Livestock Science 170 (2014) 22–29

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Estimations of genomic linkage disequilibrium and effective population sizes in three sheep populations

Fuping Zhao [a,1], Guangkai Wang [a,1], Tao Zeng [b], Caihong Wei [a], Li Zhang [a], Huihua Wang [a], Shuzhen Zhang [a], Ruizao Liu [a], Zhen Liu [a], Lixin Du [a,*]

[a] National Center for Molecular Genetics and Breeding of Animal, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing 100193, China
[b] College of Animal Science and Technology, Sichuan Agricultural University, Ya'an 625014, China

ARTICLE INFO

Article history:
Received 27 February 2014
Received in revised form
15 August 2014
Accepted 17 October 2014

Keywords:
Linkage disequilibrium
Persistence of phase
Effective population size
Sheep

ABSTRACT

As one of the most important measures in population genetics, linkage disequilibrium (LD) is the basis of effective population size estimation, genomic selection, genome-wide association study and quantitative trait locus mapping. Here, we characterized the pattern of LD in three sheep populations (Sunite, German Mutton Merino, and Dorper) and subsequently estimated persistence of phase among them using the Illumina Ovine SNP50k Chip. In addition, we calculated the corresponding effective population sizes using the metric r^2 estimated. The results showed that the extent of LD was highest in Dorpers, followed by German Mutton Merinos and Sunites. Average r^2 values were 0.13, 0.20, and 0.22 at adjacent marker distances of 57.28, 59.01, and 59.93 kb for Sunite, German Mutton Merino, and Dorper sheep, respectively. The average r^2 values at the same genomic distance in German Mutton Merino and Dorper sheep were similar and higher than that in Sunite sheep, and the rate of LD decay over distance was faster than the rates in the other two sheep breeds. The correlation of phase for different breed comparisons ranged from 0.784 to 0.825. The effective population sizes in Sunite, German Mutton Merino, and Dorper sheep 7 generations ago were approximately 207, 74, and 67, respectively. We also detected a bottleneck in the effective population size for Sunite sheep that was consistent with the timing for the origin of the domestication of Chinese sheep. Our findings provide a theoretical reference for future genome marker-assisted evaluation and selection in sheep.

1. Introduction

Linkage disequilibrium (LD), also called allelic association, refers to the non-random association of alleles at two or more loci in a population (Hill and Robertson, 1968). Across the whole genome, LD information is critical for genomic selection, genome-wide association studies (GWAS), and quantitative trait loci (QTL) mapping for complex traits, which depend on the LD between markers and causative variants. Recently, many studies have been published characterizing the pattern of LD across the whole genome in both humans (Conrad et al., 2006; De La Vega et al., 2005) and animals (Badke et al., 2012; Bohmanova et al., 2010; de Roos et al., 2008; Garcia-Gamez et al., 2012; Harmegnies et al., 2006; Qanbari et al., 2010). Their results showed that LD is widespread in the genome, likely because of limited sample sizes and complicated population histories. With the rapid

* Corresponding author. Tel.: +86 10 62819997.
E-mail address: lxdu@263.net (L. Du).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.livsci.2014.10.015
1871-1413/

development of chip array-based genotyping techniques, high density single nucleotide polymorphism (SNP) chip arrays have become cheaper and thus more cost effective. Therefore, SNP markers have been extensively applied to estimate the genome-wide LD within a population as well as the persistence of phase between two populations, which is also one of the most important parameters in population genetics. A persistence of phase can explain why the linkage between a marker and a QTL detected in one population is not always validated in another. Thus, persistence of phase influences the reliability of genomic prediction in multiple breeds. Moreover, knowing the persistence of phase for two populations across a range of genomic distances, we can predict the required marker density for QTL mapping, GWAS or genomic selection.

The LD information can also be used to infer the ancestral effective population size (N_e) (Hayes et al., 2003; Sved, 1971). The N_e can not only explain how populations evolved but also improve the understanding and modeling of the genetic architecture underlying complex traits (Reich and Lander, 2001; Tenesa et al., 2007). When N_e is small, genetic variants are limited within a population, which influences the genetic gain for breeding programs in animals.

