



Genetic diversity and population structure in five Inner Mongolia cashmere goat populations using whole-genome genotyping

Tao Zhang^{1,2}, Zhiying Wang^{1,3,4}, Yaming Li^{1,3,4}, Bohan Zhou^{1,3,4}, Yifan Liu^{1,3,4}, Jinquan Li^{3,4}, Ruijun Wang^{1,3,4}, Qi Lv^{1,3,4}, Chun Li⁵, Yanjun Zhang^{1,3,4,*}, and Rui Su^{1,3,4,*}

* Corresponding Authors:

Yanjun Zhang

Tel: +86-13190601108,

Fax: +86-0471-430-9178,

E-mail: imauzyj@163.com

Rui Su

Tel: +86-15947510736,

Fax: +86-0471-430-9178,

E-mail: suruiyu@126.com

¹ College of Animal Science, Inner Mongolia Agricultural University, Hohhot, Inner Mongolia Autonomous Region, 010018, China

² Grassland Research Institute, Chinese Academy of Agricultural Sciences, Hohhot, Inner Mongolia Autonomous Region, 010018, China

³ Inner Mongolia Key Laboratory of Sheep and Goat Genetics Breeding and Reproduction, Hohhot, Inner Mongolia Autonomous Region, 010018, China

⁴ Key Laboratory Of Mutton Sheep and Goat Genetics And Breeding, Ministry of Agriculture And Rural Affairs, Hohhot, Inner Mongolia Autonomous Region, 010018, China

⁵ College of Animal Science and Technology, Inner Mongolia Minzu University, Tongliao, Inner Mongolia Autonomous Region, 028000, China

ORCID

Tao Zhang

<https://orcid.org/0009-0008-6101-6002>

Yaming Li

<https://orcid.org/0009-0003-9392-6776>

Bohan Zhou

<https://orcid.org/0009-0006-3598-5606>

Yifan Liu

<https://orcid.org/0009-0003-4368-7324>

Jinquan Li

<https://orcid.org/0000-0001-9159-5916>

Rui Su

<https://orcid.org/0000-0002-1513-8921>

Ruijun Wang

<https://orcid.org/0000-0003-3477-1105>

Qi Lv

<https://orcid.org/0000-0002-1513-8921>

Chun Li

<https://orcid.org/0000-0002-2436-8366>

Zhiying Wang

<https://orcid.org/0000-0001-8710-319X>

Yanjun Zhang

<https://orcid.org/0000-0002-5763-9118>

Submitted Oct 17, 2023; Revised Dec 3, 2023;

Accepted Jan 26, 2024

Objective: As a charismatic species, cashmere goats have rich genetic resources. In the Inner Mongolia Autonomous Region, there are three cashmere goat varieties named and approved by the state. These goats are renowned for their high cashmere production and superior cashmere quality. Therefore, it is vitally important to protect their genetic resources as they will serve as breeding material for developing new varieties in the future.

Methods: Three breeds including Inner Mongolia cashmere goats (IMCG), Hanshan White cashmere goats (HS), and Ujimqin white cashmere goats (WZMQ) were studied. IMCG were of three types: Aerbas (AEBS), Erlangshan (ELS), and Alashan (ALS). Nine DNA samples were collected for each population, and they were genomically re-sequenced to obtain high-depth data. The genetic diversity parameters of each population were estimated to determine selection intensity. Principal component analysis, phylogenetic tree construction and genetic differentiation parameter estimation were performed to determine genetic relationships among populations.

Results: Samples from the 45 individuals from the five goat populations were sequenced, and 30,601,671 raw single nucleotide polymorphisms (SNPs) obtained. Then, variant calling was conducted using the reference genome, and 17,214,526 SNPs were retained after quality control. Individual sequencing depth of individuals ranged from 21.13× to 46.18×, with an average of 28.5×. In the AEBS, locus polymorphism (79.28) and expected heterozygosity (0.2554) proportions were the lowest, and the homologous consistency ratio (0.1021) and average inbreeding coefficient (0.1348) were the highest, indicating that this population had strong selection intensity. Conversely, ALS and WZMQ selection intensity was relatively low. Genetic distance between HS and the other four populations was relatively high, and genetic exchange existed among the other four populations.

Conclusion: The Inner Mongolia cashmere goat (AEBS type) population has a relatively high selection intensity and a low genetic diversity. The IMCG (ALS type) and WZMQ populations had relatively low selection intensity and high genetic diversity. The genetic distance between HS and the other four populations was relatively high, with a moderate degree of differentiation. Overall, these genetic variations provide a solid foundation for resource identification of Inner Mongolia Autonomous Region cashmere goats in the future.

