DNA-based paternity analysis and genetic evaluation in a large, commercial cattle ranch setting[1]

A. L. Van Eenennaam,*[2] R. L. Weaber,§ D. J. Drake,* M. C. T. Penedo,* R. L. Quaas,† D. J. Garrick‡, and E. J. Pollak†

*University of California, Davis 95616; §University of Missouri, Columbia 65211; †Cornell University, Ithaca, NY 14850; and ‡Colorado State University, Fort Collins 80523

ABSTRACT: Deoxyribonucleic acid-based tests were used to assign paternity to 625 calves from a multiple-sire breeding pasture. There was a large variability in calf output and a large proportion of young bulls that did not sire any offspring. Five of 27 herd sires produced over 50% of the calves, whereas 10 sires produced no progeny and 9 of these were yearling bulls. A comparison was made between the paternity results obtained when using a DNA marker panel with a high (0.999), cumulative parentage exclusion probability (P_E) and those obtained when using a marker panel with a lower P_E (0.956). A large percentage (67%) of the calves had multiple qualifying sires when using the lower resolution panel. Assignment of the most probable sire using a likelihood-based method based on genotypic information resolved this problem in approximately 80% of the cases, resulting in 75% agreement between the 2 marker panels. The correlation between weaning weight, on-farm EPD based on pedigrees inferred from the 2 marker panels was 0.94 for the 24 bulls that sired progeny. Partial progeny assignments inferred from the lower resolution panel resulted in the generation of EPD for bulls that actually sired no progeny according to the high-P_E panel, although the Beef Improvement Federation accuracies of EPD for these bulls were never greater than 0.14. Simulations were performed to model the effect of loci number, minor allele frequency, and the number of offspring per bull on the accuracy of genetic evaluations based on parentage determinations derived from SNP marker panels. The SNP marker panels of 36 and 40 loci produced EPD with accuracies nearly identical to those EPD resulting from use of the true pedigree. However, in field situations where factors including variable calf output per sire, large sire cohorts, relatedness among sires, low minor allele frequencies, and missing data can occur concurrently, the use of marker panels with a larger number of SNP loci will be required to obtain accurate on-farm EPD.

Key words: genetic evaluation, genetic marker, microsatellite, on-farm expected progeny difference, paternity, single nucleotide polymorphism

©2007 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2007. 85:3159–3169
doi:10.2527/jas.2007-0284

INTRODUCTION

In commercial ranch settings that utilize multiple-sire mating strategies, cattle producers have no simple way to determine the number of progeny sired by each bull or the relative performance of their progeny. Parental assignment by means of DNA testing is becoming less expensive. Marker-based pedigree assignment is now a feasible option for commercial producers and facilitates sire evaluation (Dodds et al., 2005). Traditionally, highly polymorphic microsatellite (**STR**) markers have been the choice for parentage inference (Vignal et al., 2002), but there is increasing interest in using SNP for this purpose due to their abundance, potential for automation, low genotyping error rates, and ease of standardization between laboratories (Anderson and Garza, 2006). The low resolving power of biallelic loci means that SNP panels need to include more loci than STR panels to achieve similar discriminatory power. Two SNP marker panels comprised of 32 and 37 SNP loci, respectively, have been proposed for use in bovine parentage analysis and identity testing (Heaton et al., 2002; Werner et al., 2004).

A set of marker loci can be characterized by its cumulative parentage exclusion probability (**P_E**), i.e., the

[1]Financial support for this work was provided by the California Cattlemen's Association Livestock Memorial Research Fund, California Beef Cattle Improvement Association, and special grant number 2005-34444-15725 from the USDA Cooperative State Research, Education, and Extension Service.

[2]Corresponding author: alvaneenennaam@ucdavis.edu

Received May 18, 2007.
Accepted July 26, 2007.

probability that a random individual other than a true parent from a population in Hardy-Weinberg equilibrium can be proven not to be the true parent of another randomly chosen individual. For unrelated sires, the probability of unambiguous parentage assignment is equal to P_E raised to the power of the number of nonparent candidate bulls in the breeding group (Sherman et al., 2004). Although the number of bulls in a breeding group does not directly affect P_E, the likelihood of unambiguously identifying the true sire by excluding every nonparent candidate decreases when more candidate sires are present. In herds with many natural service sires in a breeding group, panels with low P_E may result in multiple bulls qualifying as possible sires for a single calf [i.e., not being excluded as a sire (Sherman et al., 2004)]. Rather than discarding information from such individuals from sire evaluations, a calf's performance can be fractionally assigned to all qualifying bulls using likelihood scores derived from their genotypes (Weaber, 2005).

