



RESEARCH ARTICLE

Genetic diversity and population structure of sheep (*Ovis aries*) in Sichuan, ChinaMingliang Zhou¹ , Gaofu Wang² , Minghua Chen¹, Qian Pang¹, Shihai Jiang¹, Jie Zeng¹, Dan Du¹, Pinggui Yang¹ , Weisheng Wu¹, Hongwen Zhao¹**1** Sichuan Academy of Grassland Sciences, Chengdu, Sichuan, China, **2** Chongqing Academy of Animal Sciences, Chongqing, China These authors contributed equally to this work.* 79560827@qq.com

Abstract

Sichuan, China, has abundant genetic resources of sheep (*Ovis aries*). However, their genetic diversity and population structure have been less studied, especially at the genome-wide level. In the present study, we employed the specific-locus amplified fragment sequencing for identifying genome-wide single nucleotide polymorphisms (SNPs) among five breeds of sheep distributed in Sichuan, including three local pure breeds, one composite breed, and one exotic breed of White Suffolk. From 494 million clean paired-end reads, we obtained a total of 327,845 high-quality SNPs that were evenly distributed among all 27 chromosomes, with a transition/transversion ratio of 2.56. Based on this SNP panel, we found that the overall nucleotide diversity was 0.2284 for all five breeds, with the highest and lowest diversity observed in Mage sheep (0.2125) and Butuo Black (0.1963) sheep, respectively. Both Wright's fixation index and Identity-by-State distance revealed that all individuals of Liangshan Semifine-wool, White Suffolk, and Butuo Black sheep were respectively clustered together, and the breeds could be separated from each other, whereas Jialuo and Mage sheep had the closest genetic relationship and could not be distinguished from each other. In conclusion, we provide a reference panel of genome-wide and high-quality SNPs in five sheep breeds in Sichuan, by which their genetic diversity and population structures were investigated.

 OPEN ACCESS

Citation: Zhou M, Wang G, Chen M, Pang Q, Jiang S, Zeng J, et al. (2021) Genetic diversity and population structure of sheep (*Ovis aries*) in Sichuan, China. PLoS ONE 16(9): e0257974. <https://doi.org/10.1371/journal.pone.0257974>

Editor: Vikas Sharma, Sant Baba Bhag Singh University, INDIA

Received: April 30, 2021

Accepted: September 14, 2021

Published: September 28, 2021

Copyright: © 2021 Zhou et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its [Supporting information](#) files.

Funding: This work was financially supported by the Mutton Sheep Innovation Research Team of Sichuan Province: Breeding, Demonstration and Promotion of Mutton sheep in Tibetan Area (CXTD2019-14). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Modern sheep (*Ovis aries*) were domesticated in the Fertile Crescent region of the Near East approximately 8,000–9,000 years ago [1]. The comprehensive analysis by genotyping 50K single nucleotide polymorphism in 2,819 animals from 74 worldwide sheep breeds revealed that they had maintained a relatively high genetic diversity in comparison with other main farm animals [2]. China has relatively abundant sheep genetic resources composed of at least 15 officially recognized breeds and many local populations [3]. Although Chinese sheep breeds may be classified geographically and morphologically classified into three groups, namely Mongolian, Kazakh, and Tibetan, molecular data from 10 representative indigenous sheep breeds

showed that they might actually be divided into two genetic groups [4], that is, the thin (Tibetan group) and the fat (Mongolian and Kazakh group) types. Overall, a large number of sheep breeds are widely distributed in a diverse range of environments with considerable genetic and morphological variation.

According to our previous investigation, there are at least eight local breeds of sheep in Sichuan province, China, and all of them have particular morphological characteristics and geographical distribution. For example, Butuo Black sheep has are geographically distributed in Butuo county of Liangshan Yi Autonomous Prefecture of southwestern Sichuan; the average adult male and female body weights are 53.51 ± 8.44 kg and 39.08 ± 7.63 kg, respectively [5]. Mage is a recently recognized and small dual-purpose (wool/meat) sheep breed in Derong county of Ganzi Tibetan Autonomous Prefecture. In addition to these local breeds, many exotic sheep breeds, such White Suffolk, have been raised because of their high production performance. However, to the best of our knowledge, there has been no previous study that systematically investigated the genetic diversity and population structures of these local sheep breeds in Sichuan.

Given the rapid advances in high-throughput DNA sequencing technologies [6], it has become EASIER and more cost-effective to identify a large number of genome-wide SNPs that could be used to study genetic diversity and population structures. Among these technologies, specific-locus amplified fragment sequencing (SLAF-seq) could produce the evenly distributed SNPs through whole genome at a very low cost [7]. In the present study, we also employed the SLAF-seq to investigate the genetic diversity and population structures of five sheep breeds distributed in Sichuan, including three local pure populations.