Keywords: Cashmere Goat; Genetic Diversity; Inner Mongolia Autonomous Region; Population Structure

INTRODUCTION

Archaeological and genetic evolutionary evidence suggests that goats were domesticated in the early Neolithic period (10,000 years ago), and were the first domesticated ruminants [1,2]. Goats are mainly distributed in developing and underdeveloped countries, such as Mongolia, Iran, India, and Afghanistan, and including China, which has abundant goat genetic resources. China is ranked first in the number of cashmere goats raised, produc-

tion, quality and cashmere trade [3,4]. Regions with relatively high cashmere production are concentrated in Inner Mongolia, Liaoning, Xinjiang, Qinghai, Gansu, Henan, and Tibet. Inner Mongolia cashmere goats are well-known for their high cashmere production and quality, and large body size. They are the largest Chinese cashmere goat population and have played a crucial role in goat breeding in the country. Additionally, cashmere goats are key protected animals in China and are strictly prohibited from export. Inner Mongolia has three cashmere goat breeds named by the National Genetic Resources Commission including Inner Mongolia, Hanshan White and Ujimqin White. Among them, there are three types of Inner Mongolia cashmere goats including: Aerbas (AEBS), Erlangshan (ELS), and Alashan (ALS) [5,6]. As a charismatic species, goats play an important role in the animal husbandry economy, especially in remote mountain areas. Due to a lack of genetic resource conservation, the genetic diversity of many goat breeds is continuously declining. Goat genetic resource diversity is related to national ecological and agricultural security. Protecting cashmere goat genetic resources and ensuring population genetic diversity will help protect special traits, maintain working cashmere goat breeds and promote sustainable development of the modern cashmere goat industry [7].

Genetic diversity generally refers to the sum of genetic variation of different individuals within a breed, usually with single nucleotide polymorphism (SNP) as the minimum unit [8]. At the earliest stage, genetic markers are obtained by restricting fragment length polymorphisms and microsatellite polymorphisms [9]. For now, SNP markers are the most common in the genome, characterized by large number and high density. More than 80% of genome genetic variation is caused by SNPs. Numerous studies have analyzed genetic diversity and population structure based on whole genome re-sequencing or SNPs chip data [10-12].

Although many studies on species genetic diversity and conservation of genetic resources have been conducted, there are relatively few on goat genetic diversity based on high-depth whole genome re-sequencing data, and none on Inner Mongolia cashmere goats. Therefore, for the first time, we determine the genetic diversity and population structure among native goat breeds in the Inner Mongolia Autonomous

Region using whole-genome re-sequencing genotyping. This is vitally important for protecting the local cashmere goat populations.

MATERIALS AND METHODS

Sample collection

DNA samples were obtained from five populations of Inner Mongolia cashmere goats lodged in the sheep&goats Genetic Resource Bank of the Inner Mongolia Agricultural University, including Inner Mongolia cashmere goats (Aerbas type, AEBS; Erlangshan type, ELS; Alashan type, ALS), Hanshan white cashmere goats (HS) and Ujimqin white cashmere goats (WZMQ). The samples were from nine adult does in each population, all of them from the original breeding farm (Table 1). All DNA samples were measured for concentrations and optical density (OD) values using a nucleic acid protein analyzer. Concentrations were >50 ng/μL and OD values between 1.6 and 1.9, so all samples were used in subsequent analyses.

Whole genome resequencing, quality control and single nucleotide polymorphism calling

DNBSEQ-T7 PE150 was used to construct sequencing libraries for each genomic DNA sample. Reads quality obtained by re-sequencing was controlled. Reads containing adaptor sequences and with N base content >10% were removed. Additionally, when the number of bases with a quality value <5 in any read exceeded 50% of the whole read, the paired reads were removed. BWA (Burrows-Wheeler Alignment) software (parameter MEM-T4-K32-m) was used to align the clean reads to the assembled high-quality reference genome in our laboratory (Youngmok and Dongsu, in prep.). Replicate reads were removed using rmbup commands in Samtools software. The Samtools mpileup module (-q 1 -C 50 -t SP-t DP-m 2) was used to load all sample alignment files, and then the bcftools (-mv-f GQ) call module was used to resolve [13]. 30,601,671 raw SNPs were obtained. The vcfutils varFilter module (-Q 20-d 2-D 100000) was used for quality control, and the original mutation sites of each sample were finally obtained. SNPs with $7 < \text{depth} \leq 138$, $\text{miss} < 0.1$ and $\text{maf} > 0.05$ were filtered and screened, provid-

Table 1. Characteristics of the Inner Mongolia cashmere goats sampled from different populations

Name	Sex	No.	Stock
AEBS	Female	9	Inner Mongolia Yiwei White Cashmere Goat Co., Ltd
ELS	Female	9	Inner Mongolia Mizhen International Trade Co., Ltd
ALS	Female	9	Alxa Left Banner Goat Breeding Farm
HS	Female	9	Chifeng Bahrain Right Banner Hanshan Goat Breeding Farm
WZMQ	Female	9	Ujimqin White Goat Original Breeding Farm

AEBS, Inner Mongolia cashmere goats (Aerbas type); ELS, Inner Mongolia cashmere goats (Erlangshan type); ALS, Inner Mongolia cashmere goats (Alashan type); HS, Hanshan white cashmere goats; WZMQ, Ujimqin white cashmere goats.

ing about 17,214,526 high-quality SNPs (Table 2). Then, ANNOVAR software was used to perform functional annotation on the detected gene variants [14].

