



OPEN

DATA DESCRIPTOR

# *De novo* transcriptome assembly database for 100 tissues from each of seven species of domestic herbivore

Yifan Wang<sup>1,2,3</sup>, Yiming Huang<sup>1,4,8</sup>, Yongkang Zhen<sup>2,8</sup>, Jiasheng Wang<sup>2,8</sup>, Limin Wang<sup>1</sup>, Ning Chen<sup>1</sup>, Feifan Wu<sup>2</sup>, Linna Zhang<sup>4</sup>, Yizhao Shen<sup>5</sup>, Congliang Bi<sup>6</sup>, Song Li<sup>3</sup>, Kelsey Pool<sup>7</sup>, Dominique Blache<sup>7</sup>, Shane K. Maloney<sup>7</sup>, Dongxu Liu<sup>4</sup>, Zhiquan Yang<sup>4</sup>, Chuang Li<sup>2</sup>, Xiang Yu<sup>2</sup>, Zhenbin Zhang<sup>2</sup>, Yifei Chen<sup>2</sup>, Chun Xue<sup>2</sup>, Yalan Gu<sup>2</sup>, Weidong Huang<sup>2</sup>, Lu Yan<sup>2</sup>, Wenjun Wei<sup>2</sup>, Yusu Wang<sup>2</sup>, Jinying Zhang<sup>2</sup>, Yifan Zhang<sup>2</sup>, Yiquan Sun<sup>2</sup>, Shengbo Wang<sup>4</sup>, Xinle Zhao<sup>4</sup>, Chengfang Luo<sup>4</sup>, Haodong Wang<sup>4</sup>, Luoyang Ding<sup>2,7</sup>✉, Qing-Yong Yang<sup>1,4</sup>✉, Ping Zhou<sup>1</sup>✉ & Mengzhi Wang<sup>2</sup>✉

Domesticated herbivores are an important agricultural resource that play a critical role in global food security, particularly as they can adapt to varied environments, including marginal lands. An understanding of the molecular basis of their biology would contribute to better management and sustainable production. Thus, we conducted transcriptome sequencing of 100 to 105 tissues from two females of each of seven species of herbivore (cattle, sheep, goats, sika deer, horses, donkeys, and rabbits) including two breeds of sheep. The quality of raw and trimmed reads was assessed in terms of base quality, GC content, duplication sequence rate, overrepresented k-mers, and quality score distribution with FastQC. The high-quality filtered RNA-seq raw reads were deposited in a public database which provides approximately 54 billion high-quality paired-end sequencing reads in total, with an average mapping rate of ~93.92%. Transcriptome databases represent valuable resources that can be used to study patterns of gene expression, and pathways that are related to key biological processes, including important economic traits in herbivores.

## Background & Summary

Herbivores have been critical for food security since the late Holocene and remain an important contributor to that security<sup>1</sup>. There are around four billion domestic herbivores on the planet, representing animals from diverse species and breeds, with different species and breeds adapted to, and therefore farmed, in different environments, even marginal lands<sup>2</sup>. The species or breed of domestic herbivore that is found in an agricultural ecosystem is adapted to that system because of its physiological adaptation to that ecosystem in terms of characteristics such as growth rate, mature size, reproductive strategy, and/or digestive system (Table 1). However, some complex phenotypes, such as specific sexual behaviours, are common to different species of domestic herbivores. For example, “flehmen” behaviour, the typical behaviour of a male exposed to the urine of a female, is shared by many herbivores including the seven species used here<sup>3</sup>. Conversely, the timing of the breeding

<sup>1</sup>State Key Laboratory of Sheep Genetic Improvement and Healthy Production, Xinjiang Academy of Agricultural Reclamation Sciences, Shihezi, 832000, P. R. China. <sup>2</sup>College of Animal Science and Technology, Yangzhou University, Yangzhou, 225009, P. R. China. <sup>3</sup>College of Life Science, Guizhou University, Guiyang, 550025, P. R. China. <sup>4</sup>Hubei Key Laboratory of Agricultural Bioinformatics and Hubei Engineering Technology Research Center of Agricultural Big Data, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, P. R. China. <sup>5</sup>College of Animal Science and Technology, Hebei Agricultural University, Baoding, 071033, P. R. China. <sup>6</sup>College of Life Science, Linyi University, Linyi, 276005, P. R. China. <sup>7</sup>UWA Institute of Agriculture, The University of Western Australia, Perth, WA, 6009, Australia. <sup>8</sup>These authors contributed equally: Yiming Huang, Yongkang Zhen, Jiasheng Wang. ✉e-mail: [luoyang.ding@uwa.edu.au](mailto:luoyang.ding@uwa.edu.au); [yqy@mail.hzau.edu.cn](mailto:yqy@mail.hzau.edu.cn); [zhpxqf@163.com](mailto:zhpxqf@163.com); [mzwang@yzu.edu.cn](mailto:mzwang@yzu.edu.cn)