Materials and methods

Ethics statement

Sample collection and study purposes were approved by the Research Committee of Sichuan Academy of Grassland Sciences, Sichuan, China.

Breeds and animals

We collected a total of 69 blood samples from five different sheep breeds in the Sichuan province (Fig 1), consisting of (1) three local pure breeds, namely Butuo Black (BT, N = 9), Jialuo (JL, N = 8), and Mage (LT, N = 19), (2) a composite Liangshan Semifine-wool sheep (LS, N = 19), and (3) the recently introduced White Suffolk (WS, N = 14). To ensure good representation of diversity, we verified that the selected animals within each breed had no genetic relationships with those within other breeds for three generations, according to their pedigree records. Morphological characteristics were carefully checked to avoid possible hybrid relationships with other breeds.

SLAF-seq

Genomic DNA was extracted from the blood samples using Axy-Prep Genomic DNA Mini-prep Kit (Axygen Bioscience, Tewksbury, MA, USA) and evaluated for quality using agarose gel electrophoresis. According to the recommended pipeline, ~1 µg genomic DNA was incubated at 37°C with *EcoRI* (New England Biolabs [NEB], Ipswich, MA, USA), T4 DNA ligase (NEB), ATP (NEB), and adapter. After heat-inactivating ligation reactions at 65°C, the samples were digested with *AluI* at 37°C. Polymerase chain reaction was performed using diluted restriction-ligation samples, dNTP, Taq DNA polymerase (NEB), and *EcoRI*-primer containing barcode1. The purified samples were subjected to the fragment selection for ~450 bp using

