

# Pervasive introgression facilitated domestication and adaptation in the *Bos* species complex

Dong-Dong Wu<sup>1,2,3,10\*</sup>, Xiang-Dong Ding<sup>4,10</sup>, Sheng Wang<sup>4,10</sup>, Jan M. Wójcik<sup>5,10</sup>, Yi Zhang<sup>4,10</sup>, Małgorzata Tokarska<sup>5</sup>, Yan Li<sup>6</sup>, Ming-Shan Wang<sup>1,2</sup>, Omar Faruque<sup>7</sup>, Rasmus Nielsen<sup>8\*</sup>, Qin Zhang<sup>9\*</sup> and Ya-Ping Zhang<sup>1,2,3\*</sup>

**Species of the *Bos* genus, including taurine cattle, zebu, gayal, gaur, banteng, yak, wisent and bison, have been domesticated at least four times and have been an important source of meat, milk and power for many human cultures. We sequence the genomes of gayal, gaur, banteng, wisent and bison, and provide population genomic sequencing of an additional 98 individuals. We use these data to determine the phylogeny and evolutionary history of these species and show that the threatened gayal is an independent species or subspecies. We show that there has been pronounced introgression among different members of this genus, and that it in many cases has involved genes of considerable adaptive importance. For example, genes under domestication selection in cattle (for example, *MITF*) were introgressed from domestic cattle to yak. Also, genes in the response-to-hypoxia pathway (for example, *EGLN1*, *EGLN2* and *HIF3a*) have been introgressed from yak to Tibetan cattle, probably facilitating their adaptation to high altitude. We also validate that there is an association between the introgressed *EGLN1* allele and haemoglobin and red blood cell concentration. Our results illustrate the importance of introgression as a source of adaptive variation and during domestication, and suggest that the *Bos* genus evolves as a complex of genetically interconnected species with shared evolutionary trajectories.**

Taurine cattle (*Bos taurus taurus*, abbreviated to 'taurine' hereafter) and zebu (*Bos taurus indicus*) are the most common large domesticated ungulates, and are both considered members of the same species (*Bos taurus*). Another three species from the *Bos* genus, gayal (*Bos frontalis*), banteng (*Bos javanicus*) and yak (*Bos grunniens*), have also been independently domesticated by humans. The gaur (*Bos gaurus*), American bison (*Bison bison*) and European bison, that is, wisent (*Bison bonasus*) are the only three extant *Bos* species that are not domesticated. Recent genome-sequencing data of taurine, zebu, gayal, wisent and yak have greatly improved our understanding of the evolution of these large ungulates and have helped elucidate the complex process of domestication and adaptation in these species<sup>1–8</sup>.

However, many aspects of this complex genus remain the subject of considerable speculation. For example, the phylogenetic relationship among species within the *Bos* genus is controversial, and different studies have reported different topologies<sup>9–11</sup>. In particular, the phylogenetic position of gayal is still under debate. The gayal has been claimed to be either an independent species<sup>12,13</sup> or a domestic form of gaur<sup>13</sup>, and it has been suggested to be a hybrid descending from crosses of wild gaur and domestic cattle<sup>14,15</sup>. Also, through human-mediated migration, taurine and zebu now occupy almost all areas of the world, and may, through introgression, have helped facilitate domestication of regional species

(for example, gayal, banteng and yak). Similarly, regional species (for example, gaur, gayal, banteng, yak, wisent and bison) may have contributed to the gene pool of domestic cattle and zebu thereby helping them to adapt to local environments. The degree to which introgression and admixture has been a driving force in the domestication and formation of this species complex is still an unanswered question.

Here, we use large-scale genomes to determine the phylogeny and evolutionary history of these species within the *Bos* genus and show that the threatened gayal is an independent species or subspecies. We show that there has been pronounced introgression among different members of the *Bos* genus, and that it in many cases has involved genes of significant adaptive importance.

## Results and discussion

**Phylogenetic analysis.** We generated five whole genomes for gayal, gaur, banteng, American bison and wisent at high depth (~20×), and combined these data with three previously published high-depth genomes from taurine, zebu and yak (Supplementary Table 1). We first used the high-depth whole genomes to estimate the long-term trend in effective population size for these species using the pairwise sequentially Markovian coalescent (PSMC) model<sup>16</sup> and found a dramatic decline in population sizes of all these extant species more than ~20,000–50,000 years ago, coinciding with the

<sup>1</sup>State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China. <sup>2</sup>Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China. <sup>3</sup>Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. <sup>4</sup>Key Laboratory of Animal Genetics, Breeding and Reproduction, Ministry of Agriculture, National Engineering Laboratory for Animal Breeding, College of Animal Science and Technology, China Agricultural University, Beijing, China. <sup>5</sup>Mammal Research Institute Polish Academy of Sciences, Białowieża, Poland. <sup>6</sup>State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming, China. <sup>7</sup>Department of Animal Breeding and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh. <sup>8</sup>Department of Integrative Biology, University of California Berkeley, Berkeley, CA, USA. <sup>9</sup>College of Animal Science and Technology, Shandong Agricultural University, Taian, China. <sup>10</sup>These authors contributed equally: Dong-Dong Wu, Xiang-Dong Ding, Sheng Wang, Jan M. Wójcik, Yi Zhang. \*e-mail: [wudongdong@mail.kiz.ac.cn](mailto:wudongdong@mail.kiz.ac.cn); [rasmus\\_nielsen@berkeley.edu](mailto:rasmus_nielsen@berkeley.edu); [qzhang@cau.edu.cn](mailto:qzhang@cau.edu.cn); [zhangyp@mail.kiz.ac.cn](mailto:zhangyp@mail.kiz.ac.cn)

time of last glacial period and the spread of modern humans across Eurasia and the Americas<sup>17</sup> (Supplementary Fig. 1).