Species	Worldwide value (US\$ millions) <sup>a</sup>	Adult Live weight (kg) <sup>b</sup>	Reproduction <sup>b</sup>			Type of digestive system <sup>b</sup>	Agricultural use <sup>b</sup>				
			Sexual maturity (month)	Reproductive strategy	Litter size		Meat	Milk	Fibre	Hide	Power
<i>Horse</i>	60	230–900	Female: 30 Male: 32	Seasonal breeder long day	1	Hindgut fermenter	•	•			•
<i>Donkey</i>	53	180–225	Female: 18 to 24 Male: 9 to 24	Seasonal breeder long day	1	Hindgut fermenter	•	•			•
<i>Cattle</i>	1,529	500–1300	Female: 12–18 Male: 12	Year round	1	Ruminant	•	•		•	•
<i>Goat</i>	1,111	10–110	Female: 13 Male: 23	Seasonal breeder short day	1–3	Ruminant	•	•	•	•	
<i>Sheep</i>	1,284	20–200	Female: 18 Male: 30	Seasonal breeder short day	1–3	Ruminant	•	•	•	•	
<i>Deer</i>	5	60–200	Female: 16–18 Male: 16–18	Seasonal breeder short day	1	Hindgut fermenter	•			•	
<i>Rabbit</i>	0.172	1.5–2.5	Female: 8 Male: 8	Year round	1–14	Hindgut fermenter	•		•		

**Table 1.** Summary of facts on the seven species included in the present study. <sup>a</sup>Worldwide value from FOASTAT (<https://www.fao.org/faostat/>, accessed on 16 August 2023), except for deer<sup>23</sup>. <sup>b</sup>Data on adult live weight, reproduction, digestive system and agriculture uses were obtained from ABW (Animal Biodiversity Web) University of Michigan's animal diversity web (<https://animaldiversity.org/>) and AnAge Database of Animal Ageing and Longevity (<https://genomics.senescence.info/species/index.html>). Both were accessed on 16 August 2023.

season can be quite different between species. A response to photoperiod drives sheep to breed during winter (short-day breeders) while the same photoperiodic signal causes horses to breed during summer (long-day breeders), and some other species breed all year round (Table 1). Breeds within a species generally occupy different ecological niches because they have been selected, either naturally or artificially, to survive and reproduce within that particular environment. A further understanding of the interaction between genetics and the environment on the expression of phenotype is essential to improve the management of domestic herbivores.

