Unraveling the Genetic Population Structure of Mongolian Indigenous Cattle Breeds Using Whole Genome Sequencing Data

Rugang Tian[1(✉)], Hojjat Asadollahpour Nanaie[2], Yuan Li[1], Xiao Wang[1], Meng Zhao[1], Hui Li[1], Hao Zhang[1], and Jianghong Wu[3(✉)]

[1] Institute of Animal Husbandry, Inner Mongolia Academy of Agricultural and Animal Husbandry Sciences, Hohhot 010031, China
tiannky@163.com

[2] Key Laboratory of Animal Genetics, Breeding and Reproduction of Shaanxi Province, College of Animal Science and Technology, Northwest A&F University, Yangling 712100, China

[3] College of Animal Science and Technology, Inner Mongolia Minzu University, Tongliao, China
wujianghonglong@126.com

Abstract. In the present study, we generated complete genome sequence data from 45 Mongolian cattle. Together with published sequence data from worldwide cattle populations, we explored the genetic diversity and population structure of worldwide cattle breeds. Our findings revealed clustering of cattle populations into three major groups, including commercial (Red- Angus, Hereford, Holstein and Jersey), Chinese (Mongolian and Tibetan) and other native cattle (including African cattle and Rashoki breed from Iran) breeds. The results from admixture analysis revealed evidence of shared genetic ancestry between different cattle populations. Furthermore, our findings showed a markedly higher level of linkage disequilibrium (LD) across all genomic distances in commercial breeds (specially Holstein cattle) compare to other native cattle groups. Our results provide valuable insights into the architecture of Mongolian Indigenous cattle breeds and their genomic relationship with other cattle populations. Pomelo.

Keywords: Mongolian cattle · Whole-genome · Population structure · Linkage disequilibrium

1 Introduction

Domestic animals have been evolved genetic adaptations to a variety of environments and serve as ideal models for genetic studies involving assessment of genetic variation and population structure [1]. Cattle are one of the most important farm animals that domesticated from wild aurochs (Bos primi genius) about 10,000 years ago [2]. By supplying protein and fat for human populations, they became the most valuable livestock in agricultural production system [3]. Recently, remarkable advances in high-throughput sequencing (HTS) technologies have revolutionized researchers' ability to study the

E. Widodo et al. (Eds.): ICESAI 2022, ABSR 28, pp. 36–42, 2023.
https://doi.org/10.2991/978-94-6463-116-6_6

genomic structure, function, and evolution at an unprecedented scale [4, 5]. By applying this technology, dissecting the genetic relationship between different populations can be conducted more quickly and economically. Since the first bovine genome sequence was published in 2001 from a Hereford cow [6], researchers have used the technology to study the genetic architecture of different cattle populations from different geographical regions. The first cattle complete genome resequencing was conducted using one individual sample (Germany Fleckvieh bull) in 2009 [7], which after mapping to the bovine reference genome, produced thousands of small indels (~ 115,000) and millions single nucleotide polymorphisms (SNPs).

Mongolian cattle is a breed of cattle in China that has been originated from the Inner Mongolian grasslands. This breed has been well adapted to the climate conditions of this area (such as arid and cold environments) and shows a high degree of adaptation to local disease [8, 9]. They have been herded for centuries by nomads and are well-known for their valuable meat quality. Despite the fact that this cattle breed is one of the most commonly distributed indigenous breeds in the north of China [10], there are lack of genomic studies which have focused on this breed and its genomic relationship with other native and commercial cattle. In the current, we systematically examined the genetic relationship between Mongolian cattle and worldwide cattle populations, using whole genome sequencing data. Our study provides valuable insights into the genetic architecture of Mongolian indigenous cattle, which up to now has not been genetically investigated.