Here, we report on field data using a 28 SNP panel to assign paternity to calves in a commercial herd that employed a multiple-sire breeding pasture. A comparison was made between genetic evaluations obtained when using a powerful STR "gold standard" marker panel to assign paternity and those obtained when using the 28 SNP panel with a comparatively low P_E in combination with software designed to fractionally assign the performance of calves to multiple qualifying bulls. Additionally, simulations were performed to model the effect of loci number, minor allele frequency (**MAF**), and number of offspring per bull on the accuracy of genetic evaluations based on parentage determinations derived from SNP data.

MATERIALS AND METHODS

This study was conducted on a commercial farm using animals that are owned by the cooperator and standard animal husbandry practices for blood collection.

Cattle Population and DNA Collection

Blood or semen samples were obtained from 8 Angus and Hereford AI or 27 natural-service herd sire candidates run as a single cohort in a multiple-sire breeding pasture, and their 625 yearling progeny. Artificial insemination was performed on a subset of cows before their exposure to the herd sires. The DNA was isolated from semen using a standard phenol-chloroform extraction method, and from FTA cards (Whatman Inc., Florham Park, NJ) according to manufacturer's instructions. The herd sires included 4 paternal half-sib groups: 3 groups consisting of 2 sires, and 1 group consisting of 5 sires. Herd sires ranged from 1 to 8 yr of age at the time of breeding. All bulls had passed a breeding soundness examination by a licensed veterinarian before the breeding season.

Genetic Testing and Parentage Assignment

The STR genotyping based on 23 STR named ADCYC, BM203, BM888, BM1818, BM1824, BM2113, BM4107, BM4208, BRN, CYP21, ETH10, ETH152, ETH225, INRA23, OarFCB5, OarFCB193, RM006, RM067, SPS115, TGLA94, TGLA122, TGLA126, and TGLA227 was performed at the Veterinary Genetics Laboratory (**VGL**), University of California, Davis. These markers are routinely used at VGL for parentage verification of cattle. Primer sequences for the STR are available in public databases and can be obtained through marker name queries using "Marker Search" at the US Meat Animal Research Center's cattle genome Web site (http://www.marc.usda.gov/genome/genome.html; last accessed August 2007) and the "Request on Loci" at BOVMAP database (http://locus.jouy.-inra.fr/cgi-bin/bovmap/intro2.pl; last accessed August 2007). The PCR reactions with fluorescence-labeled primers were carried out according to VGL standard protocols, and amplicons were resolved by capillary electrophoresis on ABI 3730 sequencers (Applied Biosystems, Foster City, CA). The fragment size analysis software STRand (http://www.vgl.ucdavis.edu/informatics/STRand/; last accessed August 2007) was used for genotyping.

The SNP-based genotyping using 28 SNP (GenBank accessions AY761135, AY773474, AY776154, AY841151, AY842472, AY842473, AY842474, AY842475, AY849380, AY850194, AY851162, AY851163, AY853302, AY853303, AY856094, AY857620, AY858890, AY860426, AY863214, AY914316, AY916666, AY919868, AY929334, AY937242, AY939849, AY941204, AY942198, AY943841; http://www.ncbi.nlm.nih.gov/sites/entrez?db=Nucleotide; last accessed August 2007) derived from the Heaton et al. (2002) paper was performed by a commercial genotyping company (Genaissance, Duluth, GA).

Genotyping results from these 2 sets of analyses were run through Cervus (http://www.fieldgenetics.com/pagesaboutCervus_Overview.jsp; last accessed August 2007; Marshall et al., 1998), to determine the number of alleles per loci or MAF in the case of the biallelic SNP panel. Samples that contained DNA from more than 1 animal were removed from STR analysis before the data were run through the program. This program was also used to estimate number of alleles, observed and expected heterozygosity (assuming Hardy-Weinberg equilibrium), polymorphic information content (PIC; Botstein et al., 1980), goodness-of-fit Hardy-Weinberg equilibrium test, and loci exclusion probabilities for the situation where genotypes were available for putative parents of one sex but the other parent was unknown [Excl(1)] or both parental genotypes were available and one of the parents is known with certainty [Excl(2)]. The cumulative parentage exclusion probability (**P_E**) for the 2 marker sets was calculated according to Jamieson and Taylor (1997).

