



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

P. M. VanRaden,\*<sup>1</sup> 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.

<sup>1</sup>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 Lstiburek, 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 MGS 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 MGS 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

methods of VanRaden et al. (2011) and the programs genosim and findhap (version 2). The haplotypes were in 577 segments of  $\leq 75$  markers each, the same as applied to actual data for each breed. On each chromosome, segments were chosen to have an equal number of markers rather than an equal physical length, and the same marker list was used across breeds.

The pool of potential MGS candidates for both simulated and actual data had 50K genotypes, must have been  $\geq 5$  yr older than the animal and not a half sibling, and included a total of 12,152 genotyped Holstein, 2,265 Jersey, and 1,605 Brown Swiss bulls. Additional restrictions imposed in the HAP method were that the MGS candidate must not be a sib or son of the MGD or the MGS of the MGD. Similarly, the MGGS candidate must not be a sib or son of the maternal great-granddam (MGGD) or the MGS of the MGGD. Those edits prevented close male relatives in more recent generations or in the maternal ancestry of the MGD or MGGD from appearing as the MGS or MGGS in question.

Actual genotypes were used for further tests by choosing a subset of animals that had confirmed sires as described by Wiggans et al. (2012), and then introducing 5% errors in the pedigree or removing the MGS pedigree connection for 5% to determine if the MGS could be corrected or recovered by genotyping. In the altered pedigree, MGS connections were removed by setting the dam's sire, rather than the animal's dam, to zero, and pedigree errors were introduced by replacing the dam's recorded sire with another genotyped sire randomly chosen within the same birth year.

The population of 4,134 Holsteins, 552 Jerseys, and 142 Brown Swiss with altered pedigrees were used to validate the HAP method of MGS confirmation and discovery. Genotypes used to validate the duo and trio methods were identical except that 4,631 Holsteins, 649 Jerseys, and 166 Brown Swiss were included for the duo method and 4,620 Holsteins, 659 Jerseys, and 160 Brown Swiss were included for the trio method because some animals had multiple genotypes (3K and 50K), but imputed dam genotypes were excluded. In addition, very few animals were included in one study but not the other because their genotype usability changed when records were added to the database. Animals were genotyped using the Illumina BovineSNP50 BeadChip or the Illumina Bovine3K BeadChip. Numbers of usable SNP were 42,503 from the 50K chip and 2,683 from the 3K chip. The trio method was further tested by first imputing from 3K to 50K and then comparing all 42,503 SNP of the animal and MGS, instead of only the 2,683 called SNP. The test included 1,475 animals with 3K genotypes, nearly the same as used in the test of the HAP method except for some differences in genotype

usability. Low-density chips with more SNP have since replaced the 3K, so current results with those chips would be better than these tests with the 3K chip.

The duo method calculated a conflict percentage (% Conflict) for each animal with each potential MGS using the following formula:

$$\% \text{ Conflict} = \frac{\#\text{Conflict}_{\text{Hom}}}{\#\text{Tested}} \times 100,$$

where  $\#\text{Conflict}_{\text{Hom}}$  is the number of opposite homozygous SNP genotypes between the 2 animals, and  $\#\text{Tested}$  is the number of SNP tested where both the animal and potential MGS have homozygous genotypes.

The trio method calculated a conflict percentage (% Conflict) for each animal with each potential MGS using the following formula:

$$\% \text{ Conflict} = \frac{\#\text{Conflict}_{\text{Hom}} + \#\text{Conflict}_{\text{Het}}}{\#\text{Tested}} \times 100,$$

where  $\#\text{Conflict}$  is the number of opposite homozygous SNP genotypes between the animal and potential MGS,  $\#\text{Conflict}_{\text{Het}}$  is the number of conflicting heterozygous SNP after determining which allele is maternal from the homozygous SNP genotype of the sire, and  $\#\text{Tested}$  is the number of SNP tested where both the animal and potential MGS have homozygous genotypes plus SNP where the sire and MGS have homozygous genotypes and the animal is heterozygous. Only the called genotypes were used in the duo and trio methods, not the imputed genotypes. The potential MGS candidates were then ranked based on lowest % Conflict value. The duo and trio methods were considered to have produced a confirmation if the potential MGS with the lowest % Conflict value was the reported pedigree MGS. Overall accuracy was the percentage of the tested animals where the reported pedigree MGS matched the best candidate MGS.