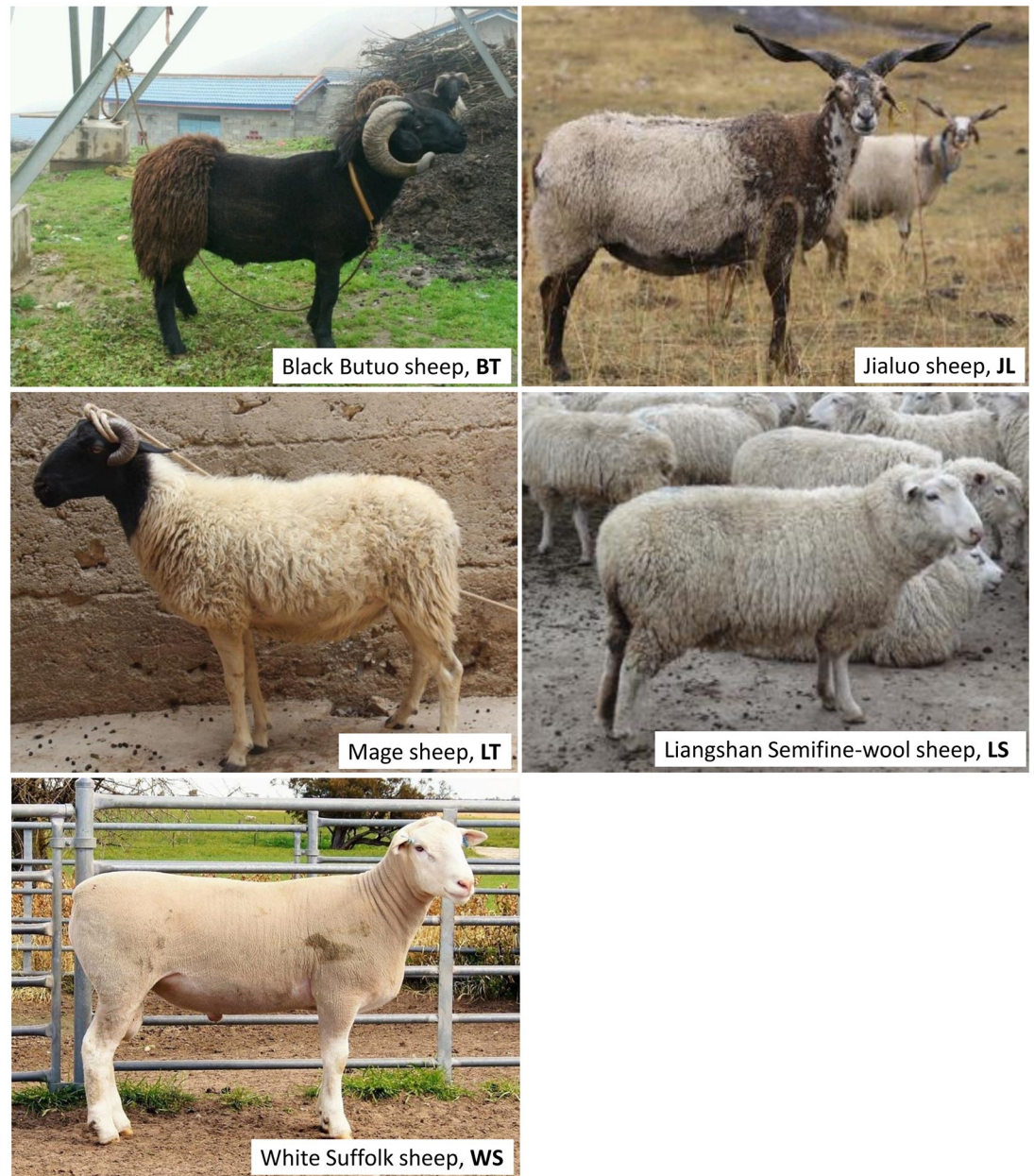


Fig 1. The five sheep breeds in Sichuan province.

<https://doi.org/10.1371/journal.pone.0257974.g001>

a Gel Extraction Kit (Qiagen, Hilden, Germany). These fragment products were then subjected to polymerase chain reaction using Phusion Master Mix (NEB) and Solexa Amplification primer mix to add barcode2. The sequencing libraries were obtained using the Illumina HiSeq platform, and 100 bp paired-end reads were generated (Biomarker, Beijing, China).

SNP calling

For the raw sequencing reads, the Q_{phred} value-based error rate and GC content were analyzed. Using the fastp tool [8], low-quality reads were filtered out if they belonged to any one of the

following three types: (1) reads containing adaptor sequences, (2) reads containing unambiguous bases of N more than 10% of the total length, and (3) reads containing low-quality bases ($Q < 5$) more than 50% of the total length. If any member of a paired read was marked as low quality, both pairs were discarded. After these steps, we obtained clean reads for subsequent analyses. Reads were aligned to the reference genome of sheep (Oar_rambouillet_v1.0 retrieved from NCBI) using BWA-MEM algorithm and default parameters in the BWA tool [9]. Picard tools (v1.134; <http://broadinstitute.github.io/picard/>) were employed to remove duplicate reads and sort the Sequence Alignment/Map files. Subsequently, SNPs were called and genotyped using GATK software (v3.7) according to GATK Best Practices recommendations [10–12].

Data analysis

All clean SNPs were individually annotated using the genome annotation file in GTF format and analysis of variance tool [13]. We analyzed the read coverage, chromosomal distribution, nucleotide diversity (π), observed heterozygosity (H_o), and expected heterozygosity (H_e), Hardy–Weinberg equilibrium, and pairwise Identity-by-State (IBS) distance using the VCF tools and Plink (Danecsek, 2011). The polymorphism information content (PIC) and Wright's fixation index (F_{ST}) of the breeds were computed using the PopSc toolkit [14].

Results

We obtained 507 million raw paired-end reads in total and 494 million clean reads after quality filtering (S1 Table). The average rates of alignment and unique alignment of reads to the reference genome sequence was 96.53% and 93.07%, respectively. After quality filtering, a total of 327,845 SNPs were detected among all the 69 samples, and all of them had an average read coverage of 23.75X. According to the positional annotation, all these SNPs were located within intergenic regions (9.068%), exons (0.826%), introns (36.229%), ncRNA transcript regions (36.560%), and 5'/3' untranslated regions (UTRs; 0.771%). With respect to the whole genome, all SNPs were evenly distributed among the 27 chromosomes, with the highest and lowest counts in chromosomes 1 ($N = 34,523$) and 24 ($N = 5,819$), respectively (Fig 2A). The overall transition/transversion ratio was 2.559 with 7,303,562 transitions and 2,854,530 transversions detected (Fig 2B). The frequency distribution of minor alleles among all the clean SNPs is shown in Fig 3.

Based on all the clean SNPs, we calculated the genetic diversity for individual and pooled breeds (Table 1). The overall nucleotide diversity was 0.2284 for all the five breeds, with the highest and lowest diversity observed in Mage (0.2125) and Butuo Black (0.1963) sheep, respectively. All the five breeds had H_o lower than H_e , and the largest and smallest differences were observed in Jialuo (0.1507 vs. 0.2212) and Liangshan Semifine-wool (0.1918 vs. 0.2132) sheep, respectively. The PIC ranged from 0.1395 in Jialuo sheep to 0.2492 in Liangshan Semifine-wool sheep, with the average value of 0.2217 among the five breeds.

