Gene 673 (2018) 211–216

Contents lists available at ScienceDirect

Gene

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

Research paper

Selection and effectiveness of informative SNPs for paternity in Chinese Simmental cattle based on a high-density SNP array

Tianliu Zhang[a,1], Liping Guo[a,1], Mingyan Shi[b], Lingyang Xu[a], Yan Chen[a], Lupei Zhang[a], Huijiang Gao[a], Junya Li[a,*], Xue Gao[a,*]

[a] Laboratory of Molecular Biology and Bovine Breeding, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, PR China
[b] College of Life Science, Luoyang Normal University, Jiqing Road, 471934 Luoyang, PR China

ARTICLE INFO

Keywords:
Paternity assignment
Informative SNPs
Chinese Simmental
BovineSNP770K BeadChip

ABSTRACT

Incorrect paternity assignment in cattle can significantly influence the accuracy of genetic evaluation. Recent advances in high-throughput technology have facilitated the identification of single nucleotide polymorphism (SNP) markers and their applications for filiation and individual identification. We genotyped 1074 bulls from a reference population of Chinese Simmental cattle for genomic selection using a BovineSNP770K BeadChip. Among them, a total of 136 bulls were randomly selected to design a suitable low-density SNP panel for paternity testing in Simmental cattle. Our results showed that 50 SNPs were determined to be the most informative markers in parental testing, with an accuracy of 99.89% for CPE (cumulative probability of exclusion) in the unknown female parent case. The 50 highly informative SNP markers were distributed across 25 chromosomes, and the mean intermarker distance per chromosome was 26.72 Mb. The average minor allele frequency (MAF), expected heterozygosity (HE), and polymorphic information content (PIC) values were 0.3748, 0.4998, and 0.4818, respectively. Finally, the 50 identified SNPs were used to estimate paternity for the remaining 938 of 1074 bulls from 23 farms. Our results revealed that 76.75% of the 938 bulls were assigned parentage to the pedigree sires with 95% confidence, and the rate of pedigree record mistakes ranged from 9.52%–39.29% in different herds. Our study is the first attempt to provide valuable insights into the extraction of informative markers through the application of high-density SNP chips for paternity testing in Chinese Simmental cattle.

1. Introduction

Individual identification and parentage control are essential for consumer protection and efficient management of beef cattle. Incorrect paternity assignment in cattle can significantly affect the rates of genetic gain. A 10% error rate in paternity determination reduces genetic gain by 4.3% per year and cumulative genetic gain by 3.5% after 20 years (Israel and Weller, 2000). Misidentification should thus be controlled to < 8% to ensure > 1% genetic gain each year (Banos et al., 2001). However, in practice, the rate of incorrect paternity ranges from 4% to 23%. Ron et al. obtained paternity rejection rates of 5.2% and 6.25% in the Israeli Holstein population based on 173 cows and 244 elite cows, respectively (Ron et al., 1996; Ron et al., 2003). Weller et al. reported a paternity misidentification rate of 11.7% in the same Israeli Holstein population, which was based on genotyping 6040 cows using 87 microsatellite markers (Weller et al., 2004). The estimated rate of sire misidentification was ~10% in the UK dairy cattle population (Visscher et al., 2002) and ~7% in the German Angeln dairy cattle population (Sanders et al., 2006). Previous studies suggested that the average rate of pedigree error ranges from 11.83% to 20.9% in some large-scale Holstein dairy farms in China (Chu Qin et al., 2011; Guo Gang et al., 2012).

In the past decade, short tandem repeat markers (microsatellites) have been successfully used in bovine identification and parentage testing (Ozkan et al., 2009; Qiu et al., 2012; Radko et al., 2005; Wang et al., 2014). However, recent advances in high-throughput DNA sequencing, computer software, and bioinformatics have facilitated the identification of single nucleotide polymorphism (SNP) markers from amplified segments of genomic DNA. Notably, high-throughput genotyping technologies have been widely used in cattle (Kwok, 2001). The BovineSNP50K and BovineSNP770K BeadChips (Illumina, Inc., San Diego, CA, USA), which include 54,001 and 774,660 SNPs, were

Abbreviations: SNP, single nucleotide polymorphism; MAF, minor allele frequency; HE, expected heterozygosity; PIC, polymorphic information content; PE, the probability of exclusion; CPE, the cumulative probability of exclusion; LOD, the likelihood-odds ratio

* Corresponding authors at: Yuanmingyuan West Road 2#, Haidian District, Institute of Animal Science, Beijing 100193, PR China.
E-mail addresses: xulingyang@caas.cn (L. Xu), zhanglupei@caas.cn (L. Zhang), jl1@iascaas.net.cn (J. Li), gaoxue76@126.com (X. Gao).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.gene.2018.06.054
Received 17 April 2018; Received in revised form 8 June 2018; Accepted 18 June 2018
Available online 19 June 2018
0378-1119/

developed for genomic selection (Legarra et al., 2009; Misztal et al., 2009; VanRaden, 2008) and to validate paternity in cattle (Weller et al., 2010). However, previous simulation studies have shown that 60–100 SNPs may allow accurate pedigree reconstruction, even in situations involving thousands of potential mothers, fathers, and offspring (Anderson, 2012; Anderson and Garza, 2006). If the SNPs are selected, fewer SNPs are needed for paternity identification, and the minor allele frequency (MAF) and polymorphic information content (PIC) values will be higher (Baruch and Weller, 2008; Van Eenennaam et al., 2007).