The HAP method of MGS confirmation and discovery was tested using the altered pedigree and the imputed genotypes, including 106 dams that were not genotyped but imputed from  $\geq 4$  progeny. Haplotype segments were phased to identify paternal and maternal contributions using methods of VanRaden et al. (2011). A match percentage (% Match) was calculated for each animal with each potential MGS using the following formula:

$$\% \text{ Match} = \frac{\#\text{Matches}}{\#\text{Segments}} \times 100,$$

where  $\#\text{Matches}$  is the number of exact matches between the maternal haplotype of the animal and either

maternal or paternal haplotype of the potential MGS, and #Segments is the count of total haplotype segments, which was 577. The potential MGS candidates were then ranked based on highest % Match value. The HAP method was confirmed if the potential MGS with the highest % Match value was the true pedigree MGS. The same process of removing known paternal contribution information was then extended another generation to validate MGS relationships. The maternal haplotypes were separated into those contributed by the MGS and the MGD, and only the latter were compared with the potential MGS candidate haplotypes.

Success rates with simulated data may be overestimated because the true MGS was included in the pedigree during imputation. Success with actual data may be underestimated because the reported MGS in the original pedigree was assumed to be correct and would not be recovered by genotyping if not the true MGS. If success rates from simulated and actual data are similar, then the overestimation and underestimation may both be small.

## RESULTS AND DISCUSSION

Estimates of haplotype sharing using simulated data indicated that the true ancestors can often be discovered by the HAP method. Percentages of shared haplotypes for the pedigree MGS and the top potential MGS were compared. If the true MGS was genotyped, the animal shared  $44 \pm 7\%$  with it, compared with  $28 \pm 5\%$  with the highest other potential MGS. The difference was  $16 \pm 6\%$ . The true MGS was rated as the top candidate 99.4%, the second place candidate 0.4%, and third or lower 0.2%. If the true MGS was not genotyped, the animal shared  $15 \pm 10\%$  with the top potential MGS compared with  $13 \pm 8\%$  with the second potential MGS. Genotyping of MGS was not random in either the simulated or actual data: the most popular MGS were genotyped and had many close relatives in the discovery pool, whereas less popular MGS were often not genotyped and had fewer close relatives. That is why the top other candidate averaged 28% if the MGS was genotyped but only 15% if not.

To avoid suggesting a MGS if the true MGS was not genotyped, the top candidate was suggested only if it shared  $\geq 35\%$  of haplotypes with the animal and  $\geq 1.15$  times the second place candidate. Only 1 MGS was falsely discovered for the 8,266 Holsteins where the true MGS was not reported or not genotyped. To avoid deleting valid MGS, the reported MGS was considered incorrect only if it shared  $< 25\%$  with the animal. Only 10 of the 101,020 genotyped true MGS of Holsteins would have been deleted with this threshold. False

discovery and false deletion rates were also  $< 0.05\%$  in Jerseys and Brown Swiss, and in MGS detection.

If the true MGS was genotyped, the animal shared  $19 \pm 5\%$  with it compared with  $13 \pm 3\%$  with the highest other potential MGS. The difference was  $6 \pm 3\%$ . The true MGS was rated as the top candidate 95.5%, the second place candidate 1.8%, and third or lower 2.7%. If the true MGS was not genotyped, the animal shared  $10 \pm 4\%$  with the top potential MGS compared with  $9 \pm 4\%$  with the second potential MGS. To avoid suggesting a MGS if the true MGS was not genotyped, the top candidate was suggested only if it shared  $\geq 20\%$  with the animal and  $\geq 1.2$  times the second place candidate. The reported MGS was considered a probable error if it shared  $< 5\%$  with the animal.

