Yi et al. BMC Genomics (2020) 21:108
https://doi.org/10.1186/s12864-020-6495-2

BMC Genomics

RESEARCH ARTICLE Open Access

Whole-genome sequencing of wild Siberian musk deer (*Moschus moschiferus*) provides insights into its genetic features

Li Yi[1†], Menggen Dalai[2*†], Rina Su[1†], Weili Lin[3], Myagmarsuren Erdenedalai[4], Batkhuu Luvsantseren[4], Chimedragchaa Chimedtseren[4*], Zhen Wang[3*] and Surong Hasi[1*]

Abstract

Background: Siberian musk deer, one of the seven species, is distributed in coniferous forests of Asia. Worldwide, the population size of Siberian musk deer is threatened by severe illegal poaching for commercially valuable musk and meat, habitat losses, and forest fire. At present, this species is categorized as Vulnerable on the IUCN Red List. However, the genetic information of Siberian musk deer is largely unexplored.

Results: Here, we produced 3.10 Gb draft assembly of wild Siberian musk deer with a contig N50 of 29,145 bp and a scaffold N50 of 7,955,248 bp. We annotated 19,363 protein-coding genes and estimated 44.44% of the genome to be repetitive. Our phylogenetic analysis reveals that wild Siberian musk deer is closer to Bovidae than to Cervidae. Comparative analyses showed that the genetic features of Siberian musk deer adapted in cold and high-altitude environments. We sequenced two additional genomes of Siberian musk deer constructed demographic history indicated that changes in effective population size corresponded with recent glacial epochs. Finally, we identified several candidate genes that may play a role in the musk secretion based on transcriptome analysis.

Conclusions: Here, we present a high-quality draft genome of wild Siberian musk deer, which will provide a valuable genetic resource for further investigations of this economically important musk deer.

Keywords: Wild Siberian musk deer (*Moschus moschiferus)* genome, De novo assembly, Genetic features, Musk secretion

Background

Musk deer (*Moschus,* Moschidae) are small hornless Pecora ungulates, occurring commonly at mountains and forests of central Asia, belong to Cetartiodactyla, Ruminantia [1, 2]. At present, musk deer comprise seven species, including Anhui musk deer (*M. anhuiensis*), forest musk deer (*M. berezovskii*), Alpine musk deer (*M. chrysogaster*), black musk deer (*M. fuscus*), Himalayan musk deer (*M. leucogaster*), Kashmir musk deer (*M. cupreus*) and Siberian musk deer (*M. moschiferus*) [3–5]. This species is shy, timid, cautious, sensitive, crepuscular and nocturnal, and likes to be alone and does not live in groups [6, 7]. Musk deer inhabits a fairly fixed area throughout its life and rarely changes [1]. Musk deer are famous for secretion musk from the musk gland (only in males), which with specific odor and color, and appear to serve for attracting the females and mark territory [8–10]. Moreover, its secretion is widely used in traditional medicines and perfume industries since the fifth century, because of its unique fragrance and its significant anti-inflammatory and anti-tumor roles, as well as its effects on the human central nervous and cardio-cerebral-vascular systems [11–15]. The musk is regarded as one of the most valuable of all animal scents, even more, expensive than gold [16]. However, the population of musk deer has dramatically decreased due to illegal

* Correspondence: mengendalai@sina.com; Chi.chimedragchaa@yahoo.com; zwang01@sibs.ac.cn; surong@imau.edu.cn
†Li Yi, Menggen Dalai and Rina Su contributed equally to this work.
[2]Affiliated Hospital of Inner Mongolia Medical University, Hohhot 010050, China
[4]Institute of Traditional Medicine and Technology, Ulaanbaatar, Mongolia
[3]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[1]Inner Mongolia Agricultural University Key /Laboratory of Clinical Diagnosis and Treatment Technology in Animal Disease, Ministry of Agriculture and Rural Affairs , Hohhot 010018, China

poaching for their meat and musk, exploitation of natural resources, trade, infrastructure construction, fast urbanization [16–19]. Therefore, six species being listed as endangered and one as vulnerable by the International Union for Conservation of Nature (IUCN 2017) [20]. All of them are also listed in Category I of the State Key Protected Wildlife List of China [21].

In recent years, there has been significant progress in the studies of musk deer ecology, taxonomy, evolution history by paleontological, morphological, ecological and ethological and molecular analysis [22–40]. The musk composition and secretory mechanism of musk have been explored by various aspects, including microsatellite, mtDNA marker, and transcriptome sequencing data [41–46]. Besides, the gut microbial communities have been illustrated by metagenome sequencing [9, 47, 48]. Unfortunately, genomic resources of the species are rarely limited. Recent work has provided the first complete genome sequence of the forest musk deer [49]. Siberian musk deer is one of the seven species, widely occurs in Korea, Mongolia, Russia, China, Kazakhstan, Kyrgyzstan, Nepal, and Vietnam [50]. However, the population size of Siberian musk deer is dwindling rapidly by the same reasons as other musk species, and they have been categorized as Vulnerable on the IUCN Red List [51]. As a result of the extinction crisis of Siberian musk deer and economic and medical value of its musk, understanding the genetic basis and features, environment adaptions, and the musk secretion mechanism is necessary. However, the whole-genome sequencing of Siberian musk deer has not been performed, and their potential value has yet to be discovered.

In this study, we perform high-quality whole-genome sequencing of three wild Siberian musk deer (WSMD) from Mongolia, and transcriptome sequencing of one mixture of tissue from a naturally died female WSMD. These genomic and transcriptome analyses provide evidence of Siberian musk deer genetic features and musk secretion.

Results

Genome sequencing, assembly, and evaluation

Genomic DNA extracted from a female WSMD was subjected to shotgun sequencing using the Illumina Hiseq Xten platform. We prepared 19 pair-end libraries spanning several insert sizes (from 250 bp to 10 kb, Additional file 1: Table S1) to generate short pair-end reads. A total of 326.64 Gb (102.97× coverage) raw data were generated from all constructed libraries, from which 283.22Gb of clean data was obtained after removal of low-quality reads, duplicates, adaptors, and reads with more than 10% N bases. The genome assembly was estimated to be approximately 3.10Gb using K-mer = 41 analysis [52], which was slightly bigger than that of the forest musk deer (2.72Gb) [49]. The assembly consisted of 13,344 scaffolds (≥1 kb) with an N50 of 7, 955,248 bp and 165,764 contigs with an N50 of 29,145 bp (Table 1). The genome-wide proportion of G + C was 41.96% (Additional file 1: Table S2). By mapping the short-fragment libraries to the assembled genome with BWA mem (v0.7.12), 98% reads were mappable (93.16% properly paired), indicating a highly accurate assembly (Additional file 1: Table S3).

Subsequently, we used Benchmarking Universal Single Copy Orthologs BUSCO (BUSCO, V2.0) [53] to assess the completeness of the genome assembly. BUSCO results showed that 93.30% of the 4104 mammalian single-copy orthologues were complete (Additional file 1: Table S4). Furthermore, we downloaded the musk gland and heart RNA-sequencing data (SRA accession: SRR2098995, SRR2098996, and SRR2142357) of forest musk deer from the National Center for Biotechnology Information (NCBI) and mapped to the genome assembly using STAR [54]. The alignment coverage of expressed sequences was ranged from 35 to 75% in the genome assembly. These assessments indicated that our assembly with a high level of completeness. Hence, a high-quality assembly of WSMD is provided here, rendering it a valuable source for studying genome structure and evolution.

Genome comparison of Siberian musk deer and forest musk deer

We compared the genome assembly of the Siberian musk deer and forest musk deer recently reported by Fan et al. [55] (Additional file 3: Table S17). The continuity of our assembly was remarkably increased compared with that of the forest musk deer genome assembly, particularly in regard to the scaffold N50 (7.95 vs 2.85 Mb) and scaffold number (13,344 vs 79,206). We then aligned the two genome assemblies using mummer4 [56]. At least 2.16 Gb (80.16%) of our assembly could be aligned with that of the forest musk deer, most of which (2.13 Gb) were one-to-one alignment (Additional file 3: Table S17). The average identify of the alignments was 98.74%, suggesting close relationship between the two species.