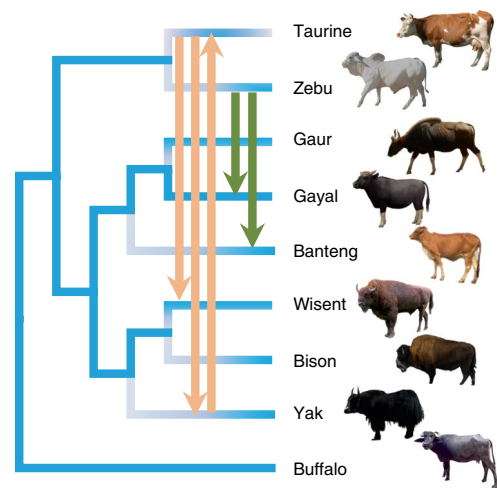
We next inferred a concatenated genome tree using neighbour-joining and maximum-likelihood methods<sup>18</sup> for the autosomes using the RAxML program<sup>18</sup>. The topology ((taurine, zebu), (((gaur, gayal), banteng), (bison, yak))), henceforth labelled Topology A, was both the neighbour-joining and maximum-likelihood topology (Fig. 1 and Supplementary Fig. 2). We also performed ASTRAL analysis<sup>19</sup> and MP-EST analysis<sup>20</sup> using autosome sequences and found additional support for Topology A with high bootstrap support (Supplementary Fig. 3).

However, we also noticed that many segments of the genome supported alternative topologies (Supplementary Figs. 4–10). For example, the best-supported topology for the X chromosome was (((taurine, zebu), (gaur, gayal)), banteng), (bison, yak)) (Supplementary Figs. 2 and 3), henceforth labelled Topology B. We also used only protein-coding sequences to infer phylogeny by RAxML and MP-EST, and obtained different topologies depending on the choice of window size for the MP-EST analysis (Supplementary Figs. 2 and 11). Similar to previous studies<sup>3,21</sup>, the mitochondrial phylogeny was inconsistent with the nuclear DNA phylogeny (Supplementary Fig. 12), probably due to genetic introgression or incomplete lineage sorting.

To further investigate these patterns, we split the genomes into 100 kb non-overlapping windows, resulting in 25,127 and 1,489 windows for the autosome and X chromosome, respectively, and estimated maximum-likelihood trees for all windows separately<sup>18</sup>. The most common topology is Topology A; however, even this topology is only found in 13.42% and 13.97% of the autosomal and X chromosomal segments, respectively (Supplementary Figs. 4 and 5). Similar results were found when using 1 Mb non-overlapping sliding windows (Supplementary Figs. 6 and 7). The observation that the majority topology is so rare, despite the presence of plentiful phylogenetic signals, suggests that incomplete lineage sorting and/or introgression is driving the phylogenetic pattern in these species.

We find that the tree-length on average is longer for Topology A than for Topology B (Supplementary Fig. 2e), an observation compatible with the hypothesis that Topology A reflects the majority species tree (Fig. 1), but that other topologies, such as Topology B, occur, at least in part, due to introgression. To further elucidate whether introgression might help explain the observed topological incongruence, we performed a series of ABBA-BABA tests, and indeed found evidence of pervasive introgression between different species (Supplementary Table 2).

We further inferred the divergence time by the coalescent hidden Markov model (CoalHMM)<sup>22</sup> using autosome sequences, and found that gayal and gaur diverged ~994,000 years ago (Supplementary Figs. 13–15). Considering the relatively long divergence time, and that the karyotype of gayal is  $2n = 58$  (where  $n$  is the haploid number of chromosomes), while gaur is  $2n = 56$  (ref. <sup>15</sup>), we argue that the gayal should be treated as a separate subspecies or even species, not a domestic form of gaur, and is probably domesticated from an extinct subspecies or species. These results will help facilitate future exploration of genetic diversity and evolution of gaur and gayal, which are distributed across southern East Asia and will, in particular, help focus the conservation efforts of gayal. One caveat is that the our estimates of divergence times tend to be lower than those previously reported in treetime (<http://www.timetree.org/>), which were from many studies inferred from molecular data or from the fossil record<sup>23</sup>. One possible explanation is a bias in our estimates due to our procedure of variant calling by mapping of sequencing reads of other species to the cattle reference genome. This procedure could miss some variants in highly divergent regions. Another possible explanation is introgression, which also would attenuate the estimates of divergence times. De novo assembled genomes for each species could help distinguish between these explanations.



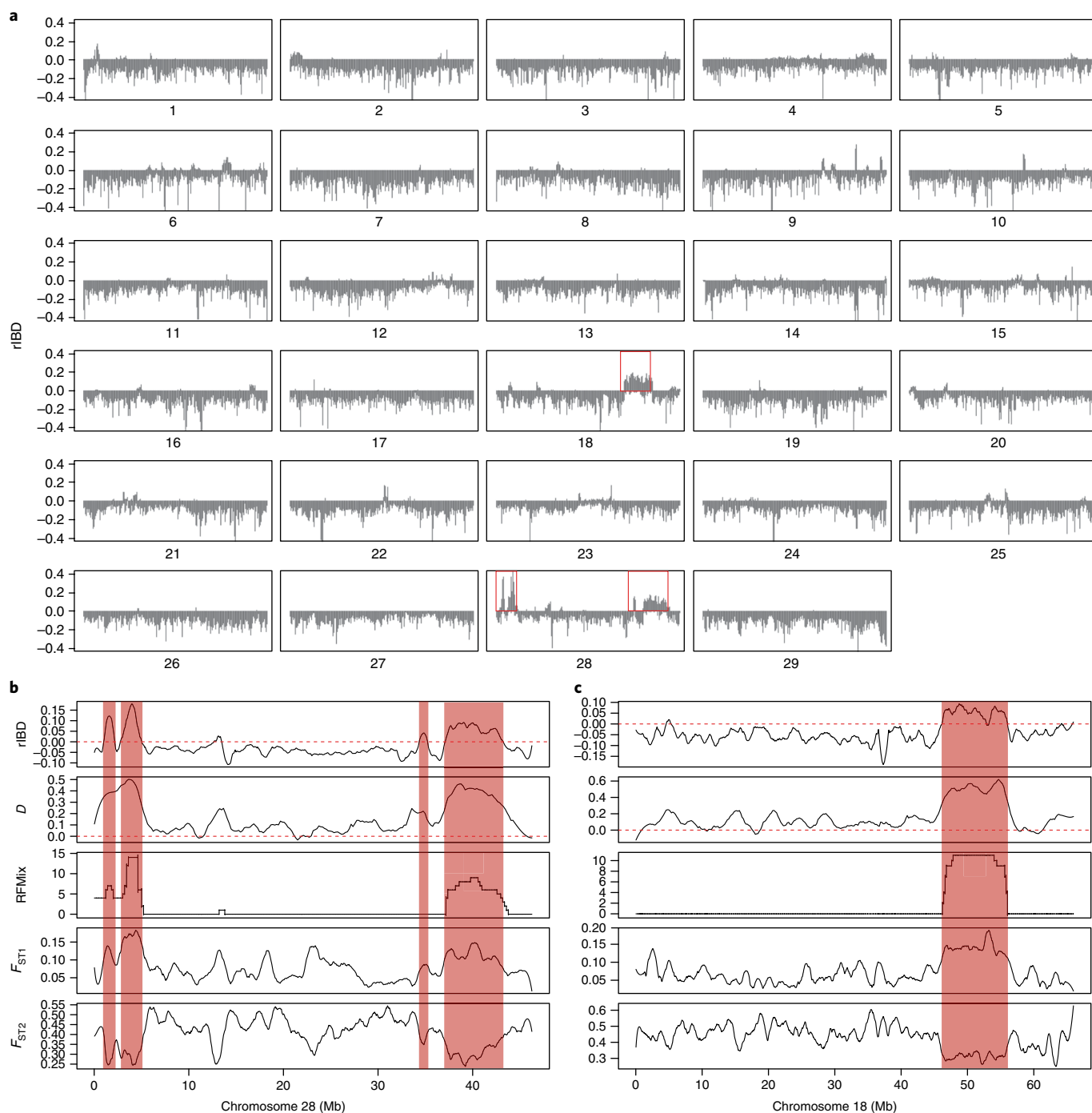
**Fig. 1 | Phylogenetic tree and genetic introgression of species in the *Bos* genus.** Only introgressions reported in this study are highlighted.