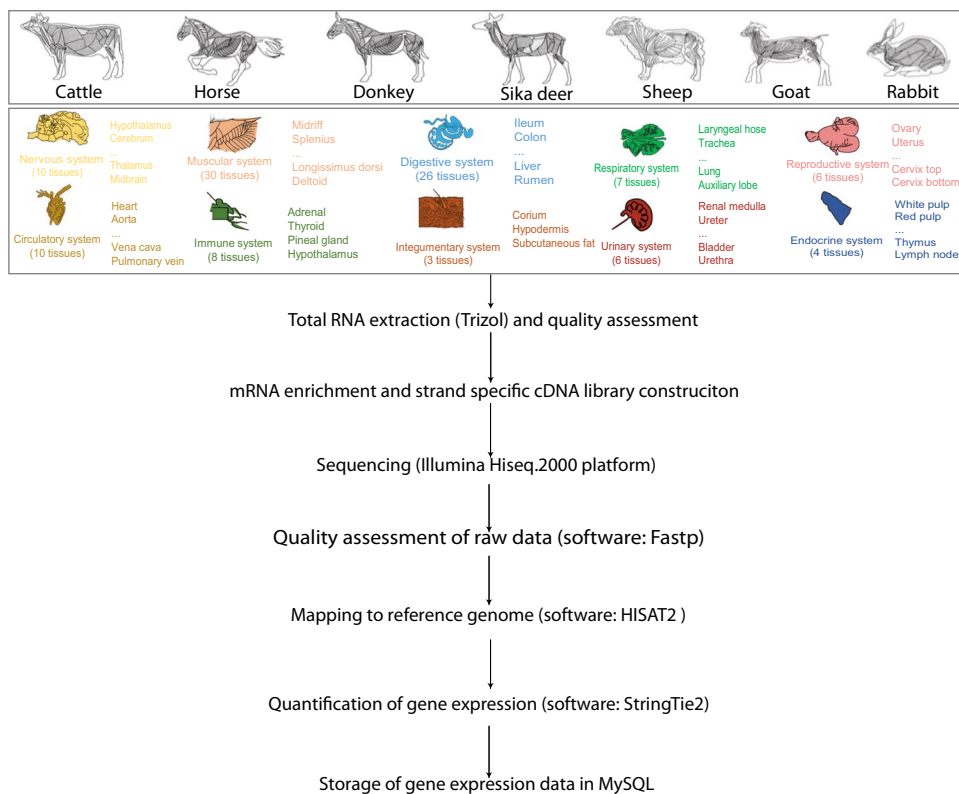
While the study of variation or mutation in the genome is informative, most complex phenotypes, such as behaviour, are encoded by an interacting suite of numerous genes, potentially involving hundreds of genes<sup>4</sup>. Comparative transcriptomics is a powerful tool that can be used to explore the interaction between genotype and environment in the context of biological adaptation and the potential limits of a species<sup>4</sup>. Using a combination of genomics and transcriptomics was first proposed by the Functional Annotation of Animal Genomes (FFANG) Consortium to identify the genes that code for proteins that are involved in critical production traits in farm animals<sup>5</sup>. Transcriptomic atlases were generated for common domestic animals such as pigs, cattle, sheep, and chickens<sup>6–10</sup> as well as other species of herbivores such as donkey<sup>11</sup>, water buffalo<sup>12</sup>, and goat<sup>13</sup>. While these published transcriptomic atlases of herbivores are valuable, they are published in independent databases, with no standard protocols for sample collection, RNA sequencing, or data processing. While the transcriptomic database for cattle includes 51 tissues<sup>8</sup>, most of the other published transcriptomic atlases contain a relatively small number of tissues (ranging from 12 to 25). Moreover, while the transcriptomic data from a particular tissue is informative about the transcriptome of that particular organ or tissue, it is not necessarily representative of the entire transcriptome of every organ or tissue. Therefore, it is essential to obtain samples from specific regions within organs that are functionally heterogeneous, such as the brain<sup>7,8</sup>. Similarly, it could be argued that a complete transcriptome should include representative tissues from each of the ten main organ systems, and that the collection and analysis protocol should be the same for each. To our knowledge, none of the transcriptome atlases mentioned above analysed samples from different regions of heterogeneous tissues or organs from each of the ten main organ systems using a standard protocol<sup>7,8</sup>.

With the current study we aimed to generate a public raw RNA-seq database that will facilitate the comparison of gene expression profiles between different tissues from seven species of domestic herbivores that have important economic value (Fig. 1). The seven species were the horse, the donkey, the rabbit, the deer, the goat, cattle, and sheep. Two breeds of sheep, the Hu and the Han sheep, were included because of their economic relevance and specific geographic distributions in China (Fig. 2a). The transcriptome was analysed in organs and tissues from the ten organ systems, and regions within heterogeneous organs (Fig. 1). This database can be used primarily to further identify the genetic foundations of economically significant traits and open novel pathways to genetic improvement. Secondly, this database could facilitate further understanding of the evolutionary processes that have taken place in herbivores during their domestication and the interaction between genetics and environment on the expression of similar and different phenotypes.

