

1 Genomic perspectives on adaptation and conservation in the endangered long-tailed
2 goral (*Naemorhedus caudatus*).

3 Jessica A. Weber^{1†}, Oksung Chung^{2,3†}, Jungeun Kim^{2†}, JeHoon Jun^{2,3}, Hyejin Lee², Sungwoong Jho²,
 4 Yun Sung Cho³, Dae-Soo Kim⁵, Woon ki Paek⁶, Soonok Kim⁷, Hanshin Lee⁸, Semin Lee⁹, Jeremy S.
 5 Edwards¹⁰, Joseph A. Cook^{1*}, Junghwa An^{7*}, and Jong Bhak^{2,3, 4,9*}

6 ¹Department of Biology and Museum of Southwestern Biology, University of New Mexico,
 7 Albuquerque, NM 87131, United States

8 ²Personal Genomics Institute, Genome Research Foundation, Osong 28160, Republic of Korea

9 ³Geromics, Ulsan 44919, Republic of Korea

10 ⁴The Genomics Institute (TGI), Ulsan National Institute of Science and Technology (UNIST), Ulsan
 11 44919, Republic of Korea

12 ⁵Rare Disease Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon,
 13 305-333, Republic of Korea

14 ⁶National Science Museum, Daejeon, 34143, Republic of Korea

15 ⁷National Institute of Biological Resources, Incheon, 22689, Republic of Korea

16 ⁸University of California - San Diego, CA 92122, United States

17 ⁹Department of Biomedical Engineering, School of Life Sciences, UNIST, Ulsan, 44919, Republic of
 18 Korea

19 ¹⁰Department Chemistry and Chemical Biology, Molecular Genetics and Microbiology, and Chemical
 20 and Nuclear Engineering, University of New Mexico – Albuquerque, NM 87131, United States

21

1 *Correspondence: jongbhak@genomics.org; cookjose@unm.edu, safety@korea.kr

2 †Equal contributors

3 **Email addresses:**

4 JAW: weberj@unm.edu

5 OC: okokookk219@gmail.com

6 JK: jungeunkim079@gmail.com

7 JHJ: junjh0701@gmail.com

8 HL: leehj08111@gmail.com

9 SJ: sungwoong.jho@gmail.com

10 YSC: joys0406@gmail.com

11 DSK: kds2465@kribb.re.kr

12 WKP: paekwk@naver.com

13 SK: sokim90@gmail.com

14 HL: hal073@ucsd.edu

15 SL: seminlee@unist.ac.kr

16 JAC: cookjose@unm.edu

17 JSE: JSEdwards@salud.unm.edu

18 JA: safety@korea.kr

19 JB: jongbhak@genomics.org

20 **Keywords**

21 Caprini, heterozygosity, historical demography, positively selected genes

Abstract

Native to the mountains of East Asia, the long-tailed goral (*Naemorhedus caudatus*) is a vulnerable wild ungulate in the tribe Caprini. To understand key conservation issues related to fragmentation and subsequent endangerment of this montane species, we sequenced and analyzed the genome of a long-tailed goral to explore historical demography, contemporary levels of genetic diversity, and potential immune response. When compared to ten additional mammalian reference genomes, we identified 357 positively selected genes (PSGs) in the long-tailed goral and 364 PSGs in the Caprini lineage. Gene Ontology analyses showed statistical enrichment in biological processes related to immune function and in genes and pathways related to blood coagulation. We also identified low levels of heterozygosity (0.00114) in the long-tailed goral along with decreases in population size relative to other species of Caprini. Finally, we provide evidence for positive selection on the muscle development gene *myostatin* (*MSTN*) in the Caprini lineage, which may have increased the muscle development and climbing ability in the caprine common ancestor. Low effective population size and decreased heterozygosity of the long-tailed goral raise conservation concern about the effects of habitat fragmentation, over harvest, and inbreeding on this vulnerable species.

Introduction

The long-tailed goral (*Naemorhedus caudatus*), also known as the Amur goral, is a small goat-like ungulate that has been classified as Vulnerable by the International Union for Conservation of Nature (IUCN) and Endangered by the South Korean Ministry of Environment ¹. While bovid taxonomy has been historically challenging to resolve, gorals are

1 now recognized as members of the Caprini tribe of the subfamily Antilopinae, within the
2 diverse and widespread Bovidae family. Caprines are thought to have originated in the late
3 Miocene roughly 11 million years ago (MYA) in the mountainous archipelago between the
4 Mediterranean and Paratethys Seas ². About that time, the caprine common ancestor evolved
5 shortened metacarpals ideally suited for climbing, allowing it to colonize new ecological
6 niches associated with steep mountains that were not occupied by other large herbivores ². As
7 a result, caprines rapidly radiated across the mountain ranges of the Northern Hemisphere and
8 now include 62 species in 14 genera ³.

9 The genus *Naemorhaedus* is comprised of six species of gorals that inhabit the
10 mountains of Asia ^{3,4}. Long-tailed gorals are adept climbers that inhabit steep rocky slopes ¹
11 and are found in the rugged mountainous forests and shrublands of the Korean peninsula,
12 eastern Russia (Primorsky and Khabarovsk Territories), and northeastern China ⁵. In Korea,
13 the long-tailed goral ranges throughout the Taebaek and Sobaek mountains, spanning
14 elevations from sea level to 2,000 m. They are diurnal browsers that feed on grasses, shoots,
15 leaves, nuts, and fruits and occur in small groups of 4-12 individuals with a group home
16 range of roughly 40 hectares ¹. Previous studies of goral populations in Korea have shown a
17 preference for steep slopes as a defense against predators, though they were observed to
18 descend to lower habitats during periods of heavy winter snow when vegetation is
19 unavailable at higher elevations ⁶⁻⁸. Various serious diseases have been described in both wild
20 and captive populations of long-tailed gorals ⁹, and deaths due to taeniasis parasites
21 (tapeworms), pneumonia, gastroenteritis, and hepatitis have been reported in captive
22 individuals ¹⁰. Surveys of a Russian population of long-tailed gorals identified more than 30
23 species of helminth parasites ¹¹⁻¹³, which result in massive infections that can cause the death
24 of the host, as occurred in 1,937 when all ten long-tailed gorals sent from the Lazovsky

1 Nature Reserve to the Moscow Zoo died within eight months of their arrival ¹¹. In addition,
2 extremely high gastrointestinal helminth parasite loads were also reported in closely related *N.*
3 *goral* captive individuals, and mixed endoparasite infections were seen in the majority of the
4 wild animals examined ¹⁴.

5 The long-tailed goral was internationally classified as Vulnerable in 2008 due to
6 over-exploitation by hunters and habitat destruction and degradation ^{1,15}, and is protected in
7 the Republic of Korea as both an endangered wild species and as a designated national
8 monument through the Cultural Heritage Administration ¹⁵. Despite its protected status, the
9 long-tailed goral continues to be poached for meat and traditional medicine ¹. Its range has
10 significantly decreased over the last century, with the majority of individuals now occupying
11 the Russian and Korean coast of the Sea of Japan ⁴. Deforestation from timber operations,
12 along with increased human encroachment and agricultural expansion, have fragmented
13 populations and heightened concerns of inbreeding depression and reduced effective
14 population size ¹⁶⁻¹⁸. Population surveys in 2002 found a significant decline in the South
15 Korean population with a total of 690-784 extant individuals ¹⁹. Microsatellite studies of
16 South Korean long-tailed goral populations have shown populations along the Taebaek
17 Mountains to be genetically distinct with lower genetic diversity in wild individuals from the
18 southern group relative to the northern group ²⁰. Furthermore, Russian surveys in the early
19 20th century estimated roughly 2,000 individuals, which have since been reduced to about 900
20 individuals ^{4,21}. The species is believed to be almost extinct in China ^{22,23}. Across its range, it
21 is now missing from many areas of its original occurrence; and only six subpopulations
22 currently have more than 100 individuals, while many of the other localities only contain a
23 few animals ⁴. Taken together, the recent range loss and population declines raise heightened
24 concerns about the future of this species.