Genetic diversity analysis

The stacks software population module was used to analyze the expected heterozygosity (H_e), observed heterozygosity (H_o) and proportion of polymorphic sites (Pn%) of the filtered high-quality SNPs [15]. The Plink software genome method was used to obtain the ratio of Identity by Descent (PI_HAT) [16]. The SNPs average inbreeding coefficient (F) was performed using a python script ($F = 1 - H_o/H_e$).

Principal component analysis, phylogenetic tree construction and population structure analysis

The GCTA software `gcta64 --grm cashmeregoats_grm --pca 3 --out cashmeregoats_pca` command was used to perform principal component analysis (PCA) of SNPs for the five populations [17]. The three largest eigenvectors generated by PCA analysis were taken as the principal axes, and the R plot function was used to draw PCA plots [18]. The population genetic distance of the filtered high-quality SNPs was calculated, and then the TreeBest software NJ method was used to construct the SNPs phylogenetic tree. Finally, the population phylogenetic tree was drawn on the itol website (<https://itol.embl.de/>). The admixture software admixture method was used to analyze the SNPs population structure [19], and the barplot function in R software was used to plot the SNPs

Table 2. Single nucleotide polymorphism annotation results of five Inner Mongolia cashmere goat populations

Category	Number of SNPs
Upstream	91,647
Exonic	
UTR3	70,383
UTR5	12,741
UTR5;UTR3	17
Stop gain	670
Stop loss	123
Synonymous	62,948
Non-synonymous	46,878
Unknown	256
Intronic	4,859,122
Splicing	232
Downstream	94,018
Upstream/downstream	4,633
Intergenic	11,970,858
ts	12,126,852
tv	5,087,674
ts/tv	2.383
Total	17,214,526

SNPs, single nucleotide polymorphisms; UTR, untranslated region; upstream/downstream, the upstream of a gene and downstream of another gene; ts, transitions; tv, transversions; ts/tv, the ratio of transitions to transversions.

lineage composition.

Linkage disequilibrium decay analysis

Linkage disequilibrium (LD) decay was analyzed using the PopLDdecay (version 3.40) software PopLDdecay method, and the LD decay map was plotted using the Plot_MultiPop method [20].

Population genetic differentiation

Different sub-populations are affected by environments or by artificial selection, resulting in an increase or decrease in genotype frequency of certain loci. Therefore, at this locus, the magnitude of the difference between sub-populations is higher than that between normal sub-populations. Significantly different SNPs are considered to have been selected. Generally, F_{st} is used to indicate the size of the difference between sub-populations, with higher F_{st} indicating greater sub-population differences. We used vcftools software to calculate F_{st} between sub-populations [21], and also to determine θ_π which indicates population nucleotide diversity fractions on SNPs. Usually, a lower polymorphism indicates a higher degree of selection.

RESULTS

Whole genome resequencing

After library construction and sequencing with DNBSEQ-T7 PE150, raw data were produced for each individual (Supplementary Table S1). The sequencing data of 45 individuals in the five goat populations showed that Q20 and Q30 were >98% and $\geq 93\%$, respectively. Guanine and cytosine (GC) content was between 42% and 44%, which indicated that sample quality was good and samples could be used for subsequent analysis. The re-sequencing data of the five populations were compared with the reference genome. The mapping rate was between 96.83% to 99.95%, average sequencing depth was 28.5 \times , and range of individual sequencing depth was 21.13 \times to 46.18 \times (Supplementary Table S2). The number of raw SNPs was about 30,601,671, and after filtering and screening, 17,214,526 high-quality SNPs were finally obtained.

SNP call and annotation showed that 91,647 SNPs were found in the 1 kb region upstream of the genes, and 194,016 SNPs were found in the exon region, including 670 stop gain, 123 stop loss, 62,948 synonymous mutations, and 46,878 non synonymous mutations. The intron region contained 4,859,122 SNPs, and the gene interval 11,970,858 SNPs. The downstream 1 kb region contained 94,018 SNPs and 232 SNPs at the splice site mutation (Supplementary Table S1).

Genetic diversity analysis

Genetic diversity parameters were estimated using filtered

high quality SNPs (Table 3). Proportion of polymorphic loci (P_N) values of the five populations ranged from 79.28% to 88.57%, with the largest in the ALS population ($P_N = 88.57$) and the smallest in the AEBS population ($P_N = 79.28$). The AEBS population inbreeding coefficient was relatively high (0.1348), but for the WZMQ, ELS, and ALS populations it was low (F_{is} from 0.0731 to 0.0773). The HS inbreeding coefficient was the highest among the five populations. Overall, selection intensity in AEBS was relatively high, while in HS it was low. PI_HAT reflects the degree of genetic relationship between two individuals. Among the five populations, the highest and lowest PI_HAT were 0.1021 and 0.010, for the AEBS and ALS populations, respectively. MAF of these five populations ranged from 0.1859 to 0.1941. The observed heterozygosity (H_o) and expected heterozygosity (H_e) of the five populations were relatively low, ranging from 0.2755 to 0.2886, and from 0.2554 to 0.2721, respectively. Observed heterozygosity was highest in ELS, and lowest in HS, while expected heterozygosity, was lowest in AEBS, and highest in

ALS.