Paternity based on STR was determined by comparing the genotypes of all 35 potential sires against each calf's genotype. An exclusion was recorded when a bull and a calf had no allele(s) in common at a locus. Sire assignments based on STR were made in 2 rounds of analyses. First, sire(s) was assigned to a calf if there were no exclusions. Second, for remaining calves, a sire was assigned if he was the only bull with a 1-locus exclusion. Paternity was denoted unknown if no bull met either criterion.

Paternity based on SNP was assigned with the SireMatch software (E. J. Pollak, Cornell University), which uses a likelihood-based method to compute a probability that a putative sire is the true sire given genotypes of the calf, the dam, and all putative sires. In the current study, the dam's genotype was not collected and so population genotype frequencies computed from the genotypes of all bulls and calves were used. Genotype mismatches were not permitted; one excluding locus disqualified a bull. In cases where 2 or more bulls qualified with no exclusions as potential sires for a given calf (for both STR and SNP panels), probabilities from SireMatch were used to either categorically assign each calf to the single most probable sire or fractionally assign calves to all compatible (no exclusions) bulls for the genetic evaluations.

Genetic Evaluation

A genetic evaluation of 583 weaning weight records from progeny of 27 herd bulls and 8 AI sires was carried out using a sire model equation, $\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$, where $\mathbf{y}$ represented adjusted weaning weight observed from a single cohort, $\mathbf{b}$ was a vector of sex effects (bull, steer, heifer), $\mathbf{u}$ was a vector of direct sire progeny differences, $\mathbf{X}$ was an incidence matrix relating weaning weight observations to their sex, $\mathbf{Z}$ was an incidence matrix relating calves with weaning weights to their potential sires using parentage probabilities ascertained from either SNP or STR markers, and $\mathbf{e}$ was a vector of residuals. In contrast to the usual $\mathbf{Z}$ matrix that contains a single nonzero element of unity in each row, in the column corresponding to the sire of the calf represented by that row, this incidence matrix included as many nonzero elements in each row as there were potential sires for the calf. The sum of all the nonzero elements in any row was always 1. The number of nonzero elements in any column was the actual number of progeny that were assigned to that sire with any nonzero probability, whereas the total of any column was the equivalent number of summed progeny assigned to that sire (i.e., the sum of the fractional probabilities). The vector $\mathbf{u}$ included all known male ancestors of the calves (i.e., sires and paternal grandsires). This enabled straightforward computation of the inverse of the numerator relationship matrix and accounted for the half-sib relationships that existed between some sires. Weaning weights were adjusted for age at weaning and dam age according to BIF guidelines (BIF, 2002) using the computer software CattlePro (Bowman Farm Systems Inc., Cynthiana, KY), and the heritability of direct weaning weight was assumed to be 0.25. Records from calves with uncertain paternity have reduced genetic variation, and so the residual variance was inflated for these animals in order that the assumed phenotypic variance was identical regardless of paternity probabilities. The resulting mixed model equations were solved directly, and BIF accuracies (BIF, 2002) were computed from diagonal elements of the inverse coefficient matrix as if the assumed paternity was exact. Identical procedures were used in the analysis of simulated data.

Simulation Studies

Two studies were undertaken using simulated markers to investigate the influence of number of markers, MAF, and the number of offspring per bull. The first study quantified the probabilities a calf in this experiment would be assigned a single sire on the basis of its genotype. The second study determined the impact of uncertain paternity on the accuracy of EPD estimated from field data.

A more comprehensive SNP panel representative of the field information was simulated by creating 2 additional SNP markers with identical MAF to each of the 28 actual SNP. This generated a set of 84 realistic loci that was used to investigate P_E for marker panels ranging in size from 4 to 84 loci. An ideal panel with maximal exclusion rate was also created with the same number of markers but with a MAF of 0.5 at each locus. The probability of a calf having a unique sire pedigree assignment [**P(unique sire assignment)**] was computed for various panel sizes in increments of 4 loci, for both the realistic and maximal hypothesized panels. This probability was computed as: P(unique sire assignment) = $(P_E)^{n-1}$, where n = the number of possible sires, set at 27 to correspond to the number of natural service sires in the field study.