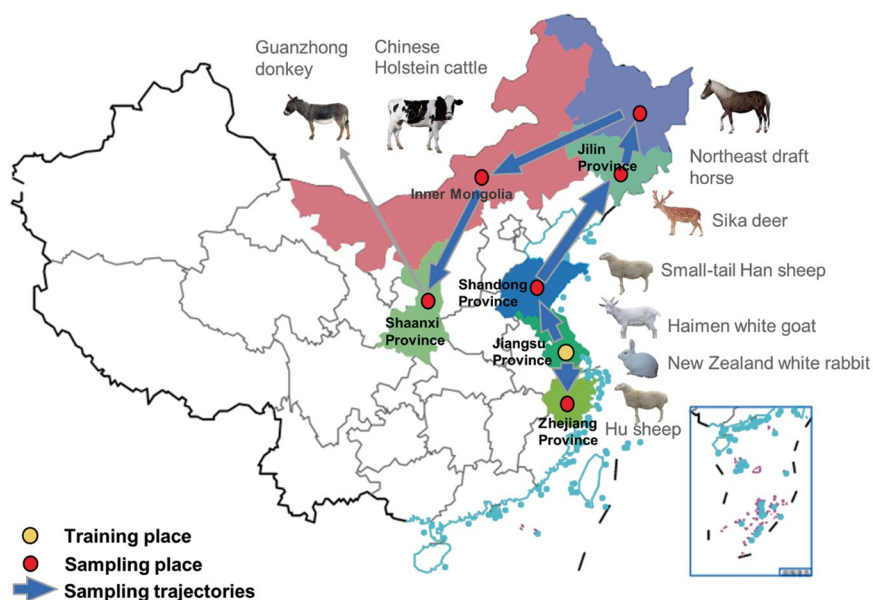
## Methods

All animal experiments were carried out in accordance with the ARRIVE guidelines and were approved by the Animal Ethics Committee of Yangzhou University under approval number RA202203-046. The experimental processes involved in the data collection and quality validation are shown in Fig. 1.

**Data collection.** The RNA-seq data were obtained using transcriptome sequencing from between 100 and 105 tissue samples that were dissected from two healthy juvenile female individuals of eight domestic herbivores, representing seven species and including two breeds of sheep (Table S1). The seven species were Chinese Holstein cattle



**Fig. 1** Schematic overview of the data collection and the quality validation of the herbivore RNA-seq database.



**Fig. 2** An overview of the sampling location and geographic distributions of the horse, the donkey, the rabbit, the deer, the goat, cattle, and sheep in China.

(*Bos taurus*) samples), Haimen white goat (*Capra hircus*), Northeast draft horse (*Equus caballus*), Guanzhong donkey (*Equus asinus*), sika deer (*Cervus nippon*), New Zealand white rabbit (*Oryctolagus cuniculus*), and two breeds of the domestic sheep (*Ovis aries*), the Hu sheep and the Small-tail Han sheep. The tissue samples were selected to cover the ten organ systems (Table S1): integumentary system, muscular system, digestive system, respiratory system, urinary system, reproductive system, circulatory system, immune system, nervous system, and endocrine system<sup>14</sup>. Tissue samples were collected from all organs and regions within heterogenous organs within 15 minutes after death by overdose of propofol (H20051843, Guangdong Jiabo pharmaceutical Co. Ltd. Qingyuan China).

Species	Breed	Number of samples	Number of reads	Aligned ratio (%)	Number of genes detected in present study	Proportion of protein-coding to non-coding genes	Reference genome	Ref	Annotation database
Horses	Northeast draft horse	204	5,665,395,789	96.06	33,091	1.76	EquCab3.0	24	<i>Equus caballus</i> RefSeq Annotation Release 103 <sup>25</sup>
Donkeys	Guanzhong donkey	202	7,166,740,391	91.7	38,516	1.28	ASM1607732V2	26	<i>Equus asinus</i> Annotation Release 101 <sup>27</sup>
Cattle	Chinese Holstein cattle	210	7,182,117,084	95.69	34,801	1.52	ARS-UCD1.2	28	<i>Bos taurus</i> Annotation Release 106 <sup>29</sup>
Goats	Haimen white goat	208	7,082,253,582	96.4	28,920	2.48	ARS1	30	<i>Capra hircus</i> Annotation Release 102 <sup>31</sup>
Sheep	Hu sheep	210	6,918,363,418	96.25	33,800	1.69	ARS-UI_Ramb_v2.0,	32	<i>Ovis aries</i> Annotation Release 102 <sup>33</sup>
				90.41	29,298	2.34	Oar_v4.0	34	<i>Ovis aries</i> Annotation 104 <sup>35</sup>
Sheep	Small-tail Han sheep	198	6,489,163,398	95.18	33,800	1.69	ARS-UI_Ramb_v2.0,	32	<i>Ovis aries</i> Annotation Release 102 <sup>33</sup>
				90.17	29,298	2.34	Oar_v4.0	34	<i>Ovis aries</i> Annotation 104 <sup>35</sup>
Sika deer	Sika deer	210	6,378,047,294	92.85	*	*	MHL_v1.0	36	*
Rabbits	New Zealand white rabbit	200	7,237,287,013	86.94	29,347	2.24	OryCun2.0	37	<i>Oryctolagus cuniculus</i> Annotation Release 101 <sup>38</sup>