1 To better understand genome evolution and the demographic history of the long-
2 tailed goral, and to provide a foundation for future conservation studies, we generated and
3 analyzed the whole genome sequence (WGS) of a wild individual from the northern Taebaek
4 Mountains in the Republic of Korea. We compared the genetic diversity of the long-tailed
5 goral to other close species of Caprini and then performed demographic modeling to infer
6 historical population structure. We also compared the long-tailed goral genome to the
7 reference genomes of seven other bovids (goat, sheep, mouflon, Tibetan antelope, bison,
8 cattle, and yak) and three distantly related mammals (human, pig, and horse) to characterize
9 environmental adaptations present in the long-tailed goral and the caprine common ancestor.

10 **Results and Discussion**

11 **Long-tailed goral whole genome sequencing**

12 A blood sample of a wild long-tailed goral from the northern range of the Taebaek Mountains
13 in the Republic of Korea (Fig. 1) was collected through the Association of Korean Goral
14 Conservation. Genomic DNA was sequenced using the Illumina HiSeq2000 platform, from
15 which we obtained 430 million (M) paired reads (Additional file 1: Table S1). The reads were
16 then mapped to the closely related sheep (*Ovis aries*) reference genome. In total, 94% of the
17 WGS reads were aligned to the sheep genome and yielded a 27x depth of coverage. Using the
18 aligned reads, a variant analysis identified 55 M single nucleotide variants (SNVs) between
19 the long-tailed goral and the sheep. To quantify the genomic effects of possible inbreeding
20 depression resulting from habitat fragmentation and isolation in this species, we examined the
21 proportion of heterozygous SNVs. In total, the variant composition included 52,621,471
22 (94.71%) homozygous and 2,936,977 (5.29%) heterozygous SNVs. We also analyzed the
23 number of homozygous and heterozygous SNVs for each chromosome (Additional file 1:
24 Table S2). The average homozygous and heterozygous SNVs for each chromosome were

21.12/Kb (20.12-22.34) and 1.29/Kb (0.93-1.86), respectively. Higher number of SNVs was observed in autosomes than in the sex chromosome (Chr. X). We also analyzed the homozygous SNV hotspots with 0.5M windows. Hot spots were defined by Z-score > 2 (Additional file 2: Figure S1, pink) and Z-score > 3 (red). Most of the hotspots were distributed on the chromosome ends. The heterozygous SNVs were replaced by “N” and the homozygous SNVs were used to construct a 2.61 Gb consensus sequence, which we used to predict 18,690 genes with 98.6% of the coding sequence (CDS) regions having a read depth greater than or equal to five.

9 **Evolutionary relationships of the long-tailed goral inferred by phylogenomic analysis**

For comparative evolutionary analyses, ten mammalian reference genomes were downloaded and analyzed to infer the evolutionary history of the long-tailed goral. Included were four additional members of the Caprini tribe in the Antilopinae subfamily (goat, sheep, mouflon, and Tibetan antelope), three members of another bovid subfamily, Bovinae, in the tribe Bovini (bison, cow, and yak), and three distantly related mammals (horse, pig, and human). To reveal evolutionary relationships among the orthologous genes, we constructed orthologous gene clusters with these ten reference mammalian genomes. A total of 28,661 ortholog clusters were identified, with 7,533 (26.3%) clusters shared by all ten of the reference mammals. Of these, 5,279 (70.1%) were also single-copy long-tailed goral orthologs. Divergence times were calculated using 656,303 fourfold degenerate sites in the single-copy orthologous gene families (Fig. 2). Phylogenomic analyses showed a deep divergence of the Bovidae from the Suidae roughly 57.9 million years ago (MYA), consistent with previous findings²⁴. Divergence of the Antilopinae and Bovinae subfamilies within Bovidae was dated to 14.1 MYA and is close to previous estimates (17.6-15.1 MYA)²⁵. The Tibetan antelope is the sole species in the genus *Pantholops* and its classification has

historically been disputed²⁶. Recent genetic work has reassigned it from Antilopini, where it was traditionally placed, to the Caprini tribe where it is sister to all other caprine species^{3,5,27,28}. While our phylogeny did not include genomes from other Antilopinae tribes, we estimated that the Caprini (goat, sheep, mouflon, and Tibetan antelope) shared a most recent common ancestor (MRCA) 6.38 MYA. That date is notable because it is much younger than the proposed 14 MYA split between the Caprini and Antilopini²⁵, lending genomic support to the previous genetic-based assignment of *Pantholops* to Caprini.

Positively selected genes in the long-tailed goral and caprines

To gain insight into potential environmental adaptations that evolved in the long-tailed goral and in the Caprini common ancestor, we analyzed the 5,279 single-copy orthologous genes across our eleven species to infer genes that experienced positive selection in these lineages. In the long-tailed goral, we identified 218 and 153 PSGs by applying the branch and branch-site models, respectively (Additional file 1: Table S3 and S4). Gene Ontology (GO) enrichment analyses of the long-tailed goral PSGs against the Caprini species *Ovis aries* (sheep) as a background revealed nine biological processes related to immune function (Additional file 1: Table S5 and S6). In addition, six KEGG pathways related to immune function were found to be statistically enriched in the long-tailed goral (Additional file 1: Table S7 and S8). Previous analyses of evolutionary rate differences across mammalian genomes have reported rapid evolution in immune function, particularly in genes related to the innate immune response²⁹.

A critical component of the immune response, Type 2 T helper cells (T_H2) are believed to have evolved to combat extracellular parasites and are the primary immune response to helminthiasis³⁰⁻³³, which has been shown to cause mortality among captive gorals¹⁴. Among the T_H2 PSGs identified in the long-tailed goral, *interleukin 5 (IL5)* is a T_H2

1 cytokine known to play critical roles in the immune response to extracellular parasitic
2 infections. Numerous studies of IL5 mouse knockouts have shown this cytokine to be critical
3 to both primary and secondary immune responses to helminthiasis from a variety of
4 nematode species³⁴⁻³⁶. Also of interest, the PSG *C-C motif chemokine ligand 25 (CCL25)*
5 has been previously shown to play a critical role in the immune response to *Trichuris*
6 (whipworm) by targeting lymphocytes to the small intestine³⁷. Infections by both
7 intracellular and extracellular pathogens can also cause tissue damage; either from the
8 antimicrobial Type 1 T helper (T_H1) response which results in collateral tissue damage, or by
9 physical trauma caused by helminths entering, exiting, and migrating through the host. As a
10 result, wound healing and angiogenesis are critical components of the host response to
11 infection³⁸. In addition to the increased wound healing mediated by over-expression of *IL5*³⁹,
12 several other long-tailed goral PSGs are involved in these pathways and processes. Among
13 them is the *colony-stimulating factor 2 (CSF2)*, which promotes the maturation of
14 macrophages in lung and gut epithelium and is essential for the recruitment of innate immune
15 cells to damaged tissues⁴⁰. We also observed seven other long-tailed goral PSGs in additional
16 pathways involved in wound healing, including: Notch (*ITCRA*), WNT (*CSNK2B*, *FOSL1*,
17 and *RUVBL1*), and NF-κB signaling (*BCL10*, *CSNK2B*, and *TNFAIP3*). Finally, we identified
18 numerous PSGs involved in the T_H1 response to viral and bacterial infections (Additional file
19 1: Table S6-S9). In particular, interleukin 12 (IL-12) is a cytokine that induces T_H1
20 differentiation, and the ‘positive regulation of interleukin-12 production’ GO biological
21 process (GO:0032735) was identified as statistically enriched (p-value: 0.01). Furthermore,
22 the bactericidal/permeability increasing fold containing family A member 1 (BPIFA1) was
23 positively selected, which is known to be expressed in the upper airways and plays a critical
24 role in the inhibition of biofilm formation by pathogenic gram-negative bacteria such as
25 *Pseudomonas aeruginosa* and *Klebsiella pneumoniae*⁴¹.