Repetitive sequences and gene annotation

Using a combination of homology-based (Ruminant and mammal) and de novo methods, we identified transposable elements (TEs) and other repetitive elements in the WSMD genome. We estimated 44.44% of our genome to be composed of repetitive elements using a combination of homology-based and de novo approaches (Additional file 1: Table S6). The de novo method identified 38.60% of the genome as repetitive, whereas the homology-based method predicted more (44.27 and 43.67%, respectively). The repeat element landscape of WSMD mostly consists of retrotransposons,

Table 1 Statistics of the genome assembly (The minimum size of contigs for reporting is 1 Kb)

Statistics	Size of contigs (bp)	Size of scaffolds (bp)	Number of contigs	Number of scaffolds
Total	2,435,924,293	2,703,175,379	165,764	13,344
Max	498,578	35,164,634	–	–
N50	29,145	7,955,248	23,516	93
N60	22,936	6,419,411	32,950	130
N70	17,477	4,597,695	45,098	179
N80	12,390	3,185,516	61,590	250
N90	7170	1,717,083	87,009	365

including long interspersed elements (LINES), short interspersed elements (SINES) and long terminal repeats (LTRs). Among them, LINES represented the most predominant type of repeat sequences, occupying 30.37% of the genome, while the other repeat elements (SINE and LTR) comprised 4.78 and 4.42%, respectively. DNA transposons were particularly rare, forming only 2.27% of the genome.

Gene annotation of the WSMD genome was conducted using several approaches, including ab initio, homology-based and transcript-based methods (Additional file 1: Table S4, Additional file 1: Table S8, and Table S9). Gene models generated from all the methods were integrated by EVM (EvidenceModeler) to build a consensus gene set for the WSMD genome. The final gene set is a union of a gene predicted by Genewise and supplemented with EVM that removed the genes only predicted by ab initio. In total, 19,363 non-redundant protein-coding genes were annotated in the WSMD genome (Additional file 1: Figure S1 and Table S4), which is less than the predicted gene numbers of forest musk deer (24,352 genes) [49]. The BUSCO evaluation showed that 99.1% of genes were identified as complete and fragmented, with genes that were considered missing in the gene set. The BUSCO results showed that our gene predication was more complete (Additional file 1: Table S4). Alongside this, we also provide the length of genes in Additional file 1: Table S8.

Evolutionary analysis and phylogeny

Compared with protein-coding genes of nine other species (goat, sheep, cattle, white-tail deer, pig, horse, dog, human and mouse), we found 17,336 orthologous of WSMD that were shared by at least one species (Additional file 1: Table S11), and 14,936 orthologous shared by human, cattle, white-tailed deer and WSMD. There were 167 gene families specific for WSMD (Fig. 1a). Further, we constructed a phylogenic tree using MEGA based on fourfold degenerate codon sites extracted from single-copy orthologous genes identified by TreeFam (Additional file 1: Table S10 and Fig. 1b). The phylogenic tree was indicated that the WSMD and the Cattle were within a subclade, which was most likely derived from a common ancestor ~ 22 Ma ago (Mya) (Fig. 1b).

Gene gains and losses are one of the primary contributors to functional changes [57]. To obtain greater insight into the evolutionary dynamics of the genes, we determined the expansion and contraction of the gene orthologue clusters among these ten species. We found 27 gene families were expended, whereas 208 gene families were contracted in WSMD (Fig. 1b), which might indicate that losses of function might have an important role in functional evolution. The expanded genes were significantly enriched to several pathways associated with fat digestion and absorption, glycerolipid metabolism, and amino acid metabolism (Additional file 1: Figure S3). The contracted gene families were enriched in pathways related to the sensory system, immune system and infectious diseases (Additional file 1: Figure S4). The corresponding GO terms were shown in Additional file 1: Table S13 and Additional file 1: Table S14.

Positive selection genes and functional enrichment

To observation of positively selected genes (PSGs) in the WSMD genome raises the question of what signatures of selection are to be found in the extant genomes. A total of 184 PSGs were identified by the branch-site likelihood ratio test, and then mapped them to KEGG pathways and GO categories (Fig. 3b and Additional file 1: Table S15). It was shown that those PSGs are enriched in 8 pathways associated to metabolism (amino sugar and nucleotide sugar metabolism, and lysine degradation), cellular processes (peroxisome and p53 signaling pathway), organismal systems (insulin secretion, pancreatic secretion, mineral absorption and bile secretion), and environmental information processing (cGMP-PKG signaling pathway) (Fig. 3b). GO classification showed that those PSGs are enriched in these functional categories, including cellular components (Cell part, Cell, Intracellular, Intracellular part, Organelle, Membrane-bounded organelle, Cytoplasm, and Intracellular orangelle), biological processes (Cellular process, Single-organism process, single-organism cellular process, and metabolic process) and molecular functions (binding and protein

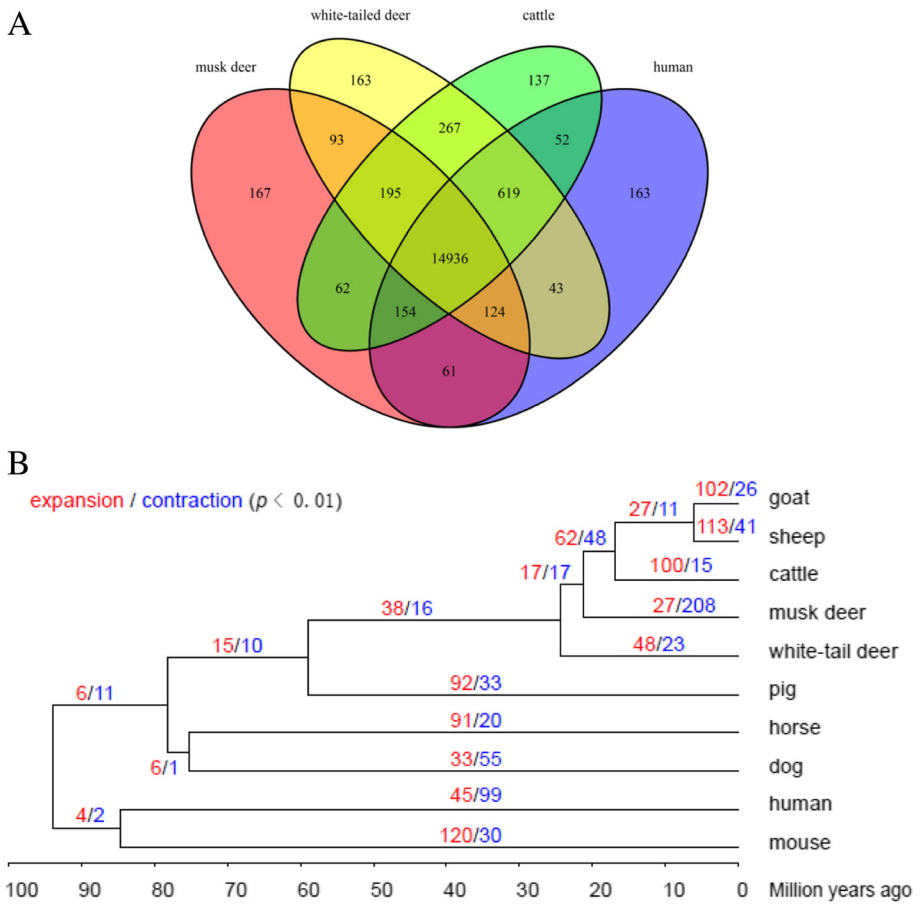

Fig. 1 **a** The Venn diagram shows the number of orthologs shared among musk deer and other representative mammals. **b** Phylogeny and gene family size evolution. The phylogenetic tree is constructed based on four-fold degeneration sites among single-copy orthologs with the neighbor-joining method. The timelines indicate inferred divergence times among the species based on the molecular clock. The number of significantly expanded (red) and contracted (blue) gene families (branch-specific *p*-value < 0.01) are shown at each branch

binding)(Additional file 1: Table S15). Musk deer is a nocturnal mammal with sensitive hearing, smell, and sight for its locating food and avoiding predators in darkness [6, 58]. We found 12 PSGs (*ATR, EYA1, NEK4, XRCC1, TRIP12, CNOT8, TOPBP1, PLA2R1, ZFYVE26, UIMC1, MCM10*, and *FBXO18*) were involved in DNA damage and repair categories. This finding possibly avoids the Siberian musk deer from the DNA damage caused by UV radiation and hypoxia in high-altitude environments. Thirty-five PSGs were involved in stress response categories. Among 35 PSGs, 7 genes also associated with the nervous system. In addition, we also observed 2 PSGs (*NR0B2* and *MED25*) distributed in retinoid X receptor binding (GO:0046965, corrected *p*-value = 0.0033).