Accuracy of discovering the true MGS and MGS from actual data was also very high using the HAP method. The pedigrees that were switched to an incorrect MGS were detected as mistakes for 99.7% of Holstein, 100% of Jersey, and 100% of Brown Swiss MGS. The HAP method was more accurate for MGS discovery than the trio method because it used additional markers, especially for animals genotyped at lower density. Accuracy of discovering the correct MGS was 98% from 50K genotypes, 94% from 3K genotypes, and 92% for dams whose genotypes were imputed from  $\geq 4$  progeny. The HAP and trio methods both achieved higher accuracy than the duo method by including information from the paternal genotype (Table 1).

With the duo method, accuracy was 65% for 50K and 52% for 3K. By adding sire information, the accuracy of the trio method was 97% from 50K but only 78% from 3K. The percentage of SNP conflicts was greater with 3K than with 50K genotypes in the actual data (Wiggans et al., 2012), but 3K and 50K genotypes were of equal quality in the simulated data. Across chip types, the 3 breeds had similar accuracies of 61, 60, and 65%, respectively, for Holsteins, Jerseys, and Brown Swiss with the duo method; 95, 91, and 94% with the trio method; and 97, 95, and 97% with the HAP method (Table 2).

Accuracy of MGS discovery improved to 92% across breed using genotypes imputed from 3K to 50K compared with 78% using only the 3K SNP. Accuracy was not as high from imputed genotypes as the 97% from 50K genotypes because of a higher number of parent-progeny conflicts in the original 3K genotypes coupled with reduced SNP accuracy from imputation, and was not as high as the 94% for the HAP method from 3K data. The HAP and trio methods were most accurate for animals with 50K genotypes that had fewer closely related candidate MGS. If the true MGS was not genotyped, the HAP method had fewer false discover-

**Table 1.** Percentages of correct maternal grandsires discovered using individual markers without sire information (duo), individual markers with sire information (trio), or haplotype (HAP) methods by chip type [BovineSNP50 (~50,000 SNP), Bovine3K (~3,000 SNP), or imputed]<sup>1</sup>

Method	BovineSNP50		Bovine3K		Imputed dams	
	No.	% Discovered	No.	% Discovered	No.	% Discovered
Duo	3,542	65	1,818	52	—	—
Trio	3,620	97	1,733	78	—	—
HAP	3,197	98	1,455	94	106	92

<sup>1</sup>BovineSNP50 and Bovine3K chips by Illumina Inc. (San Diego, CA).

ies because of the additional limits introduced to avoid suggesting MGS with few shared haplotypes.

Accuracy of MGGS discovery was 92% for Holsteins, 95% for Jerseys, and 85% for Brown Swiss (Table 2) when the HAP method was extended another generation to great-grandsires. With high inbreeding, the HAP method for MGGS discovery could not easily distinguish the MGS contribution from the MGD contribution. In this case, removal of the MGS contribution can accidentally remove many haplotypes that were contributed by the MGD, causing fewer matches to be found between the MGD and the MGGS haplotypes.

Accurate pedigrees improve quality control and reliability of genomic evaluations. Discovery of missing ancestors allows mating programs to avoid inbreeding more easily in herds with incomplete pedigrees. If dams are not genotyped, confirmation or discovery of MGS assists in resolving switched samples and other coding mistakes. If the MGS is correct but the sire is wrong, the dam's service sire may have been recorded incorrectly. If the sire and MGS are both wrong, the DNA samples may have been switched or the calf may have been switched with another at birth. The latter may occur more often in large herds if many cows share the same maternity pen and are not attended full time or their calves are not tagged immediately. Because most ancestor bulls have been genotyped, sires and more distant ancestors such as MGS and MGGS can be accurately confirmed, corrected, or discovered for most dairy animals using the percentages of haplotypes that they share with the candidate males.

Choice of MGS confirmation and discovery methods is a balance between convenience, speed, and accuracy.

The duo method is simplest because a confirmed sire is not required. The duo and trio methods can be performed as soon as a genotype is received because no imputation is required. Comparing 50K imputed genotypes with genotypes of all MGS candidates requires more computation than comparing only the 3K subset. Therefore, MGS validation and discovery is done initially using raw genotypes as a first-stage screening tool, and final candidates are selected using the HAP method. The HAP method has greater accuracy because it includes low-density genotypes imputed to the 50K SNP and tests for patterns of SNP instead of each SNP individually. The HAP method can be extended to earlier generations.