To investigate adaptive evolution in the caprine common ancestor, we also analyzed the amino acid substitutions that were exclusively present in all members of the Caprini tribe (goat, sheep, mouflon, Tibetan antelope, long-tailed goral). From these, a total of 235 and 154 PSGs were detected using branch and branch-site models, respectively (Additional file 1: Table S9 and S10). Similar to the long-tailed goral, GO enrichment analyses of the Caprini PSGs showed numerous biological processes, molecular functions, and KEGG pathways related to both adaptive and innate immune function (Additional file 1: Table S11-S14). Of the PSGs in the innate response, the cytokines *interleukin 4 (IL-4)* and *interleukin 33 (IL-33)* are essential components of the T_H2 immune response to helminths⁴²⁻⁴⁸. Additional positively selected genes consist of other cytokines and chemokines, including *TNF*, *CXCL13*, and *CD34*, which play critical roles in the recruitment of immune cells to injured tissues⁴⁹. Furthermore, IL-17F recruits dermal innate immune cells during the inflammation and cell proliferation phases of wound response⁵⁰; and fibroblast growth factor1 (FGF1) is known to stimulate angiogenesis and tissue repair and to induce the proliferation of keratinocytes⁵¹, which are required for the final stage of wound healing.

Non-synonymous amino acid substitutions possibly confer selective advantages

The codon models for detecting positive selection may fail to detect genes that are positively selected if there are too few amino acid substitutions across a given gene⁵². To overcome this limitation, we examined function-altering amino acid changes in single-copy orthologues to provide a more robust examination of the environmental adaptations of the long-tailed goral and Caprini. By aligning the 5,279 single copy orthologous genes from the eleven mammalian genomes, we investigated the amino acid substitutions among these species. We identified 3,802 unique amino acid substitutions (uAAS) in the long-tailed goral from 2,061 proteins (Additional file 1: Table S15). We then estimated the functional effects of these

1 unique long-tailed goral substitutions using the programs PolyPhen2⁵³ and PROVEAN⁵⁴. In
2 total, PolyPhen2 predicted 1,582 (41.61%) amino acids as “possibly or probably damaging”
3 and PROVEAN predicted 1,583 (41.63%) amino acids as “deleterious”. Of these, 1,341
4 (65.07%) proteins contain function-altering uAAS specific to the long-tailed goral as
5 predicted by at least one program, and 967 long-tailed goral uAAS in 701 genes were
6 predicted as function-altering substitutions by both programs. Similarly, we identified 6,543
7 uAAS unique to the Caprini lineage in a total of 2,744 proteins (Additional file 1: Table S16).
8 Among them, 1,859 (28.41%) and 2,131 (32.08%) uAAS were predicted as function-altering
9 by the PolyPhen2 or PROVEAN, respectively, in a total of 1,594 (58.09%) proteins with
10 1,004 uAAS in 674 proteins predicted by both programs.

11 To investigate the biological significance of these substitutions, we identified the
12 KEGG pathways associated with the function-altering uAAS. In total, 187 (14.04%) proteins
13 with unique long-tailed goral substitutions and 447 (51.68%) proteins with Caprini specific
14 uAAS were mapped to 103 and 275 KEGG pathways, respectively (Additional file 1: Table
15 S17). KEGG enrichment analyses of these two gene sets revealed a number of pathways
16 related to immune function, RNA transport and biosynthesis, DNA repair, and cellular
17 metabolism (Additional file 1: Table S18 and S19). Of particular interest in the Caprini are
18 ‘Cytokine-cytokine receptor interaction’, ‘NF-κB signaling pathway’, ‘Apoptosis’,
19 ‘Sphingolipid signaling pathway’, and ‘Complement and coagulation cascades’ (Table 1,
20 Additional file 2: Figure S1). The pathway involved in ‘Cytokine-cytokine receptor
21 interaction’ included 38 proteins containing unique Caprini uAAS. Of these, six additional
22 tumor necrosis factor (TNF) family members were included over the three identified by the
23 PSG KEGG enrichment (TNF, TNFRSF8, TNFRSF21). They were comprised of four TNF
24 ligand superfamily members (TNFLSF4, 9, 10, and 15) and three TNF receptors superfamily

members (TNFRSF1A, 9, 19). In addition, we identified five cytokines involved in hematopoiesis: the ciliary neurotrophic factor (CNTF), granulocyte colony-stimulating factor 3 (CSF3), interleukin 9 (IL9), and two interleukin receptors (IL23R, IL15RA); five chemokines: four C-C motif ligands (CCL4, CCL27, CCR1, CCR9) and one C-X-C motif ligand (CXCL9); and other minor cytokines: interferon kappa (IFNK), interleukin 26 (IL26), and interleukin 1 receptor type 1 (IL1R1). Also included as an enriched KEGG pathway related to immune function was ‘Complement and coagulation cascades’ (Additional file 1: Table S19). Our KEGG enrichment analyses of Caprini PSGs previously identified this pathway as enriched with four genes (*KNG1*, *MBL2*, *TFPI*, *CFI*) (Additional file 1: Table S13). Added to this, our functional analysis yielded an additional four proteins for complement cascade (C1QA, C1QC, C5AR1, and C8B) and seven proteins related to blood coagulation (SERPINC1, F5, F2RL2, PLAUR, PLAU, PLAT, and FGG), which may be an adaptation to steep and rugged habitats.

Structural analysis of myostatin protein provides evidence for increased muscle development in Caprini

We observed nine Caprini-specific uAAS in MSTN, encoding the growth differentiation factor myostatin (Table 2). This protein is known to negatively regulate skeletal muscle mass by binding to the activin receptor type IIB (ACVR2B)^{55,56}, and humans and other mammals with loss of function mutations that block the activity of myostatin have significantly more muscle mass than those with functional proteins⁵⁷⁻⁶³. To assess the impact of the Caprini uAAS on the function of this protein, we constructed a structural model of myostatin interacting with the receptor ACVR2B (Fig 3) and estimated the cumulative stability changes of the Caprini myostatin uAAS by calculating the change in binding free energy ($\Delta\Delta G_{Bind}$) (Table 2). Among nine function-altering uAAS identified in Caprini, three amino acids

1 physically interact with ACVR2B. Our simulations showed that the total binding energy was
2 slightly increased by substituting these three Caprini uAAS. The resulting decrease in binding
3 affinity between myostatin and ACVR2B in the Caprini may contribute to the increased
4 muscle mass that has allowed this group to successfully inhabit steep, mountainous terrains;
5 and warrants future experimental testing.