Genomic diversity and demography inference

To understand the genetic diversity and demographic history in Siberian musk deer, we sequenced two additional WSMD (one male:s190119001, and one female: s180119002) genome generated a total of 78.27Gb raw data, and for each individual nearly 98% of reads mapped to the reference genome assembly with 8.83× average coverage (Additional file 1: Table S3). We performed single-nucleotide polymorphism (SNP) calling and identified 4.81 million (M) SNPs from three individuals, and the Ts/Tv ratio for SNPs was 1.84 (Additional file 1: Table S11). For each individual, 2,420,974, 2,002, 344 and 2,337,725 heterozygous single-nucleotide polymorphisms (SNPs), respectively, along the assembled Siberian musk deer genome (Additional file 1: Table S11).

Historical fluctuations in effective population size (N_e) for the three individuals were constructed with the help of the Pair-wise Sequentially Markovian Coalescent (PSMC) model [59], three genomes returned concordant PSMC population trajectories that with three declines and two expansions (Fig. 2). The three genomes returned concordant PSMC population trajectories, suggesting no population structure in the species. The first decline in N_e was inferred to have occurred approximately 0.70 Mya, coinciding with the Naynayxungla glaciation (0.78–0.50Mya),

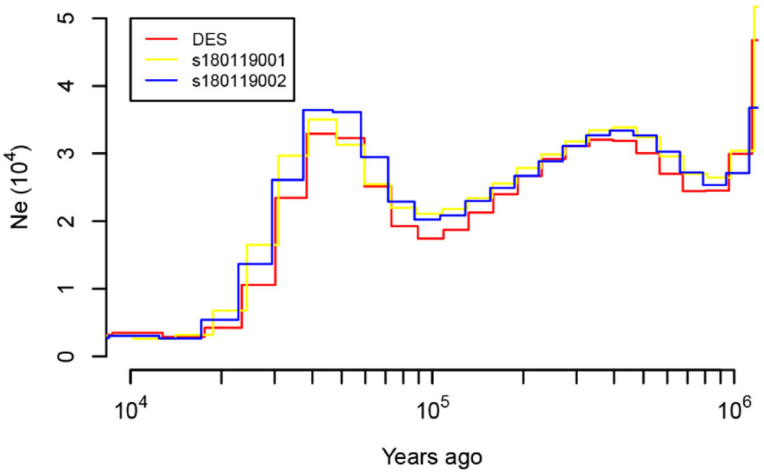

Fig. 2 Historical effective population size inferred by PSMC. Each line represents one individual. The result is scaled using a generation time of 5 years and a mutation rate of 1.1 × 10–8 per site per generation

which was the most extensive glaciation during the Quaternary Period [60–62]. After the first decline, the N_e for Siberian musk deer recovered and peaked at ~ 0.30 Mya, during the Penultimate glaciation (0.30–0.13 Mya) [60–62]. The cold-climate interval and rising sea level at this stage could have contributed to a population expansion because an increase in grassland was likely under such environmental conditions [63].

The second declines occurring between 0.20 to 0.09 Mya, was detected towards and end of the interglacial period (0.13–0.07 Mya), which presented environmental conditions similar to that of the present [64]. The uplift of the Tibetan Plateau, which caused aridification, and desertification that was dramatically enhanced in the middle Pleistocene age, which reduced the habitat of the musk deer, resulting in a decline of population size [40, 65]. The Siberian musk deer population size then recovered again between 0.05–0.03 Mya during the greatest lake period (0.03–0.04 Mya) because the glaciations were less extended, weather became warm and the forest had expanded that could have contributed to the population expansion [60–62]. Subsequently, a sharp decline in N_e for Siberian musk deer coincided with the extreme cooling climate during the last glaciation (~ 20,000 years ago), it is likely that Siberian musk deer suffered from the effects of climate change, over-hunting, and habitat loss.

RNA sequencing of mixture tissue

To evaluate the genome completeness, gene annotation and excavating genes related to musk secretion, we sequenced the transcriptome of a mixture tissue (including liver, kidney, lung, heart, skin, and stomach) which collected from a female Siberian musk deer. The Illumina high-throughput next-generation RNA sequencing resulted in 22,927,488 raw reads generated from a mixture of tissue. After removing low-quality sequences, a total of 17,323,786 clean reads were generated. Over 68% of clean reads mapped to the assembly using STAR, suggesting that the majority of transcribed genes are present (Additional file 1: Table S9). After the cufflinks assembly generated 44,271 genes and 61,96 isoforms (Additional file 1: Table S12). Another notable result is that approximately 56% of the counted reads were mapped to exonic regions of a unique gene, and a small proportion of reads (5.8%) were defined as unannotated, which probably contain novel genes and exons (Additional file 1: Table S12).

Differentially expressed genes and functional enrichment analysis

We explored the differences among the transcriptomes among the musk gland, heart, and mixture tissue. A total of 189 genes were identified to be upregulated differentially expressed genes (DEGs) in the musk gland, as compared with the same genes in heart and mixture tissues (FDR < 0.05, $\log_2$-fold change < − 5) (Fig. 3a). There were 78 DEGs that were specifically expressed in the musk gland.

The Go annotation classified the DEGs into 3 categories: molecular functions (MF), cellular components (CC) and biological processes (BP) (Additional file 2: Table S16). Molecular functions included genes mainly involved in binding (112genes, GO:0005488) and protein binding (81genes, GO:0030414). Genes related to cellular components (CC) were primarily cell (136 genes, GO:0005623), cell part (135 genes, GO:0044464), intracellular (117 genes, GO:0005622), intracellular part (112 genes, GO:0044424), organelle (106genes, GO:0043226) and membrane-bounded organelle (102genes, GO:0043227). In addition to the largest proportion of cell-related components, the organelle occupies an

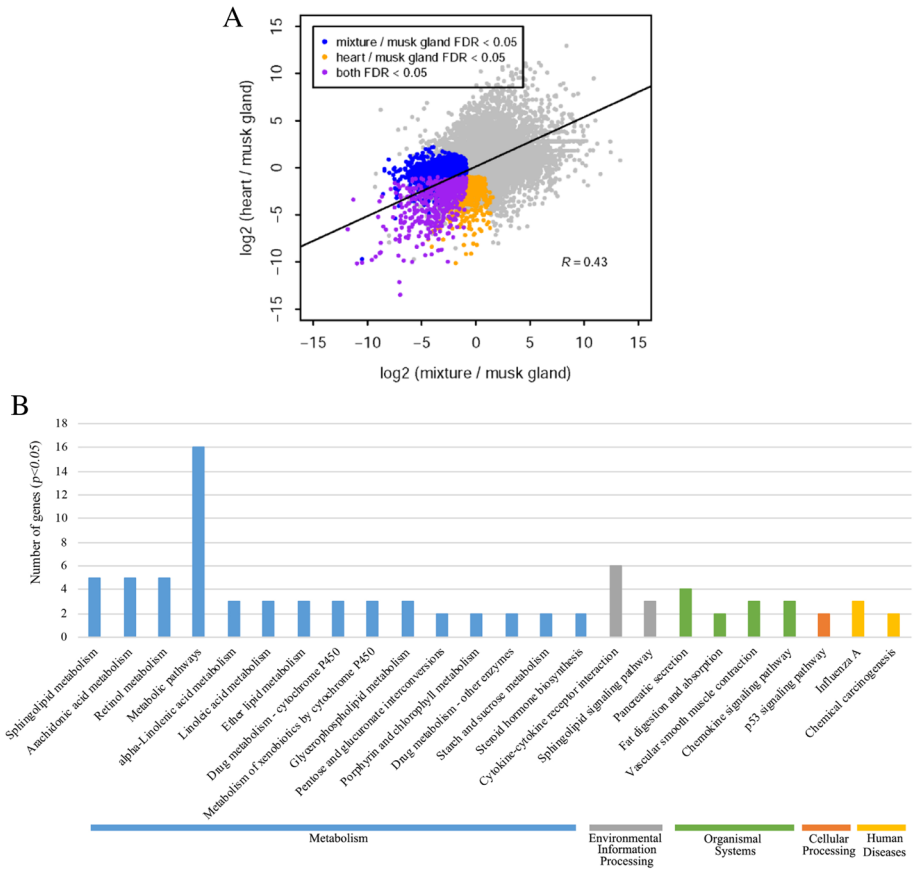

Fig. 3 **a** Log_2-fold change in normalized counts between the mixture tissue and musk gland, as well as between the heart and a musk gland. The points represent genes, and genes with significant over-expression (FDR < 0.05) in the musk gland are colored. A cutoff of log_2-fold change < − 5 in both comparisons is also applied to screen genes with high expression specifically in the musk gland. **b** KEGG pathway enrichment of DEGs in the Siberian musk deer. The x-axis shows the KEGG functional categories, while eh the number of genes in each category is plotted on the y-axis

important proportion. This result indicates that the molecular components involved in the physiological activities of the siberian musk deer are not only concentrated in cells but also widely distributed in organelles, and play an important role. In the biological process part (BP), a total of 814 terms (7148 genes) are involved, of which the single-organism process (120 genes, GO:0044699) accounts for the largest proportion, followed by metabolic process (98 genes, GO:0008152) and cellular process (118 genes, GO:0008152). Also, it also includes response to the stimulus (71 genes, GO:0050896), cellular response to stimulus (50 genes, GO:0051716), and many categories related to metabolism. This result is consistent with the biological characteristics of the siberian musk deer, which can especially explain its survivability under extreme conditions and its obvious response and alertness to external stimuli [19, 40, 66]. The distribution of GO annotations in different functional categories indicated a substantial diversity of DEGs.

