Whole-genome resequencing reveals domestication and signatures of selection in Ujimqin, Sunit, and Wu Ranke Mongolian sheep breeds

Hanning Wang¹,a, Liang Zhong²,a, Yanbing Dong¹, Lingbo Meng¹, Cheng Ji¹, Hui Luo¹, Mengrong Fu¹, Zhi Qi¹,*, and Lan Mi¹,*

Objective: The current study aimed to perform whole-genome resequencing of Chinese indigenous Mongolian sheep breeds including Ujimqin, Sunit, and Wu Ranke sheep breeds (UJMQ, SNT, WRK) and deeply analyze genetic variation, population structure, domestication, and selection for domestication traits among these Mongolian sheep breeds.

Methods: Blood samples were collected from a total of 60 individuals comprising 20 WRK, 20 UJMQ, and 20 SNT. For genome sequencing, about 1.5 μg of genomic DNA was used for library construction with an insert size of about 350 bp. Pair-end sequencing were performed on Illumina NovaSeq platform, with the read length of 150 bp at each end. We then investigated the domestication and signatures of selection in these sheep breeds.

Results: According to the population and demographic analyses, WRK and SNT populations were very similar, which were different from UJMQ populations. Genome wide association study identified 468 and 779 significant loci from SNT vs UJMQ, and UJMQ vs WRK, respectively. However, only 3 loci were identified from SNT vs WRK. Genomic comparison and selective sweep analysis among these sheep breeds suggested that genes associated with regulation of secretion, metabolic pathways including estrogen metabolism and amino acid metabolism, and neuron development have undergone strong selection during domestication.

Conclusion: Our findings will facilitate the understanding of Chinese indigenous Mongolian sheep breeds domestication and selection for complex traits and provide a valuable genomic resource for future studies of sheep and other domestic animal breeding.

Keywords: Domestication; Metabolism; Sheep; Selection; Whole-genome Resequencing

INTRODUCTION

Domestication reshaped animal morphology, physiology, and behavior, providing the opportunity to study the molecular processes triggering evolutionary change. Animal domestication has been widely studied to understand the animal genetic and phenotypic alterations in various species including pig [1], cattle [2], and sheep [3,4]. Sheep (Ovis aries) is one of the most important livestock that provide meat, fur, and sheep milk for humans. The history of Chinese sheep domestication is traced back over 5,000 years [3,4]. Chinese sheep originated from the Mongolian Plateau about 5,000 to 5,700 years ago. Afterwards, they might spread in the upper and middle reaches of the Yellow River about 3,000 to 5,000 years ago (3 to 5 kya), and reach the Qinghai-Tibetan and Yunnan-Kweichow plateaus about 2 to 2.6 kya [5]. Up to now, about 42 native sheep breeds have been harbored in China [6].

Mongolian sheep evolved from the wild Argali sheep in Central Asia area, migrated to northern parts of China, and were introduced into other provinces including Gansu, Shan-
Jiergalangtu Sumu, Abaga Banner, Inner Mongolia. Ujimqin Wu Ranke breeding sheep base in the Bayan oobo Gacha, Wu Ranke 4-year-old female breeding sheep were from the 60 individuals comprising 20 WRK, 20 UJMQ, and 20 SNT. Blood samples. Blood samples were collected from a total of went to the local breeding sheep farm and collected sheep To obtain blood from purebred WRK, UJMQ, and SNT, we Sample collection, DNA extraction, and sequencing