6 **Genetic diversity and population history**

7 Deeply sequenced genomes allow for the identification of heterozygous positions and an
8 estimation of the historical population structure ⁶⁴. The nucleotide diversity for Caprini and
9 Bovidae species was calculated by dividing the number of heterozygous sites by the genome
10 size (Additional file 1: Table S20). The nucleotide diversity of the long-tailed goral was
11 significantly lower (0.00113) than the mouflon (0.00379) and slightly higher than the
12 formerly endangered Tibetan antelope (0.00079), which underwent a severe population
13 bottleneck in the 1980's ⁶⁵. Interestingly, the nucleotide diversity of the long-tailed goral was
14 also lower than the domestic sheep (0.00298) and goat (0.00189), which should be depressed
15 from inbreeding associated with artificial selection (Additional file 1: Table S20). Effects,
16 both environmental bottleneck and artificial selection, enable to decreased nucleotide
17 diversities shown in Tibetan antelope and sheep and goat. A long-tailed goral, we applied in
18 this study, lived in wild. Therefore, lower nucleotide diversity of the long-tailed goral was
19 signature of inbreeding associated artificial selection. To further compare genetic diversity
20 and infer the demographic histories of our five Caprini species, we used the Pairwise
21 Sequentially Markovian Coalescent (PSMC) model ⁶⁶. The population size of all of the
22 Caprini species except for the long-tailed goral fluctuated over the past million years, while
23 the goral showed an overall decline (Fig 4). Notably, the sheep and mouflon maintained a
24 similar pattern until about 100,000 years ago, and the effective population sizes of these

species both peaked approximately 1.0 MYA. The two domesticated Caprini (sheep and goat) showed secondary expansions at 100 KYA and 70 KYA, respectively, which were tightly linked with atmospheric surface air temperature (T_{suf}); while the effective population size for the long-tailed goral and Tibetan antelope did not significantly increase around these times. In contrast, the effective population size of the long-tailed goral slightly increased about 30 KYA and the Tibetan antelope increased substantially, followed by declines in both species; though these date estimates occur at the edge of the accuracy range for this test. In conjunction with the observed population decline in this species, these findings suggest a possible need to review long-tailed goral's IUCN conservation status and stress the importance of ongoing conservation and monitoring efforts.

Conclusions

We produced a whole genome sequence assembly of the endangered long-tailed goral by mapping 87 Gb of next-generation sequencing data to the closely related sheep genome. Using these data and ten publically available reference genomes, we produced a phylogeny dating the Antilopinae/Bovinae divergence to 14.1 MYA. Consistent with previous observations of serious infectious diseases and high parasite loads in long-tailed goral populations, evolutionary analyses of the long-tailed goral genome and the Caprini lineage identified positively selected genes in enriched GO and KEGG pathways related to immune function and blood coagulation; highlighting the importance of ongoing evolution in both the adaptive and innate immune responses in these groups. Together with previous mortality reports, these findings emphasize the need for additional studies to characterize the role of infectious disease and parasitism across extant populations of *N. caudatus*; which can be used to better inform management and conservation efforts. We hope that genome will stimulate comparative genomic studies aimed at investigating host-parasite co-evolutionary immune

1 responses, potentially leading to targeted treatment strategies. We also identified the muscle
2 development gene *MSTN* to be positively selected in the Caprini lineage, and discovered
3 three amino acid substitutions that decrease the binding affinity between myostatin and its
4 receptor ACVR2B. We speculate that these substitutions may have played a role in increasing
5 muscle mass and, by proxy, climbing ability of the caprine common ancestor and warrants
6 further functional investigation. Finally, we discovered lower levels of heterozygosity in the
7 long-tailed goral than in other Caprini species, as well as a historical trend of population
8 decline. These findings reinforce concerns over habitat fragmentation and genetic isolation in
9 this species, and suggest that the IUCN status of the long-tailed goral should be re-evaluated.
10 This foundational genome should stimulate future conservation genetic studies of this
11 enigmatic endangered species.

12 **Methods**

13 **Whole genome sequencing and alignment**

14 A blood sample from a long-tailed goral was acquired from the Association of Korean Goral
15 Conservation under the Cultural Heritage Administration 217 (Korea) Permit. Genomic DNA
16 was extracted using a QiAamp DNA Mini kit (Qiagen, CA, USA) and quantified using an
17 Infinite F200 Pro NanoQuant (TECAN, Männedorf, Germany). Libraries were prepared
18 using an Illumina TruSeq kit (Illumina, CA, USA) according to the manufacturer's
19 instructions with an average insert size of 388 bp. 100 bp paired-end sequencing was
20 completed using an Illumina HiSeq2000 with v. TrueSeq Kit v3 chemistry at the Korea
21 Research Institute of Bioscience and Biotechnology (KRIBB). The raw sequence data were
22 deposited in the Sequence Read Archive (SRP098771). The reads were aligned to the sheep
23 genome assembly Oar_v4.0 using BWA-MEM (ver. 0.7.10) with default options⁶⁷. The
24 rmdup command of SAMtools (ver. 0.1.19) was used to remove PCR duplicates⁶⁸, the GATK

(ver. 3.3) IndelRealigner algorithm was used to realign reads ⁶⁹, and SNVs were called using the mpileup command of the SAMtools with the default option followed by the view command with the -cg option. Finally, consensus sequences were generated using vcf2fq with -d 5 options in the SAMtools package.

Orthologous gene clustering and phylogenomic analysis

Reference genomes for the goat (*Capra hircus*, CHIR_1.0), sheep (*Ovis aries*, Oar_v4.0), mouflon (*Ovis aries musimon*, Oori1), Tibetan antelope (*Pantholops hodgsonii*, PHO1.0), bison (*Bison bison bison*, Bison_UMD1.0), cattle (*Bos taurus*, Bos_taurus_UMD_3.1.1), yak (*Bos mutus*, BosGru_v2.0), horse (*Equus caballus*, EquCab2.0), pig (*Sus scrofa*, Sscrofa10.2), and human (*Homo sapiens*, GRCh38.p7) were downloaded from NCBI. To compare the long-tailed goral genes with protein sequences from these reference genomes, we constructed orthologous gene clusters using OrthoMCL 2.0.9 ⁷⁰. To construct a phylogenetic tree, we used the 656,303 fourfold degenerate sites in the single-copy orthologous genes. A maximum likelihood phylogeny was then generated using RAxML ⁷¹ with the GTRGAMMA evolutionary model and 100 bootstrap replicates. And then, the divergence time was calculated using MEGA7 program ⁷² with maximum likelihood statistics and a Jukes-Cantor evolutionary model, along with the previously determined phylogenetic tree topology. The calibration times of horse-pig divergence (78.4 MYA) and human-horse divergence (97.5 MYA) were taken from the TimeTree database (<http://www.timetree.org>) ⁷³.

Identification of positively selected genes (PSGs)

The PRANK program was used to align the orthologous genes ⁷⁴. The codon-based alignment was performed with this program, and the region containing the heterozygous SNVs was replaced with 'NNN'. And the branch and branch-site models were used to identify PSGs. The

1 CODEML program from the PAML 4.5 package was used to estimate the rates of
 2 synonymous (d_S) and nonsynonymous substitutions (d_N), and the general selective pressures
 3 acting on all species were estimated by applying the d_N/d_S ratio (ω) for the branch model ⁷⁵.
 4 PSGs were detected using the branch-site model using the d_N/d_S ratio along each branch and
 5 applying a conservative 10% false discovery rate (FDR) for the likelihood ratio tests (LRTs)
 6 ⁷⁶. To analyze the long-tailed goral PSGs, we set the genes of long-tailed goral as the
 7 foreground branch and those genes in other species were set as background. The foreground
 8 branch for the caprines included all five of the Caprini species (long-tailed goral, goat, sheep,
 9 mouflon, Tibetan antelope), with the six remaining species set as background. The DAVID
 10 bioinformatics resources tool was used to identify GO and KEGG pathway enrichment of the
 11 PSGs and uAAS ⁷⁷.

