The Tibetan antelope (*Pantholops hodgsonii*) is endemic to the extremely inhospitable high-altitude environment of the Qinghai-Tibetan Plateau, a region that has a low partial pressure of oxygen and high ultraviolet radiation. Here we generate a draft genome of this artiodactyl and use it to detect the potential genetic bases of highland adaptation. Compared with other plain-dwelling mammals, the genome of the Tibetan antelope shows signals of adaptive evolution and gene-family expansion in genes associated with energy metabolism and oxygen transmission. Both the highland American pika, and the Tibetan antelope have signals of positive selection for genes involved in DNA repair and the production of ATPase. Genes associated with hypoxia seem to have experienced convergent evolution. Thus, our study suggests that common genetic mechanisms might have been utilized to enable high-altitude adaptation.
The Tibetan antelope (TA; Pantholops hodgsonii) is a large, endemic artiodactyl that lives at elevations of 4,000–5,000 m on the Tibetan Plateau. This habitat has a low partial pressure of oxygen (PO2) and a high level of ultraviolet radiation. Non-native animals such as humans that visit such high-altitude regions may experience life-threatening acute mountain sickness. In contrast, the TA, which has survived for millions of generations on the plateau, can run at up to 80 km per hour for several hours in this low oxygen condition. These observations indicate that the TA must have evolved exceptional mechanisms to adapt to this extremely inhospitable habitat, yet the genetic bases of such adaptations remain unknown. Sequecing the TA genome will undoubtedly facilitate the discovery of potential molecular mechanisms of high-altitude adaptation.

Herein, using the next-generation, massively parallel sequencing technology—Illumina Genome Analyser—we report our generation and assembly of a draft genome for the TA. Further, we compare the TA genome to that of the American pika (Ochotona princeps), which is also native to high altitudes, to study the potential genetic basis for high-altitude living.

Results
Shotgun sequencing and de novo assembly. Genomic DNA extracted from a male TA at Kekezéli National Nature Reserve, Qinghai-Tibetan Plateau was subjected to shotgun sequencing using the Illumina short paired-end sequencing platform. We prepared 19 paired-end libraries spanning several insert sizes (from 165 to 20 Kb, Supplementary Table S1) to generate short pair-end reads. A total of 187 Gb of sequence data was generated for paired-end read lengths of 45 and 75 bp.

Generated reads were assembled using a pipeline designed for short paired-end reads. We improved the quality of assembly by adding 90 bp pair-end reads of a 165 bp insert size for contig assembly, and 45 bp pair-end reads of the 20 Kb insert size for scaffold assembly. Our final assembly had an N50 contig size of 18.6 Kb and an N50 scaffold size of 2.76 Mb (Supplementary Table S2).

We evaluated genomic coverage by sequencing a set of 5,547 expressed sequence tags (ESTs) followed by BLAT. About 95% of the ESTs mapped to the assembly (Supplementary Table S3). While considering the remaining 5% of unmapped ESTs that had an equivalent gene distribution, we estimated that the genome size was 2.75 Gb, which was slightly smaller than that of the cow (2.92 Gb) (ref. 6). We then used SOAPALIGNER to realign all usable sequencing reads onto an assembly to evaluate the single-base accuracy of the assembled genome. The peak sequencing depth was 53X and >97% of the assembled sequences were covered by over 20 reads (Supplementary Fig. S1).

The GC content pattern of the TA was similar to that of the cow, horse, human and mouse. Only a minor fraction of the TA genome had GC content < 20% (0.01%) or > 80% (0.2%) (Supplementary Fig. S2). Thus, the de novo genomic assembly for the TA covered high reads and it was not strongly affected by GC-biased non-random sampling. Accuracy of the scaffolds was evaluated by aligning them with the 26 available sheep BAC clones and the cow genome. Fine genomic synteny was detected (Supplementary Fig. S3). The trinucleotide class of simple sequence repeats in TA was usually < 12 bp (Supplementary Fig. S4a) and this was similar to that of the cow (Supplementary Fig. S4b). Whole-genome alignment of TA with the cow and human showed 635 Mb of shared sequences. Of the remainder, 499 Mb were shared between TA and cow, which was much higher than that between TA and human (107 Mb) or between cow and human (72 Mb) (Supplementary Fig. S5). We identified 2.2 million heterozygous single-nucleotide polymorphisms (SNPs) along the assembled TA genome and found two peaks in the distribution of SNP density (Fig. 1).

