3]*, Faysal Bibi[4], Yongzhi Yang[5], Jinhuan Wang[6], Wenhui Nie[6], Weiting Su[6], Guichun Liu[1,7], Qiye Li[8], Weiwei Fu[2], Xiangyu Pan[2], Chang Liu[1], Jie Yang[1], Chenzhou Zhang[1], Yuan Yin[1], Yu Wang[2], Yue Zhao[2], Chen Zhang[1], Zhongkai Wang[1], Yanli Qin[1], Wei Liu[7], Bao Wang[7], Yandong Ren[7], Ru Zhang[1], Yan Zeng[7], Rute R. da Fonseca[9,10], Bin Wei[2], Ran Li[2], Wenting Wan[1,7], Ruoping Zhao[7], Wenbo Zhu[1], Yutao Wang[11], Shengchang Duan[12], Yun Gao[12], Yong E. Zhang[13,14,15], Chunyan Chen[13,14], Christina Hvilsom[16], Clinton W. Epps[17], Leona G. Chemnick[18], Yang Dong[19,20], Siavash Mirarab[21], Hans Redlef Siegismund[9], Oliver A. Ryder[18,22], M. Thomas P. Gilbert[23,24], Harris A. Lewin[25], Guojie Zhang[7,8,15,26]†, Rasmus Heller[9]†, Wen Wang[1,7,15]†

The ruminants are one of the most successful mammalian lineages, exhibiting morphological and habitat diversity and containing several key livestock species. To better understand their evolution, we generated and analyzed de novo assembled genomes of 44 ruminant species, representing all six Ruminantia families. We used these genomes to create a time-calibrated phylogeny to resolve topological controversies, overcoming the challenges of incomplete lineage sorting. Population dynamic analyses show that population declines commenced between 100,000 and 50,000 years ago, which is concomitant with expansion in human populations. We also reveal genes and regulatory elements that possibly contribute to the evolution of the digestive system, cranial appendages, immune system, metabolism, body size, cursorial locomotion, and dentition of the ruminants.

Ruminantia is an important group of terrestrial herbivores, including at least 200 extant species (*1*) spanning six families: Tragulidae, Antilocapridae, Giraffidae, Moschidae, Cervidae, and Bovidae. The most species-rich of these families is Bovidae, which encompasses at least 143 species (*2*, *3*), including important livestock animals (such as cattle, buffalo, yak, sheep, and goat) (*4*, *5*). Ruminants possess several distinct and characteristic anatomical hallmarks, such as a multichambered stomach and cranial appendages (headgear). The acquisition of the rumen in the ruminants and of the omasum in pecorans (all ruminants except the Tragulidae) allow these animals to use plant material with a higher efficiency than other herbivorous mammals, such as equids (*6*–*8*). This effect is believed to be associated with the evolutionary success of these animals in terms of diversity, abundance, and geographic range (*9*). Ruminants have also evolved extreme morphological diversity, ranging in body weight from <2 kg to >1200 kg (*10*, *11*), and a wide variety of distinct behavioral and physiological traits. Ultimately, these adaptations in ruminants likely explain the remarkable abundance of these animals among domestic animals.

Despite their biological prominence and value to human civilization, much remains to be learned about the ruminants. For example, the phylogeny of ruminants is far from resolved, and inconsistencies remain even at the family level (*12*–*15*). Moreover, the genetic basis underlying many of their characteristic traits remains unknown. To fill in our gaps in knowledge, we performed de novo assembly of the genomes of 44 ruminant species, representing 36 genera that span all six families. In combination with five previously published bovid genomes, two published cervid genomes (*16*–*22*), and recently updated fossil information (*15*), we constructed a time-calibrated phylogenetic tree of the group, analyzed species population histories, and investigated the genomic evolution of these species. Our results not only provide data for understanding the origin and evolution of this important mammalian group and their particular traits but also have implications for placing ruminant livestock genomic resources into an evolutionary context and for conserving ruminant biodiversity.

Results

Genome sequencing, assembly, and annotation

We used Illumina sequencing technology (*23*) to generate more than 40 trillion base pairs (Tbp) of raw data and then de novo assembled genomes for 44 ruminant species (Table 1 and tables S1 and S2). Eleven of the species are listed as "vulnerable" or worse on the International Union for Conservation of Nature (IUCN) Red List (table S1). Forty of the genomes were assembled into large scaffolds (Table 1 and tables S3 and S4) with a series of mate-paired large-insert libraries, whereas four genomes—common eland (*Taurotragus oryx*), mountain nyala (*Tragelaphus buxtoni*), bongo (*Tragelaphus eurycerus*), and oribi (*Ourebia ourebi*)—were only assembled to the contig level because of degenerated samples but still qualified for most comparative genomic analyses (Table 1 and table S4). In addition, we improved the quality of the genome assembly of the black muntjac deer (*Muntiacus crinifrons*; contig N50 = 2.4 Mbp)

[1]Center for Ecological and Environmental Sciences, Northwestern Polytechnical University, Xi'an 710072, China. [2]Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China. [3]Department of Special Animal Nutrition and Feed Science, Institute of Special Animal and Plant Sciences, Chinese Academy of Agricultural Sciences, Changchun 130112, China. [4]Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Invalidenstrasse 43, 10115 Berlin, Germany. [5]State Key Laboratory of Grassland Agro-Ecosystem, School of Life Sciences, Lanzhou University, Lanzhou, 730000, China. [6]Kunming Cell Bank, State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. [7]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. [8]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China. [9]Section for Computational and RNA Biology, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark. [10]Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark. [11]College of Life and Geographic Sciences, Kashgar University, Kashgar 844000, China. [12]Nowbio Biotechnology Company, Kunming 650201, China. [13]Key Laboratory of Zoological Systematics and Evolution and State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. [14]University of Chinese Academy of Sciences, Beijing 100049, China. [15]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China. [16]Copenhagen Zoo, Frederiksberg, Denmark. [17]Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR 97331, USA. [18]San Diego Zoo Institute for Conservation Research, Escondido, CA 92027, USA. [19]Yunnan Research Institute for Local Plateau Agriculture and Industry, Kunming 650201, China. [20]State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan Agricultural University, Kunming 650201, China. [21]Department of Electrical and Computer Engineering, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. [22]Evolution, Behavior, and Ecology, Division of Biology, University of California, San Diego, La Jolla, CA 92093, USA. [23]EvoGenomics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [24]Norwegian University of Science and Technology, University Museum, 7491 Trondheim, Norway. [25]Department of Evolution and Ecology and the UC Davis Genome Center, University of California, Davis, CA, 95616, USA. [26]Section for Ecology and Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark.
*These authors contributed equally to this work.
†Corresponding author. E-mail: wwang@mail.kiz.ac.cn (W.W.); rheller@bio.ku.dk (R.H.); guojie.zhang@bio.ku.dk (G.Z.)

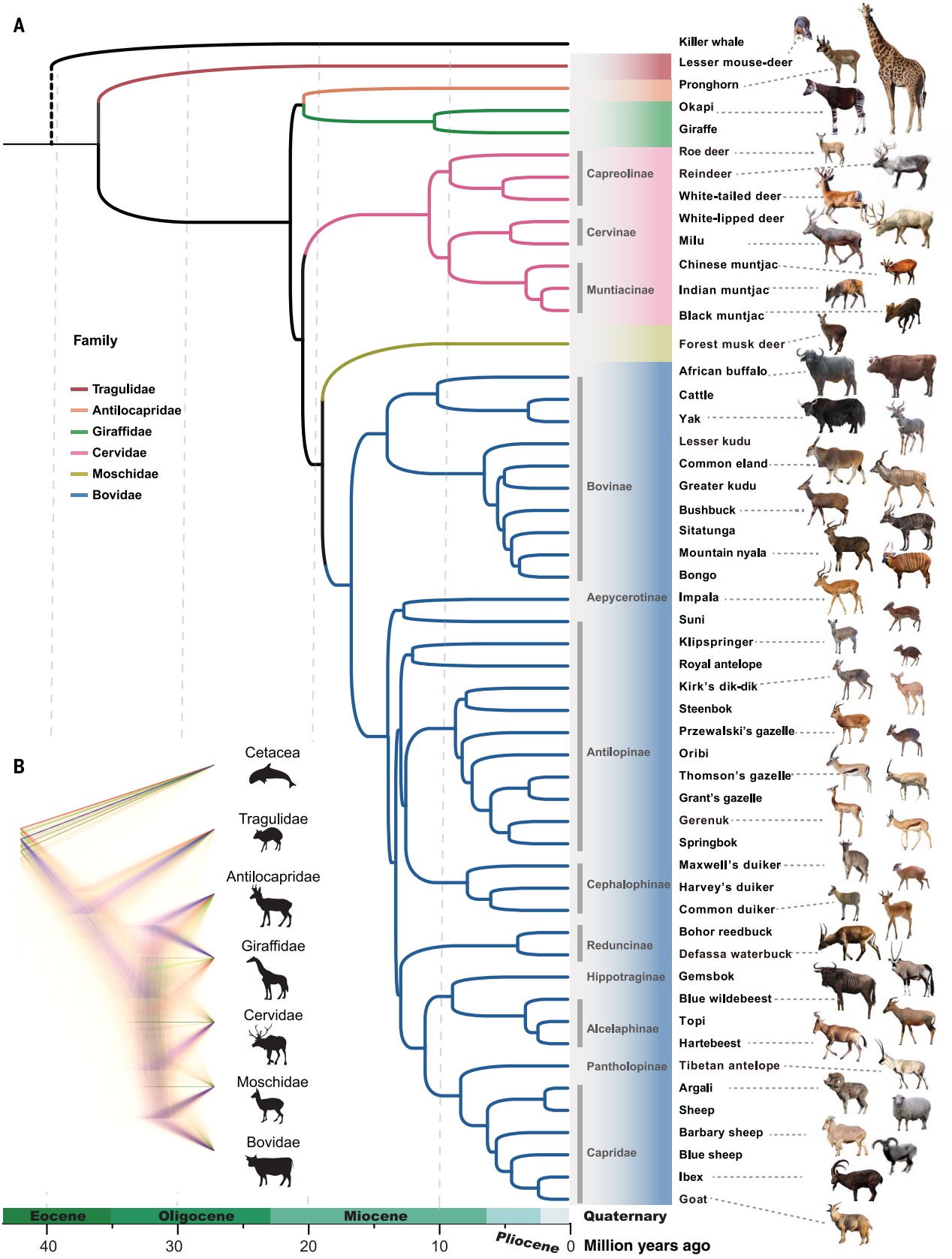


Fig. 1. Phylogeny of ruminants. (**A**) The maximum likelihood phylogenetic tree from whole-genome sequences of 51 ruminant species and 13 fossil calibrations. To compute the node supports, 200 bootstraps were used, and all nodes have 100% support. The origin and credit of the portraits of different species are listed in table S54. (**B**) Prevalent discordance among 10,000 random WGTs was observed across different families of ruminant.

using long reads generated from PacBio Single Molecule, Real-Time (SMRT) sequencing. This improved assembly of a Cervidae species facilitated synteny and other comparative analyses within the ruminants. Three genomes—reindeer (*Rangifer tarandus*), milu deer (*Elaphurus davidianus*), and Marco Polo sheep (*Ovis ammon*)—were released in data notes before this study (*24*–*26*).

To evaluate the quality of these genome assemblies (fig. S1 and tables S5 to S7), we performed BUSCO and synteny analyses and observed high BUSCO (*27*) scores (average 87%, from 75 to 93%) (table S6) and synteny continuity (table S7), showing that the majority of the assemblies were of high quality for downstream comparative analyses. The assembled genomes ranged in size from 2.52 to 3.25 Gbp, which was mostly consistent with the cytological *C* values (*28*) and *k*-mer–based estimations (table S5), indicating relatively complete genome coverages for all species.

After masking repeats (table S8), we used both de novo and homology-based gene prediction to annotate the genomes. The protein sequences of human (Ensembl 87 release), cattle (Ensembl 87 release), and sheep (Ensembl 87 release) were used as templates for homology-based gene prediction. The final annotated gene numbers range from 19,304 to 25,753 (table S9) among different species, with variations driven primarily by the quality of the assembled genomes. The gene length and exon length distributions were similar to those of other mammals (fig. S2 and table S9). We also identified 221,166 ruminant-specific conserved nonexonic elements (RSCNEs) by comparing all the ruminant genomes to 12 mammalian outgroup species (fig. S3 and table S10). These RSCNEs occupy ~0.61% of the genomes (~16.5 Mbp) (table S11) (*23*). A user-friendly, publicly available genome browser database was established (http://animal.nwsuaf.edu.cn/code/index.php/Ruminantia) for visualization of the genomic and transcriptomic data presented in this study.