**Analysis of genetic introgression.** We next generated another 98 genomes from 22 gayal from China, India and Bangladesh, 59 taurine and zebu from India and China, 7 bali cattle domesticated from banteng, 10 wisent, sequenced at an average of 5×, and combined with 39 previously published genomes from 6 yaks and 33 European and Korean taurine (Supplementary Table 3) to investigate potential pattern and consequences of introgression. To estimate the history and pattern of introgression in these species, we used TreeMix to estimate an admixture tree<sup>24</sup>. Introgression events were inferred from domestic cattle to bison and wisent (Fig. 1 and Supplementary Figs. 17 and 18), corroborating previous work<sup>25</sup>. In addition, genetic introgression was also determined from zebu to gayal, and from zebu to bali cattle. These signals of genetic introgression were supported by *D*-statistic analyses (ABBA-BABA tests)<sup>26,27</sup> (Supplementary Table 4).

As admixture between wisent and European cattle has been investigated in previous studies<sup>25</sup>, we focus our analyses on introgression between zebu and bali cattle and between zebu and gayal. We used the relative frequency of identical-by-descent (rIBD) statistic<sup>28</sup> to infer regions of the introgression sliding window (window size, 20 kb; step size, 10 kb) of the genome. A positive rIBD value suggests potential introgression between zebu and gayal or bali cattle (Supplementary Figs. 19 and 20). We further estimated phylogenetic trees of all individuals for each genomic segment to corroborate the signal of genetic introgression, and found that the discordance rate is 0.38% and 18.8% using rIBD analysis to identify genetic introgression between zebu and gayal and between zebu and bali cattle, respectively (Supplementary Tables 5 and 6). The introgression segments validated by phylogenetic analysis contain in total ~32.42 Mb and ~25.61 Mb, with 451 and 328 protein-coding genes, between zebu and gayal and between zebu and bali cattle, respectively.

Categorization based on expression profiles also shows a similar enrichment in the nervous system and immune system (Supplementary Figs. 19 and 20). Gene-enrichment analysis identified some genes involved in immunity and the neural system, although different methods identified different enriched categories (Supplementary Tables 7 and 8).

We examined population differentiation between gayal and bali cattle for each introgressed region to evaluate whether there is any evidence of positive selection. We identified a few introgression regions with elevated levels of population differentiation (Supplementary Tables 5 and 6). Among the ten regions showing introgression from zebu to gayal and with a high level