2 Materials and Methods

2.1 Sampling, Quality Checking and SNP Calling

A total of 45 whole-blood samples were collected from native cattle individuals in Mongolian. DNA genomic was extracted from blood cells using the phenolchloroform procedure. The concentration and quality of extracted DNA was evaluated by both NanoDrop spectrophotometer and agarose gel electrophoresis (1.5%). The high-quality sequencing reads for all animals were produced using the Illumina Hiseq 2500 platform (read length of 125 bp). In addition to this dataset, we also collected publicity available whole genome sequence data from both commercial (Red- Angus, Hereford, Holstein and Jersey) and indigenous (Rashoki, Ankole, N'Dama, Kenana and Tibetan) cattle breeds (ncbi.nlm.nih.gov). The quality of fastq raw data was verified using FastQC software (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). In order to remove ambiguous bases (N), the Illumina paired-end reads were preprocessed using Trimmomatic software. All filtered reads from the current study and previously published data were aligned against to the latest bovine reference genome (ARS-UCD1.2) assembly with Burrows Wheeler Aligner (BWA mem, V 0.7.16) software [11]. SAMtools software [12] was used for discarding un-mapped and non-unique reads, and also for converting mapping results to the binary format (BAM). PCR duplicates were marked and removed using Picard Toolkit (http://broadinstitute.github.io/picard/). Base quality score recalibration and local indel realignment were conducted using Genome Analysis Toolkit (GATK) software package [13]. Finally, all variants were discovered using the GATK HaplotypeCaller module.

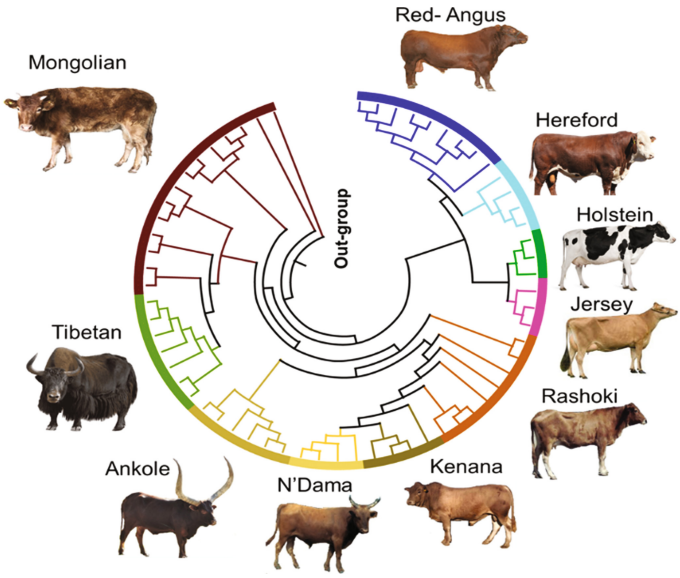

Fig. 1. Phylogenetic tree was built based on maximum-likelihood (ML) method.

2.2 Population Structure and Admixture Analyses

Using Fasttree software (version 2.1.11), a maximum likelihood phylogeny tree was established based on the whole genome sequence data from all the studied cattle individuals [14]. The tree was then visualized using the online tool iTOL (V6; https://itol.embl.de/). A likelihood model-based approach, available in ADMIXTURE software, was used to understanding the degree of genetic admixture between different populations, with an ancestor population (K) size ranging from 2 to 6 and 10,000 iterations for each run [15]. In addition, the decay of linkage disequilibrium (LD) was calculated by plink software [16], as squared correlation of allele frequencies (r^2) for different genetic distances (1, 3, 5, 15, 60 and 100 Kb) SNP pairs.

3 Results and Discussion

Analysis of population genetic structure can provide valuable information about the origin and composition of breeds, thus, enhance our understanding about the domestication process of certain populations [17]. In this study, to assess the genetic relationships of cattle populations from a genome-wide perspective, a phylogenetic tree for all individuals was constructed using ML method. Based on this tree (Fig. 1), all 92 individuals could be classified into three major groups, representing commercial (Red- Angus, Hereford, Holstein and Jersey), Chinese indigenous (Mongolian and Tibetan) and other native cattle breeds (including African cattle and Rashoki breed from Iran). These results indicated that evolutionarily, the Mongolian breed has a close kinship with Tibetan cattle population from China.