Among all the 69 samples, 122,816 SNPs were revealed to be in Hardy–Weinberg equilibrium, with P -value < 0.05 . The pairwise F_{ST} values among the five breeds are shown in Fig 4. The most significant genetic divergence was observed between Butuo Black and White Suffolk sheep ($F_{ST} = 0.0789$), whereas the lowest F_{ST} was 0.0474 (Mage vs. Jialuo sheep). Among the three local pure breeds of Butuo Black, Jialuo, and Mage sheep, the pairwise genetic divergences slightly ranged from 0.0656 to 0.0774 in terms of F_{ST} value. PCA-based clustering was performed based on the pairwise IBS distances (Fig 5). It was clearly revealed that all individuals of Liangshan Semifine-wool, White Suffolk, and Butuo Black sheep were clustered together

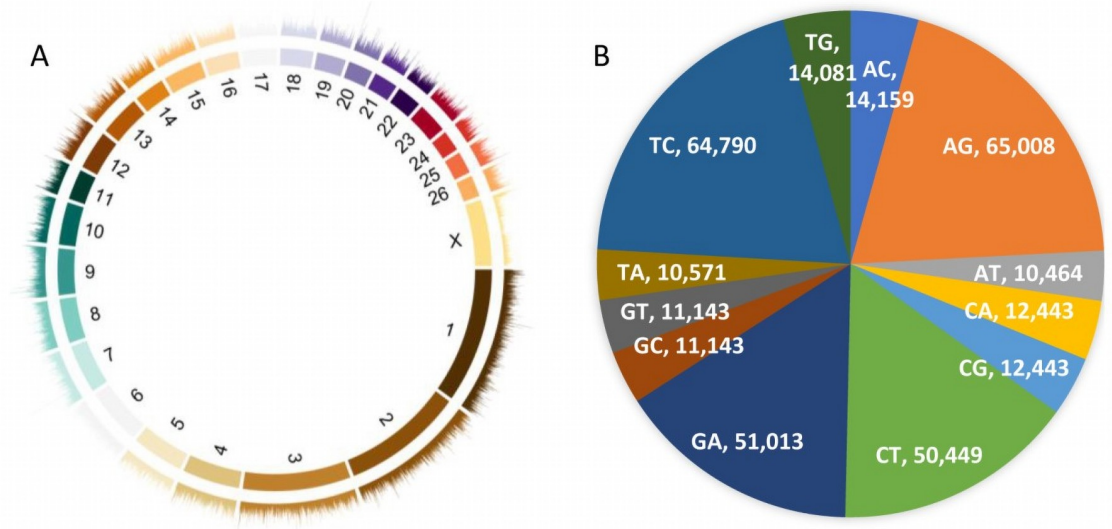


Fig 2. Distribution density across 27 chromosomes (A) and substitution types (B) of clean single nucleotide polymorphisms (SNPs) in this study.

<https://doi.org/10.1371/journal.pone.0257974.g002>

and obviously separated from each other. However, all individuals of Jialuo and Mage sheep were clustered together and could not be distinguished from each other.

Discussion

Given the climate changes and increasing threats to intensively selected composite animal breeds, it has become urgent to effectively conserve local genetic resources of domestic animals. In particular, considerable efforts have been made in field to investigate novel and valuable local breeds of sheep in Sichuan, China [5]. This province has abundant genetic resources of sheep and other farm animals, mainly due to its highly variable environmental conditions [3]. Further, sheep breeds in Sichuan have unique morphological characteristics and geographic distribution. However, the genetic diversity and population structures have not been systemically investigated yet, especially at the genome-wide level. In this study, we successfully employed the high-throughput sequencing technology SLAF-seq and found a large number of evenly distributed genome-wide and high-quality SNPs in five sheep breeds, including three local pure breeds, one composite breed (Liangshan Semifine-wool sheep), and one recently introduced exotic breed (White Suffolk) [7]. Based on these molecular markers, the genetic diversity and population structures of the sheep breeds were evaluated. To the best of our knowledge, this is the first report on the genetic diversity of Sichuan sheep breeds based on genome-wide SNPs.