**Table 2.** Descriptive summary of the newly generated profiles of transcriptome sequencing that were based on 110 organs/tissues from each of eight species/breeds and the annotation from databases that are available in the public domain. \*The annotation for MHL\_v1.0 contains a total of 21,449 protein-coding genes and the exon number per gene was 9.29<sup>18,19</sup>.

**RNA extraction and sequencing.** Total RNA was extracted from each tissue sample using Trizol (Qiagen, Germany) and from each blood sample using QIAamp RNA Blood Mini Kit (Qiagen). Contamination from genomic DNA was removed using DNase (New England Biolabs, USA). The quality and concentration of RNA was assessed using the Agilent 2100 RNA 6000 Nano Kit (Agilent Technologies, Germany). Construction of the transcript library, clustering, and sequencing were performed at the Novogene Bioinformatics Institute (Wuhan, China). Briefly, a total of 3 µg total RNA of each sample was purified using poly-T oligo-linked magnetic beads (Invitrogen, USA). The RNA strands were subsequently fragmented with divalent cations in NEB First Strand Synthesis reaction buffer (NEB, USA), followed by the synthesis of first and second strand cDNAs. The cDNAs were sequenced using the Illumina HiSeq. 2000 platform, generating 100 bp paired-end reads.

**Quality control and processing of RNA-seq data.** Fastp software was used to check the read quality of the RNA-seq and to filter out any low-quality reads<sup>15</sup>. Single-end and paired-end reads were then mapped to the reference genome using HISAT2 (v2.2.1) (Table 2)<sup>16</sup>, and the results were sorted and converted to BAM format using SAMtools<sup>17</sup>. The reference genome of deer, downloaded from RGD v2.0<sup>18,19</sup>, and the reference genomes of other species, from the NCBI database, were used for gene annotation. Data on gene expression were quantified using StringTie<sup>20</sup>. The expression level of each gene was normalized as both fragments per kilobase of exon per million mapped fragments (FPKM) and as transcripts per kilobase of exon model per million mapped reads (TPM). The sample-sample relationship in each of the seven species of herbivore was assessed with a Pearson correlation coefficient test using the gene expression data, and the results are reported in the supplementary figures.