Principal component analysis, phylogenetic tree and population structure

Principal component analysis indicated that the genetic distance between HS and the other four populations was relatively high, and it was clearly separated from other populations. The genetic distances between ALS and WZMQ, and AEBS and ELS were relatively close (Figure 1A). The phylogenetic tree showed that the five populations could be divided into two branches, with the ALS population closest to the root. The HS and AEBS populations were separated based on the WZMQ and ELS populations, respectively (Figure 1B).

Population structure was used to analyze genetic differences between sub-populations (Figure 2). There was an increasing trend of cross-validation error values according to k-values, and cross-validation error values ranged from 0.5572 to 0.9042. The maximum likelihood value was 2, indicating that the five cashmere goat populations should be

Table 3. Estimates of genetic diversity parameters of five Inner Mongolia cashmere goat populations

Breed	Number	P_N	PI_HAT	MAF	H_e	H_o	F_{is}
AEBS	9	79.28	0.1021	0.1859	0.2554	0.2806	0.1348
ELS	9	85.50	0.0702	0.1915	0.2660	0.2886	0.0773
ALS	9	88.57	0.0109	0.1941	0.2721	0.2842	0.0731
HS	9	80.98	0.0398	0.1878	0.2590	0.2755	0.1404
WZMQ	9	87.02	0.0411	0.1928	0.2697	0.2878	0.0737

P_N , proportion of polymorphic marker; PI_HAT, the proportion of identity by descent; MAF, minor allele frequency; H_e , expected heterozygosity values; H_o , observed heterozygosity values; F_{is} , inbreeding coefficients; AEBS, Inner Mongolia cashmere goats (Aerbas type); ELS, Inner Mongolia cashmere goats (Erlangshan type); ALS, Inner Mongolia cashmere goats (Alashan type); HS, Hanshan white cashmere goats; WZMQ, Ujimqin white cashmere goats.

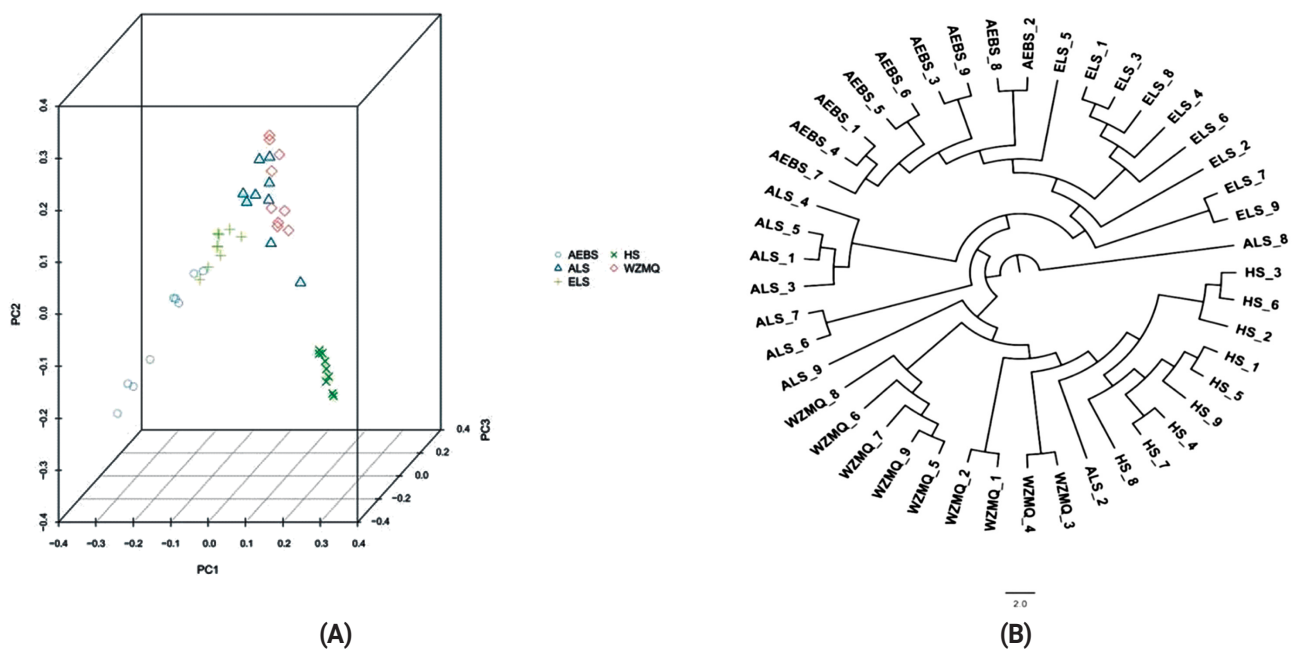


Figure 1. (A) Multidimensional scaling plot depicting the relationships among the five populations. PC1, first component; PC2, second component. (B) Phylogenetic tree.