To predict the number of genes in the TA genome, we used both evidence-based and ab initio methods. Considering the finest gene set of human and the closet phylogenic distance of cow to TA, 22,389 human protein-coding genes and 20,892 protein-coding cow genes were projected for the TA genome, and we predicted 16,995 and 18,957 gene models, respectively. This method obtained conserved gene models among mammals only. To predict species-specific genes, we used GENSCAN, AUGUSTUS and GLIMMERHMM with model parameters trained on human and cow genomes, and predicted 53,608, 51,944 and 23,402 gene models, respectively. We integrated all five gene-sets to obtain a final gene-set of 21,426 protein-coding genes (Supplementary Table S4). The length distribution of genes, length of coding sequences (CDS), exon length, intron length, CDS GC ratio and distribution of exon number per gene for TA were compared with those of cow, horse, human and mouse. No obvious difference in any of these measures was seen for the TA (Supplementary Figs S6 and S7).

Genome and genome evolution. Compared with protein-coding genes of eight other mammals (cow, human, mouse, rat, chimpanzee, macaque, horse and dog), we detected 12,077 orthologous clusters that were shared between all the organisms. Further, 357 orthologous clusters were shared only between the TA and cow; these may have represented ruminate-specific genes (Fig. 2). Gene Ontology (GO) function categories were shown in Supplementary Table S5. Transposable elements comprised ~37% of the TA genome. The majority of known repeats present in the cow were found in the TA genome and the total number of elements in each major category was similar between the two genomes (Supplementary Table S6). Among major repeat families (Supplementary Table S7), the non-LTR (long terminal repeat) retrotransposons accounted for > 80% of the repeats and two, major, long interspersed elements (RTE-BovB and L1) were active.

Ancestral homologous synteny blocks (aHSBs) for the common ruminant ancestor of the cow and TA were reconstructed by identifying shared homologous synteny blocks (HSBs). The human genome was used as the outgroup to determine if the
The genome of the TA was compared with the draft genome of the American pika, another high-altitude species. Positively selected genes in the TA and pika were detected by assigning branches leading to the American pika and TA as forward branches (Supplementary Data sets 1 and 2). They were enriched in ATPase and DNA repair categories (Table 1). The TA and pika shared 76 genes that showed signals of positive selection. GO analysis identified 12 genes involved in the regulation of angiogenesis \( P = 0.018 \), folic acid and derivative biosynthetic processes \( P = 0.036 \), as well as DNA repair \( P = 0.046 \).

Among the 247 hypoxia genes examined, seven showed significant signals of convergent/parallel evolution \( P<0.05 \): ADORA2A, CCL2, ENG, PIK3C2A, PKLR, ATP12A and NOS3. As the American pika genome was a draft production, we also resequenced these seven genes in the Tibetan pika \( Ochotona curzoniae \), which occurs on the highest plateau in the world\(^{11} \); we detected convergent/parallel evolved amino acid sites in both species of pika and the TA.

Segmental duplication (SDs) is a common means of increasing gene copy number. We identified 4,640 duplicated fragments that represented recent SDs (\( > 90\% \) identity, \( > 1 \text{ Kb} \) length) in the TA genome via self-alignment. These comprised 22.6 Mb of sequences, which was less than that seen in the cow (Supplementary Table S10). The TA and cow shared the majority of the SDs. TA-specific SDs were related to energy metabolism such as NAD- and NADH-binding (GO:0051287) and ATP synthesis (GO:0006754 and GO:0015986) (Supplementary Table S11).

Gain and loss of genes in gene families could have had a role in adaptive evolution. The size of each gene family was determined by comparing the TA gene set with those of nine other mammals (cow, dog, horse, human, chimpanzee, macaque, opossum, mouse and non-Artiodactyla laurasiatherians (horse and dog, orange colour) on the basis of a representative gene in at least one of the grouped species.}

Potential molecular mechanism of high-altitude adaptation.