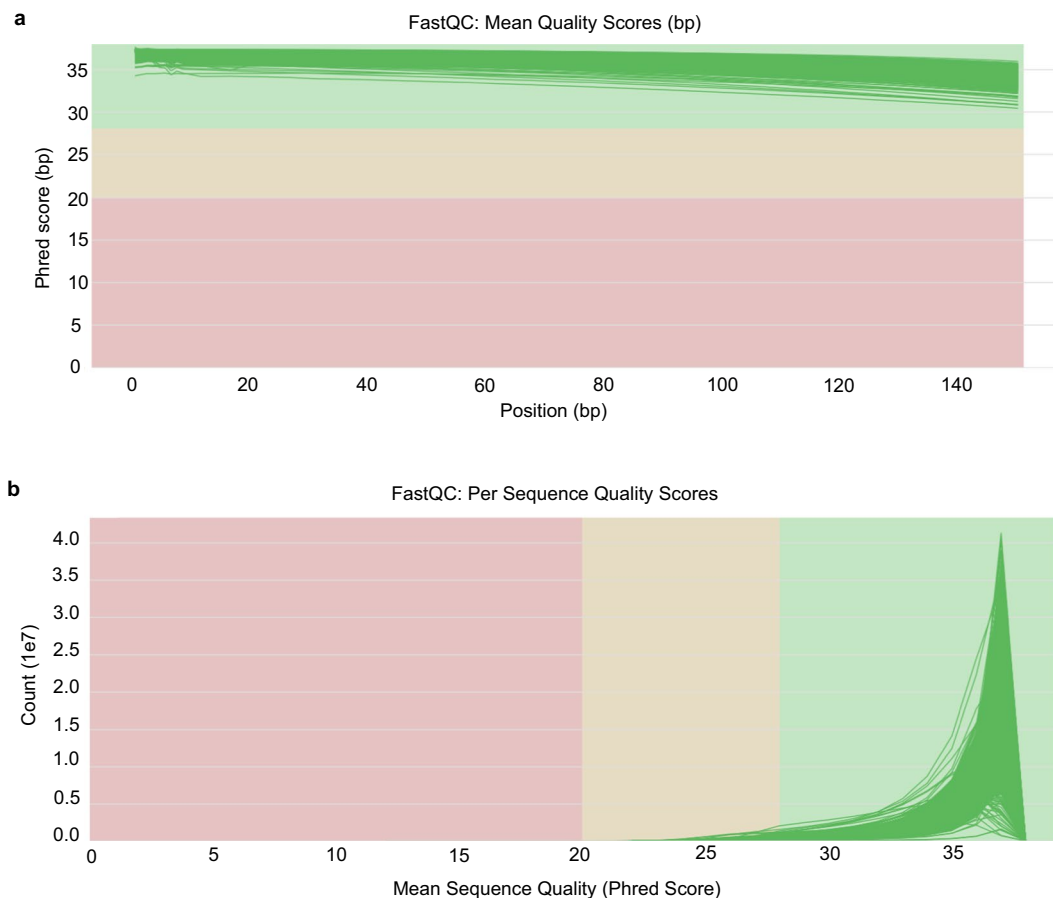
### Data Records

Raw RNA-Sequence reads of 1,642 tissues of seven species of herbivores (an average read mapping rate of 93.92%, Table 2) have been deposited in the NCBI Sequence Read Archive (SRA) database under the NCBI project (<https://www.ncbi.nlm.nih.gov/bioproject/>) with an accession number of PRJNA1017964<sup>21</sup>.

A hierarchical coding system was used to assign a unique code to each sample (Table S1) similar to Medical Subject Headings (MeSH)<sup>22</sup>. Each code comprises four levels. Level 1 indicates the organ system and starts with the letter “A” followed by a number between 01 and 10, representing the 10 organ systems. Level 2 comprises of two digits that represent the organ within the organ system. Level 3 comprises of three digits that represent the tissue within the organ. Level 4 comprises of two digits that represent the region within the tissue. The same codes are used with each species to facilitate the retrieval and analysis of both data generated for the present project and public RNA-Seq data.

### Technical Validation

**Quality control of tissue collection.** The quality of the tissue that were collected for RNA sequencing in this study was optimised by using a very experienced team of neurobiologists, animal physiologists, veterinarians, and biologists to perform rapid sampling of the 100 to 105 tissues from each of the seven species and two breeds. Each sampling session was performed by a team of at least 15 people, so that all the 100–105 tissue samples were snap frozen in liquid nitrogen within 15 minutes after the animal was declared dead.



**Fig. 3** Sequence quality assessment metrics of filtered and trimmed RNA-Seq data using MultiQC. **(a)** The mean quality scores across each base position, **(b)** quality score distribution over all sequences.