Wang and Wu (2018) used 12 microsatellite markers and investigated the genetic diversity of six sheep breeds in Sichuan [15], which revealed a PIC of 0.559–0.612 and H_e of 0.610–0.670. Using 30 microsatellite markers, we previously found that the genetic diversity of Butuo Black sheep was 0.6257 for PIC and 0.6269 for H_e [16]. In the present study, we observed that Butuo Black sheep had the lowest genetic diversity in comparison with the other four breeds, including the composite breed of Liangshan Semifine-wool sheep and the exotic breed of White Suffolk sheep. The results suggest relatively high genetic diversity of sheep breeds in Sichuan, which is consistent with a recent publication investigating the national sheep breeds in China [17]. We also observed that, in addition to local pure breeds, the

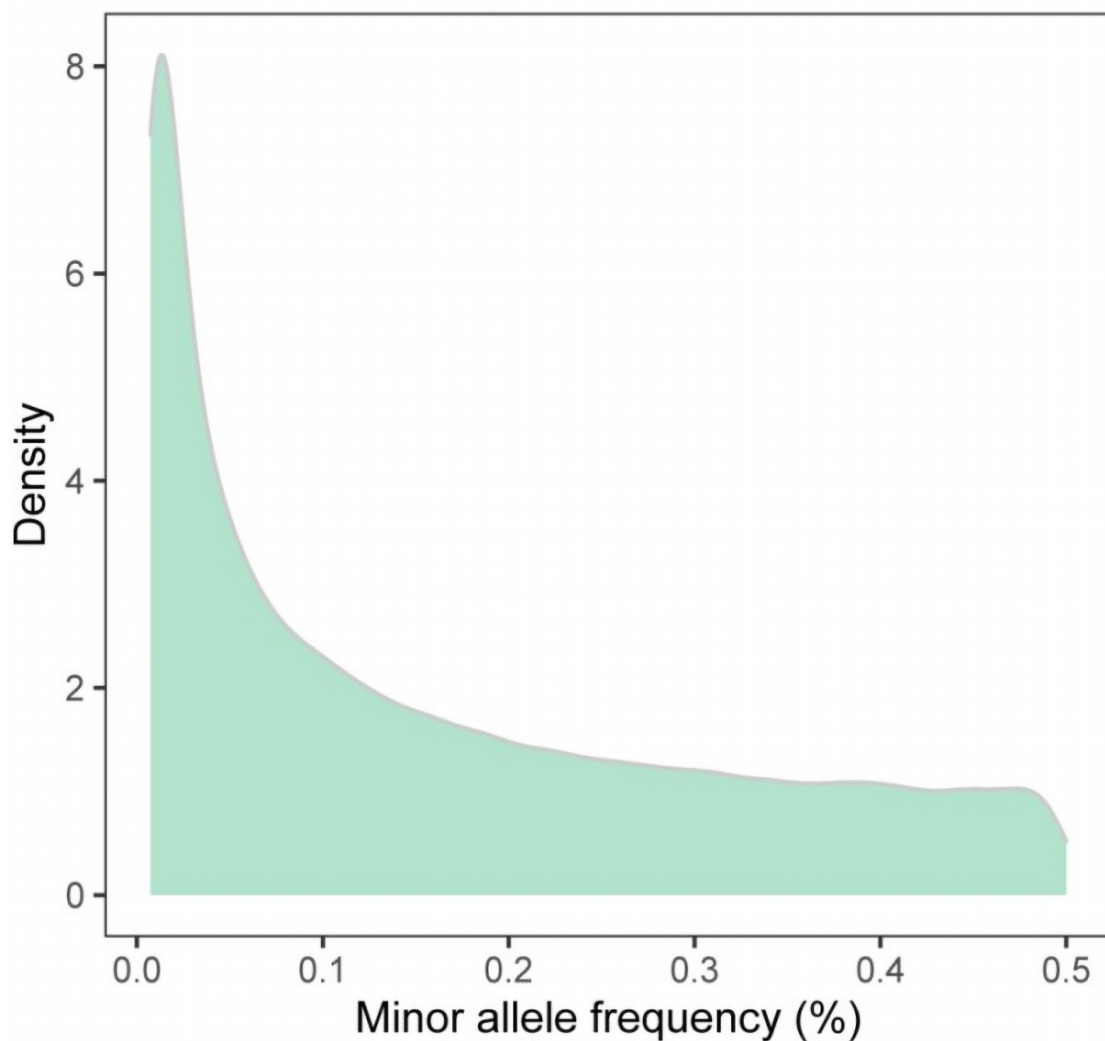


Fig 3. Distribution density of minor allele frequencies.

<https://doi.org/10.1371/journal.pone.0257974.g003>

Table 1. Genetic diversity among the five sheep breeds.