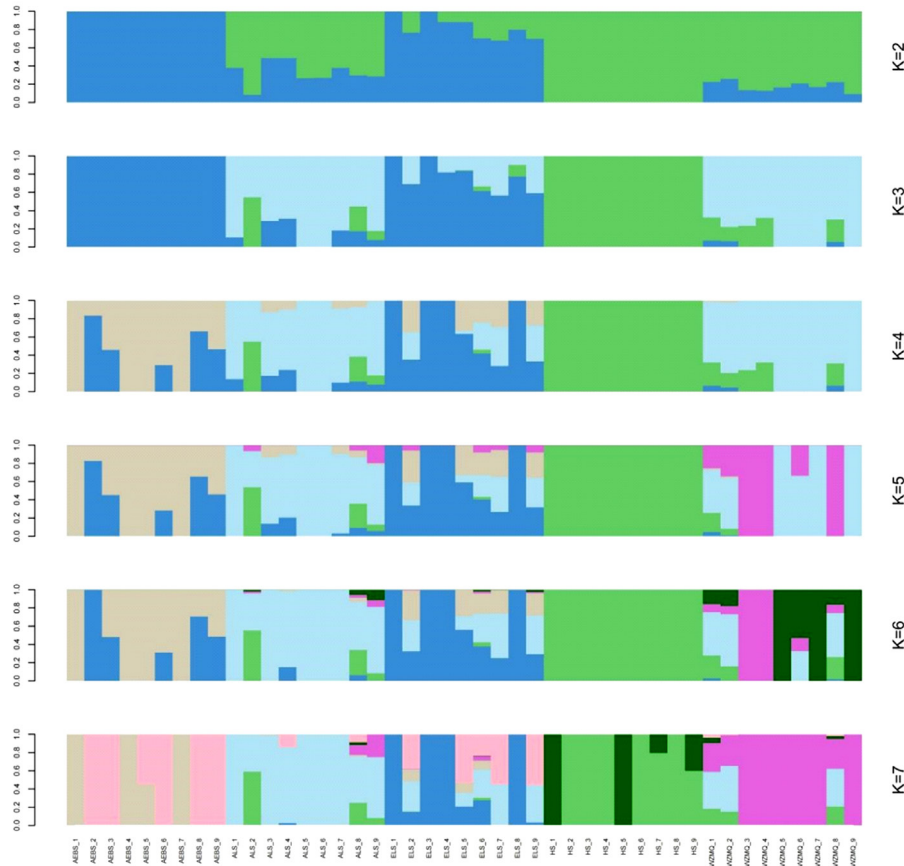


Figure 2. Admixture analysis of cashmere goat populations for a range of K-values (2 to 7).

divided into two subgroups. When $K = 2$ and $K = 3$, the AEBS and HS populations were relatively independent, and there was a certain degree of genetic hybridization among the ALS, ELS, and WZMQ populations. When $K = 4, 5,$ and 6 , the HS group was relatively independent, while other populations had a certain degree of confusion. Overall, the five goat populations in the Inner Mongolia Autonomous Region should be divided into two subpopulations, consistent with the principal components and phylogenetic tree analyses.

Estimation of population genetic differentiation parameters

Variations in the genetic differentiation parameter (F_{st}) represent the magnitude of difference between two populations (0 to 1), where the larger the F_{st} value, the greater the difference. A moderate degree of genetic differentiation was found between HS and ELS, HS and AEBS, AEBS and ALS, and AEBS and WZMQ. The degree of differentiation between AEBS and HS was relatively high, while between the other populations was relatively low and not obvious (Figure 3).

The nucleotide polymorphism value (θ_{π}) represents the degree of nucleotide polymorphism among individuals in

the same population, and the higher the value of θ_{π} , the greater the polymorphism. Nucleotide diversity values of the

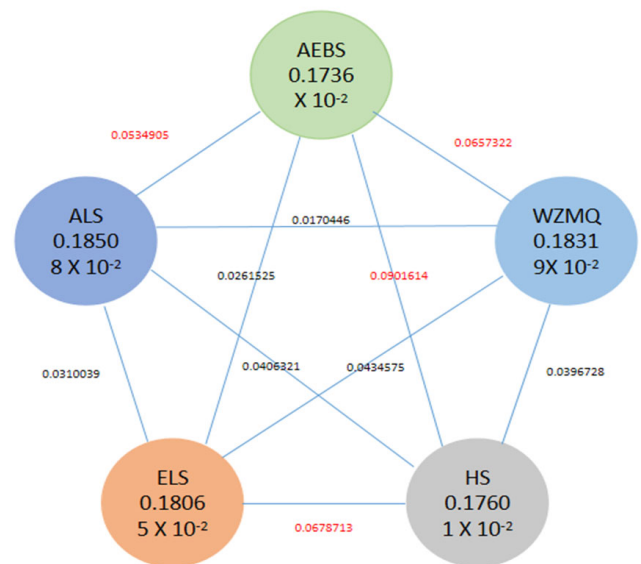


Figure 3. Population genetic differentiation parameters and nucleotide diversity. F_{st} , the connection value between two groups; θ_{π} , the corresponding circle within the population.

five populations from high to low were ALS>WZMQ>ELS>HS>AEBS, but there was no significant difference in genetic diversity among the five populations (Figure 3).