We identified the biochemical pathways based on the DEGs detected in FMD. The KEGG annotation of the DEGs suggested that they were distributed in 24 pathways related to metabolism (59 genes), environmental information processing (9 genes), organismal systems, celluar processing (12 genes), and human diseases (5 genes), (Fig. 3b). Among the identified functional categories of metabolism, metabolic pathways (16 genes) were highly represented, followed by sphingolipid metabolism (5 genes), arachidonic acid metabolism (5 genes), and retinol metabolism (5 genes). In the environmental information processing, mainly has the cytokine-cytokine receotor interaction and sphingolipid signaling pathway. Organismal systems included functions mainly involved in pancreatic secretion, fat digestion and absorption,vascular smooth muscle contraction and chemokine signaling pathway. About human diseases involved in Influenza A and chemical carcinogenesis.

Genes related to musk secretion

To obtain greater insight into the mechanisms of musk secretion, it was crucial to understanding their metabolic processes and the corresponding pathways and genes. Thus, we screened the GO terms and KEGG pathways

associated with the musk compounds and metabolism (Fig. 3b and Additional file 2: Table S16). There were 21 DEGs that were closely involved in related pathways and terms, including steroid biosynthesis and transport (map 00140, GO:0015918 and GO:0036314), terpenoid and diterpenoid metabolic process (GO:0006721 and GO:0016101), hormone response and metabolic process (GO:0009725, GO:0034754, GO:0010817 and GO:0042445), cholesterol transport (GO:0030301) and cytochrome P450 metabolism pathway (map 00980). Among them, *UGT1A4* and *SULT2B1*was annotated in the steroid hormone biosynthesis (map 00140). *UGT1A4* is regarded as the main enzyme that catalyzes N-glucuronidation of various endogenous compounds (eg., steroids and thyroid hormones, fatty acids, bile acids, and bilirubin), as well as of xenobiotics including drugs and foreign compounds [66–68]. *SULT2B1* is a member of the large cytosolic sulfotransferase superfamily that is engaged in the synthesis and metabolism of steroids [69]. It further belongs to the *SULT2* family of enzymes that are primarily involved in the sulfoconjugation of neutral steroids and sterols [70]. It further belongs to the *SULT2* family of enzymes that are primarily involved in the sulfoconjugation of neutral steroids and sterols [70]. Steroid biosynthesis is catalyzed by a suite of enzymes including members of the cytochrome P450 (CYP), short chain dehydrogenase (SDR), and aldo-keto reductase (AKR) superfamilies [71]. CYP2B6, a member of CYP groups of enzyme, was annotated in cytochrome P450 metabolism pathway that participated in the metabolism of arachidonic acid, lauric acid and steroid hormones including testosterone, estrone and 17β-estradiol [72, 73]. It might hint that these genes played significant roles in musk formation and secretion.

Discussion

In this study, we performed a draft genome of wild Siberian musk deer using next generation sequencing technology. The final assembly of WSMD genome is 3.10 Gb with a contig N50 of 29,145 bp and a scaffold N50 of 7,955,248 bp, accounting for about 87.98% of the whole genome with coverage over 30x. Compared with the genome of the forest musk deer, the present assembly of WSMD has larger genome size, contig N50 and scaffold N50 lengths [49]. The results came from BWA mem, BUSCO and STAR analyses indicated that our assembly with high level of accuracy and completeness, and enough for the following analyses.

We observed that TEs occupied 44.44% of the whole assembly, which was lower than those of cattle (45.14%) and human (46.07%), but larger than those of pig (38.66%), mouse (40.53%) (Additional file 1: Table S7) and forest musk deer (42.05%) [49]. A total of 19,363 non-redundant protein-coding genes was annotated in WSMD genome, which was less than the predicted gene numbers of forest musk deer (24,352 genes) [49]. Moreover, we constructed a phylogenic tree was indicated that the WSMD and the Cattle were within a subclade, which was most likely derived from a common ancestor ~ 22 Ma ago (Mya). Moschidae shows a mixture of Bovidae and Cervidae characteristics [74, 75] so that its phylogenetic status has been strongly debated. The taxonomy of Moschidae as a separate family has been elucidated by the combination of paleontological, morphological, ecological and ethological and molecular analysis [22–32]. However, Moschidea is a sister group of Bovidae or of Cervidae, has obtained different results in different analyses [28, 31–34]. Previous studies on phylogenetic analysis based on whole-genome sequences revealed that forest musk deer as more closely related to Bovidae than to Cervidae, which is consistent with the results of the present study [35, 36, 76]. Historically, the fossil records and some molecular phylogenetic studies regarded Siberian musk deer WSMD as the primitive species in Moschus [25, 37, 38]. However, the divergence time between WSMD and cattle was latter than the time (~ 27.3Mya) at which forest musk deer divided with Bovidae [39]. Pan et al. (2015) have also reported that Siberian musk deer occurs latter than Alpine musk deer branches on the phylogenetic tree based on complete mtDNA analysis [40]. These results were suggested that Siberian musk deer was not the most primitive musk deer.

To adapt to environments of the high mountain forests, Siberian musk deer may have been formed some characteristics under natural selection. It is worth noting that musk deer has sensitive smell and hearing to locating food in darkness. Therefore, it is interesting to uncover evolutionary evidence for its adaptation by comparative analysis. By comparison with nine other species, we found 27 gene families were expended, whereas 208 gene families were contracted in WSMD. Studies have shown that due to the small body size and small appetite musk deer could not get enough food in one time to obtain more energy [77]. Therefore, musk deer often choose high-energy and digestible good, especially in the cold winter and spring when the food is scarce [78]. We found that the expansion gene families were significantly enriched in energy metabolism pathways and GO terms which might help Siberian musk deer to optimize their energy storage and production in the forest. The contraction gene families were most prominent in olfactory transduction pathway (Additional file 1: Figure S4). It might be attributed possibly to musk deer adaptation to the cold and high-altitude environment (1000-4200 m) where food sources and odorants are limited and diffused slowly, and the interactions between odorants and receptors weakened

[79, 80]. Similar results have been obtained in some high plateau animal genome studies, such as avian [81], wild boars [82], hot-spring snake [83] and Tibetan chicken [84]. Moreover, we observed 12 PSGs and 2 PSGs were involved in DNA damage and retinoid X receptor binding categories, respectively. These categories seem to be help musk deer living at high altitudes avoid high levels of ultraviolet radiation and forage in darkness. The previous study based on the forest musk genome has identified eight PSGs genes enriched in the phototransduction pathway and retinol metabolism pathways [35]. Our results and theirs did not have overlap candidate genes. Taken together, these results provide evidence for musk deer to adapt to the environments. In addition, the demographic historical pattern was similar with sheep [85], panda [86], bear [87] and Yak [88], suggesting that global glaciations and severely cold climates at this time had substantial evolutionary impact on the population size of terrestrial mammals [89].