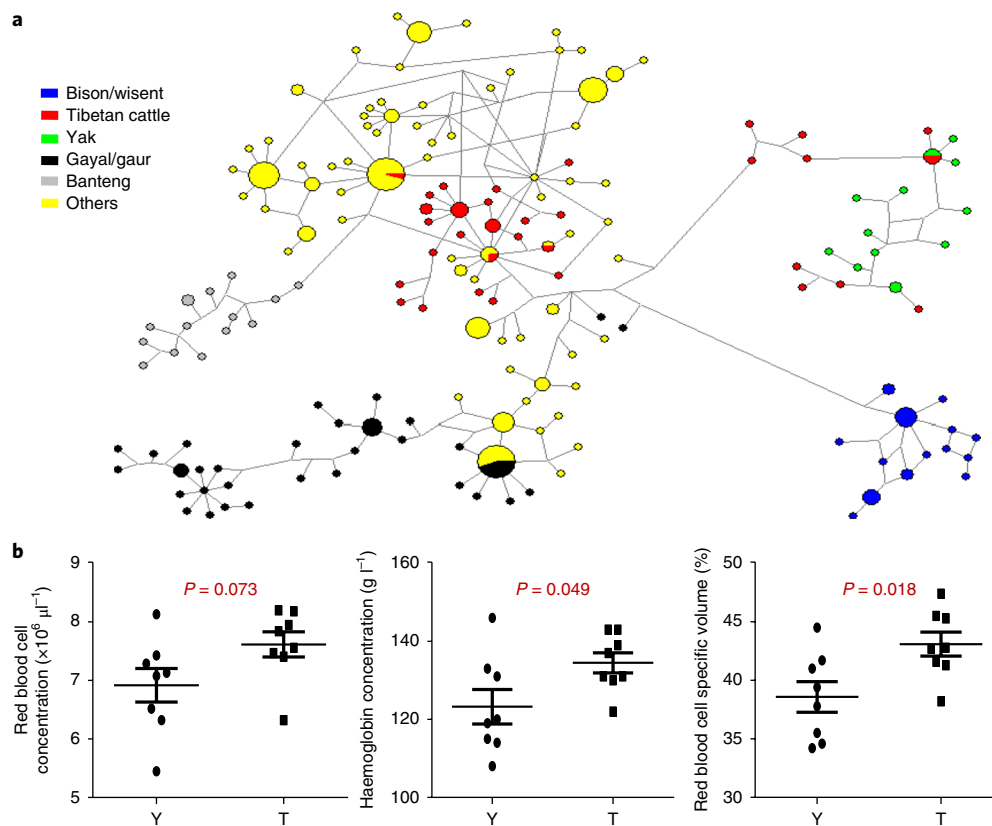


**Fig. 2 | Distribution of regions in the genome where introgression occurred between Tibetan cattle and yak. a**, Landscape of rIBD values across genomes. Three large regions on chromosomes 18 and 28 are marked by red lines. **b, c**, Values of rIBD,  $D$  (ABBA-BABA), RfMix,  $F_{ST1}$  (between Tibetan cattle and zebu),  $F_{ST2}$  (between Tibetan cattle and yak) across chromosome 28 (**b**) and chromosome 18 (**c**). A detailed description of these parameters is given in the Methods section. The significant regions are marked by red shading.

of population differentiation (Supplementary Table 5), we note that one region (chromosome 5: 100300001–100770001) contains a cluster of genes (*CLEC1A*, *CLEC9A*, *CLEC1B*, *CLEC2A*, *KLR2*, *CLEC12A* and *CLEC12B*) encoding members of the C-type lectin/C-type lectin-like domain (CTL/CTLD) superfamily (Supplementary Fig. 21). The human orthologous region containing these genes is on the short arm of chromosome 12, includes CTL/CTLD genes expressed in natural killer, dendritic and endothelial cells, and plays an important role in the immune system<sup>29</sup>. It may suggest a possible role for immune-

and defence-related adaptation, as also suggested by the enrichment analyses.

The highest introgression peak between zebu and gayal contains only one gene, *SYN3* (synapsin III) (Supplementary Fig. 19a), and a phylogenetic analysis supports genetic introgression of *SYN3* from zebu to gayal (Supplementary Fig. 19c). The introgressed *SYN3* shows a high level of population differentiation, compatible with positive selection in gayal (Supplementary Table 5). *SYN3* is a member of the synapsin gene family, which are a family of neuron-specific synaptic vesicle-associated phosphoproteins that have been implicated in



**Fig. 3 | Genetic introgression of *EGLN1* from yak to Tibetan cattle.** **a**, Network of *EGLN1* haplotypes indicates its introgression from yak to Tibetan cattle. **b**, Comparison of three haematologic parameters (red blood cell concentration, haemoglobin concentration and red blood cell specific volume) between Tibetan cattle carrying yak-like *EGLN1* haplotypes (Y) and other Tibetan cattle (T). The foreleg venous blood was collected to measure physiological parameters. Only male individuals were used for the comparison. Error bars show the standard error of the mean. Student's *t*-test was employed to compute the *P* values.

synaptogenesis and in the modulation of neurotransmitter release<sup>30</sup>. *Syn3*-knockout mice display perturbations in conditioned fear<sup>31</sup>. This makes it a particularly relevant domestication gene, as domesticates manifest a reduced fear of humans<sup>32,33</sup>. Similarly, the region (chromosome 17: 74800001–74820001) showing introgression from zebu to Bali cattle and potential positive selection in Bali cattle with a high level of population differentiation contains *SEPT5* and *GPIBB* (Supplementary Fig. 22). *Sept5*-knockout mice exhibit decreased anxiety-related behaviour<sup>34</sup>. Previous studies have reported that many genes involved in the nervous system have evolved under artificial selection, which has been tightly coupled to the behavioural shift of domestic animals compared with their wild progenitors<sup>35–38</sup>. The zebu has been domesticated for much longer than gayal and Bali cattle and it is possible that introgression of neural genes from zebu to gayal and Bali cattle, might have facilitated successful domestication of the latter. In support of this, a recent study<sup>5</sup> reported that introgression between domestic cattle and yak is significantly enriched in genes involved in brain development and function.

**Genetic introgression between yak and Tibetan cattle.** We next investigated the pattern of introgression between yak and Tibetan cattle using rIBD (Fig. 2a). The phylogenetic trees for each genomic segment were concordant with the inferences made using rIBD. In total, ~30.7 Mb, containing 541 protein-coding genes, exhibited signals of introgression between Tibetan cattle and yak (Supplementary Table 9). Noticeably, there are several large regions on chromosomes 18 and 28 with large rIBD values, exhibiting a very strong signal of introgression between yak and Tibetan cattle (Fig. 2a). Admixture signals in these regions were further supported by

ABBA-BABA tests (Fig. 2b). We also used a third method, RFBmix, for local-ancestry inference<sup>39</sup>, which provided additional support for introgression between Tibetan cattle and yak (Fig. 2b). Finally, we used a method from ref. 40 (see Methods) to exclude the possibility of incomplete lineage sorting of these long regions ( $P \approx 0$ ). These regions displayed a high level of population differentiation between Tibetan cattle and zebu, but a lower level of population differentiation between yak and Tibetan cattle (Fig. 2b,c), corroborating introgression from yak to Tibetan cattle.

We divided the two regions on chromosome 28 into different windows and calculated rIBD peaks and RFBmix values for each of them (Fig. 2b). Phylogenetic analysis strongly supported the direction of introgression from yak to Tibetan cattle (Supplementary Figs. 23–26). In particular, the second window contains *EGLN1*, which is a gene in the hypoxia inducible factor (HIF) pathway and one of two key genes for hypoxia adaptation of Tibetan humans<sup>41,42</sup>. Network construction of *EGLN1* haplotypes supported introgression from yak to Tibetan cattle (Fig. 3a). We further examined the phenotypic consequence of introgression of *EGLN1*, and found that the Tibetan cattle carrying yak-like *EGLN1* haplotypes exhibited on average substantially lower haemoglobin concentrations and red blood cell counts than other Tibetan cattle (Fig. 3b). This pattern is consistent with that observed for *EPAS1* in Tibetan humans, which favours attenuation of the induced maladaptive phenotypic changes in response to hypoxia, such as elevated haemoglobin concentration (polychythemia)<sup>42,43</sup>.