MATERIALS AND METHODS

Sample collection, DNA extraction, and sequencing

To obtain blood from purebred WRK, UJMQ, and SNT, we went to the local breeding sheep farm and collected sheep blood samples. Blood samples were collected from a total of 60 individuals comprising 20 WRK, 20 UJMQ, and 20 SNT. Wu Ranke 4-year-old female breeding sheep were from the Wu Ranke breeding sheep base in the Bayan oobo Gacha, Jiergalangtu Sumu, Abaga Banner, Inner Mongolia. Ujimqin 2-year-old female breeding sheep were from the Ujimqin breeding sheep farm of East Ujimqin Banner Dongxing Livestock comprehensive development base in Amuguleng Gacha, Wuqiyasi town, East Ujimqin Banner, Inner Mongolia. Sunite 2-year-old female breeding sheep were from the Sunite breeding sheep farm of Sonid Left Banner in Bayan Erdene Gacha, Dalai Sumu, Sonid Left Banner, Inner Mongolia. For genome sequencing, about 1.5 μg of genomic DNA was used for library construction with an insert size of about 350 bp. Pair-end sequencing were performed on Illumina NovaSeq platform (Illumina, San Diego, CA, USA), with the read length of 150 bp at each end. The resequencing data have been submitted to the SRA database in NCBI with SRA: SRP338713.

Alignment and variation calling

Raw reads performed quality control with Trimmomatic (v0.36) [18] and filtered high quality reads were aligned against the sheep reference genome (Oar_v3.1) using BWA-MEM (version 0.7.12) with default parameters [19]. Single nucleotide polymorphisms (SNPs) and insertion and deletion (INDELS) were called using the Unified Genotyper implemented in GATK (v4.1.2.0) and subsequently filtered using the hard filtering process recommended by GATK [20]. For each locus, the genotype with maximum posterior probability was picked as the genotype for that locus. Variations were annotated with the ENSEMBL Variant Effect Predictor (VEP v98) [21].

Genome-wide association study

Variation and phenotype data were analyzed by PLINK, and phenotypic information including demographic information, age, sex, and breeds, which from a group of individuals. The variants with a minor allele frequency of less than 0.05, missing call frequencies greater than 0.1 and Hardy-Weinberg equilibrium exact test p-value less than 0.00001 were excluded. Then pairwise genome-wide association analysis was performed between three breeds of sheep with PLINK, p-value was adjusted by Bonferroni method. Genome-wide association study (GWAS) results visualized with R package rMVP [22].

Population genetics analysis

The neighbor-joining tree was constructed with filtered variations for all samples based on the nucleotide p-distance matrix then visualized using R package ggtree. Principal component analysis (PCA) analysis was performed with PLINK v1.9 and the first 2 eigenvectors were used for visualizing the results. Population structure was assessed by ADMIXTURE v1.3.0 with default parameters [23]. Furthermore, we used PSMC method to estimate the changes in the effective population (Ne) of sheep over the last 10,000 years [24]. The parame-
ters were set as follows: -N30 -t15 -r5 -p ‘4+25×2+4+6’. An average mutation rate (l) of 2.5×10⁻⁸ per base per generation and a generation time (g) of 3 years based on previous study [17]. Bootstrapping was performed 100 times for each sample.

Selective sweep analysis
Tests for selective sweeps analysis was performed based on Fst (Fixation index), pairwise single-site nucleotide diversity (θ̂n) and Tajima’s D between three breeds of sheep via vcftools (v0.1.15) [25]. θ̂n ratios (θ̂nSNT/θ̂nUJMQ, θ̂nSNT/θ̂nWRK, and θ̂nUJMQ/θ̂nWRK) of those three breeds were calculated with a sliding window approach (100 kb windows with 10 kb increments), and the θ̂n ratios were log₁₀-transformed. Tajima’s D value was also calculated with a sliding-window approach (100 kb windows with 10 kb increments) to assess the selective signature for target genes [26]. The gene ontology (GO) functional and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses of genes with significant signals in GWAS and selective sweep analysis were performed in R with GOstats package [27].

RESULTS
Genetic variation
WRK, SNT, and UJMQ are local sheep breeds, which are mainly distributed in Abaga Banner, Sonid Left and Right Banners, East and West Ujimqin Banners, respectively. In our study, we sequenced 20 WRK, 20 SNT, and 20 UJMQ from local grazing breeding sheep populations in three regions of Inner Mongolia (Figure 1A). There was an average of 10× coverage per individual (~33.75 Gb of high-quality paired-end sequence data) after filtering and quality control, resulting in 13.50 billion mappable reads across 60 sheep.