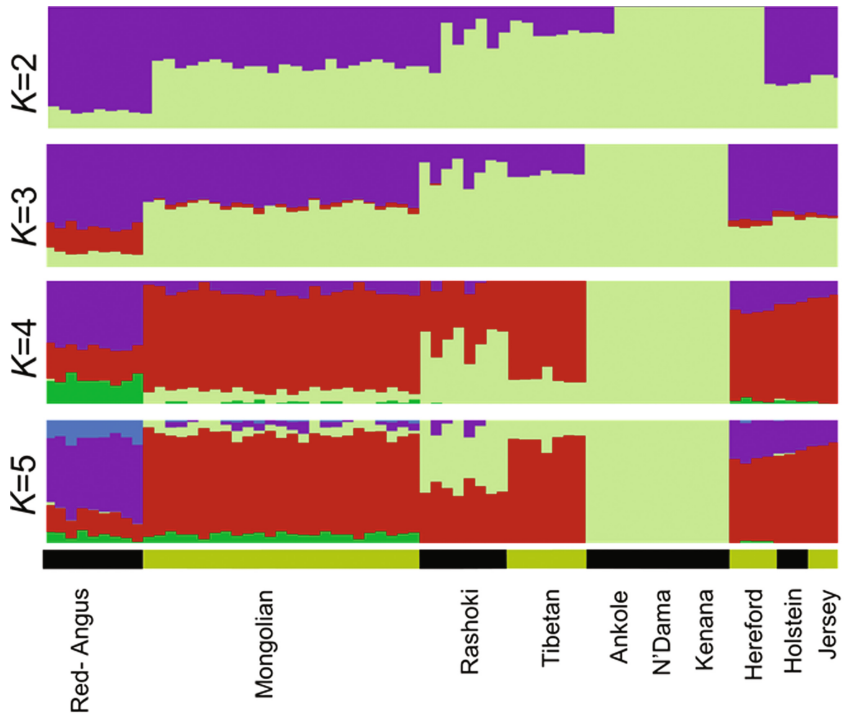

Fig. 2. ADMIXTURE model-based clustering analysis on the subset of 92 cattle individuals. Results from $k = 2$ to $k = 5$ are represented. The group names are at the bottom.

The ADMIXTURE results were generally consistent with the phylogenetic tree (Fig. 2). The $K = 2$, with the lowest cross-validation (CV) error, splits the all African breeds from other native and commercial cattle populations. The $K = 4$, distinguishes samples from Red- Angus from other commercial cattle. Furthermore, plot with $K = 5$ cluster, suggests that both Mongolian and Red- Angus cattle may derive from genetic admixture between at least four different populations. Our result is in accordance with previous study by Chen et al. [18], which shows Chinese Tibetan cattle and Northeast Asian breeds are belong to a distinct East Asian clade. Furthermore, we found that all African cattle breeds cluster together and showed clear differentiation from other Chinses and commercial breeds, that is supported from previous studies that have compared these breeds with other cattle groups from different geographical regions [19, 20].

We next using the classic r2 calculator estimated the patterns of LD decay, as a function of the genetic distance in the genome. This information can be used to measure haplotype diversity, demographic processes, population history and also breeding systems in a population [21] (Fig. 3). Our finding revealed that at shorter marker distances ($\leq$1 Kb) the r2 values were the highest for all studied groups (ranged from ~0.54 to 0.74, for Rashoki and Holstein, respectively) with a gentle decline with increasing distance between SNPs and then stable trend (>60 Kb). Among the breeds included in this study, a higher level of LD across all physical genomic distances was detected in