Resolving the ruminant phylogeny

The lack of a fully resolved phylogeny for ruminants (*12*–*15*) still hinders an understanding of the evolution of ruminant diverse phenotypes. In particular, the phylogenetic positions of the Antilocapridae and Moschidae families have been strongly debated, and so have the relationships among subfamilies within the diverse Bovidae family [(*12*), review]. Furthermore, although "dwarf antelopes" have previously been grouped in the tribe Neotragini, recent molecular studies suggest that these animals are polyphyletic (*14*, *15*). These controversies are probably attributable to convergent evolution (challenging morphology-based approaches) and incomplete lineage sorting (ILS) in conjunction with the short internal branches in the ruminant radiation (challenging genetic inference) (*12*).

To resolve these phylogenetic challenges, we first estimated a whole-genome phylogenetic tree with ExaML under the GTR+GAMMA model (*23*) using the killer whale genome (*29*) as an outgroup. In total, 373 Mbp of orthologous syntenic sequences were obtained from whole-genome alignment by using the goat as a reference genome (*19*), yielding a whole-genome tree with 100% bootstrap support for all nodes (Fig. 1A). We also performed phylogenetic analyses with other genome partitions, including 6406 orthologous protein-coding genes identified in the 51 ruminant species and the killer whale, the fourfold degenerate sites in these genes, conserved nonexonic elements (CNEs), and complete mitochondrial genomes (mtDNA). With the exception of the mtDNA tree, the topologies of all other trees were identical with that of whole-genome tree (Fig. 1A and figs. S4 to S10).

Although the phylogenetic tree constructed with concatenated nuclear genome sequences (nDNA tree) was highly supported, phylogenetic discordance was pervasive across genomic regions (Fig. 1B). To further assess genome-wide tree discordance, 10,000 random genomic 1-Kbp windows at least 50 Kbp distant from each other were extracted. These 1-Kbp windows had enough segregating sites to generate window-based gene trees (WGTs) of high resolution (fig. S11). Although all individual WGTs differed from the species-level nDNA tree topology, 21.3% of WGTs exhibited the same family-level topology as that of the concatenated nDNA tree (table S12). This type of tree incongruence is usually caused by ILS, which has been widely observed in phylogenies of many animal groups—such as African cichlids (*30*), birds (*31*), and great apes (*32*, *33*)—and is known to be exacerbated by the effects of gene tree estimation error (*34*, *35*).

To ensure that the reconstructed phylogeny is robust in the presence of ILS, we performed several analyses with the coalescent-based phylogenetic method ASTRAL (*36*) using the 10,000 WGTs. We did not use entire genes as the unit of gene

Table 1. Assembly statistics of 44 ruminant species.

Species	Common name	Scaffold N50 (bp)	Contig N50 (bp)	Species	Common name	Scaffold N50 (bp)	Contig N50 (bp)
Tragulus javanicus	Lesser mouse-deer	243,250	6286	*Redunca redunca*	Bohor reedbuck	438,845	17,874
Antilocapra americana	Pronghorn	1,463,792	61,696	*Syncerus caffer*	African buffalo	2,316,376	11,115
Okapia johnstoni	Okapi	3,620,116	58,892	*Gazella thomsoni*	Thomson gazelle	1,581,717	36,935
Giraffa camelopardalis	Giraffe	3,197,404	22,538	*Tragelaphus strepsiceros*	Greater kudu	520,720	16,623
Muntiacus muntjak	Indian muntjac	1,398,591	10,925	*Nanger granti*	Grant's gazelle	520,131	6041
Muntiacus reevesi	Chinese muntjac	1,253,719	68,151	*Sylvicapra grimmia*	Common duiker	541,191	5209
Elaphurus davidianus	Milu	3,040,530	32,708	*Aepyceros melampus*	Impala	343,699	51,379
Rangifer tarandus	Reindeer	1,059,113	91,805	*Madoqua kirkii*	Kirk's dik-dik	489,835	26,372
Muntiacus crinifrons	Black muntjac	NA	1,458,913	*Oreotragus oreotragus*	Klipspringer	340,335	13,019
Gervus albirostris	White-lipped deer	3,567,448	22,599	*Antidorcas marsupialis*	Springbok	698,575	10,778
Moschus chrysogaster	Forest musk deer	2,509,225	57,721	*Tragelaphus imberbis*	Lesser kudu	1,774,691	6545
Oryx gazella	Gemsbok	1,583,972	18,132	*Tragelaphus spekii*	Sitatunga	78,973	5410
Litocranius walleri	Gerenuk	3,128,641	47,546	*Philantomba maxwellii*	Maxwell's duiker	390,552	4423
Damaliscus lunatus	Topi	1,172,125	25,829	*Raphicerus campestris*	Steenbok	474,033	5764
Ammotragus lervia	Barbary sheep	1,263,981	18,541	*Neotragus pygmaeus*	Royal antelope	365,736	5931
Pseudois nayaur	Blue sheep	2,076,308	23,854	*Alcelaphus buselaphus*	Hartebeest	11,258	4274
Ovis ammon	Argali	5,734,776	45,638	*Capra ibex*	Ibex	15,190,720	24,835
Kobus ellipsiprymnus	Defassa waterbuck	782,102	20,722	*Neotragus moschatus*	Suni	957,022	8233
Procapra przewalskii	Przewalski's gazelle	5,152,914	20,018	*Taurotragus oryx*	Common eland	no scaffold	1262
Connochaetes taurinus	Blue wildebeest	3,511,341	46,638	*Tragelaphus buxtoni*	Mountain nyala	no scaffold	1309
Cephalophus harveyi	Harvey's duiker	365,462	39,715	*Tragelaphus euryceros*	Bongo	no scaffold	1980
Tragelaphus scriptus	Bushbuck	890,554	9965	*Ourebia ourebi*	Oribi	no scaffold	1259

tree construction because genes can span over several recombination blocks in the genome (*37*). The topology of the obtained ASTRAL coalescent tree (fig. S12) is identical to the nDNA tree (Fig. 1A). We also used another coalescent-based phylogenetic method, MP-EST (*38*), to carry out additional phylogenetic analyses using both the 10,000 windows as well as the 6406 orthologous genes and obtained an identical topology as those of the ASTRAL and nDNA trees (fig. S13), further supporting the robustness of the inferred ruminant phylogeny. To minimize negative impacts of low tree resolution in the WGTs, we contracted low-supported branches [below 3, 10, and 20%, bootstrap support, as suggested in (*36*)] and still observed the same topology as that of the nDNA tree in all cases (fig. S14).

We further tested whether gene flow could have contributed to the observed topological discordances. The results from the ABBA-BABA test (*39*), the D_{FOIL} software (extend from 4 taxa to 5 taxa) (*40*), and admixturegraph (*41*) consistently suggested some possible ancient gene flows among Giraffidae, Cervidae, Bovidae, and Moschidae (fig. S15 and table S13). The strongest gene flow signal between Giraffidae and Cervidae-Bovidae-Moschidae coincides with an overrepresentation of this topology in the WGTs relative to the expectation of ILS, which requires that for a set of four species trees, alternative two phylogenetic topologies other than the species tree have equal proportion (fig. S16). By contrast, PhyloNet (*42*) and PhyloNetworks (*43*) were sensitive to model and parameter choices and did not generate consistent results (figs. S17 and S18). It is plausible that the high parameter space of the phylogenetic model precluded these methods from performing complete explorations of the parameter space in our large-time-scale phylogenomic data.

Our new ruminant tree supports a sister-group relationship for Antilocapridae and Giraffidae, representing the oldest branch among the extant pecoran families. This is different from the mtDNA-based phylogenies, which placed Antilocapridae as an outgroup to all other pecorans (figs. S9 and S10) (*12*, *14*, *15*). The sister relationship of these two groups has been proposed before (*44*, *45*) but now received 100% bootstrap support by the nDNA tree and a local posterior probability (*46*) of 1 in the ASTRAL tree and had 23% higher frequency of WGTs than those of the alternative topologies (fig. S16). Another contentious issue is on the placement of Moschidae, which was previously placed at the base of Pecora (*47*–*49*) or as a sister group to Bovidae (*50*) from morphological data. In some mtDNA studies, Moschidae was proposed as a sister group of Cervidae (*51*) or a sister taxon to Bovidae (*14*, *52*, *53*). Our nDNA tree confirmed that Moschidae is a sister group of Bovidae, and the frequency of WGTs further support this conclusion (fig. S16).

Our results also resolved some controversial subfamily relations in Bovidae, such as the

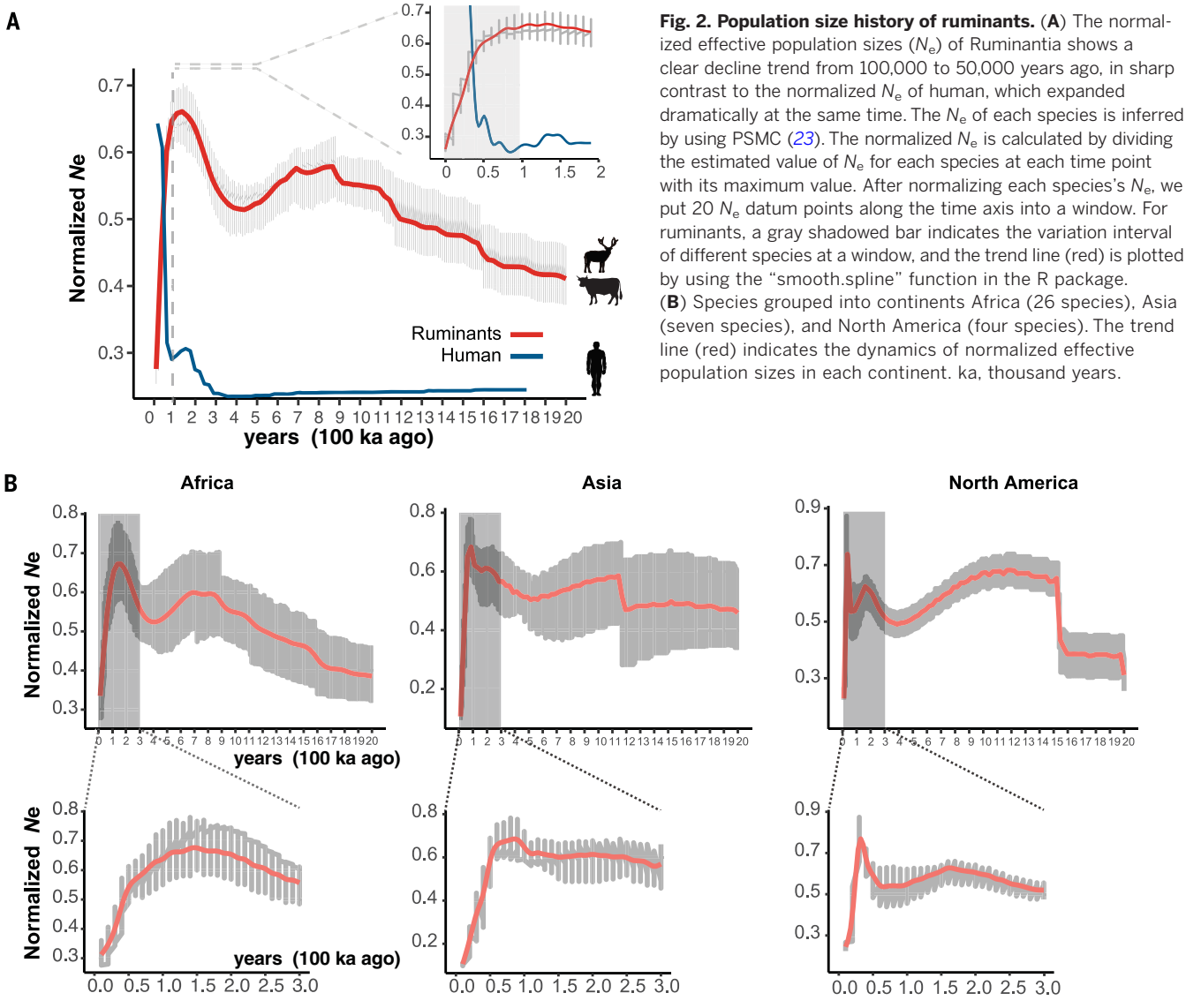


Fig. 2. Population size history of ruminants. (**A**) The normalized effective population sizes (N_e) of Ruminantia shows a clear decline trend from 100,000 to 50,000 years ago, in sharp contrast to the normalized N_e of human, which expanded dramatically at the same time. The N_e of each species is inferred by using PSMC (*23*). The normalized N_e is calculated by dividing the estimated value of N_e for each species at each time point with its maximum value. After normalizing each species's N_e, we put 20 N_e datum points along the time axis into a window. For ruminants, a gray shadowed bar indicates the variation interval of different species at a window, and the trend line (red) is plotted by using the "smooth.spline" function in the R package. (**B**) Species grouped into continents Africa (26 species), Asia (seven species), and North America (four species). The trend line (red) indicates the dynamics of normalized effective population sizes in each continent. ka, thousand years.

placement of the Reduncinae (12). With the whole-genome tree, Reduncinae were confirmed to be a sister group to the ancestral lineage of Caprinae, Alcelaphinae, and Hippotraginae. This topology is also supported by the ASTRAL tree and has a slightly higher WGT frequency than those of alternatives (fig. S19). We further confirmed the previously ambiguous sister-group relationship between impala (*Aepyceros melampus*) and suni (*Neotragus moschatus*) (54, 55) and found no phylogenetic support for the "waste-bucket" Neotragini tribe (56). These findings mostly agree with the topology from Decker *et al.* (57), which used single-nucleotide polymorphism (SNP) chip data but lacked samples from Moschidae and Tragulidae. Furthermore, our results provided higher bootstrap supports for the deep nodes and also refined some species positions within Bovidae—the position of bushbuck (*Tragelaphus scriptus*) and mountain nyala (*Tragelaphus buxtoni*).