As we known that musk deer is famous for secreting musk. Musk is a secreted external hormone or information compound that is stored in musk scent glands of the males of species within the family Moschidea [90]. Like those produced by muskrat (*Ondatra zibethicus L.*) and small Indian civet (*Viverricula indica*), the musk that musk glands of males secrete during the rutting season is not only an important pheromone for attracting females and mark territory, but also precious materials for pharmaceutical and perfume industries [91, 92]. Chemical analysis indicated that musk contains various ingredients, such as muscone, steroid compounds (cholestanol, cholesterol, and a number of the androstane derivatives), macrocyclic ketone, waxes, muscopyridine, and hydroxymuscopyridines, etc. [9, 10]. Fan et al. (2018) [93] has reported that testosterone and estradiol may play a major role in determining musk composition during the early stage of musk secretion but not during the course of musk maturation, which suggests that musk secretion may be promoted by increases in sex hormones in June. Other studies have shown that testosterone plays an important role in the seasonal development of musk glands [92], and oxytocin may regulate the function of muskrat scented glands by the locally expressed receptors [94]. Studies based on transcriptome [42, 43] and genetic analysis [35] have shown that a considerable number of genes involved in musk metabolism pathways, such as steroid biosynthesis, flavone and flavonol biosynthesis, terpenoid backbone biosynthesis, aldosterone-regulated sodium reabsorption was played a significant role in musk secretion. In this study, we identified 21 up-regulation DEGs which were closely associated with metabolism and response of steroid, terpenoid and hormone which were coincident with the previous reports [35, 42, 43]. Although there have been several studies on the secretion of musk, the genetic mechanisms of musk secretion are still poorly understood. Thus, further studies are needed to explore the musk secretion.

Conclusion

Siberian musk deer once inhabited most of Asia, but today they are sharply declining and being endangered status due to overharvesting, natural disaster, and diseases. In this study, we report the first whole genome sequencing, assembly, and annotation of the wild Siberian musk deer. Comparative genomic analyses characterized genetic diversity, the population structure of Siberian musk deer, and even the genetic features associated with energy metabolism and adaptations in cold and high-altitude environments. The candidate genes identified in this study may be useful for understanding the mechanism of musk secretion. Collectively, the draft genome will provide a valuable resource for studying essential developmental processes in the musk deer, investigation evolution and providing the molecular breeding of this economically important species.

Methods

Sample collection, DNA and RNA isolation

Whole blood samples from two female and one male WSMD (DES, s190119001, s180119002, respectively) living at the Siberian musk deer breeding farm in Gachuurt village (45 km from Khan Khentii Strictly Protected Area), Khentii aimag, Mongolia, was collected during a routine veterinary examination. A mixture tissue sample (including liver, kidney, lung, heart, skin, and stomach) was collected from a female Siberian musk deer the naturally died. The genomic DNA was extracted from blood samples with Qiagen DNA blood and tissue kit (Qiagen, Velencia, USA), and the total RNA was isolated from mixture tissue using TRIzol reagent (Qiagen, Hilden, the Netherlands) following the manufacturer's protocols.

All collected samples were approved by the Mongolian Musk Deer Breeding Center and completed with the help of staff.

Genome sequencing and assembly

The whole genome shotgun strategy based on the Illumina HiSeq Xten platform was used to sequence the genome of one female WSMD (DES). In total, 19 paired-end libraries with insert sizes of 250 bp, 450 bp, 2 kb, 5 kb, and 10 kb were constructed and sequenced with the 2 × 150 bp mode. (Additional file 1: Table S1). For libraries with insert sizes > 1 kb, the DNA fragments were circularized by self-ligation. The raw reads were cleaned according to: 1) trimming adaptors; 2) filtering reads with N% > 0.1; 3) filtering reads with low-quality (score < 5) bases% > 0.2. Duplicated reads were also filtered. To check

the quality of the libraries, the reads were mapped to an assembly of a close species (*Cervus elaphus*) with BWA mem (v0.7.12) [95] to re-estimate the insert sizes. The genomic sequence was assembled de novo by AllPaths-LG (v52488) [96]. Gaps were filled by short-fragment libraries with GapCloser (v1.12) -p 25 -l 150 in SOAPdenovo2 [97]. The consistency was evaluated by re-mapping the short-fragment libraries to the assembled genome with BWA mem (v0.7.12) [95] and then summarized with Picard (v2.3.0, https://broadinstitute.github.io/picard/) (Additional file 1: Table S3). The completeness was evaluated by BUSCO (v2.0) [53], based on 4104 universal single-copy orthologs in mammalian set (Additional file 1: Table S4). 262 nuclear sequences belonging to Moschus were fetched from Genbank and aligned to the assembled genome with exonerate (v2.2.0) [98] est2genome (Additional file 1: Table S5).

Genome comparison of Siberian musk deer and forest musk deer

The genome of forest musk deer assembled by Fan et al. was downloaded from (http://gigadb.org/dataset/100411). We performed whole-genome alignment between the Siberian musk deer and forest musk deer assembly using mummer4 (nucmer -l 100 -c 500 --maxmatch) [56]. The alignment was filtered with minimum alignment length of 5 Kb (delta-filter -l 5000), and the difference was summarized using dnadiff.

Genome annotation

Transposable elements (TEs) in the genome were identified by RepeatMasker (v4.0.6) -s -nolow (http://www.repeatmasker.org/) (Additional file 1: Table S6). TEs from both homology-based searchings against known ruminantia and mammalian sequences in Repbase (v16.10) [99], as well as de novo prediction by RepeatModeler (v1.0.8) were combined and masked. The TEs of other genomes for comparison were fetched from RepeatMasker datasets online (http://repeatmasker.org/genomicDatasets/RMGenomicDatasets.html) (Additional file 1: Table S7).

Protein-coding genes were annotated by three approaches (Additional file 1: Table S8). Firstly, AUGUSTUS (v3.0.1) [100], GENEID (v1.4.4) [101], GeneMark_ES (v2.3e) [102], GlimmerHMM (v3.0.2) [103] and SNAP (v2013-11-29) [104] were applied for ab initio scan of gene structures. Secondly, the longest protein sequences of each gene from humans, cattle, dogs, sheep, pig, mouse and goat were fetched from RefSeq and projected to the assembled genome. The rough alignment was performed by genBlastA (v1.0.1) [105], with protein coverage greater than 30%. Then precise alignment aware of gene structure on the target DNA sequences was performed by GeneWise (v2.4.1) [106]. Thirdly, RNA-Seq data of mixture tissues, as well as previously reported RNA-Seq data [42] of musk gland (SRR2098995, SRR2098996) and heart (SRR2142357) were mapped to the genome by Tophat2 (v2.1.1) [107] (Additional file 1: Table S9). The transcripts were assembled with Cufflinks (v2.2.1) [108] and merged with cuffmerge. Only transcripts with putative coding regions were preserved with TransDecoder (v3.0.1) [109]. Finally, the three gene sets were merged by EVM [110] with a weight combination (GeneWise > Cufflinks > ab initio). As evaluated by BUSCO (v2.0) [53], EVM genes with only ab initio evidence were removed (Additional file 1: Figure S1), and the remaining genes were complemented and updated with GeneWise [106] (Additional file 1: Table S4).

Gene family construction

Gene families among the musk deer and other mammals were constructed with the TreeFam pipeline [111] (Additional file 1: Table S10), as described in detail by Li et al. [112]. Protein sequences were downloaded from RefSeq, and the longest one of each gene was chosen. All-to-all pairwise blastp were performed with -e 1e-10. Local alignments were joined by solar, and the alignment length should cover at least 1/3 on both proteins. A h-score was calculated for each protein pair (p1, p2) based on the blast score: h-score = score(p1, p2)/max (score(p1, p1), score(p2, p2)). Homologous proteins were then clustered with hcluster_sg -w 5 -s 0.33 and the opossum as an outgroup. Multiple alignments for each protein cluster were performed by clustalo (v1.2.0) [113], which was translated to CDS alignment by treebest backtrans. Guided by the common tree from NCBI Taxonomy, the phylogenetic tree for each cluster was constructed by treebest best. Orthologs were inferred from the cluster with treebest nj -t dm -v. Solar, hcluster_sg, and treebest were obtained from https://sourceforge.net/p/treesoft/code/ HEAD/tree/branches/lh3/.

Genome evolution

Four-fold degeneration sites were extracted from the CDS alignment of single-copy orthologs. They were concatenated to reconstruct the species tree with the NJ method by MEGA (v7.0.18) [114]. The species tree was calibrated by MCMCtree in PAML (v4.9) [115], using the following divergence time from TimeTree [116] (2.5% lower and upper bounds): cattle-sheep (10–40 Mya), cattle-pig (40–80 Mya), cattle-horse (55–90 Mya) and cattle-human (65–150 Mya).

The evolution of gene family size was inferred by CAFE (v3.1) [117] based on the homologous clusters. For families with significant size variations (family-wide p-value < 0.01), the branches with significant expansion and contraction were selected (Viterbi p-value < 0.01).