We further identified about 45.9 million variants mainly including about 39.9 million SNPs and 4.6 million INDELS, which were not significantly different among three groups. The SNPs ranged from 10.7 to 11.8 million SNPs and the average SNPs per sample were 11.2 million (Figure 1B). The homozygosity and heterozygosity numbers were about 5.1 and 0.24 million, respectively (Figure 1C). In addition, the INDELS ranged from 11.2 to 12.7 million INDELS and the average INDELS per sample were 4.57 million (Figure 1D).

 Genome-wide association study among three sheep breeds
The GWAS was performed between three types of sheep breeds. The region of chromosome 16 (1 to 5 Mb) and chromosome 2 (65 to 67 Mb) displayed an ultra-high density of SNPs compared with other regions, which might play an important role on taxonomic and evolutionary implications for those breeds (Figure 2). A total of 468 and 779 significant (p<10⁻⁵) loci were identified from SNT versus UJMQ, and UJMQ versus WRK, respectively. However, only 3 loci were identified from SNT versus WRK, and those variations located in intron or intergenic regions which produced an insignificant effect on molecular phenotype. These results suggested that UJMQ were distinct from WRK and SNT, while SNT were very similar with WRK, similar with the maximum likelihood tree and PCA data.

Population structure and domestication
To further investigate the relationships among three sheep breeds, we performed the maximum likelihood tree based on the pairwise genetic distances and PCA based on their genomic variants among all individuals, revealing that UJMQ clustered into subgroups that were distinct from WRK and SNT, while WRK and SNT were genetically indistinguishable at an overall genomic level (Figures 3A, 3B). This finding is similar with the considerable genetic difference among Mongolian sheep including UJMQ, Tan sheep, Small-tail Han sheep, Hu sheep, and Bayinbuluke sheep [3]. We further performed the population structure which estimates individual ancestry and admixture by assuming the number of ancestry K. With K = 2, UJMQ were separated from WRK and SNT, while WRK showed a mixture with SNT. With K increasing, a clear division was found between UJMQ and WRK or SNT. There were still no obvious division between WRK and SNT, in agreement with the PCA data (Figure 3C). We also explored the demographic history of our samples to differentiate whether domestication of Mongolian sheep including UJMQ, WRK, and SNT, was the result of one or multiple events (Figure 3D). The WRK populations have exhibited concordant demographic trajectories with SNT populations with three apparent contractions, followed by a subsequent expansion since 70 kya which is towards the end of the interglacial period (130 to 70 kya). UJMQ populations differentiated from WRK/SNT populations before the last glacial maximum (LGM) around 26 to 70 kya. Last glacial period (LGP) later, all populations showed similar trajectories. In detail, from a peak of about 45,000 to 50,000 during 45 to 70 kya, the population size of WRK and SNT populations dropped to about 23,000 between 45 to 26 kya during LGP, increased to about 28,000 at about 26.5 kya, and remained stable until 10 kya. Afterwards, SNT populations showed an increase and WRK populations remained stable until 3.5 kya. Differently, from a peak of about 33,000 at 45 kya, the population size of UJMQ dropped between 35-9.5 kya. All populations decreased the population size to about 8,000 between 4-3 kya, and further dropped to a small number between 2.5-2 kya and reached to 15,000-17,500 since about 1.5 kya. Next, we tested multiple demographic models using a diffusion approximation method for the allele frequency spectrum (𝑑𝑎𝑑𝑖). The demographic model also
supported the aforementioned results, and the maximum log composite likelihood value of the optimal model were $-1,157.57$, $-1,151.21$, and $-1,055.20$ (Supplementary Figure S1).

Taken together, based on population genetic structure and demographic history analysis, we concluded that SNT and WRK populations were very similar, which were different from UJMQ populations.

**Selection for domestication**

To investigate the signature of selection for traits related to sheep domestication, we identified the genomic regions with
a high Fst and nucleotide diversity ratio ($\theta_\pi$) between the populations of UJMQ, WRK, and SNT. We measured allele frequency divergence as Fst between species in 100 kb windows with 10 kb increments (Figure 4). Based on Fst>0.25 and selecting the windows with a relatively high or low diversity ratios, i.e., low diversity in the SNT population but high in the UJMQ population. Finally, we identified 292 putative selective sweeps located within 75 genes between UJMQ and WRK, and 262 putative selective sweeps located within 94 genes between SNT and UJMQ, which were under selection during domestication (Figures 4A, 5A). As WRK and SNT were very similar based on analysis mentioned above, only 18 significant loci and 2 genes between WRK and SNT (Supplementary Figure S2).