**Quality validation of RNA-Seq data.** The sequencing depth of the tissue samples from Haimen white goat, sika deer, Guanzhong donkey, Northeast draft horse, New Zealand white rabbit, Chinese Holstein cattle, Hu sheep, and Small-tail Han sheep are reported in Table S1. The minimum and mean sequencing depths were 1.94 and 3.49 in Haimen white goat, 2.26 and 3.65 in sika deer, 2.40 and 4.37 in Guanzhong donkey, 2.32 and 3.32 in Northeast draft horse, 2.06 and 3.96 in New Zealand white rabbit, 2.14 and 3.77 in Chinese Holstein cattle, 2.10 and 3.79 in Hu sheep, and 1.24 and 3.75 in Small-tail Han sheep. For the newly-generated RNA-seq datasets, the samples with low quality RNA and cDNA library were removed from the transcriptome catalogue. The quality of raw and trimmed reads was assessed in terms of base quality, GC content, duplication sequence rate, overrepresented k-mers, and quality score distribution using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). MultiQC software was used to aggregated and analysed the quality control metrics of all samples. Here, the quality report of newly generated RNA-seq data generated by MultiQC software showed that quality scores per base were high with phred score above 30 and most of sequences were high-quality with mean quality scores above 30 (Fig. 3). Next, high-quality filtered and trimmed RNA-Seq data was aligned to the reference genome of the corresponding species (Table 2). The alignment results indicated that the average aligned ratio of goat, sika deer, donkey, horse, rabbit, cattle, sheep were 94.35%, 91.84%, 93.28%, 93.11%, 87.40%, 94.91%, 95.57%, respectively (Table S1), confirming the high quality of the RNA-seq data in this study.

### Code availability

The authors declare that no custom code was used in this study.

Received: 13 December 2023; Accepted: 2 May 2024;

Published online: 11 May 2024

### References

1. Van Neer, W. in *Droughts, food and culture: Ecological change and food security in Africa's later prehistory* 251–274 (Springer, 2002).
2. Mottet, A., Teillard, F., Boettcher, P., De'Besi, G. & Besbes, B. Domestic herbivores and food security: current contribution, trends and challenges for a sustainable development. *Animal* **12**, s188–s198 (2018).
3. Mota-Rojas, D. *et al.* Olfaction in animal behaviour and welfare. *CABI Reviews*, 1–13 (2018).
4. Harrison, P. W., Wright, A. E. & Mank, J. E. in *Seminars in Cell & Developmental Biology*. 222–229 (Elsevier).