The other strong candidate region is ~10 Mb long and is located on chromosome 18 (46125199–56073102). It harbours a significantly elevated *D* value and signals of introgression using RFBmix, and

exhibited a high level of population differentiation between Tibetan cattle and zebu, but a lower level of population differentiation between yak and Tibetan cattle (Fig. 2c). Phylogenetic analysis of the haplotypes also strongly supports that the direction of introgression is from yak to Tibetan cattle (Supplementary Fig. 27). Remarkably, the introgressed haplotypes contain two genes in the HIF pathway: *EGLN2* and *HIF3a*. Yak are thought to have inhabited the Tibetan Plateau for million of years and have numerous adaptations to high altitude, such as substantially enlarged lungs and hearts. In contrast, domestic taurine were introduced to the Tibetan Plateau by humans only a few thousand years ago<sup>44</sup>. Taurine may not have been well adapted to the extreme environment when initially brought to the Tibetan Plateau, and still usually suffers from severe pulmonary hypertension when reared there. Our results show that taurine received some of the central HIF pathway genes by introgression from the yak. This pattern mimics that observed in humans where the crucial *EPAS1* haplotype was introgressed from Denisovans to Tibetans<sup>40</sup>, and the pattern in dogs where introgression occurred from the Tibetan wolf to the Tibetan mastiff<sup>45</sup>. Remarkably, in all these three cases of humans or human-domesticated animals, a crucial step in the adaptation to high altitude appears to be introgression from related species that already are adapted to the local environment.

We also investigated whether domestic yak received alleles from domesticated cattle as part of their domestication process. We identified 5.54 Mb containing 89 protein-coding genes introgressed from domestic cattle to yak, which included the two regions reported by ref. <sup>5</sup> to be introgressed in the homozygous state in the yak reference genome (Supplementary Figs. 28 and 29). Interestingly, we also identified introgression of alleles in the microphthalmia-associated transcription factor gene (*MITF*; Supplementary Fig. 30). *MITF* is the master regulator of melanocytes, playing a key role in their development, differentiation, function and survival, and pigment synthesis<sup>46</sup>. *MITF* targets many pigment genes, such as tyrosinase (*TYR*), dopachrome tautomerase (*DCT*, also known as tyrosinase-related protein 2, or *TYRP2*) and tyrosinase-related protein 1 (*TYRP1*)<sup>46</sup>. Previous studies have shown that the genes under strongest selection during domestication in cattle are genes associated with coat colouration, and that one of the genes most strongly targeted by domestication selection is *MITF*<sup>47–49</sup>. Presumably yak have been under similar domestication selection to cattle involving coat colour, facilitating the adaptive introgression of the *MITF* alleles.

## Conclusions

Using whole-genome sequencing, we have clarified the phylogenetic relationship among extant species in the *Bos* genus. However, we discovered that the *Bos* genus is a species complex with extensive introgression among species. Also, this introgression is likely to have adaptive significance, as best illustrated by the introgression of *EGLN1* alleles from yak to cattle that lowers the haemoglobin and red blood cell concentration, an adaptive response to hypoxia-induced polychythemia seen in many other species including humans. Together with other recent studies showing evidence of adaptive introgression in many species including canids and humans, our results suggest that these species do not evolve in genetic isolation, but that introgression from other species is a very important source of new genetic variation when adapting to a new environment. The *Bos* genus is a genetically connected species complex where introgression has helped facilitate adaptation in connection with both domestication and environmental changes. This changing view of the basic properties of species has profound implications for animal breeders and conservation biologists alike. Introgression and admixture has previously been considered a detrimental process to avoid. We now know that it is an important natural process of significant importance for adaptation.

## Methods

**Samples for genome sequencing.** Animal sampling was carried out in accordance with the animal experimentation guidelines and regulations of the Kunming Institute of Zoology, and was approved by the Institutional Animal Care and Use Committee of the Kunming Institute of Zoology. The cell line of one gayal individual kept in Kunming Cell Bank of Kunming Institute of Zoology was chosen for high-depth genome sequencing, as its karyotype was validated to contain 58 chromosomes. Blood samples of another 22 gayal individuals were collected from the Gaoligong mountain region in Yunnan province in China, and India and Bengal. One gaur individual used for genome sequencing was collected from Xishuangbanna, Yunnan, and its species status were validated by morphology and mitochondrial DNA sequence. Blood samples of 8 Bali cattle were from Indonesia. The above samples were from the Animal Branch of the Germplasm Bank of Wild Species (GBOWS), Chinese Academy of Sciences. One American bison individual was sampled from Beijing Zoo. Blood samples of Chinese yellow cattle were collected from different original birthplaces in China. The bloods of the Tibetan cattle used for whole-genome sequencing were collected from Gongbogyamda region in Tibet and haematologic parameters were measured using a BC-2800Vet Auto Hematology Analyzer (Mindray).

DNA was extracted from all samples using standard phenol–chloroform extraction protocol. The analysed 11 wisent individuals were culled in the Knyszyn Forest, northeastern Poland according to the Ministry of Environment permission. Genomic DNA was extracted from soft tissues using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocols. Using the Illumina standard protocol, genomic libraries with insert size ~400 bp were constructed, and sequenced by HiSeq2000 or HiSeq2500.

**Reads mapping and variants calling.** Sequencing reads were examined and filtered by the Btrim program<sup>50</sup>. All passed quality-filtered reads were aligned to the *Bos taurus* reference genome assembly UMD3.1 using the MEM algorithm of Burrows–Wheeler Aligner (version 0.7.10)<sup>51</sup>. After sorting reads, marking duplicate reads and local realignment around indels, variants were called using the UnifiedGenotyper of the Genome Analysis Toolkit (GATK, version 3.2.2)<sup>52</sup>. Finally, the hard filters were applied to the raw single nucleotide polymorphisms (SNPs) using the GATK program according to the criteria as follows: QUAL < 30; QualByDepth (QD) < 2.0; RMSMappingQuality (MQ) < 40.0; MappingQualityRankSumTest (MQRankSum) < -12.5; ReadPosRankSumTest (ReadPosRankSum) < -8.0; HaplotypeScore > 13.0.