The GO terms and KEGG pathways were further analyzed (Supplementary Tables S1, S2). In the comparison of SNT to UJMQ, a total of 34 GO terms were significantly enriched, mainly including gamma-aminobutyric acid catabolic/metabolic process, response to cocaine/anesthetic/alkaloid, estrogen biosynthetic/metabolic process, transition metal ion homeostasis, regulation of exocytosis, and other metabolic process. A total of 3 KEGG pathways were significantly enriched and focused on metabolic pathways. From the highlighted GO and KEGG analysis, genes including 4-aminobutyrate aminotransferase (ABAT), ATPase H+ transporting V1 subunit A (ATP6V1A), hydroxysteroid 17-beta dehydrogenase 12 (HSD17B12), rabphilin 3A like (without C2 domains) (RPH3AL), carbamoyl-phosphate synthase 1 (CPS1), antizyme inhibitor 2 (AZIN2), dimethylglycine dehydrogenase (DMGDH), diacylglycerol kinase beta (DGKB),

Figure 2. GWAS among three sheep breeds. (A) Manhattan plot of analysis results. (B) The number of SNPs between UJMQ and WRK. (C) The number of SNPs between SNT and UJMQ. GWAS, genome-wide association study; SNP, single nucleotide polymorphism; UJMQ, Ujimqin sheep; SNT, Sunit sheep; WRK, Wu Ranke sheep.
and mannosidase alpha class 1C member 1 (MAN1C1) were identified as being under positive selection in \( F_{st}>0.25 \) (Figure 4B). Particularly, RPH3AL and CPS1 showed high \( F_{st} \) and \( \theta_{\pi} \) value but lower Tajima’s D values compared to neighboring regions, indicating functional importance (Figures 4C, 4D).

In the comparison of UJMQ to WRK, a total of 86 GO terms were significantly enriched, mainly including metabolite biosynthetic process, response to cGMP/cAMP, and neuron development and differentiation. KEGG pathways including Long-term potentiation, phosphatidylinositol signaling system, calcium signaling pathway, and Alzheimer’s disease, were significantly enriched. From the highlighted GO and KEGG analysis (Supplementary Tables S3, S4), genes including FA complementation group C (FANCC), glutaredoxin (GLRX), integrator complex subunit 7 (INTS7), inositol 1,4,5-trisphosphate receptor type 2 (ITPR2), poly(ADP-ribose) glycohydrolase (PARG), phosphodiesterase 1A (PDE1A), Rap guanine nucleotide exchange factor 2 (RAPGEF2), serine/threonine kinase 10 (STK10), glutamate ionotropic receptor NMDA type subunit 2B (GRIN2B), and diacylglycerol kinase bet (DGKB) were identified as being under positive selection in \( F_{st}>0.25 \) (Figure 5B). Importantly, STK10 and FANCC showed high \( F_{st} \) and \( \theta_{\pi} \) value, but lower Tajima’s D values compared to neighboring regions, suggesting that a strong selective sweep occurred in these genes (Figures 5C, 5D).

**DISCUSSION**

The availability of whole-genome resequencing contributes to exploring genetic basis of animal domestication at the genomic level. Recently, population structure, domestication, and selection in wild and improved sheep with distinct phe-
notypes have caused more attention. The Mongolian sheep has been the most widely distributed in China, mainly including Inner Mongolia areas, eastern coastal regions, and the Central Plains, due to their great adaptability to the environment and Genghis Khan’s expedition in the Yuan dynasty [28]. Different Mongolian sheep breeds including WRK, SNT, UJMQ, and Hulunbuir sheep are living in different regions of Inner Mongolian Grassland, leading to their unique