The discovered MGS and MGGS may be as accurate as or more accurate than those reported by breeders. However, inserting the corrected or discovered ancestors into pedigrees can be difficult. The pedigree error could have occurred in recording an earlier (sire of the dam) or later (dam of the animal) generation. Even when the true MGS or MGGS are discovered, the true dam or MGD may remain undiscovered if not genotyped. Pedigree files usually do not skip a generation, but could perhaps record the discovered MGS or MGGS even if the dam and MGD remain undiscovered. If the dam is reported but not genotyped and the MGS is discovered, a pedigree record linking the dam to the MGS (her sire) is now generated automatically if no previous pedigree for the dam is available.

A fourth method of ancestor discovery could be more accurate than the duo, trio, and HAP methods. Each individual is the source of unique new haplotypes caused by crossovers between the maternal and pater-

**Table 2.** Percentages of correct maternal grandsires and great-grandsires discovered using individual markers without sire information (duo), individual markers with sire information (trio), haplotype maternal grandsire (HAP MGS), or haplotype maternal great-grandsire (HAP MGGS) methods by method and breed

Method	Holstein		Jersey		Brown Swiss	
	No.	% Discovered	No.	% Discovered	No.	% Discovered
Duo	4,631	61	649	60	166	65
Trio	4,620	95	659	91	160	94
HAP MGS	4,134	97	552	95	142	97
HAP MGGS	652	92	33	95	20	85

nal haplotypes in the gametes produced by each parent. If a later individual has one of these unique haplotypes, it must have inherited it from the source ancestor or its descendants. All other potential ancestors can be ruled out. One difficulty is that the proposed MGS could be an ancestor of the other parent (the MGD) instead of being the MGS. An initial attempt to use this method was not successful, but the idea has merit and could be pursued in further studies.

Since December 2011, MGS corrections and suggestions have been provided to breed associations and genotype requesters to improve quality control of the genotyping process and accuracy of the pedigree database. Confirmation and discovery of pedigree are important to breeders of many species and will become easier in the future as the numbers of markers, individuals, and generations genotyped increase within each population.

## CONCLUSIONS

Three methods of MGS discovery were compared using simulated genotypes that were consistent with the pedigree and using actual data where 5% of pedigree connections were set to unknown and another 5% of the reported MGS were set to an incorrect MGS. The duo and trio methods were simpler to use, but the HAP method discovered the true MGS more often because haplotype imputation allows more loci and more relatives to contribute. The true MGS was detected correctly 99.5% of the time from the list of potential MGS using the HAP method and simulated data. With actual 50K genotypes, the true MGS was detected 98% of the time using the HAP method. Incorrectly reported MGS and MGGS were almost always detected. The true MGGS was detected correctly 95.5% of the time from simulated data and 95% from actual data across genotype densities and breeds. Accurate confirmation, correction, and discovery of sires, MGS, and perhaps more distant ancestors of most dairy cattle are now possible because most ancestor bulls have been genotyped.