The genome of the TA was compared with the draft genome of the American pika, another high-altitude species\(^5 \). Positively selected genes in the TA and pika were detected by assigning branches leading to the American pika and TA as forward branches (Supplementary Data sets 1 and 2). They were enriched in ATPase and DNA repair categories (Table 1). The TA and pika shared 76 genes that showed signals of positive selection. GO analysis identified 12 genes involved in the regulation of angiogenesis \( P = 0.018 \), folic acid and derivative biosynthetic processes \( P = 0.036 \), as well as DNA repair \( P = 0.046 \).

Among the 247 hypoxia genes examined, seven showed significant signals of convergent/parallel evolution \( P<0.05 \): ADORA2A, CCL2, ENG, PIK3C2A, PKLR, ATP12A and NOS3. As the American pika genome was a draft production, we also resequenced these seven genes in the Tibetan pika \( Ochotona curzoniae \), which occurs on the highest plateau in the world\(^{11} \); we detected convergent/parallel evolved amino acid sites in both species of pika and the TA.

Segmental duplication (SDs) is a common means of increasing gene copy number. We identified 4,640 duplicated fragments that represented recent SDs (\( > 90\% \) identity, \( > 1 \text{ Kb} \) length) in the TA genome via self-alignment. These comprised 22.6 Mb of sequences, which was less than that seen in the cow (Supplementary Table S10). The TA and cow shared the majority of the SDs. TA-specific SDs were related to energy metabolism such as NAD- and NADH-binding (GO:0051287) and ATP synthesis (GO:0006754 and GO:0015986) (Supplementary Table S11).

Gain and loss of genes in gene families could have had a role in adaptive evolution. The size of each gene family was determined by comparing the TA gene set with those of nine other mammals (cow, dog, horse, human, chimpanzee, macaque, opossum, mouse
and rat) obtained from Ensembl. These analyses facilitated inferences into the expansion or contraction of each family (Supplementary Fig. S9). For TA, a large fraction of lost genes involved olfactory receptors and immunity. In contrast, genes that had functions associated with mitochondrial membranes, and, thus, potentially metabolic functions, had gains (Supplementary Table S12). Many expansions of gene families were related to energy metabolism (listed in Table 2).

**Discussion**

Our final assembly of the TA genome based on next-generation sequencing technology has an N50 contig size of 18.6 Kb and an N50 scaffold size of 2.76 Mb (Supplementary Table S2). About 95% of the ESTs map to the assembly (Supplementary Table S3). This suggests that our assembly is good enough for the following comparative genome analyses. The length distribution of genes, length of coding sequences (CDS), exon length, intron length, CDS GC ratio and distribution of exon number per gene for the TA do not obviously differ from those of the cow, horse, human and mouse (Supplementary Figs S6 and S7). Correspondence in these measures indicates a high quality of annotation for gene structure in the genome of the TA.

The pattern of heterozygous SNPs along the assembled TA genome is similar to that of the panda (Fig. 1). The same as panda, this pattern may indicate a bottleneck in the TA population caused by human hunting during the past tens of years.

Compared with eight other mammals (cow, human, mouse, rat, chimpanzee, macaque, horse and dog), 12,077 orthologous clusters are shared between all the organisms, and 357 are shared only between the TA and cow. The latter cluster may represent ruminant-specific genes (Fig. 2).

Positively selected genes in TA are in the ATPase and DNA repair categories (Table 1). These categories appear to be biologically relevant to living at high altitudes. ATPase genes have a role in providing energy. DNA repair genes may need to be more efficient given exposure to high levels of ultraviolet radiation. In addition, segmental duplications (Supplementary Table S13) and expansions of gene families (Table 2) also relate to energy metabolism. Thus, positive selection and expansion of gene families involved in energy metabolism appear to have an important role for TA via efficiently providing energy in conditions of low PO2.