**Phylogenetic inference.** Phylogenetic relationships of the nine species was inferred by the maximum-likelihood method implemented in the ExaML program (v8.1.17) with 100 bootstraps<sup>18</sup>. We first used variants within coding sequences (CDSs) at the autosome and X chromosome. The protein-coding gene annotation information was downloaded from the University of California, Santa Cruz Genome Browser (<http://genome.ucsc.edu/>, bosTau6, UMD\_3.1). After removing CDSs not with lengths of multiples of 3, we obtained 19,111 and 831 CDSs for the autosome and X chromosome, respectively. Initial trees for the optimization were initially constructed using both Parimonator (v1.0.2) and the GTRCAT model in RAxML<sup>18</sup>. Phylogenetic inference was then carried out using the GAMMA model implemented in RAxML, and partitioned by the first two codon positions and third codon position for each CDS region. In addition, we obtained 100 bootstrap trees randomly generated by RAxML for each CDS. MP-EST software<sup>30</sup> was also used for phylogenetic analysis. To examine discordant coalescent histories among different genomic segments, we split the whole genomes into 100 kb non-overlapping windows and estimated maximum-likelihood trees in RAxML using the GTRGAMMA model. Windows containing > 50% missing bases in total or any individuals with > 40% missing data were removed from this analysis. After filtering, 25,127 and 1,489 windows remained for 100 kb segments on the autosomes and X chromosome, respectively, while 2,529 and 149 windows remained for 1 Mb segments. Numeration of trees and branch-length statistics were done using the ape (analyses of phylogenetics and evolution) package<sup>53</sup>. Finally, we summarized the window-based trees into a coalescent-based consensus species tree using ASTRAL (v4.11.2). The node support values were calculated from 1,000 resampling of window-based trees using the option '--gene-only -r 1000' in ASTRAL<sup>19</sup>.

**Inference of divergence time.** IM-CoalHMM, a coalescent hidden Markov model, was performed to estimate the effective population sizes and divergence times with gene flow among the bovine species<sup>54,55</sup>. The IM-CoalHMM implemented two models: an isolation model and an isolation with migration model. The IM-CoalHMM utilizes the genome alignments of two species to calculate the speciation time, average recombination rate and the ancestral effective population size with or without considering gene flow. We made use of the multi-alignment from the above phylogenetic analysis, randomly chose alleles for each heterozygous site to get haploid genomes and split them to ~1 Mb segments to perform the models independently. We discarded the windows with more than 10% missing rate or results where (1) the recombination rate was below 0.1 cM Mb<sup>-1</sup> or above 10 cM Mb<sup>-1</sup>, (2) the ancestral effective population size below 5,000 or above 1,000,000 or (3) the split time was below 5,000 years or above 10 million years.

**Demographic history.** We inferred the demographic history of nine species using the PSMC<sup>16</sup> method (<https://github.com/lh3/psmc>) based on the high-coverage genomes. The parameter setting and running procedure followed the best practice from the PSMC website. The bovine generation time ( $g$ ) = 5–7 years and neutral mutation rate per generation ( $\mu$ ) =  $2.2 \times 10^{-9}$  per base pair per year were based on a previous report<sup>56</sup>. We performed 100 bootstrapping simulations to estimate the variance of population size fluctuation. X chromosome data were excluded from the PSMC analyses due to different effective population sizes of autosomes and the X chromosome.

**Detection of genetic introgression.** To infer migration events among the nine species, we used TREEMIX v1.12 (ref. <sup>24</sup>) to construct a maximum-likelihood tree. In this analysis, some species only have one sample; therefore, the ‘-nossc’ option was added to turn off sample size correction. We grouped together 1,000 SNPs to account for linkage disequilibrium, and set yak as root. After building the maximum-likelihood tree, we respectively allowed 1–5 migration events in the tree to draw the maximum-likelihood graph and corresponding residuals. Blocks with 5,000 SNPs were applied to estimate the standard errors.

A windows-based Patterson’s  $D$  statistic (also called four-taxon ABBA-BABA test,  $D(P1, P2, P3, \text{Outgroup})$ ) with 50-kb-length non-overlapping windows was performed using the scripts from <http://datadryad.org/resource/doi:10.5061/dryad.j1rm6>.

To further examine for the genomic regions of potential introgression of three panels, (1) yak and Tibetan cattle, (2) zebu and gayal, (3) zebu and Bali cattle, we calculated the frequency of shared identical-by-descent (IBD) using the method described in a previous study<sup>58</sup>. First, all individuals were used to infer the IBD chunks using Beagle’s IBD detection algorithm<sup>57</sup>. Second, the IBD chunks between two individuals in presupposed donor and recipient population (that is, yak and Tibetan cattle, gayal and zebu, Bali cattle and zebu) or populations with more closed relationship in species tree (for example, Tibetan cattle and zebu, zebu and taurine) was extracted to count the frequency of shared IBD with a window size of 20 kb and a step size of 10 kb. Finally, the relative frequency of IBD (rIBD) was calculated by subtracting the two frequencies of the donor–recipient group and the closely related group. The windows with positive values of rIBD are potential introgressive regions. For more credibility, we only focused on the windows with top 0.01 rIBD values. We then applied a robust local-ancestry inference (RFMix<sup>55</sup>) to these three panels to screen for the potential sharing haplotype between donor and recipient. Furthermore, to validate the introgression, we calculated a window-based Patterson’s  $D$  statistic on the three panels using buffalo as outgroup and Weir and Cockerham’s pairwise  $F_{ST}$  (fixation index) between two populations using Vcftools<sup>59</sup>.

To detect the direction of introgression between zebu and gayal, between zebu and Bali cattle and between yak and Tibetan cattle, windows with significant rIBD values were merged within 1 Mb. Finally, a total of 260, 345 and 83 segmentations were obtained, respectively, and the neighbour-joining tree from the phased haplotype data for each segmentation was constructed using MEGA 6.0 (ref. <sup>60</sup>). By manually checking each tree, we could identify the direction of introgression of all segmentations (Supplementary Table 6). The introgression direction of each window with significant rIBD values is defined as the introgression direction of the segmentation the window locates.