## REFERENCES

- Ahlstrom, D. 2012. Discover your genetic ancestors. *The Irish Times*, May 3, 2012. Accessed Sept. 12, 2012. <http://www.irishtimes.com/newspaper/sciencetoday/2012/0503/1224315503586.html>.
- Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics* 172:2567–2582.
- Banos, G., G. R. Wiggans, and R. L. Powell. 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. *J. Dairy Sci.* 84:2523–2529.
- Cowell, R. G. 2009. Efficient maximum likelihood pedigree reconstruction. *Theor. Popul. Biol.* 76:285–291.
- Dechow, C. D., H. D. Norman, N. R. Zwald, C. M. Cowan, and O. M. Meland. 2008. Relationship between individual herd-heritability estimates and sire misidentification rate. *J. Dairy Sci.* 91:1640–1647.
- Dodds, K., M. Tate, and J. Sise. 2005. Genetic evaluation using parentage information from genetic markers. *J. Anim. Sci.* 83:2271–2279.
- El-Kassaby, Y. A., and M. Lstiburek. 2009. Breeding without breeding. *Genet. Res. (Camb.)* 91:111–120.
- Fisher, P. J., B. Malthus, M. C. Walker, G. Corbett, and R. J. Spelman. 2009. The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. *J. Dairy Sci.* 92:369–374.
- Gorbach, D. M., M. L. Makgahlela, J. M. Reecy, S. J. Kemp, I. Baltenweck, R. Ouma, O. Mwai, K. Marshall, B. Murdoch, S. Moore, and M. F. Rothschild. 2010. Use of SNP genotyping to determine pedigree and breed composition of dairy cattle in Kenya. *J. Anim. Breed. Genet.* 127:348–351.
- Gusev, A., J. K. Lowe, M. Stoffel, M. J. Daly, D. Altschuler, J. L. Breslow, J. M. Friedman, and I. Pe'er. 2009. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* 19:318–326.
- Hayes, B. J. 2011. Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. *J. Dairy Sci.* 94:2114–2117.
- Heaton, M. P., G. P. Harhay, G. L. Bennett, R. T. Stone, W. M. Grosse, E. Casas, J. W. Keele, T. P. Smith, C. G. Chitko-McKown, and W. W. Laegreid. 2002. Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mamm. Genome* 13:272–281.
- Heyen, D. W., J. E. Beever, Y. Da, R. E. Evert, C. Green, S. R. E. Bates, J. S. Ziegler, and H. A. Lewin. 1997. Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplexes for semiautomated parentage testing. *Anim. Genet.* 28:21–27.
- Hill, W. G., B. A. Salisbury, and A. J. Webb. 2008. Parentage identification using single nucleotide polymorphism genotypes: Application to product tracing. *J. Anim. Sci.* 86:2508–2517.
- Illumina. 2011a. BovineSNP50 Genotyping BeadChip. Accessed Aug. 3, 2011. [http://www.illumina.com/Documents/products/datasheets/datasheet\\_bovine\\_snp50.pdf](http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp50.pdf).
- Illumina. 2011b. GoldenGate Bovine3K Genotyping BeadChip. Accessed Aug. 3, 2011. [http://www.illumina.com/Documents/products/datasheets/datasheet\\_bovine3K.pdf](http://www.illumina.com/Documents/products/datasheets/datasheet_bovine3K.pdf).
- International Committee on Animal Recording. 2012. ICAR rules and guidelines for parentage testing in cattle. Accessed Sept. 4, 2012. [http://www.icar.org/Documents/Rules and regulations/Amendments 2012/Parentage Testing in Cattle.pdf](http://www.icar.org/Documents/Rules%20and%20regulations/Amendments%202012/Parentage%20Testing%20in%20Cattle.pdf).
- Kirkpatrick, B., S. C. Li, R. M. Karp, and E. Halperin. 2011. Pedigree reconstruction using identity by descent. *Lecture Notes in Computer Science* 6577:136–152.
- McClure, M., T. Sonstegard, G. Wiggans, and C. P. Van Tassell. 2012. Imputation of microsatellite alleles from dense SNP genotypes for parental verification. *Front. Genet.* 3:140.
- Pemberton, J. M. 2008. Wild pedigrees: The way forward. *Proc. Biol. Sci.* 275:613–621.
- Stormont, C., R. D. Owen, and M. R. Irwin. 1951. The B and C systems of bovine blood groups. *Genetics* 36:134–161.
- VanRaden, P. M., J. R. O'Connell, G. R. Wiggans, and K. A. Weigel. 2011. Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43:10.
- Visscher, P. M., J. A. Woolliams, D. Smith, and J. L. Williams. 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *J. Dairy Sci.* 85:2368–2375.
- Wiggans, G. R., T. A. Cooper, P. M. VanRaden, K. M. Olson, and M. E. Tooker. 2012. Use of the Illumina Bovine3K BeadChip in dairy genomic evaluation. *J. Dairy Sci.* 95:1552–1558.
- Wiggans, G. R., T. S. Sonstegard, P. M. VanRaden, L. K. Matukumalli, R. D. Schnabel, J. F. Taylor, F. S. Schenkel, and C. P. Van Tassell. 2009. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J. Dairy Sci.* 92:3431–3436.