Positively selected genes shared by the TA and pika involve the regulation of angiogenesis (P = 0.018), folic acid and derivative biosynthetic processes (P = 0.036), as well as DNA repair (P = 0.046). Folic acid, required for the synthesis and repair of DNA and the production of healthy red blood cells, aids in preventing anaemia. Therefore, gene term GO:0009396 (folic acid and derivative biosynthetic process) may reflect gene selection for both low PO2 and high ultraviolet radiation.

Seven of 247 hypoxia genes have signals of parallel/convergent sequence evolution in TA and pika. Several of these are particularly interesting because of their functional implications. For example, PKLR encodes pyruvate kinase, which catalyses the transphosphorylation of phosphoenolpyruvate into pyruvate and ATP, a rate-limiting step in glycolysis. Parallel/convergent evolution of this gene may reflect the importance of glycolysis in energy metabolism for survival in hypoxic, high-altitude environments. Another gene, NOS3 (endothelial nitric oxide synthase), is a critical mediator of cardiovascular homeostasis. It regulates the diameter of blood vessels and in doing so maintains an antiproliferative and antiapoptotic vascular environment.

Parallel/convergent evolution of this gene suggests that high-altitude adaptation involves the use of nitric oxide to regulate the diameter of blood vessels, which increases blood flow, thus allowing tissues to attain more oxygen. In extremely high environments, low PO2 can result in a precipitous reduction in O2 saturation in arterial blood. At 4,000 m altitude, PO2 of inspired air is 60% of that at sea-level. Thus, in the absence of adaptations or compensatory physiological mechanisms, O2 transport to tissues is severely compromised and this influences metabolism and the capacity to sustain physical activity.

Analyses of gene sequence convergence indicate that indigenous montane animals may use two major strategies to deal with hypoxia: (1) placing increased reliance on glycolysis; and (2) regulating blood vessel diameter, specifically through nitric oxide. High ultraviolet and especially hypoxia are the most important ecological factors restricting the viability of high-altitude animals.

---

**Table 1 | GO enrichment analysis of positive selection genes in the branch leading to Tibetan antelope and pika, respectively.**

| GO ID   | GO Term                                                      | P-value  |
|---------|--------------------------------------------------------------|----------|
| TA      |                                                              |          |
| GO:0042623 | ATPase activity, coupled                                    | 4.48E-05 |
| GO:0016887 | ATPase activity                                             | 9.55E-04 |
| GO:0033180 | Proton-transporting V-type ATPase, V1 domain                | 1.96E-03 |
| GO:0033176 | Proton-transporting V-type ATPase complex                    | 2.83E-03 |
| GO:0006281 | DNA repair                                                   | 7.11E-03 |
| GO:006974  | Response to DNA damage stimulus                              | 7.24E-03 |
| GO:0033554 | Cellular response to stress                                 | 8.30E-03 |
| GO:0033178 | Proton-transporting two-sector ATPase complex, catalytic domain | 1.89E-02 |
| GO:0045648 | Positive regulation of erythrocyte differentiation           | 2.68E-02 |
| GO:0015986 | ATP synthesis coupled proton transport                       | 3.31E-02 |
| GO:006289  | Nucleotide-excision repair                                   | 4.84E-02 |
| Pika     |                                                              |          |
| GO:0042623 | ATPase activity, coupled                                    | 1.44E-03 |
| GO:0016887 | ATPase activity                                             | 5.35E-03 |
| GO:0045765 | Regulation of angiogenesis                                  | 3.86E-02 |
| GO:0016702 | Oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, | 4.03E-02 |
| GO:0016701 | Oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, | 4.25E-02 |
| GO:0044429 | Mitochondrial part                                          | 4.92E-02 |
| GO:006281  | DNA repair                                                   | 5.33E-02 |

*Tibetan antelope.*

---

**Supplementary Information**

*Notes and references for the table.*

**Supplementary Figures**

*Images and additional data associated with the figures.*

**Supplementary Tables**

*Additional data and analyses not included in the main text.*
Native animals on the plateau that have survived there over thousands of years must have developed adaptive mechanisms to address harsh environmental stresses during their long history.