Based on the CDS alignment of single-copy orthologs, positively selected genes (PSGs) in the musk deer were identified by codeml in PAML (v4.9) [116]. Poorly aligned regions were first filtered by Gblocks (0.91b) [118]. Taking the musk deer as the foreground and other species as the background, the branch-site model (model = 2, NSsite = 2) with $dN/dS \leq 1$ (fix_omega = 1, omega = 1) and $dN/dS > 1$ (fix_omega = 0) were compared. The genes with significant $dN/dS > 1$ were identified by the likelihood ratio test ($p < 0.05$, chi-square test), and the positively selected sites (PSSs) were identified by the Bayes Empirical Bayes (BEB) analysis. To reduce the impact of defective gene annotation, genes with successive PSSs or PSSs located at the head or tail of the alignment (within 10 amino acids) were filtered. We conducted enrichment of the gene families and PSGs using KOBAS (v.3.0) [119]. Go terms and KEGG pathways with corrected p-values < 0.05 were identified as significantly enriched.

Genomic diversity and demography inference

Genomes of two additional musk deer (s190119001 and s180119002) were re-sequenced with the standard Illumina HiSeq protocol (2×150 bp). The reads were cleaned with Trimmomatic (v0.36) [120] and mapped to the assembled genome with BWA mem (v0.7.12) [1] (Additional file 1: Table S3). Duplicates were marked with Picard (v2.3.0), and Indel re-alignment was performed with GATK (v3.5) [121]. Variant calling was first performed for each sample with HaplotypeCaller -stand_call_conf 30 in the GVCF mode, which was then combined for joint genotyping with GenotypeGVCFs. SNPs were selected and filtered with VariantFiltration 'QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < − 12.5 || ReadPosRankSum < − 8.0'. Only biallelic SNPs were preserved in the following analysis (Additional file 1: Table S11). The demographic inference was performed with the PSMC model (v0.6.5) [59]. The consensus sequences for each individual were constructed with vcftools vcf-consensus (v0.1.12) [122] and transformed into the fastq format compatible with the PSMC input. Recommended parameters for the PSMC analysis were adopted, and the plot was scaled with -u 1.1e-08 -g 5 as estimated by Chen et al. [123].

RNA-Seq analysis

The RNA sequencing of mixture tissues was performed with the standard Illumina HiSeq protocol (2×150 bp). The RNA-Seq data [42] of two musk glands (SRR2098995, SRR2098996) and one heart tissue (SRR2142357) were downloaded from SRA. The raw reads were cleaned with Trimmomatic (v0.36) [120] and mapped to the assembled genome with STAR (v020201) [54] (Additional file 1: Table S12), which showed a higher mapping efficiency than Tophat2 (v2.1.1) [107]. Guided by the gene annotations, the transcripts were re-assembled with Cufflinks (v2.2.1) [108] and then merged with cuffmerge. Reads that mapped to exons were counted by HTSeq (v0.6.0) [124] and then normalized by the R package DESeq (v1.28.0) [125]. Differential expression analysis was performed with DESeq based on the negative binomial distribution (FDR < 0.05), and clustering analysis was performed with the R package NMF (v0.20.6) [126] (Additional file 1: Figure S2). Go terms and KEGG pathways were also performed by KOBAS(V3.0) with p-values < 0.05 were identified as significantly enriched.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12864-020-6495-2.

Additional file 1: Contains **Tables S1-S15** and **Figures S1-S4** with detailed results for the Figures presented in the main manuscript.

Additional file 2: Table S16. Go enrichment of DEGs in the *M. moschiferus* ($p < 0.05$).

Additional file 3: Table S17. (Comparison of the genome assembly of Siberian musk deer and forest musk deer).

Abbreviations

AKR: Aldo-keto reductase; BEB: Bayes Empirical Bayes; BP: Biological processes; CC: Cellular components; CYP: Cytochrome P450; DEGs : Differentially expressed genes; EVM : EvidenceModeler; LINES: Long interspersed elements; LTRs: Long terminal repeats; MF: Molecular functions; Mya: Million years ago; PSGs: Positively selected genes; PSMC: Pair-wise Sequentially Markovian Coalescent; PSSs: Positively selected sites; SDR: Short chain dehydrogenase; SINES: Short interspersed elements; SNP: Single-nucleotide polymorphism; TEs: Transposable elements

Acknowledgments

We would like to special thanks to all the breeders of musk deer in Mongolia.

Authors' contributions

LY, MD and RNS contributed equally to this work as first authors, wrote the manuscript, made substantial contributions to acquisition of data and analysis, interpretation of data. SH and CC are the corresponding authors and project managers in this work. MD and ZW contributed equally to this work as corresponding authors. Among them, the main contributions of authors are as follows: ME, BL and CC conducted the sample collection and biological traits analysis. LY and WLL coordinated genome sequencing, assembly and annotation. SH, ZW and MD did the comparative genome analysis, carried out the functional genomics analysis and drafted the work. RNS and LY submitted the genome sequence data, supplementary materials and substantively revised manuscript. ZW and SH edited and revised the manuscript, and did the final editing of the text, tables and figures. All authors read and approved the final manuscript.

Funding

This study was supported by grants from the National International Scientific and Technological Cooperation Project (Grant number is 2016YFE0116500), that played role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. The Youth Innovation Promotion Association CAS (2017325), the funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The dataset supporting the conclusions of this article is available on NCBI BioProject and can be accessed using the accession number PRJNA574937.

Ethics approval and consent to participate

All study procedures and animal care activities were conducted in accordance with the Bioethics Committee of the College of Veterinary Medicine in Inner Mongolia Agricultural University (12150000460029509 N) and Institute of Traditional Medicine and Technology, Ulaanbaatar, Mongolia. And we got a permit of CITES management authority of Mongolia in 2018 .

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 10 October 2019 Accepted: 14 January 2020
Published online: 31 January 2020

