J. Dairy Sci. 99:1–8
http://dx.doi.org/10.3168/jds.2015-10077

A method for single nucleotide polymorphism selection for parentage assessment in goats[1]

A. Talenti,*[2] E. L. Nicolazzi,† S. Chessa,‡ S. Frattini,* R. Moretti,‡ B. Coizet,* L. Nicoloso,* L. Colli,§ G. Pagnacco,* A. Stella,†‡ P. Ajmone-Marsan,§ G. Ptak,#‖ and P. Crepaldi*
*Dipartimento di Scienze Veterinarie e Sanità Pubblica, Università di Milano, via Celoria 10, Milano, 20133, Italy
†Fondazione Parco Tecnologico Padano (PTP), Via Einstein, Cascina Codazza, Lodi, 26900, Italy
‡Istituto di Biologia e Biotecnologia Agraria (IBBA-CNR), Consiglio Nazionale delle Ricerche, Via Einstein, Cascina Codazza, Lodi, 26900, Italy
§Istituto di Zootecnica, Università Cattolica del Sacro Cuore Via Emilia Parmense, Piacenza, 29122, Italy
#Dipartimento di Scienze Biomediche Comparate, Università di Teramo, Piazza Aldo Moro 45, Teramo, 64100, Italy
‖Institute of Animal Production, ul. Sarego 2, Krakow 31-047, Poland

ABSTRACT

Accurate pedigrees are essential to optimize genetic improvement and conservation of animal genetic resources. In goats, the use of mating groups and kidding management procedures hamper the identification of parentage. Small panels of single nucleotide polymorphisms (SNP) have been proposed in other species to substitute microsatellites for parentage assessment. Using data from the current GoatSNP50 chip, we developed a new 3-step procedure to identify a low-density SNP panel for highly accurate parentage assessment. Methodologies for SNP selection used in other species are less suitable in the goat because of uncertainties in the genome assembly. The procedure developed in this study is based on parent–offspring identification and on estimation of Mendelian errors, followed by canonical discriminant analysis identification and stepwise regression reduction. Starting from a reference sample of 109 Alpine goats with known pedigree relationships, we first identified a panel of 200 SNP that was further reduced to 2 final panels of 130 and 114 SNP with random coincidental match inclusion of 1.51×10^{-57} and 2.94×10^{-34}, respectively. In our reference data set, all panels correctly identified all parent–offspring combinations, revealing a 40% pedigree error rate in the information provided by breeders. All reference trios were confirmed by official tests based on microsatellites. Panels were also tested on Saanen and Teramana breeds. Although the testing on a larger set of breeds in the reference population is still needed to validate these results, our findings suggest that our procedure could identify SNP panels for accurate parentage assessment in goats or in other species with unreliable marker positioning.
Key words: parentage analysis, goat, single nucleotide polymorphism

INTRODUCTION

Accurate assessment of relationships between individuals in a population is one of the main requirements for a successful genetic improvement program. In goats, the accurate registration of parentage is often hampered by some widespread management practices, such as the use of mating groups, summer pasturing, and collective nurseries for kids at early stages of life. These practices usually lead to high rates of pedigree registration errors, which may involve one or both parents. Consequently, genetic progress slows down.

The use of molecular markers to determine parentage has been extensively studied in livestock (Heaton et al., 2002; Werner et al., 2004; Fisher et al., 2009; Matukumalli et al., 2009; Hayes, 2011; Heaton et al., 2014). Currently, DNA-based parentage analysis is shifting from the use of microsatellite (**MS**) to SNP markers. A panel of SNP has been recently adopted by the International Committee of Animal Recording (http://www.icar.org/) and by the International Society for Animal Genetics (**ISAG**, http://www.isag.us/) for cattle parentage testing (http://www.isag.us/Docs/Cattle-SNP-ISAG-core-additional-panel-2013.xlsx). Other comparison ring tests for parentage with a core panel of 100 SNP plus an additional panel of 100 SNP have recently been undertaken in cattle (Strucken et al., 2014). A panel of 163 SNP has also been proposed for parentage testing of sheep (Heaton et al., 2014).

The use of molecular markers for parentage analysis has been accompanied by the development of several statistical techniques for data management, principally based on exclusion, categorical, or fractional allocation and full probability (Jones et al., 2010).