Figure 4. Genomic regions with strong selective sweep signals between SNT and UJMQ. (A) Distribution of \(-\log_{10}(\theta_\pi/\theta_s)\) ratios and Fst values calculated with a sliding window approach (100 kb windows with 10 kb increments). (B) Genome-wide distribution of global Fst. (C) \(-\log_{10}(\theta_\pi/\theta_s)\) ratios and Fst values around the RPH3AL locus. Tajima’s D values around the RPH3AL locus. The red line and blue line indicate Fst and \(\theta_\pi/\theta_s\) ratios, respectively. (D) \(-\log_{10}(\theta_\pi/\theta_s)\) ratios and Fst values around the CPS1 locus. Tajima’s D values around the CPS1 locus. The green line and purple line indicate UJMQ and WRK, respectively. UJMQ, Ujimqin sheep; SNT, Sunit sheep; WRK, Wu Ranke sheep.
meat quality and flavors. However, few reports focused on comparison of these Mongolian sheep breeds at the genomic level. In previous report, Ganbold et al [29], performed phylogenetic analysis among 88 Mongolian native sheep and other breeds located in different geographic regions in Eurasia and Africa by displacement loop region in mtDNA, which

Figure 5. Genomic regions with strong selective sweep signals between UJMQ and WRK. (A) Distribution of $-\log_{10}(\theta_\pi)$ ratios and Fst values calculated with a sliding window approach (100 kb windows with 10 kb increments). (B) Genome-wide distribution of global Fst. (C) $-\log_{10}(\theta_\pi)$ ratios and Fst values around the FANCC locus. Tajima’s D values around the FANCC locus. The red line and blue line indicate Fst and $\theta_\pi$ ratios, respectively. (D) $-\log_{10}(\theta_\pi)$ ratios and Fst values around the STK10 locus. Tajima’s D values around the STK10 locus. The green line and purple line indicate UJMQ and WRK, respectively. UJMQ, Ujimqin sheep; WRK, Wu Ranke sheep.
was an ingenious method to reveal the genetic differentiation and relationship. However, the amount of data limited the accuracy for the population structure and genetic differentiation. Whole genome sequencing can provide more reliable information for further excavation with public datasets. In this study, we performed whole-genome resequencing of 60 sheep from 3 Chinese indigenous Mongolian sheep breeds including UJMQ, SNT, and WRK. To our knowledge, this is the first whole-genome study to systematically illuminate genetic variation, population structure, domestication, and selection among Chinese indigenous Mongolian sheep breeds in different regions around Xilingol Grassland in China.

We focused on the relationships among three sheep breeds and found that SNT and WRK populations were very similar at the genomic level, suggesting that SNT and WRK might be from the same Mongolian sheep population in history and be distributed in west regions, and central-north areas of Xilingol Grassland in China, respectively. UJMQ living in northeast of Xilingol Grassland were a bit different from them in population structure and domestication. At the end of the interglacial period and during LGP (about 45 to 70 kya), UJMQ population size was lower than that in SNT and WRK. Subsequently, all these populations declined at about 45 kya. SNT and WRK showed a larger reduction than UJMQ and then slightly increased, leading to the similar population size among them until the LGM. The glaciations became less extensive and grassland expansion has occurred during LGM [6,30,31], which contributed to the expansion of grassland livestock. Consistently, SNT and WRK population sizes increased at the beginning of LGM (about 26.5 kya). After the LGP, all populations showed two apparent contradictions, one being 4 to 3 kya and the other being 2.5 to 2 kya. The phenomena may have resulted from Mongolian sheep spreading in the upper and middle reaches of the Yellow River about 3 to 5 kya and reaching the Qinghai-Tibetan and Yunnan-Kweichow plateaus about 2 to 2.6 kya from the Mongolian Plateau [5]. Later, a bottleneck with the lowest population size occurred at about 2.3 to 1.5 kya and the population size increased to about 17,000 at 1.5 kya.