References

1. Nowak RM. Walker's Mammals of the World. Volume 2. 6th ed. Baltimore and London: The Johns Hopkins University Press; 1999. p. 1921.
2. Kattel B. Ecology of the Himalayan musk deer in Sagarmatha National Park, Nepal. PhD thesis. USA: Colorado State University; 1993.
3. DNPWC. Protected areas of Nepal [in Nepali]. Kathmandu: Department of National Parks and Wildlife Conservation; 2016.
4. Wilson DEM, Russell A. Handbook of the mammals of the World. Volume 2: Hoofed mammals. Barcelona: Lynx Edicions; 2011. p. 468.
5. Zhou Y, Meng X, Feng J, Yang Q. Review of the distribution, status and conservation of musk deer in China. Folia Zool. 2004;53:129–40.
6. Lee WS, Rhim SJ. Changes in distribution area of Korean musk deer (Moschus moschiferus parvipes) from 1950s to 1999 in South Korea. J Forestry Res. 2002;13(2):135–6.
7. Singh PB, Khatiwada JR, Saud P, Jiang ZG. mtDNA analysis confirms the endangered Kashmir musk deer extends its range to Nepal. Sci Rep. 2019;9:4895.
8. Xu Z, Jie H, Chen BL, Gaur U, Yang MY, Wu N, et al. De novo assembly of Chinese forest musk deer (Moschus berezovskii) transcriptome from next-generation mRNA sequencing. Peerj. 2016. https://doi.org/10.7287/peerj.preprints.2252v1.
9. Li DY, Chen BL, Zhang L, Gaur U, Ma TY, Jir H, et al. The musk chemical composition and microbiota of Chinese forest musk deer males. Sci Rep. 2016;6:18975.
10. Sokolov VE, Kagan MZ, Vasilieva VS, Prihodko VI, Zinkevich EP. Musk deer (Moschus moschiferus): reinvestigation of main lipid components from preputial gland secretion. J Chem Ecol. 1987;13(1):71–83.
11. Feng W, You Y, Yong H, Li G, Gu Q. A historical examination on the musk gland of Moschus chrysogaster. J Zool. 1981;2:33–5.
12. Jiang Z, Meng Z, Wang J. Musk market survey report. Endangered Species Scientific Commission of the People's Republic of China. Beijing; 2002.
13. Ostrowski S, Rahmani H, Ali JM, Ali R, Zahler P. Musk deer *Moschus cupreus* persist in the eastern forests of Afghanistan. Oryx. 2016;50:323–8.
14. Cao XH, Zhou YD. Progress on anti-inflammatory effects of musk. China Pharm. 2007;18:1662–5.
15. Feng QQ, Liu TJ. Progress on pharmacological activity of muscone. J Food Drug Anal. 2015;3:212–4.
16. Green MJ. The distribution, status and conservation of the Himalayan musk deer Moschus chrysogaster. Biol Conserv. 1986;35:347–75.
17. Ilyas O. Status, habitat use and conservation of Alpine musk deer (Moschus chrysogaster) in Uttarakhand Himalayas, India. J Appl Animl Res. 2014;43:83–91.
18. Yang Q, Meng X, Xia L, Feng Z. Conservation status and causes of decline of musk deer (*Moschus spp.*) in China. Biol Conserv. 2003;109:333–42.
19. Homes V. No licence to kill: the population and harvest of musk deer and trade in musk in the Russian federation and Mongolia. Brussels (BE): traffic. Europe. 2004;81. https://doi.org/10.1038/35052008.
20. IUCN. The IUCN Red List of Threatened Species. 2017. http://www.iucnredlist.org/.
21. Sheng HL, Liu ZX, editors. The musk deer in China. Shanghai: The Shanghai Scientific & Technical Publishers; 2007.
22. Webb SD, Taylor BE. The phylogeny of hornless ruminants and a description of the cranium of archaeomeryx. B Am Mus Nat Hist. 1980;167:121–57.
23. Scott KM, Janis CM. Phylogenetic relationships of the Cervidae, and the case for a superfamily "Cervoidea". In: Wemmer CM, editor. Biology and management of the Cervidae, 3–20. Washington DC.: Smithsonia Institute Press; 1987.
24. Groves CP, Wang YX, Grubb P. Taxonomy of musk-deer, genus Moschus (Moschidae, Mammalia). Acta Theriologica Sinica. 1995;15(3):181–97.
25. Su B, Wang YX, Lan H, Wang W, Zhang YP. Phylogenetic study of complete cytochrome b genes in musk deer (genus Moschus) using museum samples. Mol Phylogenet Evol. 1999;12(3):241–9.
26. Cap H, Aulagnier S, Deleporte P. The phylogeny and behaviour of Cervidae (Ruminantia, Pecora). Ethol Ecol Evol. 2002;14:199–216.
27. Li M, Hidetoshi B, Tamate H, Wei FW, Wang XM, Masuda RC, et al. Phylogenitic relationships among deer in China derived from mitochondrial DNA cytochrome b sequences. Acta Theriol. 2003;48(2):207–19.
28. Hassanin A, Douzery EJP. Molecular and morphological phylogenies of Ruminantia and the alternative position of the Moschidae. Syst Biol. 2003;52(2):206–28.
29. Kuznetsova MV, Kholodova MV, Danilkin AA. Molecular phylogeny of deer (Cervidae: Artiodactyla). Russ J Genet. 2005;41(7):742–9.
30. Fernández MH, Vrba ES. A complete estimate of the phylogenetic relationships in Ruminantia: a dated species-level super tree of the extant ruminants. Biol Rev. 2005;80:269–302.
31. Guha S, Goyal SP, Kashyap VK. Molecular phylogeny of musk deer: a genomic view with mitochondrial 16S rRNA and cytochrome b gene. Mol Phylogenet Evol. 2007;42(3):585–97.
32. Groves CP, Grubb P. Ungulate Taxonomy. Baltimore: Johns Hopkins University Press; 2011.
33. Dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang ZH. Phylogenomic datasets provide both precision and accuracy in estimating the timescale 75. Of placental mammal phylogeny. Proc. biol. Sci. 2012;279:2491–3500.
34. Bibi F. Assembling the ruminant tree: combining morphology, molecules, extant taxa, and fossils. ZittelianaB. 2014;32:197–212.
35. Zhou C, Zhang W, Wen QC, Bu P, Gao J, Wang GN, et al. Comparative genomics reveals the genetic mechanisms of musk secretion and adaptive immunity in Chinese forest musk deer. Genome Biol Evol. 2019;11(4):1019–32.
36. Cheng L, Qiu Q, Jiang Y, Wang K, Lin ZS, Li ZP, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Sci. 2019;364:1–12.
37. Su B, Wang YX, Wang QS. Mitochondrial DNA sequences imply Anhui musk deer a valid species in genus Moschus. Zool Res. 2001;22:169–73.
38. Agnarsson I, May-Collado LJ. The phylogeny of Cetartiodactyla: the importance of dense taxon sampling, missing data, and the remarkable promise of cytochrome b to provide reliable species-level phylogenies. Mol Phylogenet Evol. 2008;48:964–85.
39. Vislobokova I, Lavrov A. The earliest musk deer of the genus Moschus and their significance in clarifying of evolution and relationships of the family Moschidae. Paleontol J. 2009;43:326–38.
40. Pan T, Wang H, Hu CC, Sun ZL, Zhu XX, Meng T, et al. Species delimitation in the genus Moschus (Ruminantia: Moschidae) and its high-plateau origin. PLoS One. 2015; 10(8):e0134183.
41. Sokolov VE, Kagan MZ, Vasilieva VS, Prihodko VI, Zinkevich EP. Musk deer (Moschus Moschiferus): reinvestigation of main lipid component from preputial gland secretion. J Chem Ecol. 1987;13(1):71–83.
42. Xu ZX, Jie H, Chen BL, Gaur U, Wu N, Gao J, et al. Illumina-based de novo transcriptome sequencing and analysis of Chinese forest musk deer. J Genet. 2017;96(6):1033–40.
43. Xu ZX, Jie H, Chen BL, Gaur U, Yang MY, Wu N, et al. De novo assembly of Chinese forest musk deer (Moschus berezovskii) transcriptome form next-generation mRNA sequencing. Peer J. 2016;4:e2252v1.
44. Chen X. Studies on the genetic diversity of forest musk deer (*Moschus berezovskii*) and linkage analysis between the performance of musk productivity and AFLP markers. M. Scie. Thesis, Zhejiang University; 2007.
45. Peng H, Liu S, Zou F, Zeng B, Yue B. Genetic diversity of captive forest musk deer (Moschus berezovskii) inferred from the mitochondrial DNA control region. Anim Genet. 2009;40(1):65–72.
46. Zhao SS. Assement of genetic diversity in the captive forest musk deer (*Moschus berezovskii*) and linkage analysis between the performance of musk productivity and DNA molecular markers. D. Scie. Thesis, Zhejiang University; 2009.
47. Li YM, Hu XL, Yang S, Zhou JT, Zhang TX, QI L, et al. Comparative analysis of the gut microbiota composition between captive and wild forest musk deer. Front Microbiol. 2017;8:1705.
48. Hu XL, Liu G, Shafer ABA, Wei YT, Zhou JT, Lin SB, et al. Comparative analysis of the gut microbial communities in forest and alpine musk deer using high-throughout sequencing. Front Microbiol. 2017;8:572.