Breed	Nucleotide diversity	Observed heterozygosity	Expected heterozygosity	Polymorphism information content
BT	0.1963	0.1763	0.2090	0.1469
JL	0.2055	0.1507	0.2212	0.1395
LT	0.2125	0.1635	0.2195	0.2433
LS	0.2077	0.1918	0.2132	0.2492
WS	0.2047	0.1543	0.2131	0.1975
All	0.2284	0.1703	0.2302	0.2217

BT, Butuo Black sheep; JL, Jialuo sheep; LT, Mage sheep; LS, Liangshan Semifine-wool sheep; WS, White Suffolk.

<https://doi.org/10.1371/journal.pone.0257974.t001>

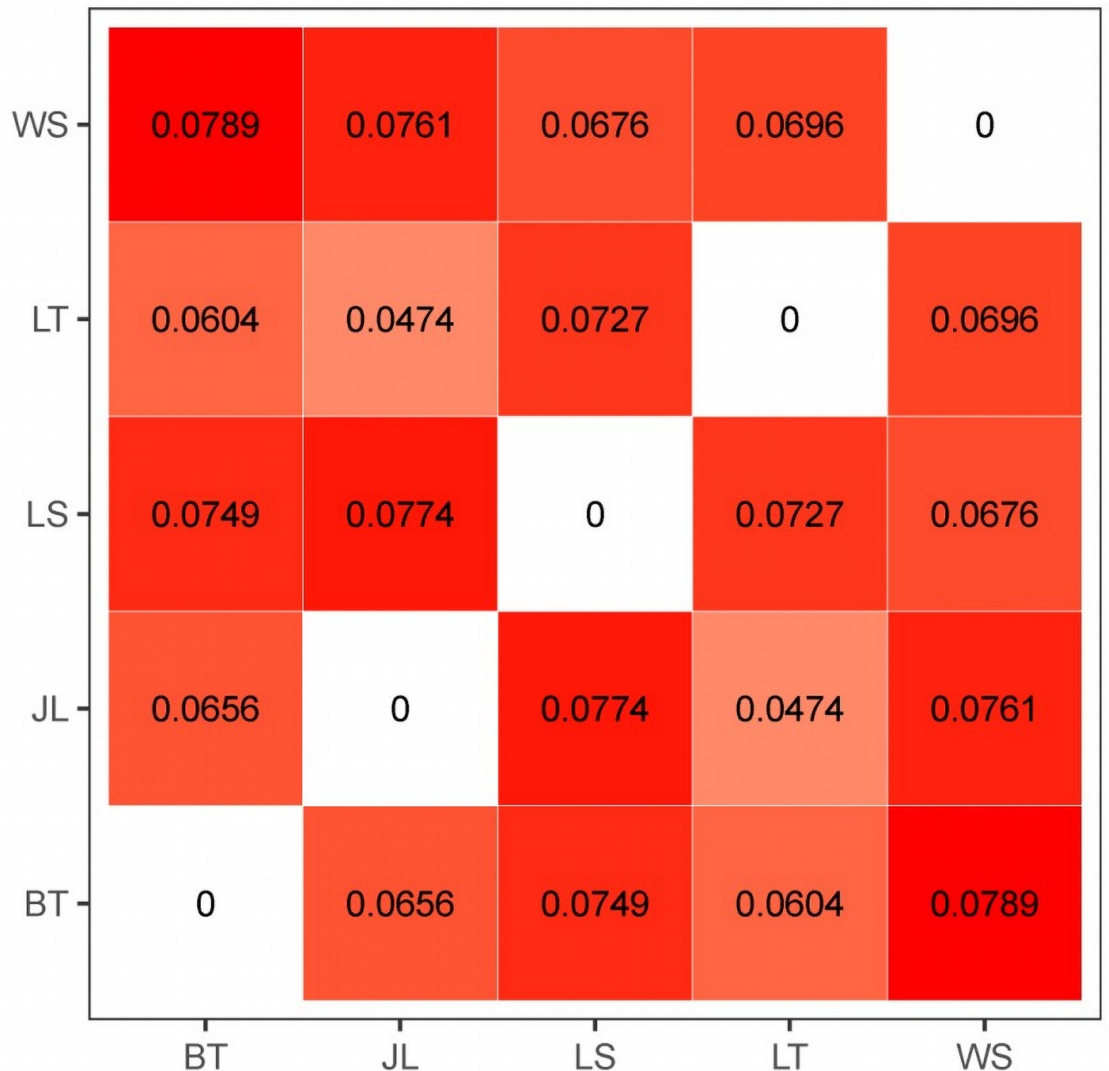


Fig 4. Pairwise wright's F_{ST} values among the five breeds.