Previous studies on Tibetan people identified some of the genetic bases for adaptation to a high-altitude environment. Considering that Tibetans arrived on the plateau only 2,750 years ago, this relatively short time scale suggests that their adaptation to the highland may be ongoing and not fully integrated genetically. Domestic yaks yield some clues for highland adaptation. Artificial selection may confuse such results. Unlike Tibetans and domestic yak, native highland animals, such as the TA and pika, have adapted to this harsh environment for millions years. Thus, the study of genomes of native highland animals should provide a more complete blueprint to the genetic mechanisms of highland adaptation. Our study obtains a draft genome for the TA and identifies common themes of positive selection involved in DNA repair, ATPase function, angiogenesis and hypoxia, and parallel/convergent sequence evolution in genes that respond to hypoxia in TA and pikas. These discoveries potentially identify common genetic mechanisms of adapting species to harsh highland environments. Unfortunately, no genome is available from a sister species of the TA that is native to low altitudes. This absence of data precludes a genome-wide comparison. Thus, we cannot exclude the possibility that some genetic differences between the TA and cow may be due simply to divergence as a function of time and not high-altitude adaptation. Additional genomes will allow testing for adaptation.

**Methods**

**Genome sequencing and de novo assembly.** We constructed 19 paired-end DNA libraries with insert sizes of about 150 bp, 500 bp, 2 kb, 5 kb, 10 kb and 20 kb. For libraries of insert size longer than 1 kb, the desired DNA fragments were circularised by self-ligation. After being randomly fragmented, fragments that crossed the ligation boundaries were then enriched using magnetic beads with biotin and streptavidin. Paired-end sequencing was done on the Illumina Genome Analyser platform following the manufacturer's instructions. Fluorescent biotin and streptavidin. Paired-end information was processed into sequences using the Illumina data processing pipeline. Genomes will allow testing for adaptation.

**Whole-genome alignment.** Pairwise whole-genome alignment among TA, cow and human genomes using the REBASE transposable element library (v.2008-08-01). Highly divergent transposable elements were identified using REPEATMASKER after aligning the genome sequence to curated transposable element-related proteins. A de novo repeat library was constructed using REPEATMODELLER.

**Comparison of sheep BAC to cow and TA genome.** Twenty-six BAC clones from sheep chromosome 20 with complete sequences were identified in the NCBI database (accession number: F985852.1- F985877.1). Following alignment to the TA genome, 16 mapped onto one TA scaffold and showed fine genomic synteny. We then mapped the TA scaffold sequence to the cow genome and obtained an alignment. The other BAC sequences, except for two sequences, mapped to other scaffolds.
using PILE and REPEATMODELER and TRF tandem repeats of period \( \leq 12 \). The cow (BTA4.0) and human (hg18) repeat-masked genomes were downloaded from [UCSC](http://genome.ucsc.edu), guided by the topology of their species tree. The human genome was set as the reference and for input pairwise alignments the human versus cow and human versus TA alignments were generated.

Whole-genome assembly comparison methods were used to identify SDs\(^6\). Self-alignment of each genome was implemented by LASTZ with parameters \( T = 2, F = 9,400 \). We defined a SD as two sequences longer than 1 kb with an identity \( > 90\% \), but \( < 95\% \) to exclude potential improperly assembled allelic variants that may possibly reside in the 'draft' genome.