We analyzed the degree of LD of SNPs within 300 kb among the five cashmere goat populations. The LD had a downward trend with increasing physical distance between SNPs (Figure 4; Table 4). Within a physical distance of 1 to 50 kb, LD decay degree from high to low was AEBS>HS>ELS>WZMQ/ALS. After 150 kb, the declining degree of LD ranges from high to low was AEBS>ELS>HS>WZMQ/ALS. Within a range of 1 to 300 kb, the AEBS population consistently maintained the highest LD and had the slowest decay rate, indicating low population genetic diversity. The difference in LD between ELS and HS was not significant. In the range of 1 to 50 kb, the decay rate of ELS was slightly higher than the HS population, and the ELS LD value was relatively low, indicating that its population genetic diversity was slightly enriched. ALS and WZMQ LD values and decay rates were almost identical over the entire distance range, both maintaining the fastest decay rate and the lowest LD, indicating the richest genetic diversity and the least artificial selection.

DISCUSSION

Genetic diversity parameter estimates indicated that proportion of polymorphism and expected heterozygosity were lowest in the AEBS population, and the proportion of ho-

Table 4. Linkage imbalance decay trend

Distance class (kb)	ALS	AEBS	HS	WZMQ	ELS
0-20	0.423	0.484	0.467	0.429	0.446
20-40	0.377	0.438	0.421	0.382	0.400
40-60	0.367	0.429	0.411	0.373	0.390
60-80	0.363	0.425	0.406	0.369	0.387
80-100	0.362	0.424	0.405	0.367	0.386
100-120	0.362	0.424	0.405	0.367	0.386
120-140	0.362	0.424	0.405	0.367	0.386
140-160	0.362	0.423	0.405	0.366	0.385

ALS, Inner Mongolia cashmere goats (Alxa type); AEBS, Inner Mongolia cashmere goats (Aerbas type); HS, Hanshan white cashmere goats; WZMQ, Ujimqin white cashmere goats; ELS, Inner Mongolia cashmere goats (Erlangshan type).

mology and average inbreeding coefficient were the highest. This may be due to the relatively large selection intensity for this population. A breeding program for this population began in 1998, and production performance has significantly increased. The average AEBS population inbreeding coefficient was 0.1348, indicating that a certain proportion of inbreeding occurred within three generations. In the ALS and WZMQ populations, the proportion of polymorphism and expected heterozygosity were relatively high, and the proportion of homology and average inbreeding coefficient were relatively low, indicating that both populations were under relatively low selection intensity. Furthermore, the aver-

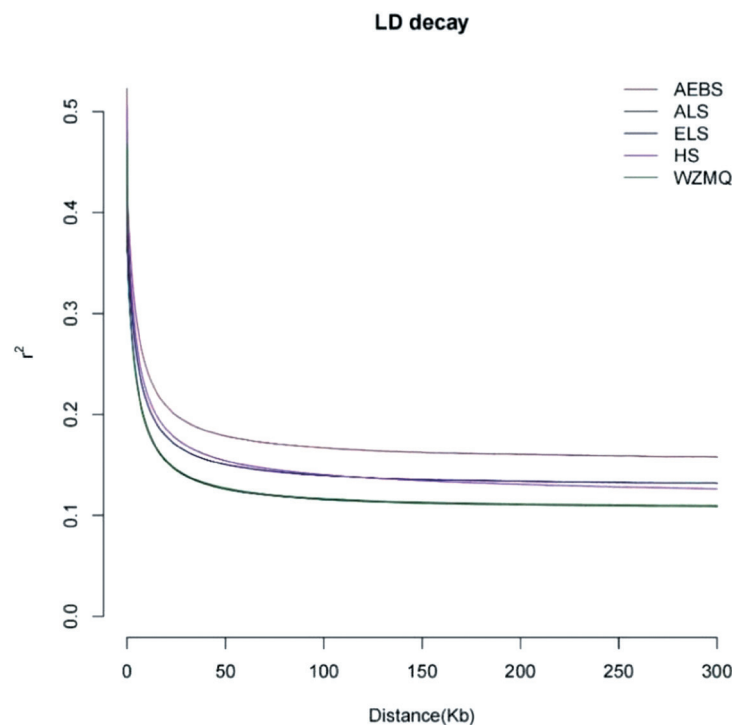


Figure 4. Genome-wide average LD decay estimated for five Inner Mongolia cashmere goat populations. The vertical axis denotes the LD coefficient r^2 and the horizontal axis denotes the distance between the genes. LD, linkage disequilibrium.

age inbreeding coefficient was only 0.07, indicating that the proportion of inbreeding within three generations was very low. As a special population, the proportion of polymorphism in the HS population was high, indicating that it was not under selection, and that the selection scheme needs to be further improved. Generally, degree of heterozygosity is positively correlated with population diversity, with higher heterozygosity indicating richer population diversity [22]. If the average population heterozygosity is >0.5 , it indicates that the population has rich genetic diversity [23]. The observed heterozygosity of the five populations estimated in this study was not significantly different, ranging from 0.2755 to 0.2886. Zhao et al [24] used microsatellite markers to estimate the expected and observed heterozygosity of four Inner Mongolia cashmere goat populations and found that estimated values were much higher than in this study. This may be related to variables such as methods, sample collection time and sample size.