<https://doi.org/10.1371/journal.pone.0257974.g004>

recently introduced and composite sheep breeds in Sichuan have a relatively high genetic diversity; this would be favorable in reducing potential inbreeding depression.

Among the five sheep breeds included in the present study, Liangshan Semifine-wool, White Suffolk, and Butuo Black sheep had obvious genetic divergences, according to pairwise IBS distance. However, we could not differentiate Jialuo and Mage sheep because all the individuals were mostly clustered together. This is inconsistent with the former microsatellite marker-based report showing that Butuo Black and Jialuo sheep have a close genetic relationship [15]; this difference might arise from the different molecular markers used in the two studies. Overall, the three local pure breeds had closer genetic relationships with each other than with the composite breed of Liangshan Semifine-wool sheep and the exotic breed of White Suffolk sheep; this would indicate their respective genetic origins. The inter-breed F_{ST} values also supported this overall population structure for the five sheep breeds studied.

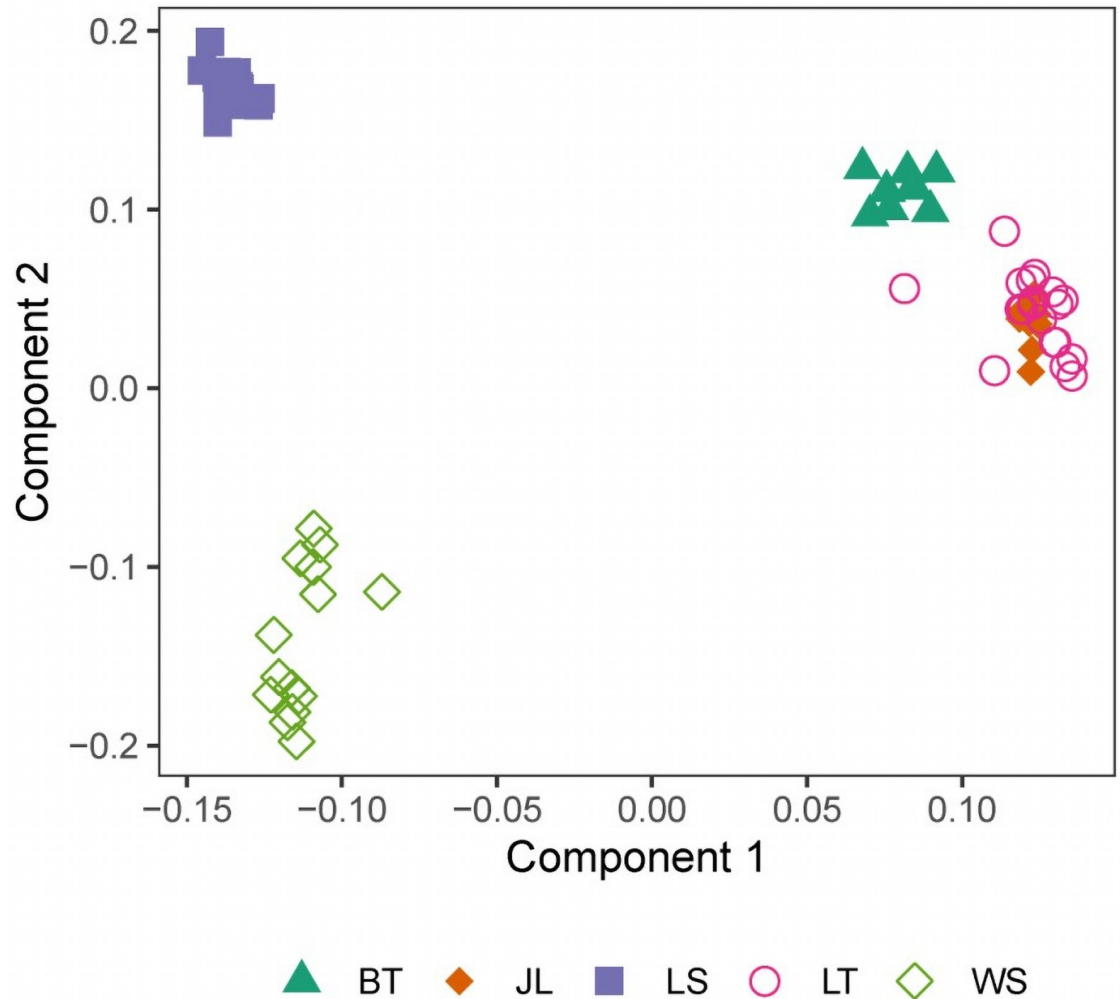


Fig 5. Sample clustering by IBS distance-based PCA plotting.