### Adaptive evolutionary analyses

We used the TREEFAM methodology\(^{27}\) to define a gene family as being a group of genes descend from a single gene in the last common ancestor. We applied a pipeline to cluster individual genes into gene families and performed phylogenetic analysis as below. (1) Data preparation: protein-coding genes from 10 mammalian species were used in this analysis; we retained the longest transcript isoform only for each gene and only considered proteins larger than 50 amino acids. (2) Pairwise relationship assignment (graph building): we performed BLASTP on all protein sequences against the database containing protein data of all of the species with an E-value cutoff of \( 10^{-5} \) and conjoined fragmented alignments for each gene pair using SOLAR (http://treesoft.svn.sourceforge.net/viewvc/treesoft/branches/dev/solar); we assigned a connection (edge) between two nodes (genes) if more than 1/3 of the region was aligned in both genes; an H-score ranging from 0 to 100 was used to weigh the similarity; two genes, G1 and G2, the H-score was defined as score \((G1/G2) / \max \text{(score(G1G1), score(G2G2))} \), (score = BLAST raw score). (3) Gene family construction: we used the average distance for the hierarchical clustering algorithm, requiring the minimum edge weight (H-score) to be \( > 5 \) and the minimum edge density (total number of edges/theoretical number of edges) to be larger than 1/3; clusters for a gene family was terminated when the presence of one or more outgroup genes was detected. (4) Phylogeny and orthology analyses: we performed multiple alignments of protein sequences for each gene family using MUSCLE\(^{28}\) and converted the protein alignments to CDS alignments using a Perl script; we built phylogenies using TREEBEST (http://treesoft.sourceforge.net/treesbest.html), which advantage of both codon-based and aa-based algorithms (nj-dn, nj-ds, nj-mm, pmml-aa and pmml-nt) and adjusted them to the topology of the species tree to form a more accurate consensus tree. We inferred orthologous and paralogous gene relationships from the gene tree.

We inferred the rate and direction of change in gene-family size for the TA, cow, dog, human, mouse, and rat. The average rate of gene turnover across the animals, which was the rate at which the size of the gene family was expected to expand or contratover time due to the gains or losses, was estimated first. Using the phylogeny, and while taking into account the topology and branch-lengths taken, we inferred changes in gene-family size and its significance.

### References

1. Schaller, G. B. *Wildlife of the Tibetan steppe* (University of Chicago Press, 1998).
2. Zhuang, H. B., Li, W. D. & Liu, Z. H. Tibetan antelope (*Pantholops hodgsoni*). *Chinese J. Zool.* 38, 74 (2003).
3. Smith, A. T. & Weston, M. L. Ochotona princeps. *Mamm. Species* 352, 1–8 (1990).
4. Li, R. et al. The sequence and de novo assembly of the giant panda genome. *Nature* 463, 311–317 (2010).
5. Kent, W. J. BLAT—The BLAST-like alignment tool. *Genome Res.* 12, 656–664 (2002).
6. Eissk, C. G. et al. Genome Sequence and Analysis Consortium. The genome sequence of *Taurinus cattus*: a window to ruminant biology and evolution. *Science* 324, 522–528 (2009).
7. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* 24, 713–714 (2008).
8. Salamov, A. A. & Solovyev, V. V. Ab initio gene finding in *Drosophila* genomic DNA. *Genome Res.* 10, 516–522 (2000).
9. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19, i225–i235 (2003).
10. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879 (2004).
11. Ge, R.-L., Kubo, K., Kobayashi, T., Sekiguchi, M. & Honda, T. Blunted hypoxic pulmonary vasoconstrictive response in the rodent *Ochotona curzoniae* (pika) at high altitude. *Am. J. Physiol. Heart Circ. Physiol.* 274, H1792–H1799 (1998).
12. Fenech, M. The role of folic acid and vitamin B12 in genomic stability of human cells. *Mutat. Res.* 475, 57–61 (2001).
13. Dimmeler, S. et al. Activation of nitric oxide synthase in endothelial cells by Akt-dependent phosphorylation. *Nature* 399, 601–605 (1999).
14. Jessen, T. H., Weber, R. E., Fermi, G., Tame, J. & Braunitzer, G. Adaptation of bird hemoglobin to high altitudes: demonstration of molecular mechanism by protein engineering. *Proc. Natl Acad. Sci. USA* 88, 6519–6522 (1991).
15. Storz, J. F. et al. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proc. Natl Acad. Sci. USA* 106, 14450–14455 (2009).
16. Xu, S. Q. et al. A mitochondrial genome sequence of the Tibetan antelope (*Pantholops hodgsoni*). *Geno. Prot. Bioinfo.* 3, 5–17 (2005).
17. Yi, X. et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329, 75–78 (2010).
18. Simonson, T. S. et al. Genetic evidence for high-altitude adaptation in Tibet. *Science* 329, 72–75 (2010).
19. Xu, S. et al. A genome-wide search for signals of high-altitude adaptation in Tibetans. *Mol. Biol. Evol.* 28, 1003–1011 (2011).
20. Qiu, Q. et al. The yak genome and adaptation to life at high altitude. *Nat. Genet.* 46, 1946–1949 (2014).
21. Li, R. et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20, 265–272 (2010).
22. Pevzner, P. A., Tang, H. & Waterman, M. S. An Eulerian path approach to DNA fragment assembly. *Proc. Natl Acad. Sci. USA* 98, 9748–9753 (2001).
23. Jurka, J. et al. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467 (2005).
24. Kent, W. J. et al. The human genome browser at UCSC. *Genome Res.* 12, 966–1006 (2002).
25. Ma, J. et al. Reconstructing contiguous regions of an ancestral genome. *Genome Res.* 16, 1557–1565 (2006).
26. Bailey, J. A. et al. Recent segmental duplications in the human genome. *Science* 297, 1003–1007 (2002).
27. Li, H. et al. TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res.* 34, D572–D580 (2006).
28. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797 (2004).