Several recent studies have investigated LD features in cattle, pigs and sheep in the last few years (Badke et al., 2012; de Roos et al., 2008; Sargolzaei et al., 2008; Uimari and Tapio, 2011; Wang et al., 2013). Miller et al. (2011) found that LD levels in domestic sheep were similar to those in wild sheep using SNP chip technology. Meadows et al. (2008) estimated the LD in five sheep populations across one chromosome (OAR18) and suggested that LD levels varied notably between different populations and that LD decayed faster in Merino and a crossbred population (Merino × Border Leicester) than in three other populations (White Faced Suffolk, Poll Dorset, and Australian sheep). A common observation in these studies is that LD in animals is larger than that reported in humans (Reich et al., 2001) due to the artificial selection and smaller effective population sizes in livestock (Amaral et al., 2008; Harmegnies et al., 2006). Many studies on estimations of effective population size have focused primarily on dairy cattle. For instance, Qanbari et al. (2010) found that the N_e in German Holstein cattle four generations ago is approximately 103. However, few studies have estimated the effective population size in sheep using SNP chip technology.

The LD pattern also provides insight into the evolutionary history of a population. The extension of LD in the genome can be used to infer ancestral effective population size, an important population parameter that helps explain how populations evolved and improves the understanding and modeling of the genetic architecture underlying complex traits. We also estimate persistence of phase between the three breeds as a measure of the relationship across these populations.

In the present study, we investigated the LD patterns of Sunite an indigenous sheep in the northern China, and German Mutton and Dorper two introduced sheep breeds, using the Illumina Ovine SNP50 BeadChip, and construct LD maps in three different sheep populations. To better understand the relationship between different breeds, we computed the persistence of LD phase between different breed combinations. We also calculated the effective population size for three sheep populations using the estimated r^2. This research will provide valuable information for future genome marker-assisted evaluation and selection in sheep.

2. Materials and methods

2.1. Ethics statement

All animal procedures strictly followed the guidelines proposed by the Chinese Council on Animal care and all protocols were approved by the Animal Care and Use Committee of Beijing, China. The approval ID and permit numbers were SYXK (Beijing) 2008-007 and SYXK (Beijing) 2008-008.

2.2. Sample collection and quality control

Sheep were randomly selected from three populations and included 69 Sunites (SUN, 57 males, 12 females) from an Inner Mongolian Sunite purebred sheep stud (Xilin Gol League, Inner Mongolia, China), 161 German Mutton Merinos (GMM, 71 males, 90 females) from an Inner Mongolian German Mutton purebred sheep stud (Xilin Gol League, Inner Mongolia, China), and 99 Dorpers (DOR, 49 males, 50 females) from Tianjin Aoqun Animal Husbandry Propriety Limited (Jinghai Country, Tianjin, China). Blood was collected from six-month-old lambs. The genomic DNA was extracted from blood samples using a TIANamp Blood DNA Kit (Tiangen Biotech Company Limited, Beijing, China). The DNA was genotyped using the Illumina Ovine SNP50 BeadChip, which contained 54,241 markers with an average probe distance of 50.9 kb.

To ensure data quality, quality control was employed according to the following procedures: (1) an individual was excluded if more than 10% of its genotypes were missing; and (2) a SNP was removed if its call rate was less than 90%. After filtering, 43,299 SNPs from 318 samples (66 Sunite sheep, 159 German Mutton sheep, and 93 Dorper sheep) remained. To exclude non-segregating SNPs from the analysis, we removed markers with a MAF below 1% within each breed separately. The numbers of markers available for SUN, GMM, and DOR were 42,616, 41,371 and 40,734, respectively. The filtering process was completed using PLINK v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/)

Missing genotypes were imputed and haplotypes were constructed for each chromosome using Beagle software package (Browning and Browning, 2009).

2.3. Estimation of LD and persistence of phase

LD can be measured by r^2 and by D' (Hill, 1974; Lewontin, 1964). Because population size has a great influence on D' but r^2 is stable (Zhao et al., 2007), we used r^2 as the measure of LD in this study. The equation is generally written as:

$$r^2 = \frac{(p_{A_1B_1} - p_{A_1} \times p_{B_1})^2}{p_{A_1} \times (1-p_{A_1}) \times p_{B_1} \times (1-p_{B_1})} \quad (1)$$

where p_{A_1} and p_{B_1} are the minor allelic frequencies at the loci A and B, and $p_{A_1B_1}$ is the frequency of the two-marker