12 **Functional effects of amino acid changes**

13 uAAS were defined as amino acid substitutions that occurred in the target species, but no
 14 other species, and no gaps were allowed at the sites. Functional effects of the unique amino
 15 acid substitutions were predicted using PolyPhen2 ⁵³ and PROVEAN v1.1 ⁵⁴ programs with
 16 the default cutoff values, and the specific amino acids of the target species were substituted in
 17 the corresponding position of the human protein. KEGG pathways analyses were carried out
 18 using KAAS (KEGG Automatic Annotation Server) ⁷⁸. As an additional functional analysis,
 19 we performed computer modeling of the myostatin protein structure using SWISS-MODEL ⁷⁹,
 20 and a 3D structure of the ACVR2B protein was obtained from the PDB database (PDB ID:
 21 2QLU) ⁸⁰. The ZDOCK program ⁸¹ was used to construct the ACVR2B and myostatin
 22 complex protein structures, and the BeAtMuSiC program ⁸² was used to evaluate the change
 23 in binding affinity between two proteins.

1 **Demographic history analysis**

2 To infer the population history of the long-tailed goral, we applied the Pairwise Sequentially
 3 Markovian Coalescent (PSMC) model for scaffolds ≥ 50 Kb in length. We performed 100
 4 rounds of bootstrapping and used 1.0×10^{-8} substitutions per site per generation as a mutation
 5 rate, and 4 years as a generation time as previously reported ⁶⁶. Nucleotide diversity in the
 6 Caprini was calculated by dividing the number of heterozygous SNVs by the total sequence
 7 length of reference genome.

8 **Availability of supporting data**

9 Whole-genome sequence data was deposited in the SRA database at NCBI with Biosample
 10 accession numbers SAMN06219541. The whole-genome sequence data can also be accessed
 11 through BioProject accession number PRJNA361026.

12 **Competing Interests**

13 The authors declare that they have no competing interests.

14 **Authors' contributions**

15 Conceived and designed the experiments: JB and WKP. Analyzed the data: OC, JK, and JAW.
 16 Study design, and sample preparation: DSK, JB, and WKP. Data interpretation: OC, JK, JAW,
 17 YSC, and JB. Wrote and edited the manuscript: OC, JAW, JK, JHJ, HL, SJ, YSC, SK, HL, SL,
 18 JSE, JAC, JA and JB.

19

1 **Figure Legends**

2 **Figure 1. Distribution of the long-tailed goral across the Korea peninsula.** The circles on
3 the map represent population densities of the long-tailed goral which were surveyed in 2002
4 by the Ministry of the Environment, Korea. The black lines represent the Taback and Soback
5 mountains which are representative habitats of the long-tailed goral. The black triangle is the
6 location of the Association of Korea Goral Conservation, where we sampled.

7 **Figure 2. Phylogenetic relationships and divergence times.** The phylogenetic topology was
8 constructed with the 5,279 single copy orthologues defined in this study using RAxML
9 (bootstrap support listed at nodes). The divergence times were estimated using four-fold
10 degenerate sites of one-to-one ortholog families using the program MEGA7.

11 **Figure 3. Caprini-specific amino acid change in the three-dimensional structure of**
12 **ACVR2B and myostatin.** The ACVR2B and myostatin are colored by green and purple,
13 respectively. The five Caprini-specific amino acids that are located on non-interface region
14 are shown in purple, and the three Caprini-specific amino acids that locate on interface region
15 are in blue.

16 **Figure 4. Demographic history of the Caprini.** Demographic history of Caprini was
17 analyzed by PSMC. g is generation time (years); μ , mutation rate per site per generation time;
18 T_{suf} , atmospheric surface air temperature; RSL, relative sea level; 10 m.s.l.e., 10 m sea level
19 equivalent.

Tables

Table 1. Immune response genes with function-altering uAASs

Pathway terms	Caprini-uAAS	Long-tailed goral-uAAS	Shared by two groups
Complement and coagulation cascades	14 (VTN, CFI, F5, SERPINC1, PLAUR, KNG1, C1QC, PLAU, PLAT, C1QA, TFPI, F2RL2, FGG, MBL2)	3 (VSIG4, C5AR1, C8B)	-
Chemokine signaling pathway	7 (SHC4, CCL4, CXCL9, LOC105372280, CCR1, CCR9, CCL27)	3 (CCR8, NCF1, STAT2, CXCR6)	4 (CCL2, CX3CR1, CXCL16)
Hematopoietic cell lineage	7 (GP1BA, CSF3, CD4, IL4, CD2, CD8B, IL1R1)	2 (IL5, DNNT)	5 (IL7R, CD34, IL1R2, FCGR1A, MS4A1)
Toll-like receptor signaling pathway	6 (LBP, CCL4, CXCL9, CASP8, SPP1, LY96)	4 (IRF5, CD40, TLR4, CD86)	1 (IL12A)
Intestinal immune network for IgA production	4 (IL15RA, IL4, MAP3K14, CCR9)	5 (CD86, IL5, HLA-DMB, HLA-DOA, CD40)	1 (PIGR)
Natural killer cell mediated cytotoxicity	7 (SHC4, TNFSF10, KLRD1, BID, CD244, PRF1, LCP2)	2 (TYROBP, FAS)	1 (IFNGR1)
Leukocyte trans-endothelial migration	4 (PECAM1, CLDN20, CLDN25, F11R)	3 (NCF1, JAM2, CLDN24)	2 (CLDN8, NCF2)
Cytosolic DNA-sensing pathway	5 (ZBP1, POLR3D, IL33, CCL4, POLR3G)	-	3 (RIPK3, MAVS, DDX58)
Antigen processing and presentation	4 (CD4, CD8B, KLRD1, LGMN)	2 (HLA-DMB, HLA-DOA)	-
T cell receptor signaling pathway	5 (CD4, LCP2, IL4, CD8B, MAP3K14)	1 (IL5)	-
Fc gamma R-mediated phagocytosis	-	2 (PLPP3, NCF1)	2 (SPHK1, FCGR1A)
Fc epsilon RI signaling pathway	3 (LCP2, IL4, FCER1A)	1 (IL5)	-
NOD-like receptor signaling pathway	2 (TNFAIP3, CASP8)	1 (PSTPIP1)	1 (CCL2)
Platelet activation	3 (LCP2, FGG, GP1BA)	-	1 (MYLK2)
B cell receptor signaling pathway	2 (CD72, BLNK)	-	-
Toll and Imd signaling pathway	1 (CASP8)	-	-

1 **Table 2. Caprini-specific myostatin amino acids**

Position of human ^a	Other animals ^b	Caprini	$\Delta\Delta\text{GBind}^c$ (kcal/mol)	Interface	PolyPhen2 ^d	PROVEAN ^d
16	I	L	N/A		benign	neutral
48	T	N	N/A		benign	neutral
113	A	V	0.25		benign	deleterious
130	Q	E	1.17	O	benign	neutral
133	G	E	0.56		benign	deleterious
148	Y	H	0.92	O	benign	neutral
247	D	E	0.64	O	benign	neutral
333	R	K	0.04		benign	neutral
366	A	G	-0.34		benign	neutral

2 The position of the human myostatin protein sequence corresponding to caprini-specific amino acids^b Other animals
3 represent non-Caprini animals analyzed in this study, including yak, bison, cattle, pig, horse, human.

4 ^c The negative sign in $\Delta\Delta\text{GBind}$ corresponds to a mutation predicted to increase the binding affinity.