Using fossil calibrations (tables S14 and S15) (15), we estimated the emergence of crown Ruminantia at 39.1 million to 32.3 million years ago (late Oligocene), and the emergence of Pecora at 23.3 million to 20.8 million years ago (Neocene) (Fig. 1A and figs. S20 to S27). The evolutionary rate in the ancestral ruminant lineage was ~1.5 × 10^{-9}, which was significantly higher than that in other mammals (Student's *t* test, $P < 0.01$) (fig. S28). Among the ruminant families, Tragulidae had the highest evolutionary rate, and Giraffidae had the lowest evolutionary rate (fig. S28). We found a significant negative correlation between evolutionary rate and log body size (fig. S29).

Pleistocene population dynamics in ruminants

We used our whole-genome dataset to investigate the demographic histories of different ruminant species using the pairwise sequentially Markovian coalescent (PSMC) method (58), which can infer changes in the effective population size (N_e) over the Pleistocene (figs. S30 and S31 and table S16). The analyses produced a species-specific demographic pattern with no clear grouping of patterns according to their habitat types or feeding types (fig. S32). This might imply that the ruminant species responded differentially to biotic and abiotic pressures associated with their different ecological niches (59). However, we found population declines for many species (25 out of the 40 species with scaffolded genome assemblies) starting from ~100,000 to 50,000 years ago (Fig. 2A), which suggests that late Pleistocene large mammal declines were much more severe than previously suspected, involving major declines in the populations of most species along with the mass extinction of large mammals at this time (60). We speculate that this community-wide decline might be at least partially attributable to human activities. This is supported by ruminant population declines coinciding with increasing human effective population size after the dispersal out of Africa during this period (Fig. 2A) (61). Intriguingly, the onset of ruminant decline coincided with the sequential population expansion of humans on different continents in this period (Fig. 2B), which would not be expected if the decline signal was a structural artefact. By contrast, ruminant population size dynamics showed no apparent correlation with the climatic changes dominated by serial glaciations during the Pleistocene (figs. S30 and S31). Taken together, these results highlight a possible role for humans in the massive decline of mammalian species in the late Pleistocene (60).

Structure and evolution of ruminant genomes

Synteny

We explored the general architecture of ruminant genomes through comparison with other high-quality mammalian genomes (23). Specifically, we compared the syntenic relationship between the goat (the most well-assembled ruminant genome) (19) and human (Primates) (62), dog (Carnivora) (63), horse (Perissodactyla) (64), pig (Suina) (65), camel (Tylopoda) (66), killer whale (Cetacea) (29), and black muntjac (Cervidae of Ruminantia). The high-quality assembly of the black muntjac (Table 1) by use of PacBio reads allowed us to assess the within-ruminant synteny by including both a Cervidae representative as well as the high-quality Bovidae representative (goat). We found that Primates and Perissodactyla have more prevalent genome rearrangements (>200 rearrangements per million base pairs) compared with that of Ruminantia (Fig. 3A and table S17). Within the Cetartiodactyla, pig (2n = 18, Suina) and camel (2n = 74, Tylopoda) experienced more genome rearrangements (~120 rearrangements per million base pairs) than the killer whale (2n = 44, Cetacea). Thus, the greater level of conserved synteny between Cetacea and Ruminantia (68 rearrangements per million base pairs) is consistent with their sister relationship within the Cetartiodactyla. Few rearrangements (19 rearrangements per million base pairs) were observed between the black muntjac (2n = 8/9) and the goat (2n = 58), suggesting that synteny has been conserved among different lineages of Ruminantia, despite large variation in chromosome numbers (table S18).

Genome size

Our ruminant genome assembly sizes ranged from 2.52 Gbp [oribi (*Ourebia ourebi*)] to 3.25 Gbp [klipspringer (*Oreotragus oreotragus*)] (table S5). The average assembled genome size of ruminants was (2.7 Gbp), which is larger than Carnivora (~2.3 Gbp), Perissodactyla (~2.4 Gbp), Suina (~2.5 Gbp), Tylopoda (~2.0 Gbp), and Cetacea (~2.4 Gbp) and smaller than Primates (~3.0 Gbp) (Fig. 3B). Analyses (23) confirm that transposable element (TE) content is the major cause of genome size variation (Fig. 3, B and C, and table S19).

When comparing the goat genome with the human, horse, pig, and killer whale genomes, we also observed and validated (23) large insertions and deletions (over 50 kbp in length) in ruminants (table S20). The largest insertion from segmental duplication in the goat contains a cluster of PAG (pregnancy-associated glycoprotein) genes, with 36 coding sequences and 32 pseudogenes (Fig. 3D). The main functions of PAGs are immune regulation and maintenance of pregnancy (67). We also found other insertions with important gene annotations in ruminants, including interferon and olfactory receptors (fig. S33 and table S21).

Evolution of genes and gene families in Ruminants

We obtained a high-confidence orthologous gene set for the full set of 51 ruminant species using camel, cat, dog, horse, human, minke whale, killer whale, and pig as outgroups (23). Using the resolved phylogeny of Ruminantia, we identified rapidly evolving genes (REGs), positively selected genes (PSGs), and newly evolved genes (Fig. 4A and tables S22 to S25) (23).

Functional enrichment analyses of the PSGs (tables S26 and S27), the REGs (table S28), and expanded gene families (table S29) in the ancestral ruminant branch all exhibit enrichment in immune functions. As many as 20 PSGs and 12 REGs are involved in the crossing of blood vessels by leukocytes, which respectively constitutes 17.9 and 10.7% of this key pathway in active immunization (Fig. 4B and tables S30 and S31) (68, 69). We also observed gene family expansions in the ruminant ancestor (table S32), including the interferon family (IFNs) (figs. S34 to S36), PAG family (fig. S34), and cathelicidin and serpin peptidase inhibitor families (figs. S34, S37, and S38), which are all involved in immune system pathways. In addition, we identified a rumen-specifically expressed newly evolved gene, *ENSBTAG00000038127* (23), which contains an immunoglobulin V-set domain, usually involved in mimicking the antibody variable domain of several diverse protein families (70). Although rapid evolution of immune genes was found in a wide range of animals, the number of PSGs from the leukocyte transendothelial pathway is highest in Ruminantia and could suggest a key role of this pathway in the evolution of ruminant immune system (table S33).

In addition to immune system–related genes, we also observed a series of PSGs, REGs, and expanded gene families in ruminants involved in lipid metabolism, glycolysis, oxidative phosphorylation, and amino acid metabolism (Fig. 4, C and D, and table S34) (23). Ruminants have a distinct digestive system, in which the main source of energy comes from volatile fatty acids (VFAs) and the rate of glycolysis is low (71). The genomic changes associated with metabolism may reflect the adaptation of the digestive system in ruminants and may therefore have played important parts in the success of the ruminants.

Genomic variations related to ruminant morphological characteristics

We were able to leverage our phylogenetic tree and genomic data to conduct evolutionary genomic analyses aimed at identifying genomic variations correlated with particular ruminant characteristics. Specifically, we investigated the evolution of the multichambered stomach, headgear, body size, cursorial locomotion, and dentition.

Evolution of the multichambered stomach

Whereas pecoran ruminants have a four-chambered stomach composed of the rumen, reticulum, omasum, and abomasum, the basal group of Ruminantia, the Tragulidae, lacks the omasum (*72*).

Using a transcriptomic dataset from 516 samples covering 50 tissues of sheep (table S35) (*23*), we found that the gene expression profiles of rumen, reticulum, and omasum are closest to that of esophagus (Fig. 5A), implying that the three stomach chambers might have originated from the esophagus, as suggested by Warner (*73*) and Xiang *et al.* (*74*). Regarding the distinct rumen organ, we found that several newly evolved genes played a role in its function. Among the 295 newly evolved genes identified in the ancestor of ruminants (Fig. 4A and fig. S39), seven were highly or specifically expressed in the rumen (fig. S40). Two genes, *PRD-SPRRII* and *TCHHL2*, are important structural genes in the rumen (*16*). Three newly evolved *serpin* genes may have inhibitory functions of different proteases or through anti-inflammatory functions (*75*, *76*), and two *KRT6A* genes are likely involved in the activation of follicular keratinocytes (*77*, *78*).

Although the abomasum is analogous to the true stomach in other mammals, it is hypothesized that the ruminant abomasum has evolved particular adaptations to digest the microbe-rich content from upstream chambers (*72*). The lysozyme *c* family, which degrades bacterial and microbial cells as members enter the abomasum to extract nutrients (*79*), has expanded to 10 or more copies in ruminants, whereas other mammals have only one or a few copies of this gene family (*79*). Our comparative genomics analyses revealed that the duplication of lysozyme *c* genes began in the ancestors of Ruminantia, continued to