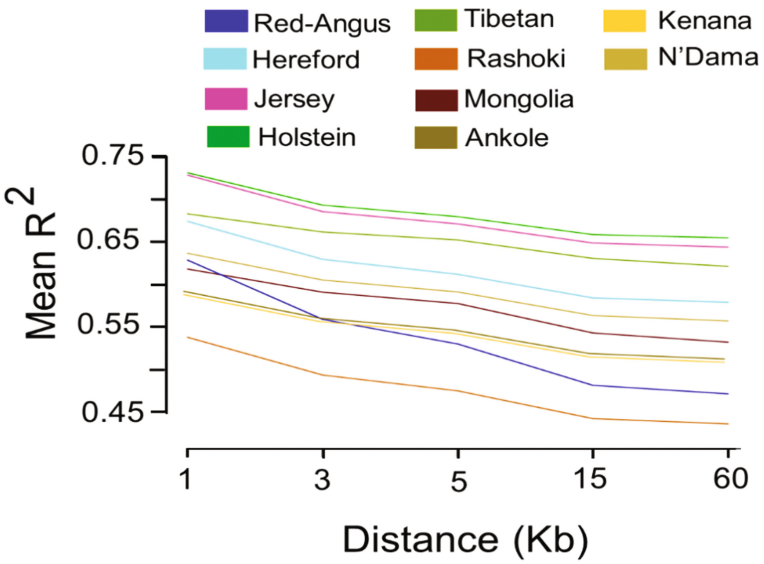

Fig. 3. The decay of linkage disequilibrium (LD) calculated as the squared correlation coefficient by pairwise physical distance in all studied cattle populations.

Holstein cattle breed. While the lower values of r2 were observed in Rashoki cattle, that seems genetically admixed with at least three other cattle populations (Fig. 2, K = 5). Focusing on studied indigenous breeds, Tibetan cattle showed higher r2 values across all genomic distances, that is genetically less mixed with cattle breeds from other groups (Fig. 2), and dwelled in northern regions of China country. In addition, we observed that cattle individuals from Mongolian and most other native cattle breeds (such as Kenana, Ankole and N'Dama) followed the same pattern of decrease in LD as the genomic distance increased, however a rapid decrease in LD over increasing genomic distance was detected in Rashoki cattle breed (Fig. 3).

The r^2 values estimated in the current study for indigenous cattle breeds were in the ranges of 0.42 to 0.67 at 60 marker pairs distance which consistent with those reported by other studies for African and commercial cattle populations [19, 20]. A relatively higher LD values in commercial cattle breeds compared with indigenous populations may be a consequence of higher level of inbreeding and extensive artificial selection for desired traits (such as milk and meat traits) in the breeding programs [22].

4 Conclusions

In the present study, we attempted to investigate the population structure and genomic relationship of different cattle populations from worldwide geographical regions. Several lines of evidence suggest that there is a high level of genetic distance between native cattle populations and commercial breeds. However, we found a relatively low genetic variability between Mongolian cattle and native cattle from Tibetan region, and also between African cattle and Rashoki breed from Iran country. In addition, we observed

a relatively high levels of LD decay in commercial cattle breeds, which may be due to artificial selection for specific traits such as milk or meat production performances. Our finding will help improve our understanding about the genetic makeup of these populations and facilitate the use of such information in future breeding scheme.

Acknowledgment. We thank the High-Performance Computing platform of Institute of Animal Husbandry, Inner Mongolia Academy of Agricultural & Animal Husbandry Sciences for providing computing resources.

Funding. This work was supported by Inner Mongolia key technologies R&D program (2020GG0069), Inner Mongolia key technologies R&D program and technological transformative project (2022YFYZ0011), Inner Mongolia major science and technology program (2021ZD0009), National program for high-level talents introduction (G2021005009L).