**Probability of incomplete lineage sorting.** We calculated the probability of incomplete ancestral lineage sorting generating the introgressed haplotype according to the method in ref. <sup>40</sup>. Details of the method can be found in the primary paper<sup>40</sup>. Briefly, let  $k$  be the introgressed haplotype length,  $r$  be the recombination rate per generation per bp and  $t$  be the length of the two species’ branch since divergence. The expected length of a shared ancestral sequence is  $L = 1/(r \times t)$ . The probability of a length of at least  $k$  is  $1 - \text{GammaCDF}(k, \text{shape} = 2, r = 1/L)$ , where GammaCDF is the Gamma distribution function.

**Functional annotation of genes.** Expression analysis of genes has been described in a previous study<sup>50</sup>. Briefly, human gene expression data (Human U133A Gene Atlas) in 84 tissues or cells were downloaded from BioGPS (<http://biogps.org>) with GEO code GSE1133. The expressions of genes in cattle were retrieved from human expression data according to their Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC) gene symbol. Gene-enrichment analyses were performed using different methods including g:profiler<sup>61</sup>, DAVID<sup>62</sup> and Gorilla<sup>63</sup>.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** All the sequences have been deposited in the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) with the accession codes PRJNA396672, PRJNA427536 and PRJNA422979.

Received: 6 November 2017; Accepted: 20 April 2018;  
Published online: 21 May 2018

## References

- Soubrier, J. et al. Early cave art and ancient DNA record the origin of European bison. *Nat. Commun.* **7**, 13158 (2014).
- Gautier, M. et al. Deciphering the wisent demographic and adaptive histories from individual whole-genome sequences. *Mol. Biol. Evol.* **33**, 2801–2814 (2016).
- Wang, K. et al. The genome sequence of the wisent (*Bison bonasus*). *GigaScience* **6**, 1–5 (2017).
- Qiu, Q. et al. Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. *Nat. Commun.* **6**, 10283 (2015).
- Medugorac, I. et al. Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. *Nat. Genet.* **49**, 470–475 (2017).
- Daetwyler, H. D. et al. Whole-genome sequencing of 234 bulls facilitates mapping of monogenic and complex traits in cattle. *Nat. Genet.* **46**, 858–865 (2014).
- Qiu, Q. et al. The yak genome and adaptation to life at high altitude. *Nat. Genet.* **44**, 946–949 (2012).
- Newman, J. H. et al. Increased prevalence of EPAS1 variant in cattle with high-altitude pulmonary hypertension. *Nat. Commun.* **6**, 6863 (2015).
- Decker, J. E. et al. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc. Natl Acad. Sci. USA* **106**, 18644–18649 (2009).
- Hassanin, A., An, J., Ropiquet, A., Nguyen, T. T. & Couloux, A. Combining multiple autosomal introns for studying shallow phylogeny and taxonomy of Laurasiatherian mammals: application to the tribe Bovini (Cetartiodactyla, Bovidae). *Mol. Phylogenet. Evol.* **66**, 766–775 (2013).
- Buntjer, J. B., Otsen, M., Nijman, I. J., Kuiper, M. T. R. & Lenstra, J. A. Phylogeny of bovine species based on AFLP fingerprinting. *Heredity* **88**, 46–51 (2002).
- Ma, G. et al. Phylogenetic relationships and status quo of colonies for gayal based on analysis of cytochrome *b* gene partial sequences. *J. Genet. Genom.* **34**, 413–419 (2007).
- Baig, M. et al. Mitochondrial DNA diversity and origin of *Bos frontalis*. *Curr. Sci.* **104**, 115–120 (2013).
- Gou, X., Wang, Y., Yang, S., Deng, W. & Mao, H. Genetic diversity and origin of gayal and cattle in Yunnan revealed by mtDNA control region and SRY gene sequence variation. *J. Anim. Breed. Genet.* **127**, 154–160 (2010).
- Mei, C. et al. Whole-genome sequencing of the endangered bovine species gayal (*Bos frontalis*) provides new insights into its genetic features. *Sci. Rep.* **6**, 19787 (2016).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- Nielsen, R. et al. Tracing the peopling of the world through genomics. *Nature* **541**, 302–310 (2017).
- Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- Mirarab, S. et al. ASTRAL: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**, i541–i548 (2014).
- Liu, L., Yu, L. & Edwards, S. V. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302 (2010).
- Verkaar, E. L. C., Nijman, I. J., Beeke, M., Hanekamp, E. & Lenstra, J. A. Maternal and paternal lineages in cross-breeding bovine species. Has wisent a hybrid origin? *Mol. Biol. Evol.* **21**, 1165–1170 (2004).
- Hobolth, A., Christensen, O. F., Mailund, T. & Schierup, M. H. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet.* **3**, e7 (2007).
- Hedges, S. B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evol.* **32**, 835–845 (2014).
- Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
- Wecek, K. et al. Complex admixture preceded and followed the extinction of wisent in the wild. *Mol. Biol. Evol.* **34**, 598–612 (2017).
- Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
- Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
- Bosse, M. et al. Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat. Commun.* **5**, 4392 (2014).
- Hofer, E., Sobanov, Y., Brostjan, C., Lehrach, H. & Döchler, M. The centromeric part of the human natural killer (NK) receptor complex: lectin-like receptor genes expressed in NK, dendritic and endothelial cells. *Immunol. Rev.* **181**, 5–19 (2001).