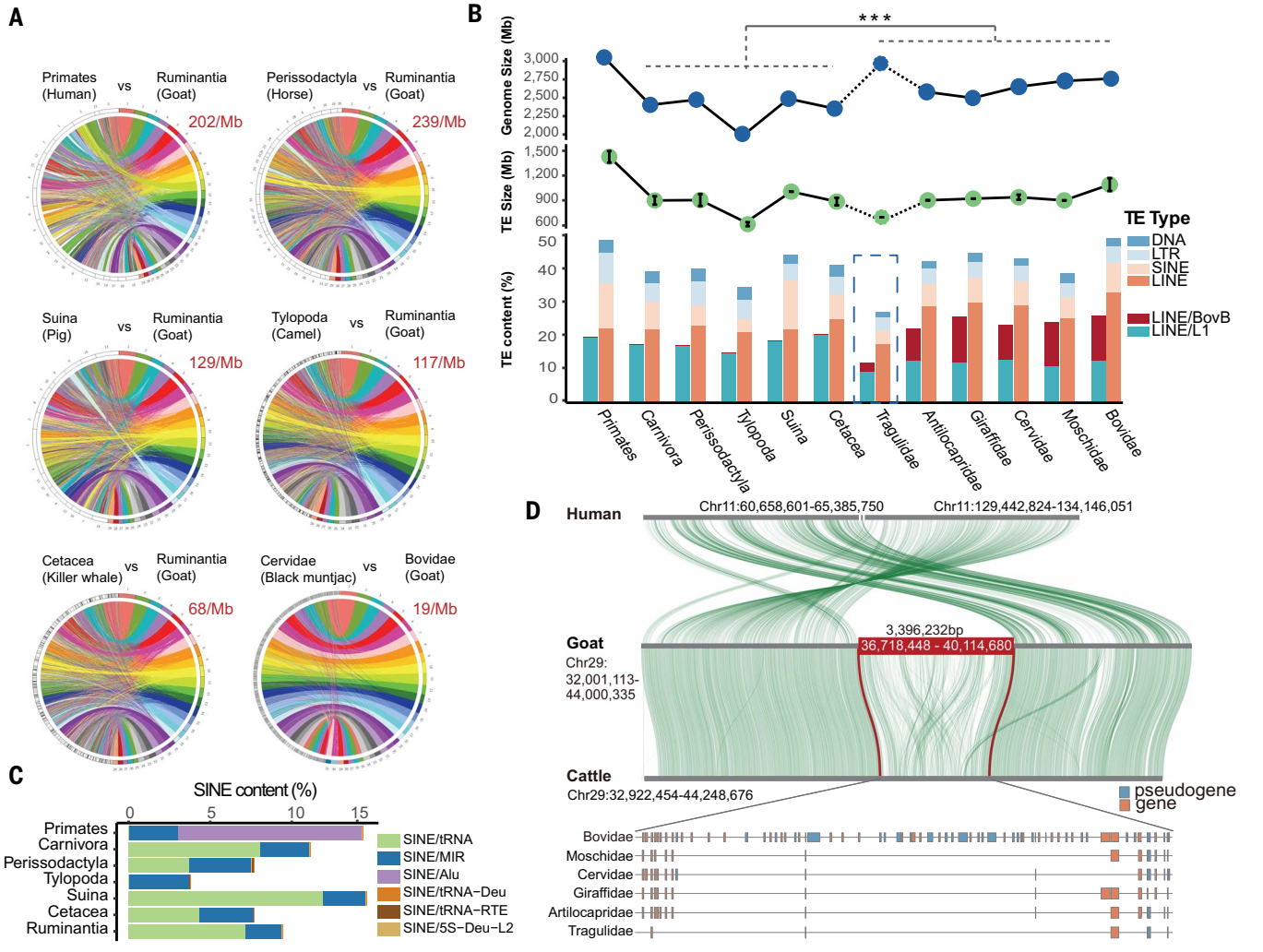


Fig. 3. Structural characteristics and evolution of ruminant genomes. (**A**) Goat (*19*) was used to represent Ruminantia-specific genomic rearrangements in comparisons with Primates (human), Perissodactyla (horse), Suina (pig), Tylopoda (camel), and Cetacea (Killer whale). Syntenic blocks are linked between genomes in a circos plot. The red number beside each circos quantifies the occurrences of rearrangement events per aligned megabase (Mb) sequence. The high-quality de novo assembly of the black muntjac was included for within-ruminant synteny inference. (**B**) The average genome sizes, TE sizes, and contents of different TE types of Primates (human, chimpanzee, gorilla, and orangutan), Carnivora (dog, cheetah, and polar bear), Perissodactyla (horse, przewalski's horse, and rhinoceros), Tylopoda (dromedary camel, bactrian camel, and alpaca), Suina (pig), and Cetacea (minke whale, killer whale, beluga whale, sperm whale, and yangtze finless porpoise). Overall, the average genome size of Ruminantia is significantly (Student's *t* test, $P < 0.01$) larger than those of Canivora, Perissodactyla, Tylopoda, Suina, and Cetacea. Tragulidae is marked with dashed lines because the genome assembly contains more gaps, which hindered the annotation of TEs. The proportions of LINE, SINE, LTR, and DNA transposons are presented in the stacked bar plot. LINE/BovB and LINE/L1 are highlighted here to present their dynamic changes among ruminants. ***$P < 0.01$. (**C**) The average contents of different SINE types are plotted in the stacked bar plot across mammalian orders and suborders of Cetartiodactyla, with different colors. (**D**) A large fragment insertion of 3,396,232 bp is observed in the goat genome, which is also validated in other ruminant families, containing a cluster of PAG genes, specifically containing 36 coding genes and 32 pseudogenes in goat.

expand along the diversification of pecoran families, and eventually culminated with the highest copy numbers in Bovidae (Fig. 4D). This accumulated evolution of lysozyme *c* gene copy number in ruminants may be associated with an improved harvesting of nutrients from the rumen microbiome.

As a newly evolved organ in the pecorans, the omasum is adjacent to the reticulum, whereas its expression profile is closest to that of the rumen. This expression overlap may be linked to the resemblance in structure and function of the rumen and omasum, which are both involved in the absorption of water, minerals, electrolytes, and VFAs, whereas the reticulum mainly has a different function as a filter for the fermentation products from the rumen (*6*). Anatomically, the omasum is composed of the same stratified squamous epithelium of mucosal-layered tissue as the rumen and reticulum, which is different from the abomasum and intestines (*6*). To further reveal the genetic basis underlying the evolution of the omasum in pecorans, we identified 75 genes that were specifically highly expressed in the omasum compared with other organs (fig. S41). Among these, one gene was newly evolved (*LOC101107119*), and another (*SCNN1D*) exhibited pecoran-specific amino acid changes relative to Tragulidae and killer whale genomes (fig. S42). *LOC101107119* is annotated as a prostaglandin F synthase 1–like gene and might have a similar function to that of the prostaglandin F synthase 1 (PGDF1), which is involved in ketone metabolism (*80*). *SCNN1D* encodes the δ subunit of the epithelial sodium channel, which mediates Na^+ reabsorption and water absorption in the digestive tract (*81*).

Transcriptional factors and regulatory elements may have played important roles in the evolution of the omasum by changing the expression patterns of their host genes to be recruited in omasum functions. We found four genes with pecoran-specific CNEs within their immediate upstream/downstream 10-kbp regions (*SIM2*, *PAX9*, *KCNK5*, and *DENND2C*) (table S36), of which PAX9 and SIM2 are important transcriptional factors. PAX9 regulates squamous cell differentiation in the esophageal epithelium (*82*). A CNE with two pecoran-specific mutations was found at the 5′ upstream of *PAX9* in pecorans, which might play a role in regulating *PAX9* expression in the omasum. *SIM2* is highly expressed in the omasum but inactive in the rumen [fragments per kilobase of exon per million fragments mapped (FPKM) < 1], and it works as a suppressor of cell proliferation in the epithelium (*83*, *84*). The differential expression patterns of *SIM2* in the omasum and rumen is consistent with the epithelial cells in the omasum not being completely renewed, as is the case in the rumen (*85*).

Evolution of headgear

The ruminant families exhibit spectacular variation in headgear morphology (*86*). To reveal the genetic basis of headgear origin in ruminants, its secondary loss in two independent lineages and the biologically exceptional, rapid regeneration ability of cervid antlers, we performed large-scale comparative transcriptomics and functional experiments in an accompanying paper (*87*). Substantial similarities in the transcriptome profiling of different headgear types (*87*) are consistent with a single origin of headgear in the pecoran ancestor. This ancestral headgear subsequently diversified into horns, antlers, ossicones, and pronghorns in the different pecoran families and was lost independently in the lineages of Moschidae and Hydropotinae perhaps because of the convergent pseudogenization of the horn-development *RXFP2* gene (*87*). These complex patterns, along with the lack of a well-resolved phylogeny, have confounded a synthesis of headgear evolution.

We further examined the 201 highly expressed genes in the headgear of both bovids and cervids (*87*) and identified 36 of these genes harboring pecoran-specific CNEs (table S36). Nine genes (*ALX1*, *VCAN*, *COL1A1*, *SATB2*, *RUNX2*, *POSTN*, *SP7*, *TNC*, and *COL4A2*) were classified in Gene Ontology (GO) categories related to bone development. *RUNX2* is a key transcriptional factor in the regulation of bone development (*88*) and has four pecoran-specific CNEs in its intronic regions (fig. S43A and table S36), potentially causing high *RUNX2* expression in the headgear sprouts. *SP7* encodes an important regulatory factor of the biomineralization and formation of bones (*89*) and has specific mutations in the pecoran 3′ untranslated region, one of which was located in the binding region of the microRNA bta-mir-145 (fig. S43B). These genes were possibly rewired for the formation of bony headgears of ruminants, laying out a hypothesis that can be tested experimentally in the future.

In addition, we explored the genetic basis of the keratinous sheath found convergently in Bovidae and Antilocapridae headgear (Fig. 5B). Transcriptomic data from horn sprouts of sheep and goat (*87*), both of which belong to Bovidae, identified seven highly expressed keratin genes: *KRT1*, *KRT2*, *KRT3*, *KRT5*, *KRT10*, *KRT14*, and *KRT84* (fig. S44A). Except for *KRT10* and *KRT14*, the above keratin genes encode Type II α-keratin proteins, suggesting an essential role for Type II α-keratin genes in the formation of the keratinous sheath of bovids. We further examined convergent amino acid substitutions between the pronghorn (*Antilocapra americana*) and bovids. Using the 12 mammalian species and other ruminant species as outgroups, we identified 106 proteins (table S37) that contain at least two convergent amino acid changes in these two ruminant families (*23*). Among these proteins, a horn-specific Type II α-keratin protein, KRT82, contained two convergently changed amino acid sites, $Ala^{82}Thr$ and $Ser^{509}Ala$, of which the $Ala^{82}Thr$ is located in the keratin head domain (Fig. 5B and fig. S44B). This suggests that Type II α-keratin may be important in forming the keratinous sheath that evolved convergently in Bovidae and Antilocapridae.

Evolution of body size in ruminants

Ruminants, especially bovids, encompass a wide body size range that spans four orders of magnitude, from as little as 2 kg to as high as 1200 kg (Fig. 5C) (*10*, *11*). We retrieved 642 genes related to the development of body size and estimated the ratios between the nonsynonymous substitution rate (d*N*) and synonymous substitution rate (d*S*) of these genes on branches that exhibit substantially increased body sizes and on branches that exhibit decreased body sizes (fig. S45). Compared with the background, these genes showed significantly higher d*N*/d*S* ratios in branches with large-sized and small-sized species (Student's *t* test, $P < 0.01$) but similar ratios in branches with medium-sized species (Fig. 5D). We observed six genes (*CXCL13*, *RNF115*, *NPNT*, *KL*, *SLC9A3R1*, and *MSTN*) that had significantly elevated d*N*/d*S* ratios along with the increasing of body size (Student's *t* test, $P < 0.01$) (table S38), and *SLC9A3R1* and *MSTN* have d*N*/d*S* values > 1, an indication of positive selection. *SLC9A3R1* affects osteogenesis by mineralizing osteoblasts, and disrupting this gene resulted in reduced body weights in mice (*90*). *MSTN* is an important gene in the regulation of muscle cell growth and differentiation (*91*), and mutations in *MSTN* affect the muscle mass of goat (*92*), sheep (*93*), and cattle (*94*, *95*) as well as other mammals (*96*, *97*). These results indicate that *SLC9A3R1* and *MSTN* might be targets of natural selection favoring increased body sizes (fig. S45) by regulating the development of bone and muscle. In reduced body size branches, five genes (*SBDS*, *BMP3*, *LRRN3*, *NFATC3*, and *SMARCAL1*) had significantly elevated d*N*/d*S* ratios (table S38), and the d*N*/d*S* ratio of *SBDS* was larger than 1. *SBDS* is an important gene in cell proliferation and is the causal gene of Shwachman-Diamond syndrome in humans, which is characterized by skeletal abnormalities and short stature (*98*). *BMP3* is a well-known gene involved in the regulation of osteogenesis in mammals (*99*). The genes mentioned above may therefore explain the body size variation in ruminants and may be relevant to livestock breeding application.

As the tallest terrestrial animal, giraffes have a distinct stature and body morphology, which likely are adaptions to their savanna habitat (*100*). Among the 366 genes related to bone development in the KEGG annotation, 115 genes had giraffe-specific mutations (table S39), including genes in the transforming growth factor–β (TGF- β), Hedgehog, Notch, Wnt, and FGF signaling pathways. Eleven genes with more than four nonsynonymous mutations may be related to the extreme elongation of body structures in giraffe because they are part of bone development pathways: TGF-β (*CHRD* and *LTBP1*), Wnt (*APC*, *APC2*, and *CREBBP*), Notch (*NOTCH1*, *NOTCH3*, and *NOTCH4*), Hedgehog (*GLI2* and *GLI3*), and FGF (*FGFRL1*) (Fig. 5C). Three such genes (*FGFRL1*, *NOTCH4*, and *CREBBP*) had been identified in a previous comparative study by using a low-coverage genome assembly of giraffe (*Giraffa camelopardalis*) (table S39) (*100*).