The HS population and the other four populations were significantly separated. However, the genetic distances between WZMQ, AEBS, ALS, and ELS were very close, indicating that genetic exchange was low [25], as confirmed by the population differentiation index. When the F_{st} value is <0.05 , it indicates a small genetic differentiation between populations, while from 0.05 to 0.15 it indicates moderate differentiation [26]. The F_{st} values between any two Inner Mongolia cashmere goat populations were relatively low, ranging from 0.017 to 0.090, indicating small or moderate differentiation. This may be due to long-term closed breeding and continuous selection in the Inner Mongolia cashmere goat population, resulting in relatively homozygous genes. Nonetheless, the Inner Mongolia cashmere goat population has been subjected to genetic resource protection and prevented from hybridizing with other populations.

The AEBS and ALS populations on the phylogenetic tree were clustered into one category, as were the HS and WZMQ populations, while the ALS population intersected these two major categories. This may be explained by geographical factors. Additionally, population structure analysis showed that $K = 2$ when the CV value was lowest, indicating that the five cashmere goat populations were divided into two subgroups, which is consistent with PCA and phylogenetic tree results. Therefore, when $K = 2$, the HS population was independent, and the other four populations had different degrees of genetic intercourse flow.

Decay rate represents population genetic diversity, the faster the decay rate, the richer the population diversity [27-30]. The AEBS population LD decay rate was the slowest, indicating that its genetic diversity was the lowest, which was consistent with the previous genetic diversity parameter estimation results, and might be related to continuous artificial and closed breeding. The WZMQ and ALS populations

had the fastest LD decay rate and higher genetic diversity, which was also confirmed by the population differentiation index.

The population nucleotide diversity value (θ_π) also reflects genetic diversity, the larger the θ_π value, the higher the population diversity. The ALS and WZMQ populations had higher genetic diversity, as indicated by θ_π values of 0.0019 and 0.0018. The AEBS population θ_π value was the lowest, a result consistent with the LD decay map.

CONCLUSION

The estimated genetic diversity parameters of five cashmere goat populations in the Inner Mongolia Autonomous Region showed that the AEBS population had the lowest polymorphic rate and expected heterozygosity, while the average coefficient of kinship and average inbreeding coefficient were the highest, indicating that this population has a high selection intensity and a certain proportion of inbreeding within three generations. Compared to other populations, the ALS and WZMQ polymorphic rate and expected heterozygosity were high, while the average coefficient of kinship and average inbreeding coefficient were relatively low, indicating relatively low selection intensity. Principal component analysis and phylogenetic tree results showed that the genetic distance between HS population and the other four populations was relatively high. Moreover, degree of differentiation between HS and other populations was relatively high, indicating a moderate degree of genetic differentiation. Among them, the HS and AEBS populations had the highest degree of differentiation, while among the AEBS, ELS, and ALS populations, it was relatively low. Linkage disequilibrium decay results showed that the LD decay trend of the five populations (AEBS, ELS, ALS, HS, and WAMQ) was basically consistent with increased distance between SNPs. The AEBS and HS population decay degree rates were relatively low, indicating relatively low genetic diversity and probably frequent artificial 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

The authors are grateful for the grants supported by Science and technology major project of Inner Mongolia Autonomous Region (2021ZD0012), Inner Mongolia Autonomous Region Science and Technology Research Project (2021GG0086), Youth Science and Technology Talent Support Plan" of col-

leges and universities in Inner Mongolia Inner Mongolia Autonomous Region (NJYT22038), Program for Innovative Research Team in Universities of Inner Mongolia Autonomous Region (NMGIRT2322), China Agriculture Research System of MOF and MARA (CARS-39).

ACKNOWLEDGMENTS

Thanks are due to Libing He for assistance with the experiments and to Oljibilig Chen for valuable discussion.

SUPPLEMENTARY MATERIAL

Supplementary file is available from: <https://doi.org/10.5713/ab.23.0424>

Supplementary Table S1. Whole genome resequencing data quality in five Inner Mongolia cashmere goats populations