<https://doi.org/10.1371/journal.pone.0257974.g005>

Conclusion

In the present study, we successfully employed SLAF-seq and constructed a reference panel of genome-wide and high-quality SNPs in five sheep breeds in Sichuan, China. Their genetic diversity and population structures were subsequently revealed, which would be helpful to facilitate our understanding and conservation of these genetic resources.

Supporting information

S1 Table. Raw sequencing data in this study.
(PDF)

Author Contributions

Conceptualization: Mingliang Zhou, Gaofu Wang, Minghua Chen, Jie Zeng.

Data curation: Mingliang Zhou, Gaofu Wang.

Formal analysis: Mingliang Zhou, Gaofu Wang, Qian Pang, Shihai Jiang, Jie Zeng.

Funding acquisition: Dan Du, Pinggui Yang.

Investigation: Mingliang Zhou, Gaofu Wang, Minghua Chen, Qian Pang, Weisheng Wu, Hongwen Zhao.

Methodology: Mingliang Zhou, Gaofu Wang, Minghua Chen, Qian Pang, Shihai Jiang.

Project administration: Mingliang Zhou, Gaofu Wang, Pinggui Yang.

Resources: Mingliang Zhou, Gaofu Wang, Pinggui Yang.

Supervision: Shihai Jiang, Pinggui Yang.

Validation: Jie Zeng, Dan Du, Weisheng Wu, Hongwen Zhao.

Writing – original draft: Gaofu Wang.

Writing – review & editing: Mingliang Zhou.

References

1. Chen SY, Duan ZY, Sha T, Xiangyu J, Wu SF, Zhang YP. Origin, genetic diversity, and population structure of Chinese domestic sheep. *Gene*. 2006; 376(2):216–23. <https://doi.org/10.1016/j.gene.2006.03.009> PMID: 16704910
2. Kijas JW, Lenstra JA, Hayes B, Boitard S, Neto LRP, San Cristobal M, et al. Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol*. 2012; 10(2):e1001258. <https://doi.org/10.1371/journal.pbio.1001258> PMID: 22346734
3. Chen WS. Report on domestic animal genetic resources in China. Beijing: Chinese Agriculture Publishers; 2004.
4. Wei C, Wang H, Liu G, Wu M, Cao J, Liu Z, et al. Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics*. 2015; 16(1):194. <https://doi.org/10.1186/s12864-015-1384-9> PMID: 25888314
5. Luo J, Liu L, Eme QZ, Zhang L. Investigation of genetic resources of Butuo Black sheep. *Herbivore Science in China*. 2019; 1:72–3 (in Chinese).
6. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet*. 2014; 30(9):418–26. <https://doi.org/10.1016/j.tig.2014.07.001> PMID: 25108476
7. Sun X, Liu D, Zhang X, Li W, Liu H, Hong W, et al. SLAF-seq: an efficient method of large-scale *de novo* SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE*. 2013; 8(3):e58700. <https://doi.org/10.1371/journal.pone.0058700> PMID: 23527008
8. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018; 34(17):i884–i90. <https://doi.org/10.1093/bioinformatics/bty560> PMID: 30423086
9. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168
10. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010; 20(9):1297–303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
11. 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; 43:491–8. <https://doi.org/10.1038/ng.806> PMID: 21478889
12. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*. 2013; 11(1110):1–33. <https://doi.org/10.1002/0471250953.bi1110s43> PMID: 25431634
13. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38(16):e164–e. <https://doi.org/10.1093/nar/gkq603> PMID: 20601685
14. Chen SY, Deng F, Huang Y, Li C, Liu L, Jia X, et al. PopSc: computing toolkit for basic statistics of molecular population genetics simultaneously implemented in web-based calculator, Python and R. *PLoS ONE*. 2016; 11(10):e0165434. <https://doi.org/10.1371/journal.pone.0165434> PMID: 27792763

15. Wang T, Wu D. Genetic diversity of sheep in Sihcuan revealed by microsatellite markers. *Chinese Journal of Animal Science*. 2018; 4:36–41 (in Chinese).
16. Zhou M, Chen M, Pang Q, Yang P, Li Y. Genetic diversity of Butuo Black sheep in Sihcuan using microsatellite markers. *Grass Science*. 2018; 2:15–7 (in Chinese).
17. Hu L, Sun W, Y. M. Study on genetic diversity and genetic structure of Tibetan sheep populations. *Acta Veterinaria et Zootechnica Sinica*. 2019; 50(6):1145–53 (in Chinese).