Adaptations to cursorial locomotion

Although early ruminants were probably forest-dwelling (*101*), many lineages have adapted to open habitats by adopting more cursorial body

plans designed for rapid and/or sustained movement in an open habitat (*102*). These include adaptations in limb morphology (*103*) and possibly also physiology. The pronghorn (*Antilocapra americana*) and members of Antilopinae such as the springbok (*Antidorcas marsupialis*) are among the fastest-running animals on Earth (*104*). Both lineages exhibit strong selection in their mitochondrial electron transport chains, although the targeted genes of selection differ between these species. In Antilopinae, two genes, *SUOX* (sulfite oxidase) and *NLN* (neurolysin), were under positive selection (χ^2 test, $P < 0.05$) (table S24). In the pronghorn, two mitochondrial electron transport chain enzyme encoding genes [*COX5A* (cytochrome c oxidase subunit 5A) and *PPOX* (protoporphyrinogen oxidase)] were subjected to positive selection (χ^2 test, P value < 0.05) (table S24). In addition, four proteins in the mitochondrial electron transport chain—NDUFA10 (reduced nicotinamide adenine dinucleotide dehydrogenase 1 α subcomplex subunit 10), SDHB (succinate dehydrogenase iron-sulfur subunit), UQCRC2 (cytochrome b-c1 complex subunit 2), and ATP5B (adenosine 5′-triphosphate synthase subunit β), functioning in oxidative phosphorylation—exhibited convergent amino acid changes (Fig. 5E and tables S40 and S41). Furthermore, all species of the bovid tribe Alcelaphini, which are adapted to open and seasonal

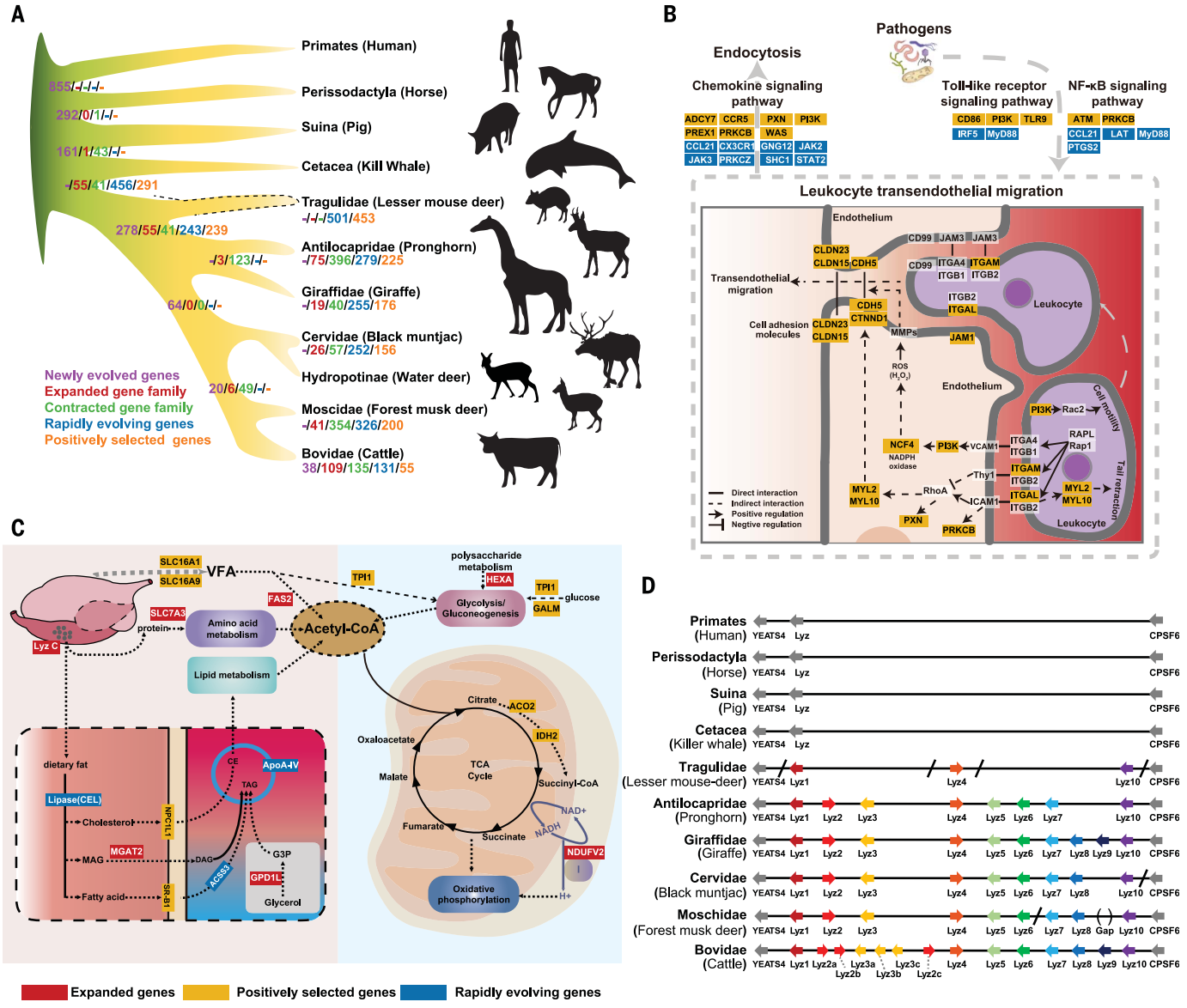


Fig. 4. Newly evolved genes, expanded and contracted gene families, REGs, and PSGs in ruminants. (**A**) Positively selected genes (PSGs), rapidly evolving genes (REGs), expanded genes, and newly evolved genes are shown along the phylogenetic tree. PSGs and REGs are identified with PAML. Newly evolved genes are identified based on the synteny relationships, with goat used as a reference. Expanded gene families are identified for the pecoran families. Tragulidae are excluded because of the relatively low quality of the assembled tragulid genome. (**B**) Diagram of the process of leukocytes crossing blood vessels, highlighting PSGs (orange) and REGs (blue). The orange rectangles indicate PSGs, and the blue rectangles represent the REGs. The solid lines represent direct interaction, and the dotted lines represent indirect interactions. (**C**) Diagram of nutrient metabolism evolution in Ruminantia displaying genes involved in nutrition metabolism. Expanded gene families are marked red, PSGs are marked yellow, and REGs are marked blue in (C) and (D). (**D**) Expansion of the lysozyme *c* gene family throughout the Ruminantia. The different expanded copies are presented with colors. Each line corresponds to the order or family in (A), and we chose the best assembled species genome sequences to draw the region. The slashes on each line indicate the ends of scaffolds in the assembled genomes.

grasslands and have cursorial locomotion, shared three specific mutations (Tyr[211]Phe, Gln[481]Arg, and Glu[622]Ala) in the *ACE* (angiotensin I converting enzyme) gene (fig. S46) and one specific mutation (Glu[97]Asp) in the *EPO* (erythropoietin) gene (fig. S47), both of which are important for endurance (*105*, *106*). These observations imply that shared or distinct molecular pathways might have played roles in convergent phenotypic evolution during the radiation of ruminants.

Evolution of dentition

Ruminants have specialized dentition patterns, lacking upper incisors and having a high prevalence of high-crowned or hypsodont teeth—a likely adaptation to abrasive diets such as grass or

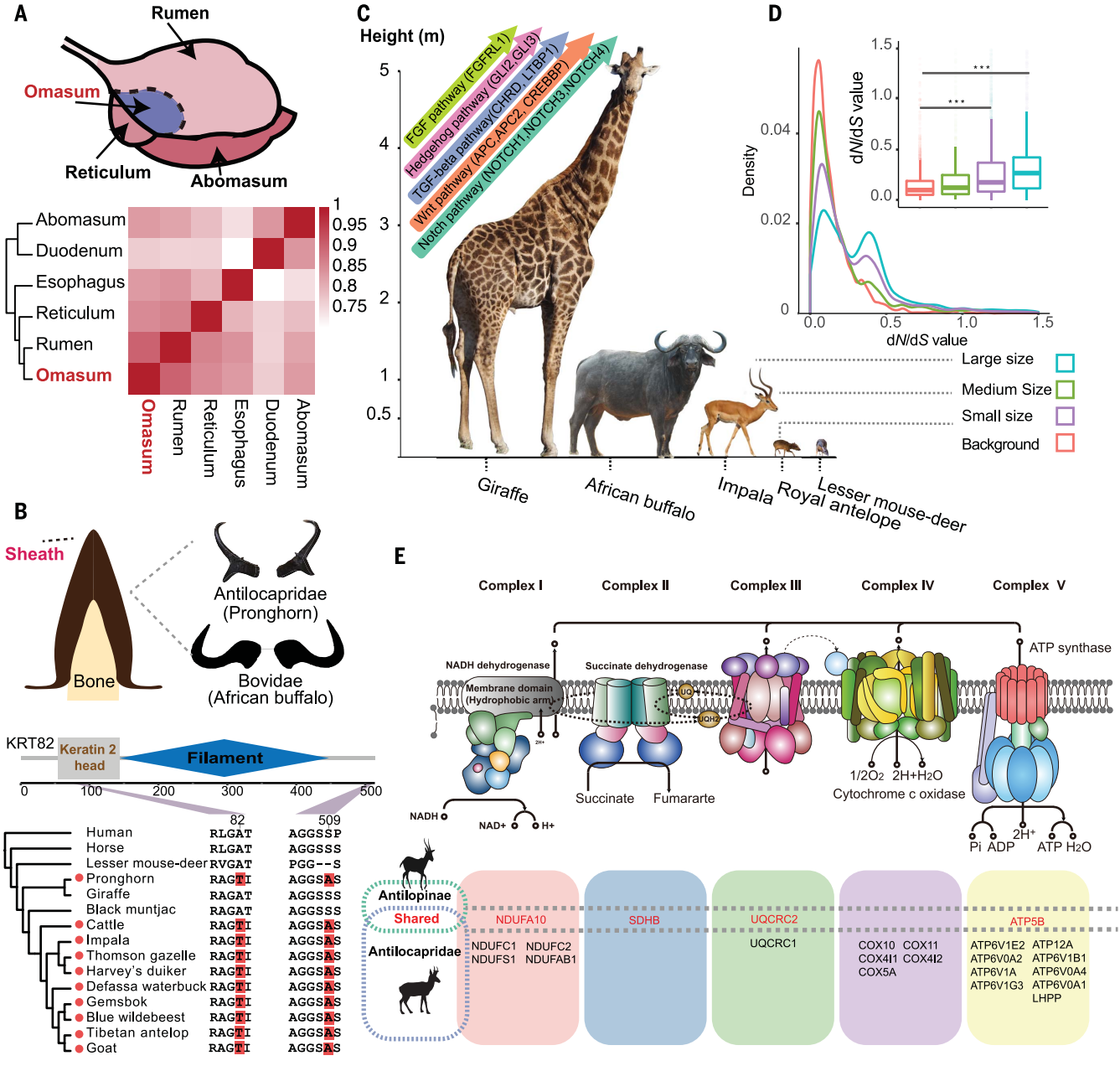


Fig. 5. Genomic features related to ruminant characteristics. (**A**) Pairwise Spearman correlations of gene expressions indicated that the omasum, rumen, and reticulum evolved as an extension of the esophagus, whereas the abomasum evolved as an extension of the duodenum. Although it is anatomically closer to the reticulum, the omasum has a more similar gene expression atlas with that of the rumen than the reticulum, which mirrors the similar functions between them. (**B**) Diagram of the convergent feature of keratinous sheath in the Bovidae and Antilocapridae. The two red amino acids indicate convergent mutations of the KRT82 between Antilocapridae and Bovidae, and the red dots indicate the included species of Antilocapridae and Bovidae. (**C**) Body size contrast among ruminants and 11 genes related with bone development that have at least four specific mutations in giraffe and are involved in TGF-β, Hedgehog, Notch, Wnt, and FGF pathways. (**D**) A total of 642 genes in GO related to the development of body size are retrieved, and we calculated the d*N*/d*S* ratios on the branches, leading to large, medium, and small body sizes, under the two-ratio branch model (model 2) of PAML. Background d*N*/d*S* ratios are calculated under one-ratio branch model (model 1) of PAML. The distribution densities of the d*N*/d*S* values are shown. A box plot of d*N*/d*S* values in different body size categories is shown. The mean d*N*/d*S* values at the branches of large body size and small body size are significantly larger than background. ***$P < 0.01$. (**E**) Antilocapridae and Antilopinae species are among the most mobile land mammals, and both of them have specific mutations in several genes of the mitochondrial electron transport chain.

feeding under abrasive conditions, as encountered in dry habitats (*107*). The distinct dentition is believed to have contributed to the evolutionary success of the ruminants by expanding their food sources (*108*). We observed that 11 genes related to dentition had RSCNEs in the 10-kbp downstream or upstream or in introns (table S42). Among these genes, *ENAM* functions in the mineralization and structural organization of enamel (*109*) and is possibly associated with enamel thickness in humans (*110*). Furthermore, *ENAM* is an important factor in the adaptation of dentition to specific diets (*111*). We observed seven RSCNEs in this gene region (fig. S48A). In addition to these RSCNEs, *ENAM* had a two–amino acid insertion and several mutations that were specific to ruminants (fig. S48, B and C). A genome-wide association analysis using the hypsodonty index of ruminants [data from (*112*)] identified one SNP significantly associated (Fisher's exact test, $P < 0.01$) with the hypsodonty index located in the binding site of the *EBF1* transcription factor in the intron of *FGFI4* (fig. S49), an important factor controlling tooth development (*113*). These genes provide candidates for future experimental studies on mammal dentition.