References

1. Andersson, L.: Domestic animals as models for biomedical research. Upsala Journal of Medical Sciences 121(1), 1–11 (2016).
2. Verdugo, M.P., Mullin, V.E., Scheu, A., Mattiangeli, V., Mattiangeli, V., Daly, K.G., et al.: Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. Science 12; 365(6449), 173–176 (2019).
3. Vigne, J.D., Peters, J., Helmer, D.: The First Steps of Animal Domestication. Oxbow Books, Oxford (2005).
4. Sharma, A., Park, J. E., Chai, H. H., Jang, G. W., Lee, S. H., Lim, D.: Next generation sequencing in livestock species- a review. Journal of Animal Breeding and Genomics, 1, 23–30 (2017).
5. Nosrati, M., Nanaei, H., Ghanatsaman, Z., Esmailizadeh, A., Whole genome sequence analysis to detect signatures of positive selection for high fecundity in sheep. Reproduction in Domestic Animals, 54(2), 358–364 (2019).
6. Elsik. C.G., Tellam, R.L., Worley, K.C.: The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science, 324, 522–528 (2009).
7. Eck. S.H., Benet-Pagès, A., Flisikowski, K., Meitinger, T., Fries, R., Strom, T.M.: Whole genome sequencing of a single Bos taurus animal for single nucleotide polymorphism discovery. Genome Biology, 10: R82 (2009).
8. Hu. L., Brito, L.F., Abbas, Z., Sammad, A., Kang, L., Wang, D., Wu, H., Liu, A., Qi, G., Zhao, M., Wang, Y., Xu, Q. Investigating the Short-Term Effects of Cold Stress on Metabolite Responses and Metabolic Pathways in Inner-Mongolia Sanhe Cattle. Animals, 11, 249 (2021).
9. Xing, Y., Qi, Y., Purev, C., Wang, S., Wang, H., Wu, K., Cao, J., Liu, C., Liu, Y., Li, L., Zhang, Y., Zhou, H. Structural variants in Mongolian originated ruminant: role in adaptation of extreme-environment. bioRxiv https://doi.org/10.21203/rs.3.rs-17431/v1 (2020).
10. Yue, X., Li, R., Liu, L., Zhang, Y., Huang, J., Chang, Z., Dang, R., Lan, X., Chen, H., Lei, C. When and how did Bos indicus introgress into Mongolian cattle? Gene, 10; 537(2), 214–9 (2014).
11. Li, H., Durbin, R. Fast and accurate short read alignment with Burrows Wheeler transform. Bioinformatics, 15, 1754–60 (2009).

12. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25: 2078–9 (2009).
13. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20, 1297–303 (2010).
14. 14 Price, M.N., Dehal, P.S., Arkin, A.P. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLoS ONE. 5(3): e9490 (2010).
15. Alexander, D.H., Novembre, J., Lange, K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19: 1655–64 (2009)
16. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C. PLINK: a tool set for whole genome association and population-based linkage analyses. American Journal of Human Genetics 81, 559–75 (2007).
17. Stift, M., Kolář, F., Meirmans, P.G. STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. Heredity, 123, 429–441 (2019).
18. Chen, N., Cai, Y., Chen, Q., et al. Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. Nature Communication, 9, 2337 (2018).
19. Kim, J., Hanotte, O., Mwai, O.A. et al. The genome landscape of indigenous African cattle. Genome Biology 18, 34 (2017).
20. Asadollahpour Nanaei, H., Dehghani Qanatqestani, M., Esmailizadeh, A. Whole-genome resequencing reveals selection signatures associated with milk production traits in African Kenana dairy zebu cattle. Genomics, 112(1), 880–885 (2019).
21. Waples, R.K., Larson, W.A., Waples, R.S. Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity, 117, 233–40 (2016).
22. Binns, M. M., Boehler, D. A., Bailey, E., Lear, T. L., Cardwell, J. M., Lambert, D. H. Inbreeding in the Thoroughbred horse. Animal Genetics, 43, 340–342 (2011).