5 ^d Programs estimating the function-altering AASs

1

2 References

- 3 1 Shackleton, D. M. *Wild sheep and goats and their relatives : status survey and conservation*
4 *action plan for Caprinae.* (IUCN, 1997).
- 5 2 Wilson, D. E. & Reeder, D. M. Vol. 1 (JHU Press, 2005).
- 6 3 Groves, C. & Grubb, P. *Ungulate taxonomy.* (Baltimore: Jons Hopkins University Press,
7 2011).
- 8 4 Hrabina, P. A new insight into the taxonomy and zoogeography of recent species of goral
9 (Nemorhaedus, Bovidae, Ruminantia) Nový pohled na taxonomii a zoogeografii recentních
10 druhů goralů (Nemorhaedus, Bovidae, Ruminantia). (2015).
- 11 5 Ropiquet, A. & Hassanin, A. Molecular phylogeny of caprines (Bovidae, Antilopinae): the
12 question of their origin and diversification during the Miocene. *Journal of Zoological*
13 *Systematics and Evolutionary Research* **43**, 49-60, doi:10.1111/j.1439-0469.2004.00290.x
14 (2005).
- 15 6 Cho, C., Kim, K. & Kwon, G. Habitat altitude and home range of the endangered long-tailed
16 goral (Naemorhedus caudatus): seasonal and monthly home range and altitude change. .
17 *Mammalia* **80**, 481-489 (2016).
- 18 7 Mishra, C. & Johnsingh, A. J. T. On habitat selection by the goral Nemorhaedus goral
19 bedfordi (Bovidae, Artiodactyla). *J of Zoology* **240**, 573-580 (1996).
- 20 8 Cho, C.-U. *et al.* Habitat use of reintroduced Long-tailed Gorals (Naemorhedus caudatus) in
21 Woraksan (Mt.) National Park in Korea. *Korean Journal of Environment and Ecology* **29**, 184-
22 191 (2015).
- 23 9 Voloshina, I. & Solomkina, N. in *Amursky Goral.* (ed Al Myslenkov) 105-118 (1992).
- 24 10 Mead, J. I. Naemorhedus goral. *Mammalia* **335**, 1-5 (1989).
- 25 11 SHULTS, R. & KADENATSII, A. Helminths of Nemorhaedus goral in the Far East. *Trudy*
26 *Gel'mintologicheskoi Laboratorii. Akademiya Nauk SSSR* **3**, 152-160 (1950).
- 27 12 Voloshina, I. & Khrustalev, A. Endoparasites: species composition, abundance and
28 pathogenicity. *Amur Goral (Collected Proceedings). Publishing House of Central research*
29 *laboratory of wildlife and reserves, Moscow*, 119-123 (1992).
- 30 13 Khrustalev, A. & Savinetsky, A. Occurrence of helminths eggs in quaternary deposits of
31 animals feces. *Parazitologiya* **26**, 122-129 (1992).
- 32 14 Rana, M. A., Ahmad, I., Jabeen, F., Naureen, A. & Shabnam, M. Comparative study of endo-
33 parasites from fecal samples of sambar (Rusa unicolor) and goral (Naemorhedus goral) in
34 captivity. *J. bio. & Env. Sci.* **6**, 399-408 (2015).
- 35 15 Oh, H. S. in *Korean Red List of threatened species* Vol. 2 (ed Sang-Bae Kim) 47 (National
36 Institute of biological Resources, 2014).

- 1 16 Roy, P. *et al.* Habitat suitability analysis of *Nemorhaedus goral*—a remote sensing and
2 geographic information system approach. *Curr. Sci.*, 685-691 (1995).
- 3 17 Lande, R. Risks of Population Extinction from Demographic and Environmental Stochasticity
4 and Random Catastrophes. *Am. Midl. Nat.* **142**, 911-927 (1993).
- 5 18 Birky, C. W. J. in *Evolution at the Molecular Level* (eds Robert K. Selander, SAndrew G. Clark,
6 & Tomas S. Whittam) (Sinauer Associates Inc., 1991).
- 7 19 Yang, B. Systematic, ecology and current population status of the goral, *Naemorhedus*
8 *caudatus*. *Korea. PhD thesis, Cheongju, Chungbuk National University, Republic of Korea*
9 (2002).
- 10 20 Choi, S. K. *et al.* Genetic diversity and population structure of the long-tailed goral,
11 *Naemorhedus caudatus*, in South Korea. *Genes & genetic systems* **90**, 31-41,
12 doi:10.1266/ggs.90.31 (2015).
- 13 21 Voloshina, I. & Myslenkov, A. Gis-based range and long-term monitoring of Amur goral in the
14 Russian Far East. *Galemys* **22**, 359-374 (2010).
- 15 22 Zhang, H. *et al.* The complete mitochondrial genome of Chinese tibetan goat (*Capra hircus*).
16 *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis* **27**, 1161-1162,
17 doi:10.3109/19401736.2014.936418 (2016).
- 18 23 Wu, J., Zhang, Y., Fu, D. & Lan, W. Comparison of home range of siberian musk deer and
19 goral in winter. *J. Northeast Forestry University* **36**, 58-60 (2008).
- 20 24 dos Reis, M. *et al.* Phylogenomic datasets provide both precision and accuracy in estimating
21 the timescale of placental mammal phylogeny. *Proceedings. Biological sciences* **279**, 3491-
22 3500, doi:10.1098/rspb.2012.0683 (2012).
- 23 25 Bibi, F. A multi-calibrated mitochondrial phylogeny of extant Bovidae (Artiodactyla,
24 Ruminantia) and the importance of the fossil record to systematics. *BMC evolutionary biology*
25 **13**, 166, doi:10.1186/1471-2148-13-166 (2013).
- 26 26 Lei, R., Jian, Z., Hu, Z. & Yang, W. Phylogenetic relationships of Chinese antelopes (sub
27 family Antilopinae) based on mitochondrial Ribosomal RNA gene sequences. *Journal of*
28 *Zoology* **261**, 227-237 (2003).
- 29 27 Gentry, A. H. Tropical Forest Biodiversity: Distributional Patterns and Their Conservational
30 Significance. *Oikos* **63**, 19-28, doi:10.2307/3545512 (1992).
- 31 28 Gatesy, J., Amato, G., Vrba, E., Schaller, G. & DeSalle, R. A cladistic analysis of
32 mitochondrial ribosomal DNA from the Bovidae. *Molecular phylogenetics and evolution* **7**,
33 303-319, doi:10.1006/mpev.1997.0402 (1997).
- 34 29 Singh, R. S., Xu, J. & Kulathinal, R. J. *Rapidly evolving genes and genetic systems*. (Oxford
35 University Press, 2012).
- 36 30 Maizels, R. M. Parasite immunomodulation and polymorphisms of the immune system.
37 *Journal of Biology* **8**, 62 (2009).
- 38 31 MacDonald, A. S., Araujo, M. I. & Pearce, E. J. Immunology of parasitic helminth infections.