Conclusion

A well-resolved phylogeny of Ruminantia with full genome data provides an opportunity to understand the evolutionary processes and genetic basis of the distinct structure and evolutionary patterns of ruminants. Our comprehensive evolutionary and comparative analyses have revealed numerous genetic variations correlated with specific traits in ruminants. This study provides valuable genomic resources as well as insights into not only the evolution and diversification of ruminants but also our understanding of mammalian biology.

Materials and methods

We sequenced and assembled 44 ruminant genomes using Illumina reads, with the black muntjac (*Muntiacus crinifrons*) further sequenced with PacBio SMART long reads. The assemblies were performed with SOAPdenovo v1.05 (*114*), Platanus v1.2.4 (*115*), and Supernova assembler (*116*), and some contigs were further scaffolded with SSPACE v3.0 (*117*) and "cross_genome" commands in the Phusion2 package (*118*). Repeat elements were annotated by combining results of Tandem Repeat Finder v4.07b (*119*), RepeatMasker v4.0.5 (*120*), RepeatModeller v1.0.4 (*120*), and LTR_FINDER v1.0.6 (*121*). Genes were annotated with homologous TBLASTN (*122*) protein searches of human (Ensembl 87 release), cattle (Ensembl 87 release), and sheep (Ensembl 87 release), combining de novo prediction of SNAP (*123*), GENSCAN v1.0 (*124*), glimmerHMM v3.0.3 (*125*), and AUGUSTUS v2.5.5 (*126*). Whole-genome alignments were constructed with LAST v885 (*127*) and MULTIZ v11.2 (*128*), using goat as the reference. The maximum likelihood phylogenic trees were constructed with RAxML v8.2.9 (*129*) and ExaML v3.0.17 (*130*) by using a general time-reversible model ("GTR+GAMMA"). ASTRAL III v5.1.1 (*36*) and MP-EST v2.0 (*38*) were used to perform coalescent species tree estimations. DiscoVista v1.0 (*131*) was used to quantify and visualize gene tree discordance for alternative topologies. Gene flows among ruminant families were performed with ABBA-BABA test (*39*), D_{FOIL} (*40*), admixturegraph v1.0.2 (*41*), PhyloNet v3.6.9 (*42*), and PhyloNetworks v0.9.0 (*43*). Time calibration was conducted with r8s v1.70 (*132*), BEAST v1.8.4 (*133*), MCMCTREE in PAML v4.8 (*134*), and Multidivtime (*135*). Demographic history reconstruction was carried out by means of PSMC analysis (*56*). PhyloFit v1.4 (*136*) and phastCons v1.4 (*137*) were used to infer conserved nonexonic elements in ruminants. PSGs and REGs were identified by use of branch and branch-site models in PAML v4.8 (*134*). The gene family expansion or contraction analysis was performed by using CAFÉ v4.7 (*138*). In-house scripts and pipelines are deposited in Zenodo (https://zenodo.org/record/2549147). Detailed methods and materials are described in the supplementary materials (*23*).

REFERENCES AND NOTES

1. IUCN, *The IUCN Red List of Threatened Species*, Version 2017-3. (2017); www.iucn.org.
2. D. E. Wilson, D. M. Reeder, *Mammal Species of the World: A Taxonomic and Geographic Reference* (JHU Press, 2005).
3. R. Heller, P. Frandsen, E. D. Lorenzen, H. R. Siegismund, Are there really twice as many bovid species as we thought? *Syst. Biol.* **62**, 490–493 (2013). doi: 10.1093/sysbio/syt004; pmid: 23362112
4. T. F. Randolph *et al*., Invited review: Role of livestock in human nutrition and health for poverty reduction in developing countries. *J. Anim. Sci.* **85**, 2788–2800 (2007). doi: 10.2527/jas.2007-0467; pmid: 17911229
5. P. H. Hemsworth, G. J. Coleman, *Human-Livestock Interactions: The Stockperson and the Productivity of Intensively Farmed Animals*. (CABI, 2010).
6. C. Janis, The evolutionary strategy of the Equidae and the origins of Rumen and Cecal digestion. *Evolution* **30**, 757–774 (1976). doi: 10.1111/j.1558-5646.1976.tb00957.x; pmid: 28563331
7. J. B. Russell, J. L. Rychlik, Factors that alter rumen microbial ecology. *Science* **292**, 1119–1122 (2001). doi: 10.1126/science.1058830; pmid: 11352069
8. M. Clauss, G. E. Rössner, Old world ruminant morphophysiology, life history, and fossil record: Exploring key innovations of a diversification sequence. *Ann. Zool. Fenn.* **51**, 80–94 (2014). doi: 10.5735/086.051.0210
9. T. J. Hackmann, J. N. Spain, Invited review: ruminant ecology and evolution: perspectives useful to ruminant livestock research and production. *J. Dairy Sci.* **93**, 1320–1334 (2010). doi: 10.3168/jds.2009-2071; pmid: 20338409
10. J. R. Castelló, *Bovids of the World: Antelopes, Gazelles, Cattle, Goats, Sheep, and Relatives*. (Princeton Univ. Press, 2016).
11. J. T. du Toit, N. Owen-Smith, Body size, population metabolism, and habitat specialization among large African herbivores. *Am. Nat.* **133**, 736–740 (1989). doi: 10.1086/284949
12. M. Hernández Fernández, E. S. Vrba, A complete estimate of the phylogenetic relationships in Ruminantia: A dated species-level supertree of the extant ruminants. *Biol. Rev. Camb. Philos. Soc.* **80**, 269–302 (2005). doi: 10.1017/S1464793104006670; pmid: 15921052
13. J. E. Decker *et al*., Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 18644–18649 (2009). doi: 10.1073/pnas.0904691106; pmid: 19846765
14. A. Hassanin *et al*., Pattern and timing of diversification of Cetartiodactyla (Mammalia, Laurasiatheria), as revealed by a comprehensive analysis of mitochondrial genomes. *C. R. Biol.* **335**, 32–50 (2012). doi: 10.1016/j.crvi.2011.11.002; pmid: 22226162
15. F. Bibi, A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla, Ruminantia) and the importance of the fossil record to systematics. *BMC Evol. Biol.* **13**, 166 (2013). doi: 10.1186/1471-2148-13-166; pmid: 23927069
16. Y. Jiang *et al*., The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**, 1168–1173 (2014). doi: 10.1126/science.1252806; pmid: 24904168
17. A. V. Zimin *et al*., A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biol.* **10**, R42 (2009). doi: 10.1186/gb-2009-10-4-r42; pmid: 19393038
18. Q. Qiu *et al*., The yak genome and adaptation to life at high altitude. *Nat. Genet.* **44**, 946–949 (2012). doi: 10.1038/ng.2343; pmid: 22751099
19. D. M. Bickhart *et al*., Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **49**, 643–650 (2017). doi: 10.1038/ng.3802; pmid: 28263316
20. R. L. Ge *et al*., Draft genome sequence of the Tibetan antelope. *Nat. Commun.* **4**, 1858 (2013). doi: 10.1038/ncomms2860; pmid: 23673643
21. C. M. Seabury *et al*., Genome-wide polymorphism and comparative analyses in the white-tailed deer (Odocoileus virginianus): A model for conservation genomics. *PLOS ONE* **6**, e15811 (2011). doi: 10.1371/journal.pone.0015811; pmid: 21283515
22. R. Kropatsch *et al*., SOX9 duplication linked to intersex in deer. *PLOS ONE* **8**, e73734 (2013). doi: 10.1371/journal.pone.0073734; pmid: 24040047
23. Materials and methods are available as supplementary materials.
24. Z. Li *et al*., Draft genome of the reindeer (*Rangifer tarandus*). *Gigascience* **6**, 1–5 (2017). doi: 10.1093/gigascience/gix102; pmid: 29099922
25. C. Zhang *et al*., Draft genome of the milu (*Elaphurus davidianus*). *Gigascience* **7**, gix130 (2018). doi: 10.1093/gigascience/gix130; pmid: 29267854
26. Y. Yang *et al*., Draft genome of the Marco Polo Sheep (*Ovis ammon polii*). *Gigascience* **6**, 1–7 (2017). doi: 10.1093/gigascience/gix106; pmid: 29112761
27. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015). doi: 10.1093/bioinformatics/btv351; pmid: 26059717
28. T. R. Gregory, *Animal Genome Size Database* (2017); www.genomesize.com.
29. A. D. Foote *et al*., Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275 (2015). doi: 10.1038/ng.3198; pmid: 25621460
30. D. Brawand *et al*., The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014). doi: 10.1038/nature13726; pmid: 25186727
31. E. D. Jarvis *et al*., Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014). doi: 10.1126/science.1253451; pmid: 25504713
32. A. Hobolth, J. Y. Dutheil, J. Hawks, M. H. Schierup, T. Mailund, Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res.* **21**, 349–356 (2011). doi: 10.1101/gr.114751.110; pmid: 21270173
33. K. Prüfer *et al*., The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527–531 (2012). doi: 10.1038/nature11128; pmid: 22722832
34. S. Mirarab, M. S. Bayzid, B. Boussau, T. Warnow, Statistical binning enables an accurate coalescent-based estimation of the avian tree. *Science* **346**, 1250463 (2014). doi: 10.1126/science.1250463; pmid: 25504728
35. S. Patel, R. Kimball, E. Braun, Error in phylogenetic estimation for bushes in the tree of life. *J. Phylogenetics Evol. Biol.* **1**, 110 (2013). doi: 10.4172/2329-9002.1000110
36. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018). doi: 10.1186/s12859-018-2129-y; pmid: 29745866
37. M. S. Springer, J. Gatesy, The gene tree delusion. *Mol. Phylogenet. Evol.* **94** (Pt A), 1–33 (2016). doi: 10.1016/j.ympev.2015.07.018; pmid: 26238460
38. L. Liu, L. Yu, S. V. Edwards, A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302 (2010). doi: 10.1186/1471-2148-10-302; pmid: 20937096
39. E. Y. Durand, N. Patterson, D. Reich, M. Slatkin, Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011). doi: 10.1093/molbev/msr048; pmid: 21325092