Received July 7, 2015.
Accepted January 25, 2016.
[1]The authors declare they have no competing interests.
[2]Corresponding author: andrea.talenti@unimi.it

A goat 53K SNP array has been recently developed by Illumina (San Diego, CA), in collaboration with the International Goat Genome Consortium (http://www.goatgenome.org/; Tosser-Klopp et al., 2014). In spite of the high potential of such a tool in goat breeding, a few technical issues still partially hamper its full exploitation. One of the main issues concerns the early stage of the reference goat genome assembly (Dong et al., 2013). One of the main parameters used to select SNP for parentage assessment (**PA**) in other species (e.g., cattle) is physical distance and linkage disequilibrium (**LD**) among markers (Strucken et al., 2014). Although recent studies have characterized overall LD in some goat breeds (Brito et al., 2015), inaccurate SNP positioning on the goat genome makes traditional methods based on physical distance less effective for PA purposes (Benjelloun et al., 2015; Bickhart et al., 2015).

This paper evaluates a new 3-step procedure to select a small number of SNP for PA in goat. Advantages of this procedure are particularly relevant for species without an accurate SNP positioning, such as the goat. Our selection procedure of informative SNP markers is independent from SNP positioning, and it is based on parent–offspring identification by assessment of Mendelian errors (**MDE**), canonical discriminant analysis, and backward stepwise regression.

At present, the official MS-based method for parentage analysis in goats is generally too expensive for this species, considering the limited economic value of a single animal (Strucken et al., 2014). As already observed in sheep, the introduction of low-density SNP panels coupled with a cost-effective DNA-based technique would allow a 40% reduction of the analysis cost compared with MS (Heaton et al., 2014). This decrease in price would likely increase the use of DNA-based parentage determination in goats and help overcome the major constraints in control of inbreeding and implementing genetic improvement strategies.

MATERIALS AND METHODS

Animal Sampling and Genotyping

Blood samples were collected from 154 animals belonging to the Alpine (n = 109) and the Saanen (n = 22) goat breeds, reared in 3 flocks in Northern Italy, and the Teramana goat breed (n = 23), reared in one flock in central Italy. Samples were collected according to the recommendations of the European Council (Council of Europe, 1986) on animal care. The DNA was extracted from whole blood using a commercial kit (NucleoSpin Blood, Macherey-Nagel, Düren, Germany) following the manufacturer's instructions. The DNA samples were genotyped using the Illumina GoatSNP50 BeadChip (Tosser-Klopp et al., 2014).

All Alpine goats (8 males and 101 females) were used as the reference data set (**REF**), whereas 22 Saanen females (**VAL1**) and 23 (2 males, 21 females) Teramana individuals (**VAL2**) were used as the validation data set.

According to pedigree data, 46 out of 109 individuals in REF formed 20 trios (father, mother, and offspring), and in total 58 animals belonged to 50 parent–offspring (**PO**) pairs.

Group VAL1 included 5 PO pairs. Group VAL2 included one trio and 5 PO pairs. Group VAL2 animals were from the only existing flock of Teramana, an endangered breed consisting of about 80 animals recorded by the National Breeder Association (http://www.assonapa.com/) that was chosen specifically for its small size and difficult PA. The latter is due to high inbreeding, which increases the number of cryptic relationships, leading to PO classification instead of full-sib.

To confirm our results, 53 animals belonging to all trios in the REF data set were also analyzed with the official MS parentage test (13 different markers: *HSC*, *ILSTS19*, *INRA005*, *INRA063*, *MAF65*, *SRCRSP5*, *SRCRSP8*, *SRCRSP24*, *ILSTS23*, *INRA023*, *MCM527*, *CSRD247*, *SRCRSP23*). This test was performed in outsourcing at the official Laboratory of the Italian Breeders Association (AIA-LGS, http://www.lgscr.it/it/chi.htm).

Data Set Preparation

The REF genotype data were quality checked according to the following thresholds: SNP call rate $\geq$0.95; minor allele frequency (**MAF**) $\geq$0.01; individual genotype call rate $\geq$0.90; and in Hardy-Weinberg equilibrium (Bonferroni corrected threshold, $P \leq 1 \times 10^{-7}$). The resulting data set was used in the initial genetic population analysis and for PA. A characterization of the animals included in the REF data set was performed by PLINK 1.07 Multidimensional Scaling (Purcell et al., 2007) to verify the absence of sampling errors. For the selection of a low-density SNP panel, all markers with MAF $\leq$0.3, unknown chromosomal assignment or placement on the sexual chromosomes were excluded.

SNP Selection and PA

The 3-step procedure to select candidate SNP for the PA panel consisted of (1) the identification of real PO by MDE, (2) the identification of informative SNP by canonical discriminant analysis, and (3) reduction of