Therefore, the objectives of this present study were to 1) identify a minimal set of informative SNPs for assignment of paternity using a BovineSNP770K BeadChip in Chinese Simmental cattle and to facilitate the design of accurate, low-cost SNP assays; 2) validate the effectiveness of the selected SNPs for assignment of paternity; and 3) perform paternity testing for young bulls using the selected SNP combination.

2. Materials and methods

2.1. Ethics statement

All animals were treated following the guidelines for experimental animals, which were established by the Council of China. Animal experiments were approved by the Science Research Department of the Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS) (Beijing, China).

2.2. Animals and SNP genotyping

The animals included in this study consisted of 136 calves with seven sires born in 2010, and 938 calves born from 2010 to 2013 with 39 sires in Ulagai, Inner Mongolia. All individuals were genotyped using an Illumina BovineHD BeadChip, which included 774,660 SNPs and was analyzed using Illumina Genome Studio software (version 2009.1, Illumina, San Diego, CA, USA). Only SNPs that met the following criteria were included for paternity verification (Cooper et al., 2013; Heaton et al., 2014):

(1) The SNPs are located on an autosome;
(2) Individual animal genotypes with call rates of > 0.95;
(3) MAF > 0.35;
(4) Deviation of heterozygote frequency from the expected Hardy-Weinberg value < 0.2;
(5) Mean intermarker distance of > 8 kb for each chromosome;
(6) All misplaced SNPs were excluded from the analysis.

2.3. Statistical methods

The measures of genetic variance, including expected heterozygosity (HE), PIC and MAF were calculated for each locus. The probability of exclusion (PE) estimates the probability of excluding a parentage relationship when the genotypes of the offspring and father are known (Marshall et al., 1998). The cumulative probability of exclusion (CPE) over unlinked loci of each individual was also calculated (Jamieson and Taylor, 1997). The natural logarithm of the likelihood-odds ratio, LOD, was used to assess the statistical confidence of that identification. A positive LOD indicates that a male is more likely to be the putative father than a male randomly drawn from the population, and a negative LOD indicates reverse. Once LOD scores are calculated for all candidate males, the male with the highest score is the putative father. All the analyses were conducted with CERVUS 3.0 based on likelihood equations that accommodate genotyping errors and hence increase the number of paternities that can be assigned at 95% (strict) confidence levels (Kalinowski et al., 2007; Marshall et al., 1998).

Table 1
Number of SNPs and distances between adjacent SNPs on each chromosome.

Chromosome	SNP number	Mean distance (bp)
1	21	7,354,475
2	16	8,029,802
3	17	6,984,484
4	17	6,996,422
5	13	9,053,982
6	12	9,595,180
7	14	7,429,312
8	12	8,042,507
9	11	8,846,757
10	12	8,447,642
11	12	8,807,058
12	12	7,902,112
13	8	9,781,722
14	8	11,082,892
15	12	7,386,769
16	11	6,649,051
17	10	7,106,908
18	10	6,448,047
19	8	7,383,361
20	10	6,816,321
21	8	7,285,076
22	7	8,215,187
23	6	8,028,097
24	8	7,314,625
25	7	5,714,290
26	8	6,247,783
27	11	3,606,057
28	5	8,936,709
29	5	9,978,805

3. Results

3.1. Identification of highly informative SNPs

A total of 311 SNPs met the selection criteria for informative diallelic markers, and the number of SNP markers on each chromosome is presented in Table 1. Among them, 8 SNPs were eliminated from further analysis because they displayed unbalanced heterozygous alleles or the HE and PIC values were significantly lower than average (Fig. 1A, B). Approximately 303 SNP markers remained. The average HE of the remaining SNP loci was 0.4899, PIC was 0.3713, and MAF was within the range of 0.4–0.5 (average: 0.4372). The CPE of the 303 SNP markers was 99.99%.

3.2. Simulation of selected SNPs for paternity testing

As the 303 polymorphic SNPs turned out to provide extremely high power to assign paternity, we also evaluated how many loci would be sufficient to ensure the assignment of paternity with 95% confidence in Chinese Simmental cattle. A simulation study was implemented with the 303 SNPs. Two strategies were used to select subsets of 100, 80, 70, 60, 50, 40, 30, 20 and 10 loci: (i) select the most polymorphic loci, and (ii) randomly select the subsets of loci. The likelihood-based parentage analysis was performed using CERVUS 3.0 software, and the parameters used were as follows: 99.67% loci typed, 1% genotyping error rate, 95% rate of strict confidence, 10,000 offspring, and the detection rate of candidate parents was set to 100%. The results showed that 40–50 loci could ensure the assignment of paternity with 95% confidence when selecting the most heterozygous SNPs. When SNPs were randomly selected, a total of 70–80 loci were required to give the same level of success (Fig. 2A). Similarly, 35–50 loci could meet a CPE of > 99% when selecting the most heterozygous SNPs; otherwise, 60–70 loci were required (Fig. 2B). The results from two comparisons also suggested that 40–50 SNP loci with an average expected heterozygosity of 0.5 were sufficient to meet the assignment of paternity at 95% confidence. This indicates that appropriate selection of highly informative SNPs led