Supplementary Table S2. Sequencing depth and coverage

REFERENCES

- Zheng Z, Wang X, Li M, et al. The origin of domestication genes in goats. *Sci Adv* 2020;6:eaa5216. <https://doi.org/10.1126/sciadv.aaz5216>
- Islam R, Liu X, Gebreselassie G, Abied A, Ma Q, Ma Y. Genome-wide association analysis reveals the genetic locus for high reproduction trait in Chinese Arbas Cashmere goat. *Genes Genomics* 2020;42:893-9. <https://doi.org/10.1007/s13258-020-00937-5>
- Li X, Su R, Wan W, et al. Identification of selection signals by large-scale whole-genome resequencing of cashmere goats. *Sci Rep* 2017;7:15142. <https://doi.org/10.1038/s41598-017-15516-0>
- Yang X, Li Z. Thoughts on current situation and development of cashmere industry in China. *Wool Text J* 2017;45:84-7. <https://doi.org/10.19333/j.mfkj.20161102511104>
- Zhang Y, Wang Z, Lei H, et al. Estimates of genetic parameters and genetic changes for fleece traits in Inner Mongolia cashmere goats. *Small Rumin Res* 2014;117:41-6. <https://doi.org/10.1016/j.smallrumres.2013.10.011>
- Su R, Gong G, Zhang L, et al. Screening the key genes of hair follicle growth cycle in Inner Mongolian Cashmere goat based on RNA sequencing. *Arch Anim Breed* 2020;63:155-64. <https://doi.org/10.5194/aab-63-155-2020>
- Toro MA, Caballero A. Characterization and conservation of genetic diversity in subdivided populations. *Philos Trans R Soc Lond B Biol Sci* 2005;360:1367-78. <https://doi.org/10.1098/rstb.2005.1680>
- Crawford A, Fassett RG, Geraghty DP, et al. Relationships between single nucleotide polymorphisms of antioxidant enzymes and disease. *Gene* 2012;501:89-103. <https://doi.org/10.1016/j.gene.2012.04.011>
- Grover A, Sharma PC. Development and use of molecular markers: past and present. *Crit Rev Biotechnol* 2016;36:290-302. <https://doi.org/07388551.2014.959891>
- Guo F, Ye Y, Zhu K, et al. Genetic diversity, population structure, and environmental adaptation signatures of chinese coastal hard-shell mussel *Mytilus coruscus* revealed by whole-genome sequencing. *Int J Mol Sci* 2023;24:13641. <https://doi.org/10.3390/ijms241713641>
- Al-Atiyat RM, Alobre MM, Aljumaah RM, Alshaikh MA. Microsatellite based genetic diversity and population structure of three Saudi goat breeds. *Small Rumin Res* 2015;130:90-4. <https://doi.org/10.1016/j.smallrumres.2015.07.027>
- Deniskova TE, Dotsev AV, Selionova MI, et al. Population structure and genetic diversity of 25 Russian sheep breeds based on whole-genome genotyping. *Genet Sel Evol* 2018; 50:29. <https://doi.org/10.1186/s12711-018-0399-5>
- Jung Y, Han D. BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics* 2022;38:2404-13. <https://doi.org/10.1093/bioinformatics/btac137>
- Meseret S, Mekonnen YA, Brenig B, et al. Genetic diversity and population structure of six ethiopian cattle breeds from different geographical regions using high density single nucleotide polymorphisms. *Livest Sci* 2020;234:103979. <https://doi.org/10.1016/j.livsci.2020.103979>
- Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. *Bioinformatics* 2017;33:2037-9. <https://doi.org/10.1093/bioinformatics/btx100>
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164. <https://doi.org/10.1093/nar/gkq603>
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol* 2013;22:3124-40. <https://doi.org/10.1111/mec.12354>
- Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75. <https://doi.org/10.1086/519795>
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76-82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Chan BKC. Data analysis using R programming. *Adv Exp Med Biol* 2018;1082:47-122. https://doi.org/10.1007/978-3-319-93791-5_2
- Montana G, Hoggart C. Statistical software for gene mapping by admixture linkage disequilibrium. *Brief Bioinform* 2007; 8:393-5. <https://doi.org/10.1093/bib/bbm035>
- Deniskova TE, Dotsev AV, Selionova MI, et al. SNP-based genotyping provides insight into the west asian origin of russian local goats. *Front Genet* 2021;12:708740. <https://doi.org/10.3389/fgene.2021.708740>

23. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics* 2011;27:2156-8. <https://doi.org/10.1093/bioinformatics/btr330>
24. Zhao YH, Wang MY, Su R, et al. The genetic diversity analysis on 4 Inner Mongolian Cashmere goats. *Hubei Nong Ye Ke Xue* 2012;8:1681-3.
25. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980;32:314-31.
26. Weir BS, Cockerham CC. Estimating f-statistics for the analysis of population structure. *Evolution* 1984;38:1358-70. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>
27. Mtileni BJ, Muchadeyi FC, Maiwashe A, et al. Genetic diversity and conservation of South African indigenous chicken populations. *J Anim Breed Genet* 2011;128:209-18. <https://doi.org/10.1111/j.1439-0388.2010.00891.x>
28. Wang X, Wang C, Huang M, et al. Genetic diversity, population structure and phylogenetic relationships of three indigenous pig breeds from Jiangxi Province, China, in a worldwide panel of pigs. *Anim Genet* 2018;49:275-83. <https://doi.org/10.1111/age.12687>
29. Zhang M, Han W, Tang H, et al. Genomic diversity dynamics in conserved chicken populations are revealed by genome-wide SNPs. *BMC Genomics* 2018;19:598. <https://doi.org/10.1186/s12864-018-4973-6>
30. Machete JB, Kgwatalala PM, Nsoso SJ, Hlongwane NL, Moreki JC. Genetic diversity and population structure of three strains of Indigenous Tswana chickens and commercial broiler using single nucleotide polymorphisms (SNP) markers. *J Anim Sci* 2021;11:515-31. <https://doi.org/10.4236/ojas.2021.114035>