To identify genomic regions that have been the selection target during domestication, we screened 94 genes between SNT and UJMQ, and 75 genes between UJMQ and WRK, based on $Fst>0.25$. Between SNT and UJMQ, 94 genes were closely associated with metabolic pathways, implying the selection of these functionalities during domestication. Two loci, CPS1 and RPH3AL, exhibited strong signs of selective sweeps presumably related to domestication. CPS1, short for Carbamoyl-Phosphate Synthase 1, provides instructions for making the enzyme carbamoyl phosphate synthetase I. Sheep CPS1 was associated with amino acid metabolism including alanine, aspartate and glutamate metabolism, and arginine and proline metabolism. It has been reported CPS1 is a liver-specific enzyme converting ammonia to carbamoyl phosphate in the first step of the urea cycle [32,33]. RPH3AL that is a Rab effector was mainly related to exocytosis and regulation of secretion/exocytosis in sheep, which was consistent with previous studies in the mouse model [34]. To better understand the molecular characteristics of genes regulating some phenotypes is important. For instance, the polymorphism of melanocortin 1 receptor (MC1R) coding region in different coat colors in Mongolian goats have been investigated [35]. In the future, we can explore the polymorphism of some genes under positive selection to investigate deeply the molecular similarities and differences between UJMQ and SNT/WRK.

Xilingol region has different grassland types, such as, meadow steppes, typical steppes, and desert steppe, distributed in different areas [36]. Each steppe includes different construction plant species [9]. For instance, construction plant species in desert steppes includes Allium polyrhizum Turcz. ex Regel and several species of Stipa. Stipa baicalensis Roshev, Bothriochloa ischaemum L. Keng, Cleistogenes mucronata Keng, L. angustum (Trin.) Pilger, Leymus chinensis Tzvel., and Folifolium sibiricum are mainly in meadow steppes. Typical steppes mainly include Cleistogenes squarrosa (Trin.) Keng, Agropyron cristatum (L.) Gaertn, Leymus chinensis Tzvel., A. intramongolica HC Fu & ZY Zhu, Artemisia frigida Willd, A. gmelinii Webb ex Stechmann, and Thymus mongolicus Ronn [9]. Importantly, west parts of Xilingol League where SNT mainly are living belong to the desert steppe. Central-north areas in which WRK are mainly living belong to the typical steppe. While, northeast areas where UJMQ are distributed belong to the typical steppe and meadow steppe [36]. Thus, SNT, UJMQ, and WRK are distributed in regions with various grassland types and feed on different plants as forages, implying different gene profiles regulating metabolite biosynthetic process, which is consistent with our findings. The SNT and WRK varieties were genetically similar throughout the study despite the difference in the main habitats of desert steppes and typical steppes, respectively. In addition to different forages, these sheep are affected by a series of factors including climate, geography, and environment during evolution, which may lead to genetically very similar results between UJMQ and SNT/WRK varieties.

Taken together, according to the whole-genome resequencing, we found that WRK and SNT populations were very similar, which were different from UJMQ populations. Genes associated with regulation of secretion, metabolic pathways, and neuron development have undergone strong selection during domestication. These results will contribute to understanding Chinese indigenous Mongolian sheep breeds domestication and selection for complex traits and provide a valuable genomic resource for future studies of sheep and...
other domestic animal breeding.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

FUNDING

We thank all the staff in the local breeding sheep farm for their participation. This work was supported by the Science and Technology Major Project of Inner Mongolia Autonomous Region of China (zdxx2018016; zdxx2018065), and Grassland talent innovative team of Inner Mongolia Autonomous Region of China (12000-12102621).

SUPPLEMENTARY MATERIAL

Supplementary file is available from: https://doi.org/10.5713/ab.21.0569

Supplementary Figure S1. Two-dimensional comparison for joint estimation between SNT and UJMQ (a), WRK and UJMQ (b), SNT and WRK (c) sheep breeds populations.

Supplementary Figure S2. Genome wide analysis of global Fst between WRK and SNT populations.

Supplementary Table S1. GO terms between SNT and UJMQ

Supplementary Table S2. KEGG between SNT and UJMQ

Supplementary Table S3. GO terms between UJMQ and WRK

Supplementary Table S4. KEGG between UJMQ and WRK

REFERENCES