- 1 *Infection and Immunity***70**, 427-433 (2002).
- 2 32 Knott, M. L. *et al.* Impaired resistance in early secondary *Nippostrongylus brasiliensis*
3 infections in mice with defective eosinophilopoiesis. *International journal for parasitology* **37**,
4 1367-1378 (2007).
- 5 33 Shim, D. S. *et al.* Protection against *Nippostrongylus brasiliensis* infection in mice is
6 independent of GM-CSF. *Immunology and cell biology***90**, 553 (2012).
- 7 34 Sasaki, O., Sugaya, H., Ishida, K. & Yoshimura, K. Ablation of eosinophils with anti-IL-5
8 antibody enhances the survival of intracranial worms of *Angiostrongylus cantonensis* in the
9 mouse. *Parasite immunology* **15**, 349-354 (1993).
- 10 35 Ovington, K., McKie, K., Matthaei, K., Young, I. & Behm, C. Regulation of primary
11 *Strongyloides ratti* infections in mice: a role for interleukin-5. *Immunology***95**, 488 (1998).
- 12 36 Daly, C. M., Mayrhofer, G. & Dent, L. A. Trapping and Immobilization of *Nippostrongylus*
13 *brasiliensis* Larvae at the Site of Inoculation in Primary Infections of Interleukin-5 Transgenic
14 Mice. *Infection and immunity***67**, 5315-5323 (1999).
- 15 37 Mullaly, S. C., Burrows, K., Antignano, F. & Zaph, C. Assessing the role of CD103 in immunity
16 to an intestinal helminth parasite. *PloS one* **6**, e19580 (2011).
- 17 38 Gause, W. C., Wynn, T. A. & Allen, J. E. Type 2 immunity and wound healing: evolutionary
18 refinement of adaptive immunity by helminths. *Nature reviews. Immunology* **13**, 607 (2013).
- 19 39 Leitch, V. D., Strudwick, X. L., Matthaei, K. I., Dent, L. A. & Cowin, A. J. IL-5-overexpressing
20 mice exhibit eosinophilia and altered wound healing through mechanisms involving prolonged
21 inflammation. *Immunology and cell biology***87**, 131-140, doi:10.1038/icb.2008.72 (2009).
- 22 40 Lavin, Y., Mortha, A., Rahman, A. & Merad, M. Regulation of macrophage development and
23 function in peripheral tissues. *Nature reviews. Immunology* **15**, 731-744, doi:10.1038/nri3920
24 (2015).
- 25 41 Liu, Y. *et al.* SPLUNC1/BPIFA1 contributes to pulmonary host defense against *Klebsiella*
26 *pneumoniae* respiratory infection. *The American journal of pathology* **182**, 1519-1531,
27 doi:10.1016/j.ajpath.2013.01.050 (2013).
- 28 42 Koyanagi, M. *et al.* Diversifying selection and functional analysis of interleukin-4 suggests
29 antagonism-driven evolution at receptor-binding interfaces. *BMC evolutionary biology* **10**, 223
30 (2010).
- 31 43 Willebrand, R. & Voehringer, D. IL-33-induced cytokine secretion and survival of mouse
32 eosinophils is promoted by autocrine GM-CSF. *PloS one* **11**, e0163751 (2016).
- 33 44 Steinfelder, S., Rausch, S., Michael, D., Kuhl, A. A. & Hartmann, S. Intestinal helminth
34 infection induces highly functional resident memory CD4+ T cells in mice. *European journal of*
35 *immunology***47**, 353-363 (2017).
- 36 45 Shimokawa, C. *et al.* Mast Cells Are Crucial for Induction of Group 2 Innate Lymphoid Cells
37 and Clearance of Helminth Infections. *Immunity***46**, 863-874. e864 (2017).
- 38 46 Coakley, G. *et al.* Extracellular Vesicles from a Helminth Parasite Suppress Macrophage

1 Activation and Constitute an Effective Vaccine for Protective Immunity. *Cell Reports* **19**, 1545-
2 1557 (2017).

3 47 Thomas, G. D. *et al.* The biology of nematode-and IL4R α -dependent murine macrophage
4 polarization in vivo as defined by RNA-Seq and targeted lipidomics. *Blood* **120**, e93-e104
5 (2012).

6 48 Yasuda, K. *et al.* Contribution of IL-33-activated type II innate lymphoid cells to pulmonary
7 eosinophilia in intestinal nematode-infected mice. *Proceedings of the National Academy of*
8 *Sciences* **109**, 3451-3456 (2012).

9 49 Kryczka, J. & Boncela, J. Leukocytes: the double-edged sword in fibrosis. *Mediators of*
10 *inflammation* **2015** (2015).

11 50 Li, Z. *et al.* Epidermal Notch1 recruits RORgamma(+) group 3 innate lymphoid cells to
12 orchestrate normal skin repair. *Nature communications* **7**, 11394, doi:10.1038/ncomms11394
13 (2016).

14 51 Werner, S. & Grose, R. Regulation of wound healing by growth factors and cytokines.
15 *Physiological reviews* **83**, 835-870, doi:10.1152/physrev.00031.2002 (2003).

16 52 Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation. *Trends in*
17 *ecology & evolution* **15**, 496-503 (2000).

18 53 Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense
19 mutations using PolyPhen-2. *Current protocols in human genetics* **Chapter 7**, Unit7 20,
20 doi:10.1002/0471142905.hg0720s76 (2013).

21 54 Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino
22 acid substitutions and indels. *Bioinformatics* **31**, 2745-2747,
23 doi:10.1093/bioinformatics/btv195 (2015).

24 55 Hill, J. J. *et al.* The myostatin propeptide and the follistatin-related gene are inhibitory binding
25 proteins of myostatin in normal serum. *The Journal of biological chemistry* **277**, 40735-40741,
26 doi:10.1074/jbc.M206379200 (2002).

27 56 Sako, D. *et al.* Characterization of the ligand binding functionality of the extracellular domain
28 of activin receptor type IIb. *The Journal of biological chemistry* **285**, 21037-21048,
29 doi:10.1074/jbc.M110.114959 (2010).

30 57 Lee, S. J. Regulation of muscle mass by myostatin. *Annual review of cell and developmental*
31 *biology* **20**, 61-86, doi:10.1146/annurev.cellbio.20.012103.135836 (2004).

32 58 Schuelke, M. *et al.* Myostatin mutation associated with gross muscle hypertrophy in a child.
33 *The New England journal of medicine* **350**, 2682-2688, doi:10.1056/NEJMoa040933 (2004).

34 59 Elkasrawy, M. N. & Hamrick, M. W. Myostatin (GDF-8) as a key factor linking muscle mass
35 and bone structure. *Journal of musculoskeletal & neuronal interactions* **10**, 56-63 (2010).

36 60 Guo, R. *et al.* Generation and evaluation of myostatin knock-out rabbits and goats using
37 CRISPR/Cas9 system. *Scientific reports* **6** (2016).

38 61 McPherron, A. C., Lawler, A. M. & Lee, S.-J. Regulation of skeletal muscle mass in mice by a

1 new TGF- β superfamily member. *Nature* **387**, 83-90 (1997).

2 62 Kambadur, R., Sharma, M., Smith, T. P. & Bass, J. J. Mutations in myostatin (GDF8) in
3 double-muscled Belgian Blue and Piedmontese cattle. *Genome research* **7**, 910-915 (1997).

4 63 Grobet, L. *et al.* A deletion in the bovine myostatin gene causes the double-muscled
5 phenotype in cattle. *Nature genetics* **17**, 71-74 (1997).

6 64 Miller, W. *et al.* Polar and brown bear genomes reveal ancient admixture and demographic
7 footprints of past climate change. *Proceedings of the National Academy of Sciences* **109**,
8 E2382-E2390 (2012).

9 65 Schaller, G. B. *Wildlife of the Tibetan steppe*. (University of Chicago Press, 1998).

10 66 Li, H. & Durbin, R. Inference of human population history from individual whole-genome
11 sequences. *Nature* **475**, 493-496, doi:10.1038/nature10231 (2011).

12 67 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
13 *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

14 68 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-
15 2079, doi:10.1093/bioinformatics/btp352 (2009).

16 69 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
17 next-generation DNA sequencing data. *Genome research* **20**, 1297-1303,
18 doi:10.1101/gr.107524.110 (2010).

19 70 Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for
20 eukaryotic genomes. *Genome research* **13**, 2178-2189, doi:10.1101/gr.1224503 (2003).

21 71 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
22 phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

23 72 Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis
24 Version 7.0 for Bigger Datasets. *Molecular biology and evolution* **33**, 1870-1874,
25 doi:10.1093/molbev/msw054 (2016).

26 73 Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence
27 times among organisms. *Bioinformatics* **22**, 2971-2972, doi:10.1093/bioinformatics/btl505
28 (2006).

29 74 Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences
30 with insertions. *Proceedings of the National Academy of Sciences of the United States of*
31 *America* **102**, 10557-10562, doi:10.1073/pnas.0409137102 (2005).