40. J. B. Pease, M. W. Hahn, Detection and polarization of introgression in a five-taxon phylogeny. *Syst. Biol.* **64**, 651–662 (2015). doi: 10.1093/sysbio/syv023; pmid: 25888025
41. K. Leppälä, S. V. Nielsen, T. Mailund, admixturegraph: An R package for admixture graph manipulation and fitting. *Bioinformatics* **33**, 1738–1740 (2017). doi: 10.1093/bioinformatics/btx048; pmid: 28158333
42. D. Wen, Y. Yu, J. Zhu, L. Nakhleh, Inferring phylogenetic networks using PhyloNet. *Syst. Biol.* **67**, 735–740 (2018). doi: 10.1093/sysbio/syy015; pmid: 29514307
43. C. Solís-Lemus, P. Bastide, C. Ané, PhyloNetworks: A package for phylogenetic networks. *Mol. Biol. Evol.* **34**, 3292–3298 (2017). doi: 10.1093/molbev/msx235; pmid: 28961984
44. O. R. Bininda-Emonds *et al*., The delayed rise of present-day mammals. *Nature* **446**, 507–512 (2007). doi: 10.1038/nature05634; pmid: 17392779
45. R. W. Meredith *et al*., Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011). doi: 10.1126/science.1211028; pmid: 21940861
46. E. Sayyari, S. Mirarab, Fast coalescent-based computation of local branch support from quartet frequencies. *Mol. Biol. Evol.* **33**, 1654–1668 (2016). doi: 10.1093/molbev/msw079; pmid: 27189547
47. S. D. Webb, B. E. Taylor, *The Phylogeny of Hornless Ruminants and a Description of the Cranium of Archaeomeryx* (AMNH, 1980).
48. C. M. Janis, J. M. Theodor, Cranial and postcranial morphological data in ruminant phylogenetics. *Zitteliana* **32**, 15–31 (2014).
49. C. M. Janis, Grades and clades in hornless ruminant evolution: The reality of the Gelocidae and the systematic position of Lophiomeryx and Bachitherium. *J. Vertebr. Paleontol.* **7**, 200–216 (1987). doi: 10.1080/02724634.1987.10011653
50. I. M. Sánchez, M. S. Domingo, J. Morales, The genus Hispanomeryx (Mammalia, Ruminantia, Moschidae) and its bearing on musk deer phylogeny and systematics. *Palaeontology* **53**, 1023–1047 (2010). doi: 10.1111/j.1475-4983.2010.00992.x
51. A. Hassanin, E. J. Douzery, Molecular and morphological phylogenies of ruminantia and the alternative position of the moschidae. *Syst. Biol.* **52**, 206–228 (2003). doi: 10.1080/10635150390192726; pmid: 12746147
52. M. dos Reis *et al*., Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. Biol. Sci.* **279**, 3491–3500 (2012). doi: 10.1098/rspb.2012.0683; pmid: 22628470
53. F. Bibi, Assembling the ruminant tree: Combining morphology, molecules, extant taxa, and fossils. *Zitteliana* **32**, 197–211 (2014).
54. C. A. Matthee, S. K. Davis, Molecular insights into the evolution of the family Bovidae: A nuclear DNA perspective. *Mol. Biol. Evol.* **18**, 1220–1230 (2001). doi: 10.1093/oxfordjournals.molbev.a003908; pmid: 11420362
55. C. Z. Yang *et al*., Phylogenetic analyses and improved resolution of the family Bovidae based on complete mitochondrial genomes. *Biochem. Syst. Ecol.* **48**, 136–143 (2013). doi: 10.1016/j.bse.2012.12.005
56. C. A. Matthee, T. J. Robinson, Cytochrome b phylogeny of the family bovidae: Resolution within the alcelaphini, antilopini, neotragini, and tragelaphini. *Mol. Phylogenet. Evol.* **12**, 31–46 (1999). doi: 10.1006/mpev.1998.0573; pmid: 10222159
57. A. García-Ruiz *et al*., Changes in genetic selection differentials and generation intervals in US Holstein dairy cattle as a result of genomic selection. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3995–E4004 (2016). doi: 10.1073/pnas.1519061113; pmid: 27354521
58. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011). doi: 10.1038/nature10231; pmid: 21753753
59. E. D. Lorenzen *et al*., Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature* **479**, 359–364 (2011). doi: 10.1038/nature10574; pmid: 22048313
60. A. D. Barnosky, P. L. Koch, R. S. Feranec, S. L. Wing, A. B. Shabel, Assessing the causes of late Pleistocene extinctions on the continents. *Science* **306**, 70–75 (2004). doi: 10.1126/science.1101476; pmid: 15459379
61. C. J. Bae, K. Douka, M. D. Petraglia, On the origin of modern humans: Asian perspectives. *Science* **358**, eaai9067 (2017). doi: 10.1126/science.aai9067; pmid: 29217544
62. J. C. Venter *et al*., The sequence of the human genome. *Science* **291**, 1304–1351 (2001). doi: 10.1126/science.1058040; pmid: 11181995
63. E. F. Kirkness *et al*., The dog genome: Survey sequencing and comparative analysis. *Science* **301**, 1898–1903 (2003). doi: 10.1126/science.1086432; pmid: 14512627
64. C. M. Wade *et al*., Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**, 865–867 (2009). doi: 10.1126/science.1178158; pmid: 19892987
65. M. A. M. Groenen *et al*., Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393–398 (2012). doi: 10.1038/nature11622; pmid: 23151582
66. H. Wu *et al*., Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun.* **5**, 5188 (2014). doi: 10.1038/ncomms6188; pmid: 25333821
67. B. P. Telugu, A. M. Walker, J. A. Green, Characterization of the bovine pregnancy-associated glycoprotein gene family—Analysis of gene sequences, regulatory regions within the promoter and expression of selected genes. *BMC Genomics* **10**, 185 (2009). doi: 10.1186/1471-2164-10-185; pmid: 19393060
68. W. A. Muller, Mechanisms of leukocyte transendothelial migration. *Annu. Rev. Pathol.* **6**, 323–344 (2011). doi: 10.1146/annurev-pathol-011110-130224; pmid: 21073340
69. T. H. Mogensen, Pathogen recognition and inflammatory signaling in innate immune defenses. *Clin. Microbiol. Rev.* **22**, 240–273 (2009). doi: 10.1128/CMR.00046-08; pmid: 19366914
70. R. D. Finn *et al*., Pfam: The protein families database. *Nucleic Acids Res.* **42** (D1), D222–D230 (2014). doi: 10.1093/nar/gkt1223; pmid: 24288371
71. R. A. Nafikov, D. C. Beitz, Carbohydrate and lipid metabolism in farm animals. *J. Nutr.* **137**, 702–705 (2007). doi: 10.1093/jn/137.3.702; pmid: 17311965
72. C. M. B. Membrive, *Rumenology* (Springer, 2016).
73. E. D. Warner, The organogenesis and early histogenesis of the bovine stomach. *Am. J. Anat.* **102**, 33–63 (1958). doi: 10.1002/aja.1001020103; pmid: 13545182
74. R. Xiang, V. H. Oddy, A. L. Archibald, P. E. Vercoe, B. P. Dalrymple, Epithelial, metabolic and innate immunity transcriptomic signatures differentiating the rumen from other sheep and mammalian gastrointestinal tract tissues. *PeerJ* **4**, e1762 (2016). doi: 10.7717/peerj.1762; pmid: 26989612
75. F. Lunardi *et al*., Overexpression of SERPIN B3 promotes epithelial proliferation and lung fibrosis in mice. *Lab. Invest.* **91**, 945–954 (2011). doi: 10.1038/labinvest.2011.1; pmid: 21403642
76. M. Gatto *et al*., Serpins, immunity and autoimmunity: Old molecules, new functions. *Clin. Rev. Allergy Immunol.* **45**, 267–280 (2013). doi: 10.1007/s12016-013-8353-3; pmid: 23325331
77. P. Martin, Wound healing—Aiming for perfect skin regeneration. *Science* **276**, 75–81 (1997). doi: 10.1126/science.276.5309.75; pmid: 9082989
78. P. Wong, P. A. Coulombe, Loss of keratin 6 (K6) proteins reveals a function for intermediate filaments during wound repair. *J. Cell Biol.* **163**, 327–337 (2003). doi: 10.1083/jcb.200305032; pmid: 14568992
79. D. M. Irwin, Genomic organization and evolution of ruminant lysozyme c genes. *Dongwuxue Yanjiu* **36**, 1–17 (2015). pmid: 25730456
80. N. Hevir, K. Vouk, J. Šinkovec, M. Ribič-Pucelj, T. L. Rižner, Aldo-keto reductases AKR1C1, AKR1C2 and AKR1C3 may enhance progesterone metabolism in ovarian endometriosis. *Chem. Biol. Interact.* **191**, 217–226 (2011). doi: 10.1016/j.cbi.2011.01.003; pmid: 21232532
81. I. Hanukoglu, A. Hanukoglu, Epithelial sodium channel (ENaC) family: Phylogeny, structure-function, tissue distribution, and associated inherited diseases. *Gene* **579**, 95–132 (2016). doi: 10.1016/j.gene.2015.12.061; pmid: 26772908
82. Z. Xiong *et al*., PAX9 regulates squamous cell differentiation and carcinogenesis in the oro-oesophageal epithelium. *J. Pathol.* **244**, 164–175 (2018). doi: 10.1002/path.4998; pmid: 29055049
83. X. Meng, J. Shi, B. Peng, X. Zou, C. Zhang, Effect of mouse Sim2 gene on the cell cycle of PC12 cells. *Cell Biol. Int.* **30**, 349–353 (2006). doi: 10.1016/j.cellbi.2005.11.012; pmid: 16530433
84. B. Laffin *et al*., Loss of singleminded-2s in the mouse mammary gland induces an epithelial-mesenchymal transition associated with up-regulation of slug and matrix metalloprotease 2. *Mol. Cell. Biol.* **28**, 1936–1946 (2008). doi: 10.1128/MCB.01701-07; pmid: 18160708
85. G. B. Penner, M. A. Steele, J. R. Aschenbach, B. W. McBride, Ruminant Nutrition Symposium: Molecular adaptation of ruminal epithelia to highly fermentable diets. *J. Anim. Sci.* **89**, 1108–1119 (2011). doi: 10.2527/jas.2010-3378; pmid: 20971890
86. E. B. Davis, K. A. Brakora, A. H. Lee, Evolution of ruminant headgear: A review. *Proc. Biol. Sci.* **278**, 2857–2865 (2011). doi: 10.1098/rspb.2011.0938; pmid: 21733893
87. Y. Wang *et al*., Genetic basis of ruminant headgear and rapid antler regeneration. *Science* **364**, eaav6335 (2019).
88. T. Komori, Regulation of bone development and maintenance by Runx2. *Front. Biosci.* **13**, 898–903 (2008). doi: 10.2741/2730; pmid: 17981598
89. C. Zhang, Transcriptional regulation of bone formation by the osteoblast-specific transcription factor Osx. *J. Orthop. Surg. Res.* **5**, 37 (2010). doi: 10.1186/1749-799X-5-37; pmid: 20550694
90. S. Shenolikar, J. W. Voltz, C. M. Minkoff, J. B. Wade, E. J. Weinman, Targeted disruption of the mouse NHERF-1 gene promotes internalization of proximal tubule sodium-phosphate cotransporter type IIa and renal phosphate wasting. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11470–11475 (2002). doi: 10.1073/pnas.162232699; pmid: 12169661
91. R. Sartori *et al*., BMP signaling controls muscle mass. *Nat. Genet.* **45**, 1309–1318 (2013). doi: 10.1038/ng.2772; pmid: 24076600
92. X. Wang *et al*., CRISPR/Cas9-mediated MSTN disruption and heritable mutagenesis in goats causes increased body mass. *Anim. Genet.* **49**, 43–51 (2018). doi: 10.1111/age.12626; pmid: 29446146
93. H. Li *et al*., Generation of biallelic knock-out sheep via gene-editing and somatic cell nuclear transfer. *Sci. Rep.* **6**, 33675 (2016). doi: 10.1038/srep33675; pmid: 27654750
94. L. Grobet *et al*., A deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat. Genet.* **17**, 71–74 (1997). doi: 10.1038/ng0997-71; pmid: 9288100
95. J. Luo *et al*., Efficient generation of myostatin (MSTN) biallelic mutations in cattle using zinc finger nucleases. *PLOS ONE* **9**, e95225 (2014). doi: 10.1371/journal.pone.0095225; pmid: 24743319
96. H. Gu *et al*., Establishment and phenotypic analysis of an Mstn knockout rat. *Biochem. Biophys. Res. Commun.* **477**, 115–122 (2016). doi: 10.1016/j.bbrc.2016.06.030; pmid: 27289021
97. Y. Bi *et al*., Isozygous and selectable marker-free MSTN knockout cloned pigs generated by the combined use of CRISPR/Cas9 and Cre/LoxP. *Sci. Rep.* **6**, 31729 (2016). doi: 10.1038/srep31729; pmid: 27530319
98. G. R. Boocock *et al*., Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat. Genet.* **33**, 97–101 (2003). doi: 10.1038/ng1062; pmid: 12496757
99. A. Daluiski *et al*., Bone morphogenetic protein-3 is a negative regulator of bone density. *Nat. Genet.* **27**, 84–88 (2001). doi: 10.1038/83810; pmid: 11138004
100. M. Agaba *et al*., Giraffe genome sequence reveals clues to its unique morphology and physiology. *Nat. Commun.* **7**, 11519 (2016). doi: 10.1038/ncomms11519; pmid: 27187213
101. D. R. Prothero, S. E. Foss, *The Evolution of Artiodactyls* (JHU Press, 2007).
102. M. Köhler, *Skeleton and Habitat of Recent and Fossil Ruminants* (F. Pfeil, 1993).
103. J. Kappelman, Morphology and locomotor adaptations of the bovid femur in relation to habitat. *J. Morphol.* **198**, 119–130 (1988). doi: 10.1002/jmor.1051980111; pmid: 3199446
104. L. Baskin, K. Danell, *Ecology of ungulates: a handbook of species in Eastern Europe and Northern and Central Asia* (Springer Science & Business Media, 2003).
105. F. Ma *et al*., The association of sport performance with ACE and ACTN3 genetic polymorphisms: A systematic review and meta-analysis. *PLOS ONE* **8**, e54685 (2013). doi: 10.1371/journal.pone.0054685; pmid: 23358679
106. E. A. Ostrander, H. J. Huson, G. K. Ostrander, Genetics of athletic performance. *Annu. Rev. Genomics Hum. Genet.* **10**, 407–429 (2009). doi: 10.1146/annurev-genom-082908-150058; pmid: 19630564
107. O. Toljagic, *Macroevolution With a Bite: Teeth Evolution and Diversification in Ruminants* (Univ. Oslo Press, 2018).
108. J. L. Cantalapiedra *et al*., Dietary innovations spurred the diversification of ruminants during the Caenozoic. *Proc. Biol. Sci.* **281**, 20132746 (2013). doi: 10.1098/rspb.2013.2746; pmid: 24352949