49. Zhenxin F, Wujiao L, Jiazheng J, Kai C, Chaochao Y, Changjun P, et al. The draft genome sequence of forest musk deer (Moschus berezovskii). GigaScience. 2018;7:1–6.
50. Tsendjav D. Mongolian Musk deer (*Moschus moschiferus Linnaeus*, 1758). Ulaanbaatar: JinstCargana Co. Ltd; 2002.
51. Nyambayar B, Mix H, Tsytsulina, K. Moschus moschiferus. The IUCN Red List of Threatened Species. 2015; T13897A61977573.
52. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2014;30(1):31–7.
53. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
55. Fan Z, Li W, Jin J, Cui K, Yan C, Peng C, et al. The draft genome sequence of forest musk deer (*Moschus berezovskii*). Gigascience. 2018;7(4). https://doi.org/10.1093/gigascience/giy038.
56. Marcais, et al. MUMmer4: A fast and versatile genome alignment system. PLoS Comput Biol. 2018;14(1):e1005944.
57. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 2005;39:121–52.
58. Qi WH, Li J, Zhanga XY, Wang ZK, Li XX, Yang CZ, et al. The reproductive performance of female forest musk deer (Moschus berezovskii) in captivity. Theriogenology. 2011;76(5):874–81.
59. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475(7357):493–6.
60. Zheng B, Xu Q, Shen YP. The relationship between climate change and quaternary glacial cycles on the Qinghai-Tibetan plateau: review and speculation. Quatern Int. 2002;97-98:93–101.
61. Lehmkuhl F, Owen LA. Late Quaternary glaciation of Tibet and the bordering mountains: a review. Boreas. 2005;34:87–100.
62. Ehlers J, Gibbard PL. The extent and chronology of Cenozoic global glaciation. Quatern Int. 2007;164–165:6–20.
63. Lorenzen ED, Nogués-Bravo D, Orlando L, Weinstock J, Binladen J, Marske KA, et al. Species-specific responses of Late Quaternary megafauna to climate and humans. Nature. 2011;479:359–36.
64. Fang X, Lu L, Yang S, Li J, An Z, Jiang PA, et al. Loess in Kunlun Mountains and its implications on desert development and Tibetan plateau uplift in West China. Sci China Ser D. 2002;45(4):289–99.
65. Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, et al. Recalibrating Equus evolution using the genome sequence of an early middle Pleistocene horse. Nature. 2013;499:74–8.
66. Tukey RH, Strassburg CP. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. Annu Rev Pharmacol Toxicol. 2000;40:581–616.
67. Kiang TK, Ensom MH, Chang TK. UDP-glucuronosyltransferases and clinical drug-drug interactions. Pharmacol Ther. 2005;106:97–132.
68. Zhou J, Tracy TS, Remmel RP. Glucuronidation of dihydrotestosterone and trans-androsterone by recombinant UDP-Glucuronosyltransferase (UGT) 1A4: evidence for multiple UGT1A4 aglycone binding sites. Drug Metab Dispos. 2010;38:431–40.
69. Javitt NB, Lee YC, Shimizu C, Fuda H, Strott CA. Cholesterol and hydroxycholesterol sulfotransferases: identification, distinction from dehydroepiandrosterone sulfotransferase, and differential tissue expression. Endocrinology. 2001;142:2978–84.
70. Yamazoe Y, Ozawa S, Nagata K, Gong DW, Kato R. Characterization and expression of hepatic sulfotransferase involved in the metabolism of N-substituted aryl compounds. Environ Health Perspect. 1994;102(Suppl 6):99–103.
71. Belyaeva OV, Kedishvili NY. Comparative genomic and phylogenetic analysis of short-chain dehydrogenases/reductases with dual retinol/sterol substrate specificity. Genomics. 2006;88(6):820–30.
72. Shou M, Korzekwa KR, Brooks EN, et al. Role of human hepatic cytochrome P450 1A2 and 3A4 in the metabolic activation of estrone. Carcinogenesis. 1997;18:207–14.
73. Mo SL, Liu YH, Duan W, Wei MQ, Kanwar JR, Zhou SF. Substrate specificity, regulation, and polymorphism of human cytochrome P450 2B6. Curr Drug Metab. 2009;10:730–53.
74. Flower WH. On the structure and affinities of the Musk-Deer (Moschus moschiferus Linn.). Proceedings of Zoological Society of London; 1875. p. 159–90.
75. Leinders JJM, Heintz E. The configuration of the lacrimal orifices in Pecorans and Tragulids (Artiodactyla, Mammalia) and its significance for the distinction between Bovidae and Cervidae. Beaufortia. 1980;30(7):155–62.
76. Wang Y, Zhang C, Wang N, Li Z, Heller R, Liu R, et al. Genetic basis of ruminant headgear and rapid antler regeneration. Science. 2019;364(6446): eaav6335.
77. Mccullough DR, Peik C, Wang Y. Home range, activity patterns, and habitat relations of Reeve's muntjacs in tai-wan. J Wildlife Manage. 2000;64(2):430–41.
78. Song YL, Gong HS, Zeng ZG, Wang XZ, Zhu L, Zhao NX. Food habits of serow. Chin J Zool. 2005;40(5):50–6.
79. Kuehn M, Welsch H, Zahnert T, Hummel T. Changes of pressure and humidity affect olfactory function. Eur Arch Otorhinolaryngol. 2008;265:299–302.
80. Singh PB, Khatiwada JR, Saud P, Jiang ZP. mtDNA analysis confirms the endangered Kashmir musk deer extends its range to Nepal. Sci Rep-Uk. 2019;9:4895.
81. Qu Y, Zhao H, Han N, Zhou G, Song G, Gao B, et al. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. Nat Commun. 2013;4:2071–9.
82. Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat Genet. 2014;45:1431–8.
83. Li JT, Gao YD, Xie L, Deng C, Shi P, Guan ML, et al. Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. P Nalt Acad Sci Usa. 2018;115(33):8406–11.
84. Wang MS, Li Y, Peng MS, Zhong L, Wang ZJ, Li QY, et al. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Mol Bio Evol. 2015;32(7):1880–9.
85. Yang J, Li WR, Lv FH, He SG, Tian SL, Peng WF, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol. 2016;33(10):2576–92.
86. Zhao SC, Zheng PP, Dong SS, Zhan XJ, Wu Q, Guo XS, et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nat Genet. 2013;45(1):67–71.
87. Miller W, Schuster SC, Welch AJ, Ratan A, Bedoya-Reina OC, Zhao FQ, et al. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc Natl Acad Sci. 2012;109:2382–90.
88. Qiu Q, Wang LZ, Wang K, Yang YZ, Ma T, Wang ZF, et al. Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions. Nat Commun. 2015;6:10283.
89. Mei CG, Wang HC, Liao QJ, Wang LZ, Cheng G, Wang HB, et al. Genetic architecture and selection of Chinese cattle revealed by whole genome resequencing. Mol Biol Evol. 2017;35(3):688–99.
90. Fan MY, Zhang MS, Shi MH, Zhang TX, Ql L, Yi J, et al. Sex hormones play roles in determining musk composition during the early stages of musk secretion by musk deer (Moschus berezovskii). Endocr J. 2018;65(11):1111–20.
91. Chen YS, Zhao WG, Zhao M, Chang ZJ, Zhang Y, Ma DW. Histological observation on musk-secreting scented gland in muskrat. Chin J Zool. 2007;42:91–5.
92. Zhang TX, Peng D, Qi L, Li WX, Fan MY, Shen JC, et al. Musk gland seasonal development and musk secretion are regulated by the testis in muskrat (ondatra zibethicus). Biol Res. 2017;50:10.
93. Fan MY, Zhang MS, Shi MH, et al. Sex hormones play roles in determining musk composition during the early stages of musk secretion by musk deer (*Moshus berezovskii*). Endocr J. 2018;65(11):1111–20.
94. Zhang FW, Liu Q, Wang ZY, et al. Seasonal expression of oxytocin and oxytocin receptor in the scented gland of male muskrat (Ondatra zibethicus). Sci Rep. 2017;7:16627.
95. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. Bioinformatics. 2010;26(5):589–95.
96. Gnerre S, Maccllum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108(4):1513–8.
97. Luo R, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1(1):18.
98. Slater GS, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;15(6):31.
99. Bao W, Kojima KK, Kohany K. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11.
100. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 2003;19(suppl 2):ii215–25.
101. Parra G, Blanco E, Guigó R. Geneid in drosophila. Genome Res. 2000;10(4): 511–5.

102. Ter-Hovhannisyan V, Lomsadze A, Chernoff Y, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 2008;18(12):1979–90.
103. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20(16):2878–9.
104. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5(1):59.
105. She R, Chu JS, Wang K, Pei J, Chen N. GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res. 2009;19(1):143–9.
106. Birney EM, Clamp DR. GeneWise and Genomewise. Genome Res. 2004;14(5): 988–95.
107. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013;14:R36.
108. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Baren MJV, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511–5.
109. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc. 2013;8(8):1494–512.
110. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 2008;9(1):R7.
111. Li H, Coghlan A, Ruan J, Coin LJ, Heriche JK, Osmotherly L, et al. TreeFam: a curated database of phylogenetic trees of animal gene families. Nucleic Acids Res. 2006;34(Database issue):D572–80.
112. Li R, Fan W, Tian G, Zhu HM, He L, Cai J, et al. The sequence and de novo assembly of the giant panda genome. Nature. 2010;463(7279):311–7.
113. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol Syst Biol. 2011;7:539.
114. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
115. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
116. Hedges SB, Dudley J, Kumar S. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics. 2006;22(23):2971–2.
117. Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 2013;30(8):1987–97.
118. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 2000;17(4):540–52.
119. Xie C, Mao XZ, Huang JJ, Ding Y, Wu JM, Dong S, et al. KOBAS 2.0: a web server for annotation and identifi- cation of enriched pathways and diseases. Nucleic Acids Res. 2011;39(Web Server issue):W316–22.
120. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–20.
121. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8.
122. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
123. Chen L, Qiu Q, Jiang Y, Wang K, Lin Z, Li Z, et al. Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science. 2019;364(6446):eaav6202.
124. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.
125. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11:R106.
126. Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010;11:367.

Publisher's Note