32 75 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and*
33 *evolution* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).

34 76 Zhang, J., Nielsen, R. & Yang, Z. Evaluation of an improved branch-site likelihood method for
35 detecting positive selection at the molecular level. *Molecular biology and evolution* **22**, 2472-
36 2479, doi:10.1093/molbev/msi237 (2005).

37 77 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large
38 gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57,

1 doi:10.1038/nprot.2008.211 (2009).
2 78 Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C. & Kanehisa, M. KAAS: an automatic
3 genome annotation and pathway reconstruction server. *Nucleic acids research* **35**, W182-185,
4 doi:10.1093/nar/gkm321 (2007).
5 79 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using
6 evolutionary information. *Nucleic acids research* **42**, W252-258, doi:10.1093/nar/gku340
7 (2014).
8 80 Han, S., Loulakis, P., Griffor, M. & Xie, Z. Crystal structure of activin receptor type IIB kinase
9 domain from human at 2.0 Angstrom resolution. *Protein science : a publication of the Protein*
10 *Society* **16**, 2272-2277, doi:10.1110/ps.073068407 (2007).
11 81 Pierce, B. G. *et al.* ZDOCK server: interactive docking prediction of protein-protein complexes
12 and symmetric multimers. *Bioinformatics* **30**, 1771-1773, doi:10.1093/bioinformatics/btu097
13 (2014).
14 82 Dehouck, Y., Kwasigroch, J. M., Rooman, M. & Gilis, D. BeAtMuSiC: Prediction of changes in
15 protein-protein binding affinity on mutations. *Nucleic acids research* **41**, W333-339,
16 doi:10.1093/nar/gkt450 (2013).

17

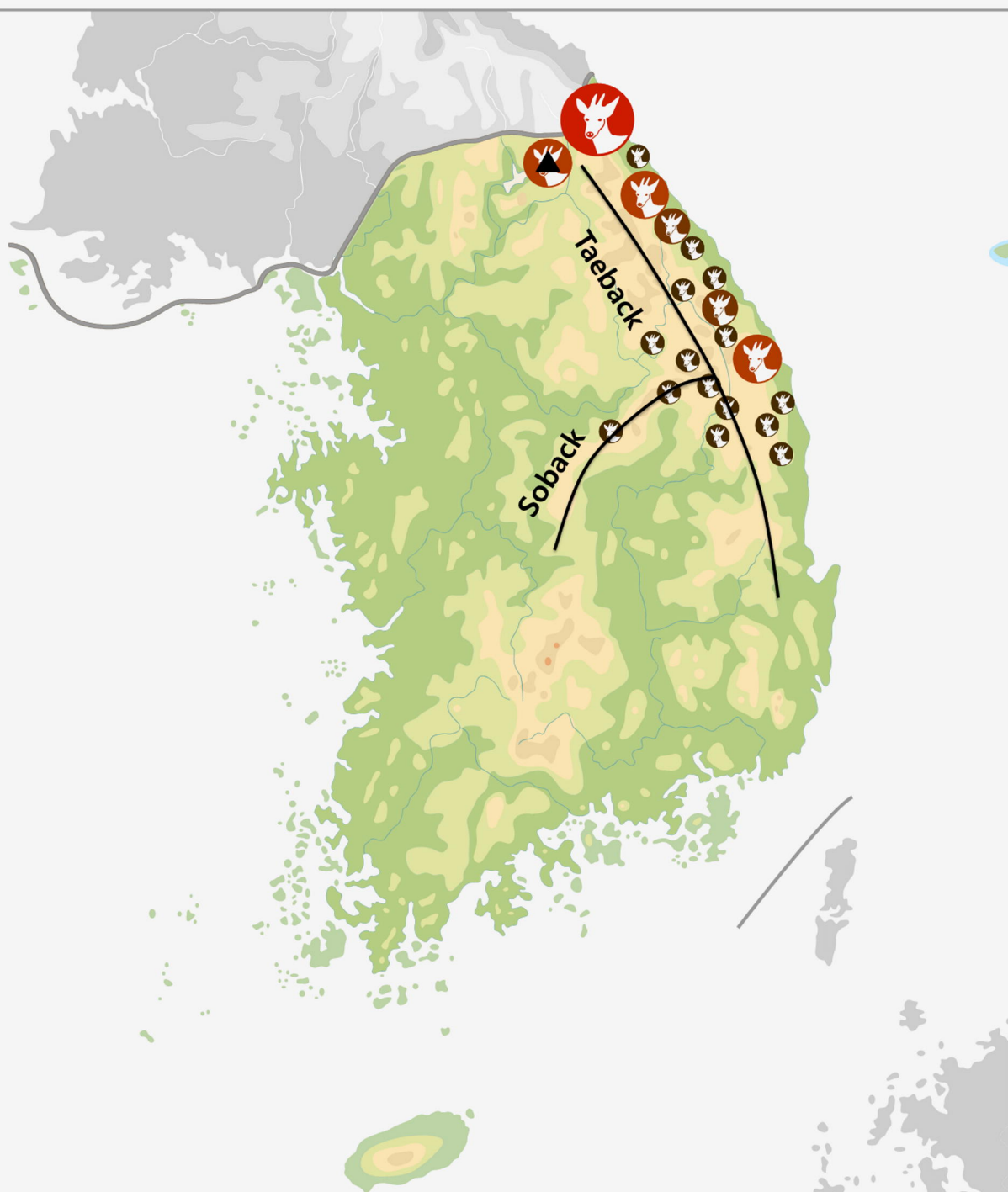
18

19 **Additional files**

20 **Additional file 1 – Table S1.** Sequencing and analysis statistics of the long-tailed goral's
21 WGS relative to the sheep genome. **Table S2.** Distribution of the homo- and hetero-SNVs for
22 long-tailed goral in each sheep chromosome. **Table S3.** PSGs of the long-tailed goral using a
23 branch model. **Table S4.** PSGs of the long-tailed goral using a branch-site model. **Table S5.**
24 GO enrichment of long-tailed goral branch PSGs. **Table S6.** GO enrichment of long-tailed
25 goral branch-site PSGs. **Table S7.** KEGG pathway enrichment of long-tailed goral branch
26 PSGs. **Table S8.** KEGG pathway enrichment of long-tailed goral branch-site PSGs (Table
27 S7). **Table S9.** PSGs of the Caprini using a branch model. **Table S10.** PSGs of the Caprini
28 using a branch-site model. **Table S11.** GO enrichment of Caprini branch PSGs. **Table S12.**
29 GO enrichment of Caprini branch-site PSGs. **Table S13.** KEGG pathway enrichment of
30 Caprini branch PSGs. **Table S14.** KEGG pathway enrichment of Caprini branch-site PSGs.
31 **Table S15.** Unique amino acid changes of the long-tailed goral. **Table S16.** Unique amino
32 acid changes of the Caprini. **Table S17.** KEGG analysis of function-altering unique amino
33 acid changes. **Table S18.** Enriched KEGG pathways using unique long-tailed goral amino

1 acid substitutions. **Table S19.** Enriched KEGG pathways using unique Caprini amino acid
2 substitutions. **Table S20.** Heterozygosity of the Caprini species.

3 **Additional file 2** - Figure S1. Hotspots of the homozygous SNVs from the long-tailed goral
4 distributed on the sheep genome. Hotspots were defined by the higher number of
5 homozygous SNVs (pink for Z-score > 2 and red for Z-score > 3) in 0.5M window bins. The
6 immune response genes with function-altering uAAS are given under the chromosomes;
7 Caprini-uAAS (yellow), long-tailed goral-uAAS (green), and uAAS observed in both (blue).



 > 200

 101-200

 10-100

 < 10