Human NCBI EntrezGene IDs were used in all analyses of gene ontology. Non-human orthologs of the human genes were retrieved from Ensembl Biomart. For conformity of functional annotation enrichment results, we used the human NCBI EntrezGene IDs to refer to both the human genes and to their putative non-human orthologs. We used DAVID\(^{34,35}\) as a functional annotation clustering tool for each combination of species and dN/dS bin to group genes with shared annotations. The algorithm assigned a significance P-value, corrected for multiple testing, to each subgroup representing a gene ontology annotation within the cluster and an enrichment score to the entire cluster. The clusters with higher enrichment scores consisted of subgroups with higher significance values, and, thus, these clusters provided an integrated view of the more significantly enriched or over-represented gene functional categories within each dN/dS bin. The enrichment score for each annotation cluster was based on the geometric mean of the P-values of the cluster's assorted annotations.
Acknowledgements

This work was supported by National Basic Research Program of China (No.2006CB504100, 2007CB411600 and 2012CB518200), Basic Science Research Foundation of Bureau of Science and Technology in Qinghai Provinces (2010-1456), Program of International S&T Cooperation of China (No.05S2012GR0195), National Nature Science Foundation of China (No.30393133), Bureau of Science and Technology of Yunnan Province, Chinese Academy of Sciences and Natural Sciences and Engineering Research Council (Discovery Grant 3148). We greatly appreciate the cooperation of Mongolian nomads of the Wu Tuo-Meren Village for their cooperation and hospitality. Without their help, capture of the antelope at this extremely high-altitude would have been impossible. We thank L. Goodman for editing the manuscript.

Author contributions

R.L.G., Q.C. and Y.Y.S. contributed equally to this work. R.L.G., Ju.W., Ji.W., Y.P.Z. and Y.Z. managed the project. L.M., X.Z., Yo.Z., Jl.X., Xin.Y., Y.L., H.B. and Y.Y. prepared the DNA sample of the Tibetan antelope. Asan, Y-f.Z., K.W., B.W., C.G. and X.O. performed sequencing. Ju.W., Q.C. and Asan designed analyses. Y.C., L.Y. and Y.H. performed genome assembly. R.H., Y.L. and M.H. performed genome annotation. JX. performed the repeat analysis. J.K., D.M.L., J.M., H.A.I., L.A. and B.C. performed comparative genomics. Y.P.Z., Y.Y.S., W.Z. and T.Z. analysed genes related to hypoxia. J.X., R.N.P. and D.A.R. performed the repeat analysis. R.L.G., Ju.W., Ji.W., Y.Z., Q.C., Y.Y.S., R.W.M. and Y.P.Z. wrote the paper. G.Z. and D.M.I. revised the paper.

Additional information

Accession codes: This Whole-Genome Shotgun project has been deposited in GenBank Genome database under the accession number AGTT00000000. The version described in this paper is the first version, AGTT01000000.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Ge, R.-L. et al. Draft genome sequence of the Tibetan antelope. Nat. Commun. 4:1858 doi: 10.1038/ncomms2860 (2013).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/