J. Dairy Sci. 96:1874–1879
http://dx.doi.org/10.3168/jds.2012-6176

Confirmation and discovery of maternal grandsires and great-grandsires in dairy cattle

P. M. VanRaden,*[1] T. A. Cooper,* G. R. Wiggans,* J. R. O'Connell,† and L. R. Bacheller*
*Animal Improvement Programs Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705-2350
†University of Maryland School of Medicine, Baltimore 21201

ABSTRACT

Selection, mating, and improvement of dairy animals have required accurate pedigrees. Genomic tools allow paternal ancestors to be easily confirmed or discovered because most sires are genotyped for many markers, but maternal ancestors are more difficult to discover because most female ancestors are not genotyped. Three methods to discover maternal grandsires (MGS) were developed and compared. Conflicts were counted one single nucleotide polymorphism (SNP) at a time between genotypes of the animal and potential MGS (duo method) or also using the sire's genotype (trio method). Alternatively, haplotypes of a potential MGS were matched to the animal's maternal haplotype, obtained by using linkage across loci (HAP method). The duo and trio methods can be performed as soon as a genotype is received because no imputation is required. The HAP method improved accuracy because genotypes with 2,683 (3K) SNP were imputed to the 45,187 (50K) SNP used for genomic evaluation. The HAP method was tested using modified pedigrees with 5% of true MGS replaced by a random genotyped bull from the same birth year and 5% of MGS set to missing for 4,134 Holsteins, 552 Jerseys, and 142 Brown Swiss that had confirmed, genotyped sires. Those same animals were used to test the duo and trio methods, except that some animals had multiple genotypes and imputed dams were excluded. Accuracy measured how often the correct MGS was selected from among 12,152 genotyped Holstein, 2,265 Jersey, and 1,605 Brown Swiss potential MGS. Accuracies were 61, 60, and 65%, respectively, with the duo method; 95, 91, and 94% with the trio method; and 97, 95, and 97% with the HAP method. Accuracy of the duo method was poor (only 52% for animals genotyped with 3K and 65% with 50K) because additional information from the paternal genotype is not used. Accuracy of the trio method was 97% with 50K but only 78% with 3K because the missing SNP were not imputed. Accuracy of the HAP method was 94% with 3K genotypes, 98% with 50K, and 92% with nongenotyped, imputed dams. When the HAP method was extended to great-grandsires, the accuracy of maternal great-grandsire discovery was 92% for 652 Holsteins, 95% for 33 Jerseys, and 85% for 20 Brown Swiss. Accuracy was even higher using simulated genotypes. Because most dairy bulls over several generations have been genotyped, percentages of haplotypes shared with candidate males can accurately confirm, correct, or discover the sires, MGS, and even more distant ancestors of most animals.

Key words: haplotype, genomics, pedigree discovery, pedigree reconstruction

INTRODUCTION

Genotypes that include more markers allow confirmation and discovery of parents and more distant ancestors such as grandparents and great-grandparents (Gusev et al., 2009; Kirkpatrick et al., 2011). Parentage in cattle has been confirmed for several decades using standard tests used by nearly all laboratories worldwide. These tests used blood group markers beginning in the 1940s (Stormont et al., 1951), microsatellite markers beginning in the 1990s (Heyen et al., 1997), and more recently hundreds or thousands of SNP. Small subsets of high-density SNP may allow confirmation of parentage across marker types by imputing the standard microsatellite alleles (McClure et al., 2012). A standard subset of approximately 100 SNP (Heaton et al., 2002) is included in almost all genotyping chips and has been accepted for international parentage confirmation (International Committee on Animal Recording, 2012). Analyses within a country or data set can improve power of detection by using all available markers instead of only the standard subset.

Discovery of parents or grandparents was not attempted with the blood group or microsatellite polymorphisms, but breeders could propose and check a small number of potential parents. Since 2008, Illumina BovineSNP50 [50,000 SNP (**50K**); Illumina, 2011a; Illumina Inc., San Diego, CA] genotypes for most re-

Received September 18, 2012.
Accepted November 22, 2012.
[1]Corresponding author: Paul.VanRaden@ars.usda.gov

cent sires and important ancestor sires have allowed unknown parents to be discovered and pedigrees to be corrected or constructed as a routine service for many dairy animals (Wiggans et al., 2009). Additional chips with more or fewer SNP are now used for genotyping a wider variety of animals. Between 4 and 14% of animals genotyped with the Illumina Bovine3K chip [3,000 SNP (**3K**); Illumina, 2011b) had incorrectly reported sires (Wiggans et al., 2012). Numbers of available SNP affect the accuracy of discovering the true sire (Dodds et al., 2005; Fisher et al., 2009; Hayes, 2011).