30. Kao, H.-T. et al. A third member of the synapsin gene family. *Proc. Natl Acad. Sci. USA* **95**, 4667–4672 (1998).
31. Porton, B. et al. Mice lacking synapsin III show abnormalities in explicit memory and conditioned fear. *Genes Brain Behav.* **9**, 257–268 (2009).
32. Per, J. Behavior genetics and the domestication of animals. *Annu. Rev. Anim. Biosci.* **2**, 85–104 (2014).
33. Driscoll, C. A., Macdonald, D. W. & O'Brien, S. J. From wild animals to domestic pets, an evolutionary view of domestication. *Proc. Natl Acad. Sci. USA* **106**, 9971–9978 (2009).
34. Suzuki, G. et al. *Sept5* deficiency exerts pleiotropic influence on affective behaviors and cognitive functions in mice. *Hum. Mol. Genet.* **18**, 1652–1660 (2009).
35. Carneiro, M. et al. The genomic architecture of population divergence between subspecies of the European rabbit. *PLoS Genet.* **10**, e1003519 (2014).
36. Li, Y. et al. Artificial selection on brain-expressed genes during the domestication of dog. *Mol. Biol. Evol.* **30**, 1867–1876 (2013).
37. Wang, G.-d. et al. The genomics of selection in dogs and the parallel evolution between dogs and humans. *Nat. Commun.* **4**, 1860 (2013).
38. The-Bovine-HapMap-Consortium Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532 (2009).
39. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).
40. Huerta-Sánchez, E. et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* **512**, 194–197 (2014).
41. Lorenzo, F. R. et al. A genetic mechanism for Tibetan high-altitude adaptation. *Nat. Genet.* **46**, 951–956 (2014).
42. Bigham, A. W. & Lee, F. S. Human high-altitude adaptation: forward genetics meets the HIF pathway. *Genes Dev.* **28**, 2189–2204 (2014).
43. Storz, J. F., Scott, G. R. & Cheviron, Z. A. Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *J. Exp. Biol.* **213**, 4125–4136 (2010).
44. Chen, F. H. et al. Agriculture facilitated permanent human occupation of the Tibetan Plateau after 3600 B.P. *Science* **347**, 248–250 (2015).
45. Miao, B., Wang, Z. & Li, Y. Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the gray wolf from the Tibetan Plateau. *Mol. Biol. Evol.* **34**, 734–743 (2017).
46. Hsiao, J. J. & Fisher, D. E. The roles of microphthalmia transcription factor and pigmentation in melanoma. *Arch. Biochem. Biophys.* **563**, 28–34 (2015).
47. Kim, J. et al. The genome landscape of indigenous African cattle. *Genome Biol.* **18**, 34 (2017).
48. Qanbari, S. et al. Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* **10**, e1004148 (2012).
49. Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J. & Goddard, M. E. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet.* **6**, e1001139 (2010).
50. Kong, Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* **98**, 152–153 (2011).
51. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
52. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
53. Popescu, A.-A., Huber, K. T. & Paradis, E. ape 3.0: new tools for distance-based phylogenetics and evolutionary analysis in R. *Bioinformatics* **28**, 1536–1537 (2012).
54. Mailund, T. et al. A new isolation with migration model along complete genomes infers very different divergence processes among closely related great ape species. *PLoS Genet.* **8**, e1003125 (2012).
55. Mailund, T., Dutheil, J. Y., Hobolth, A., Lunter, G. & Schierup, M. H. Estimating divergence time and ancestral effective population size of Bornean and Sumatran orangutan subspecies using a coalescent hidden Markov model. *PLoS Genet.* **7**, e1001319 (2011).
56. Kumar, S. & Subramanian, S. Mutation rates in mammalian genomes. *Proc. Natl Acad. Sci. USA* **99**, 803–808 (2002).
57. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
58. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).
59. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
60. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
61. Reimand, J. et al. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89 (2016).
62. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2008).
63. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).

### Acknowledgements

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No XDB13020600), National Natural Science Foundation of China (91731304, 31321002, 31272418, 31561143010), the Chinese 973 program (2013CB835204), Animal Branch of the Germplasm Bank of Wild Species (GBOWS) and the Program for Changjiang Scholar and Innovation Research Team in University (IRT\_15R621). D.-D.W. was supported by the Youth Innovation Promotion Association, Chinese Academy of Sciences.

### Author contributions

Y.-P.Z., D.-D.W., R.N. and Q.Z. lead the project, and designed and conceived the study. D.-D.W., S.W., X.-D.D. and R.N. prepared the manuscript. D.-D.W., S.W., X.-D.D., Y. Z., Y.L. and M.-S.W. performed the data analysis. J.M.W., M.T. and O.F. performed some sampling and experiments. All authors read the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41559-018-0562-y>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to D.-D.W. or R.N. or Q.Z. or Y.-P.Z.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a | Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated
- Clearly defined error bars  
*State explicitly what error bars represent (e.g. SD, SE, CI)*

*Our web collection on [statistics for biologists](#) may be useful.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data collection has been described in the Methods section.

Data analysis

All the data analyses have been described in the Methods section.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the sequences have been deposited in the SRA (<https://www.ncbi.nlm.nih.gov/sra>) with Accession ID (PRJNA396672, PRJNA427536, and PRJNA422979).



## Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences  Behavioural & social sciences  Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/authors/policies/ReportingSummary-flat.pdf](https://www.nature.com/authors/policies/ReportingSummary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We have described samples in the methods section and supplementary table 1 and 3. More than 100 genomes were generated in this study. For phylogenetic analysis of different species, one individual with high coverage is enough for each species. For population genetic analysis, like for genetic introgression, we used $\geq 5$ individuals for each population.
Data exclusions	No data was excluded as the quality of genome sequencing data was good enough for all samples.
Replication	No experiment performed.
Randomization	Grouping was based on geographical location and species, and hence we didn't need to perform randomization analysis.
Blinding	We did not need any blinding.

## Reporting for specific materials, systems and methods

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Unique biological materials
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	N/A
Wild animals	Animal sampling was carried out in accordance with the animal experimentation guidelines and regulations of the Kunming Institute of Zoology, and was approved by the Institutional Animal Care and Use Committee of the Kunming Institute of Zoology. The cell line of one gayal individual kept in Kunming Cell Bank of Kunming Institute of Zoology, was chosen for high depth genome sequencing. Blood samples of another 22 gayal individuals were collected from Gaoligong mountain region in Yunnan province in China, and India and Bengal. One gaur individual used for genome sequencing was collected from Xishuangbanna, Yunnan, and its species status were validated by morphology and mtDNA sequence. Blood samples of 8 bali cattle were from Indonesia. The above samples were from the Animal Branch of the Germplasm Bank of Wild Species (GBOWS), Chinese Academy of Sciences. One American Bison individual was sampled from Beijing Zoo. Blood Samples of Chinese yellow cattle were collected from different original birthplaces in China. The bloods of the Tibetan cattle used for whole genome sequencing were collected from Gongbogyamda region in Tibet. The analyzed 11 wisent individuals were culled in the Knyszyn Forest, north-eastern Poland according to the Ministry of Environment permission.
Field-collected samples	N/A