5. Clark, E. L. *et al.* From FAANG to fork: application of highly annotated genomes to improve farmed animal production. *Genome Biol* **21**, 285, <https://doi.org/10.1186/s13059-020-02197-8> (2020).
6. Tixier-Boichard, M. *et al.* Tissue Resources for the Functional Annotation of Animal Genomes. *Front Genet* **12**, 666265, <https://doi.org/10.3389/fgene.2021.666265> (2021).
7. Fang, L. *et al.* Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. *Genome Res* **30**, 790–801, <https://doi.org/10.1101/gr.250704.119> (2020).
8. Zhang, T. *et al.* Transcriptional atlas analysis from multiple tissues reveals the expression specificity patterns in beef cattle. *BMC Biol* **20**, 79, <https://doi.org/10.1186/s12915-022-01269-4> (2022).
9. Clark, E. L. *et al.* A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLoS Genet* **13**, e1006997, <https://doi.org/10.1371/journal.pgen.1006997> (2017).
10. Summers, K. M. *et al.* Functional Annotation of the Transcriptome of the Pig, *Sus scrofa*, Based Upon Network Analysis of an RNAseq Transcriptional Atlas. *Front Genet* **10**, 1355, <https://doi.org/10.3389/fgene.2019.01355> (2019).
11. Wang, Y. *et al.* Transcriptome Atlas of 16 Donkey Tissues. *Front Genet* **12**, 682734, <https://doi.org/10.3389/fgene.2021.682734> (2021).
12. Si, J., Dai, D., Li, K., Fang, L. & Zhang, Y. A Multi-Tissue Gene Expression Atlas of Water Buffalo (*Bubalus bubalis*) Reveals Transcriptome Conservation between Buffalo and Cattle. *Genes (Basel)* **14**, <https://doi.org/10.3390/genes14040890> (2023).
13. Muriuki, C. *et al.* A Mini-Atlas of Gene Expression for the Domestic Goat (*Capra hircus*). *Front Genet* **10**, 1080, <https://doi.org/10.3389/fgene.2019.01080> (2019).
14. Cinti, S. Anatomy and physiology of the nutritional system. *Mol Asp Med* **68**, 101–107 (2019).
15. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinform* **34**, i884–i890 (2018).
16. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907–915 (2019).
17. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinform* **25**, 2078–2079 (2009).
18. Xing, X. *et al.* The First High-quality Reference Genome of Sika Deer Provides Insights into High-tannin Adaptation. *Genomics, Proteomics and Bioinformatics* **21**, 203–215 (2023).
19. Fu, W. *et al.* RGD v2.0: a major update of the ruminant functional and evolutionary genomics database. *Nucleic Acids Research* **50**, D1091–D1099 (2022).
20. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**, 1–13 (2019).
21. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRP465528> (2023).
22. CEJBotMLA, L. Medical subject headings (MeSH). *Bull Med Libr Assoc* **88**, 265–266 (2000).
23. Daszkiewicz, T. *et al.* A comparison of the quality of the Longissimus lumborum muscle from wild and farm-raised fallow deer (*Dama dama* L.). *Small Rumin Res* **129**, 77–83 (2015).
24. Kalbfleisch, T. S. *et al.* Improved reference genome for the domestic horse increases assembly contiguity and composition. *Commun Biol* **1**, 197 (2018).
25. Santa Cruz. NCBI Equus caballus Annotation Release 103. [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Equus\\_caballus/103/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Equus_caballus/103/) (2018).
26. Wang, C. *et al.* Donkey genomes provide new insights into domestication and selection for coat color. *Nat Commun* **11**, 6014 (2020).
27. Shandong Academy of Agricultural Sciences. NCBI Equus asinus Annotation Release 101. [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Equus\\_asinus/101/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Equus_asinus/101/) (2021).
28. Pacholewska, A. *et al.* The transcriptome of equine peripheral blood mononuclear cells. *PLoS One* **10**, e0122011 (2015).
29. USDA ARS. NCBI Bos taurus Annotation Release 106. [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Bos\\_taurus/106/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Bos_taurus/106/) (2018).
30. Bickhart, D. M. *et al.* Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet* **49**, 643–650 (2017).
31. USDA ARS. NCBI Capra hircus AnnotationRelease 102. [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Capra\\_hircus/102/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Capra_hircus/102/) (2016).
32. Jiang, Y. *et al.* The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* **344**, 1168–1173 (2014).
33. International Sheep Genome Consortium. NCBI Ovis aries Annotation Release 102. [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Ovis\\_aries/102/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ovis_aries/102/) (2015).
34. Davenport, K. M. *et al.* An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. *Gigascience* **11**, giab096 (2022).
35. University of Idaho. NCBI Ovis aries Annotation Release 104. [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Ovis\\_aries/104/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Ovis_aries/104/) (2021).
36. Zhang, R., Li, Y. & Xing, X. Comparative antler proteome of sika deer from different developmental stages. *Sci Rep* **11**, 10484 (2021).
37. Carneiro, M. *et al.* Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074–1079 (2014).
38. The Broad Institute of MIT and Harvard. NCBI Oryctolagus cuniculus Annotation Release 101. [https://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/Oryctolagus\\_cuniculus/101/](https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Oryctolagus_cuniculus/101/) (2014).

## Acknowledgements

This study was supported by the funds from the State Key Laboratory of Sheep Genetic Improvement and Healthy Production (2021ZD01, 2021ZD07, 2021ZD04, NCG202232, NCG202221), the National 14<sup>th</sup> Five-Year Plan Key Research and Development Program (2021YFD1600702), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), China.

## Author contributions

Conceived and designed the experiments: Y.W., Y.Z., J.W., L.D., Q.Y., P.Z. and M.W. Sample collection and processing: Y.W., Y.Z., J.W., L.W., N.C., F.W., Y.S., C.B., L.S., C.L., X.Y., Z.Z., Y.C., C.X., Y.G., W.H., L.Y., W.W., Y.W., J.Z., Y.Z., Y.S., K.P., D.B., S.K.M. and L.D. Data analysis: Y.H., L.Z., S.W., X.Z., C.L., H.W., M.W. and Q.Y. Data validation: Y.W., Y.H., L.Z., K.P., D.B., S.W., X.Z., C.L., H.W. and Q.Y. Manuscript drafting and revision: Y.W., Y.Z., D.B., S.K.M., L.D., P.Z., Q.Y. and M.W.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03338-5>.

**Correspondence** and requests for materials should be addressed to L.D., Q.-Y.Y., P.Z. or M.W.

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

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024