109. J. C. Hu *et al*., Enamel defects and ameloblast-specific expression in Enam knock-out/lacz knock-in mice. *J. Biol. Chem.* **283**, 10858–10871 (2008). doi: 10.1074/jbc.M710565200; pmid: 18252720
110. D. M. Daubert *et al*., Human enamel thickness and ENAM polymorphism. *Int. J. Oral Sci.* **8**, 93–97 (2016). doi: 10.1038/ijos.2016.1; pmid: 27357321
111. R. P. Shellis, A. D. Beynon, D. J. Reid, K. M. Hiiemae, Variations in molar enamel thickness among primates. *J. Hum. Evol.* **35**, 507–522 (1998). doi: 10.1006/jhev.1998.0238; pmid: 9774508
112. O. Toljagic, K. L. Voje, M. Matschiner, L. H. Liow, T. F. Hansen, Millions of Years Behind: Slow adaptation of ruminants to grasslands. *Syst. Biol.* **67**, 145–157 (2018). doi: 10.1093/sysbio/syx059; pmid: 28637223
113. C. Y. Li, J. Prochazka, A. F. Goodwin, O. D. Klein, Fibroblast growth factor signaling in mammalian tooth development. *Odontology* **102**, 1–13 (2014). doi: 10.1007/s10266-013-0142-1; pmid: 24343791
114. R. Li *et al*., De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010). doi: 10.1101/gr.097261.109; pmid: 20019144
115. R. Kajitani *et al*., Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014). doi: 10.1101/gr.170720.113; pmid: 24755901
116. N. I. Weisenfeld, V. Kumar, P. Shah, D. M. Church, D. B. Jaffe, Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017). doi: 10.1101/gr.214874.116; pmid: 28381613
117. M. Boetzer, C. V. Henkel, H. J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579 (2011). doi: 10.1093/bioinformatics/btq683; pmid: 21149342
118. J. C. Mullikin, Z. Ning, The phusion assembler. *Genome Res.* **13**, 81–90 (2003). doi: 10.1101/gr.731003; pmid: 12529309
119. G. Benson, Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999). doi: 10.1093/nar/27.2.573; pmid: 9862982
120. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0 (2015); www.repeatmasker.org
121. Z. Xu, H. Wang, LTR_FINDER: An efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265-8 (2007). doi: 10.1093/nar/gkm286; pmid: 17485477
122. S. F. Altschul *et al*., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). doi: 10.1093/nar/25.17.3389; pmid: 9254694
123. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004). doi: 10.1186/1471-2105-5-59; pmid: 15144565
124. C. Burge, S. Karlin, Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997). doi: 10.1006/jmbi.1997.0951; pmid: 9149143
125. W. H. Majoros, M. Pertea, S. L. Salzberg, TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004). doi: 10.1093/bioinformatics/bth315; pmid: 15145805
126. M. Stanke *et al*., AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435-9 (2006). doi: 10.1093/nar/gkl200; pmid: 16845043
127. S. M. Kietbasa, R. Wan, K. Sato, P. Horton, M. C. Frith, Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011). doi: 10.1101/gr.113985.110; pmid: 21209072
128. M. Blanchette *et al*., Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004). doi: 10.1101/gr.1933104; pmid: 15060014
129. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). doi: 10.1093/bioinformatics/btu033; pmid: 24451623
130. A. M. Kozlov, A. J. Aberer, A. Stamatakis, ExaML version 3: A tool for phylogenomic analyses on supercomputers. *Bioinformatics* **31**, 2577–2579 (2015). doi: 10.1093/bioinformatics/btv184; pmid: 25819675
131. E. Sayyari, J. B. Whitfield, S. Mirarab, DiscoVista: Interpretable visualizations of gene tree discordance. *Mol. Phylogenet. Evol.* **122**, 110–115 (2018). doi: 10.1016/j.ympev.2018.01.019; pmid: 29421312
132. M. J. Sanderson, r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302 (2003). doi: 10.1093/bioinformatics/19.2.301; pmid: 12538260
133. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007). doi: 10.1186/1471-2148-7-214; pmid: 17996036
134. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007). doi: 10.1093/molbev/msm088; pmid: 17483113
135. J. L. Thorne, H. Kishino, Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* **51**, 689–702 (2002). doi: 10.1080/10635150290102456; pmid: 12396584
136. M. J. Hubisz, K. S. Pollard, A. Siepel, PHAST and RPHAST: Phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011). doi: 10.1093/bib/bbq072; pmid: 21278375
137. A. Siepel *et al*., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005). doi: 10.1101/gr.3715005; pmid: 16024819
138. T. De Bie, N. Cristianini, J. P. Demuth, M. W. Hahn, CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006). doi: 10.1093/bioinformatics/btl097; pmid: 16543274
139. L. Chen *et al*., Data for "Large-scale ruminant genome sequencing provides insights into their evolution and unique traits". Dryad (2019); doi:10.5061/dryad.52213gc

ACKNOWLEDGMENTS

We thank J. Zhou, L. Wong, and X. Ren from the China Biodiversity Conservation and Green Development Foundation (CBCGDF) for providing the sample of Przewalski's gazelle, A. al-Cher for help with extracting DNA from tissue, the Givskud Zoo for providing a skin sample for a bongo, R. Crowhurst for assistance with DNA extraction for the pronghorn, the members of the FANNG project for sharing their transcriptome data, and L. Nakeleh and J. Zhu for their patient instruction when doing PhyloNet analysis. **Funding:** This work was supported by the Talents Team Construction Fund of Northwestern Polytechnical University (NWPU); the Strategic Priority Research Program of CAS (XDB13000000) to W.W.; the Talents Team Construction Fund of Northwestern Polytechnical University (NWPU) and the National Program for Support of Top-notch Young Professionals to Q.Q.; the Villum Foundation Young Investigator grant (VKR023447) to R.H.; the Lundbeck Fellowship (R190-2014-2827), the Carlsberg Foundation (CF16-0663), and the Villum Investigator Grant to G.Z.; the National Natural Science Foundation of China (31822052) to Y.J.; the 1000 Talent Project of Shaanxi Province to W.W., Q.Q., and W.K.; the Fundamental Research Funds for the Central Universities (3102019JC007) to W.W. and Q.Q.; the Natural Science Foundation of China (NSFC) (31501984); the Natural science foundation from Jilin province (20170101158JC) and the Central Public-interest Scientific Institution Basal Research Fund (Y2019GH13) to Z.P.L.; the Danish National Research Foundation (grant DNRF96) to R.R.F.; the Robert and Rosabel Osborne Endowment and the U.S. Department of Agriculture Grants (538 AG2009-34480-19875 and 538 AG 58-1265-0-031) to H.A.L.; the National Science Foundation (NSF III-1845967) to S.M.; and the San Diego Zoo Global John and Beverly Stauffer Foundation to O.A.R. **Author contributions:** W.W., R.H., and G.Z. conceived the project and designed research. L.C., W.W., and K.W. drafted the manuscripts. W.W., R.H., G.Z., F.B., Y.J., H.R.S., M.T.P.G., Q.Q., H.A.L., and S.M. revised the manuscript. L.C, Z.S.L., K.W., Q.Q., C.Z.Z., Y.Y., C.L., Y.Zh., Y.G., R.Z., Y.R., R.L., and Y.Ze. performed genome assembly and genome annotation analysis. K.W., L.C., Z.S.L, S.M., and W.L. performed phylogeny analyses. J.Y., L.C., and R.L. assembled mitochondrial genomes. L.C., Z.S.L., Y.J., F.B., and B.We. performed time calibration, PSMC analyses, and interpretation of results. L.C., Z.S.L., X.P., W.F., Y.W., Y.E.Z., C.C., W.L., and B.Wa. conducted the analysis in identifying conserved elements and new genes. L.C., C.L., Z.S.L., Y.Z.Y., and Q.Q. performed gene family analysis. Q.Q., Y.Z.Y., and C.L. conducted positive selection and rapidly evolving gene analysis. L.C., Z.S.L., X.P., Y.W., Y.Zh.., B.We., C.Z., Z.W., and Y.Q. performed analysis relating to genomic features of ruminant characters. L.C., Z.S.L., C.L., and Z.P.L. performed metabolism pathway analysis. H.A.L. contributed the sequence data of gemsbok and indian muntjac. G.L., W.T.W., W.Z., and R.P.Z. prepared and extracted DNA samples. J.W., W.S., W.N., Y.T.W., C.H., C.W.E., L.G.C., O.R., R.H., R.R.F., and H.R.S. provided tissue samples. Y.D., W.T.W., S.D., and Y.G. prepared the library and sequencing. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Genome assembles and DNA sequencing data have been deposited into NCBI with project no. PRJNA438286 and can also be accessed through the browser of database we constructed: http://animal.nwsuaf.edu.cn/code/index.php/Ruminantia. Transcriptome data of newly sequenced seven sheep samples are deposited into NCBI with accession IDs SRR7755234, SRR7755240, SRR7755450, SRR7755493, SRR7755310, SRR7755366, and SRR7755370. The alignment datasets of whole genomes, mitogenomes, orthologous genes, CNEs, and 1-kbp windows and the alternative trees based on these data are deposited in the Dryad Digital Repository (*139*).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/364/6446/eaav6202/suppl/DC1
Materials and Methods
Figs. S1 to S54
Tables S1 to S54
References (*140*–*199*)

8 October 2018; accepted 16 May 2019
10.1126/science.aav6202

Science

Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits

Lei Chen, Qiang Qiu, Yu Jiang, Kun Wang, Zeshan Lin, Zhipeng Li, Faysal Bibi, Yongzhi Yang, Jinhuan Wang, Wenhui Nie, Weiting Su, Guichun Liu, Qiye Li, Weiwei Fu, Xiangyu Pan, Chang Liu, Jie Yang, Chenzhou Zhang, Yuan Yin, Yu Wang, Yue Zhao, Chen Zhang, Zhongkai Wang, Yanli Qin, Wei Liu, Bao Wang, Yandong Ren, Ru Zhang, Yan Zeng, Rute R. da Fonseca, Bin Wei, Ran Li, Wenting Wan, Ruoping Zhao, Wenbo Zhu, Yutao Wang, Shengchang Duan, Yun Gao, Yong E. Zhang, Chunyan Chen, Christina Hvilsom, Clinton W. Epps, Leona G. Chemnick, Yang Dong, Siavash Mirarab, Hans Redlef Siegismund, Oliver A. Ryder, M. Thomas P. Gilbert, Harris A. Lewin, Guojie Zhang, Rasmus Heller and Wen Wang

Science **364** (6446), eaav6202.
DOI: 10.1126/science.aav6202

Phylogeny and characteristics of ruminants

Ruminants are a diverse group of mammals that includes families containing well-known taxa such as deer, cows, and goats. However, their evolutionary relationships have been contentious, as have the origins of their distinctive digestive systems and headgear, including antlers and horns (see the Perspective by Ker and Yang). To understand the relationships among ruminants, L. Chen *et al.* sequenced 44 species representing 6 families and performed a phylogenetic analysis. From this analysis, they were able to resolve the phylogeny of many genera and document incomplete lineage sorting among major clades. Interestingly, they found evidence for large population reductions among many taxa starting at approximately 100,000 years ago, coinciding with the migration of humans out of Africa. Examining the bony appendages on the head —the so-called headgear—Wang *et al.* describe specific evolutionary changes in the ruminants and identify selection on cancer-related genes that may function in antler development in deer. Finally, Lin *et al.* take a close look at the reindeer genome and identify the genetic basis of adaptations that allow reindeer to survive in the harsh conditions of the Arctic.

Science, this issue p. eaav6202, p. eaav6335, p. eaav6312; see also p. 1130