Correct identification and known pedigrees improve genetic progress by increasing both the number of usable phenotypic records and the effective heritability (Banos et al., 2001; Visscher et al., 2002; Dechow et al., 2008). Most early studies focused only on mistakes in paternity because confirmation or discovery of maternal ancestors was difficult due to few or no markers for most female ancestors. Discovery of pedigree from DNA is becoming a very general and important research topic affecting many plants, animals, and humans across the planet (Anderson and Garza, 2006; Hill et al., 2008; Gorbach et al., 2010; Ahlstrom, 2012). In natural populations such as trees or birds, no pedigrees were recorded historically and reconstruction of the pedigree information from DNA may be a very cost-effective breeding option (Pemberton, 2008; El-Kassaby and Lstibůrek, 2009). Exact algorithms such as maximum likelihood may not be feasible for pedigree reconstruction in populations having more than 30 individuals (Cowell, 2009).

Goals were (1) to develop new methods to confirm known or discover unknown maternal grandsires (**MGS**), (2) to compare accuracy of the methods for animals genotyped with different numbers of SNP, (3) to extend the haplotype-based method to confirm or discover unknown maternal great-grandsires (**MGGS**), and (4) to describe current and potential uses for ancestor discovery.

MATERIALS AND METHODS

Ancestor confirmation and discovery methods can use genotypes from other family members in addition to those of the animal and its proposed ancestor, and can count SNP conflicts one locus at a time or count haplotypes that the animal shares with its proposed ancestor. For example, standard parent verification increases the number of informative loci by using trio testing if the other parent is genotyped. Imputation can also increase the number of informative loci by filling missing alleles before comparing the animal and ancestor genotypes or haplotypes.

Three methods were compared for MGS confirmation and discovery. The first test (duo) simply counts the number of opposite homozygotes in the animal and MGS genotypes (Wiggans et al., 2009). The second test (trio) also counts conflicts using heterozygous loci if the sire is genotyped and homozygous, because the allele contributed by the dam is then known. Conflicts are counted if the allele from the dam is A and the MGS genotype is BB or if the allele from the dam is B and the MGS genotype is AA. The third test (**HAP**) imputes genotypes for all loci and counts the haplotypes in common instead of individual SNP conflicts. The paternal haplotype is removed from the animal's genotype (similar to the trio method) to determine the maternal contribution. A match is declared if the maternal haplotype is the same as either of the 2 MGS haplotypes.

Confirmation and discovery were extended to an additional generation with the HAP method. If the MGS was confirmed or discovered, the haplotypes contributed by the MGS were removed from the animal's maternal haplotypes to determine the maternal granddam (**MGD**) contribution. Those haplotypes remaining (about one-fourth of the animal's) are checked against the MGGS haplotypes. Theoretically, this process could be repeated to confirm or discover even more distant ancestors. For the MGS test, expected percentage of haplotypes in common is $<50\%$ because crossovers generate new haplotypes within some segments, but can be $>50\%$ if MGS haplotypes are homozygous because of inbreeding. For the MGGS test, expected percentage of haplotypes in common is $<25\%$.

Discovery of male ancestors is much easier than discovery of female ancestors in dairy cattle because most important historical bulls are genotyped. For genotyped bulls with US registrations, 100% of their sires are recorded and 96% are genotyped, whereas 100% of their dams are recorded but only 47% are genotyped. For these bulls, 95% of their MGS are genotyped. For genotyped females with US identification, 93% of their sires are recorded and 87% are genotyped, whereas 81% of their dams are recorded but only 22% are genotyped. For these females, 66% of their MGS are genotyped. To quantify the success of ancestor discovery methods, both actual and simulated genotypes are useful.

Simulated genotypes always match the pedigree file, whereas actual genotypes may not match the pedigree file because of recording errors. Simulated genotypes were used in initial testing to determine thresholds of action; for example, whether to suggest an MGS or to inform the breeder that the true MGS is probably not genotyped if no potential MGS is sufficiently related to the animal. The simulated data had the same pedigree information as the 109,286 genotyped Holsteins in August 2011, with 42,000 markers for 70,667 animals and 3,000 markers for 38,619 animals. The genotypes were simulated and missing markers were imputed using the