The second study simulated individual phenotypes and genotypes for 4 to 40 markers corresponding to a known pedigree involving 20 unrelated sires, with 5 or 30 progeny for every sire. The markers had realistic MAF based on the field study or maximal exclusion probabilities by assuming a MAF of 0.5 at all loci. The 40 markers were assumed to assort independently. Exclusion probabilities for each marker set and the theoretical maximum power of exclusion, assuming equal MAF, were computed as suggested by Jamieson and Taylor (1997).

Simulation of Sire Breeding Values and Offspring Phenotype. The true breeding values were simulated for the 20 sires using a normal distribution and were then used for the simulation of progeny phenotypes. Progeny phenotypes were simulated by adding half the breeding value of the sire to a normal deviate chosen to reflect a trait with h^2 = 0.25, the value assumed in analyzing the field data.

Simulation of Sire and Offspring Genotypes. Alleles at each locus were simulated for each sire by sampling a random number from the Uniform (0, 1) distribution. If the realization was less than or equal to the MAF then the first allele was assigned, otherwise the alternate allele was assigned. Genotypes for progeny were generated by sampling a single allele from the sire pair, and a second from an unknown dam population with equal allele frequencies.

Paternity Probabilities. Paternity probabilities were assigned using a likelihood based approach analogous to the algorithm used by Sire Match. Genetic evaluations were undertaken using the same procedures described for the field data, except that an additional evaluation could be undertaken using the actual relationships that were used in simulating the data.

Statistical Analysis and Data Sampling. Accuracy of evaluations, traditionally defined as the correlation between true and estimated breeding value, were computed as Pearson product moment correlations.

Replicates. Two hundred samplings of sire breeding values and progeny genotypes and phenotypes were created for each scenario (number of markers, MAF, progeny number). For each scenario, paternity probabilities were computed, followed by genetic evaluation and computation of the accuracy of evaluation.

RESULTS AND DISCUSSION

Table 1 shows the descriptive statistics for the 2 marker panels used in the field study. For the STR panel the mean number of alleles per locus was 9, the mean PIC was 0.626, and all loci but one were in Hardy-Weinberg equilibrium. Based on the allele frequencies present in the herd, the STR marker panel had a cumulative P_E of 0.999 in the situation where genotypes were available for putative parents of 1 sex but the genotype of the other parent was unknown. A total of 533 calves (85.4%) were unambiguously assigned to 1 sire, 4 calves (0.6%) had 2 qualifying sires, and 76 (12.2%) calves had no qualifying sire (i.e., DNA from the true sire was apparently missing or there were genotyping errors). All nonsires were excluded at a minimum of 2 loci except for 1 assignment where a single highly exclusionary result at a single locus was used. Sire-calf pairs had matching alleles at every locus with the exception of 10 cases where a 1-locus mismatch was allowed. Twelve of the samples (1.8%) were removed from analyses because they amplified poorly (1) or appeared to contain DNA from more than 1 animal (11), either due to twinning or admixture during sample collection (Table 2). Five of the 27 herd sires produced over 50% of the calves. The leading digit of the sire identification number denotes the age of the bull at the time of breeding, and it can be seen that of the 10 natural-service herd bulls that sired no progeny, 9 were yearlings (Figure 1).

The biallelelic SNP loci had mean PIC of 0.35, and the calculated P_E without dam genotype was 0.956. This panel unambiguously assigned 175 (28%) calves to 1 sire with no mismatches, 420 calves (67%) to 2 or more qualifying sires (average 3 ± 2.8), and 29 (5%) of calves were found to have no qualifying sire (Table 2). At least 1 offspring was assigned to every sire in the group by the SNP panel, including 56 of 76 calves where the STR panel excluded all sires in the group. In situations where more than 1 sire qualified to a calf, the categorical assignment of the most probable sire determined by SireMatch, matched the STR panel sire assignment almost 80% of the time. Overall, the 2 methods agreed in assignment (either to an individual sire or no qualifying sire) in 75% of cases, and disagreed based either on individual bull assignment (15%) or on whether a qualifying sire was present (10%). This latter finding emphasizes the importance of having a marker panel with relatively high P_E when analyzing field data where parental genotypes, from dams or some sires, are missing (Vignal et al., 2002). It is also of interest that the presence of DNA from more than 1 animal in a sample (e.g., twin, mixed sample) was only recognized with the microsatellite-based assay where the occurrence of more than 2 alleles per locus in a single sample could be distinguished.

The 23-loci STR panel used in this study had higher exclusion power than is typically used in commercial genotyping laboratories. The median number of ISAG-recommended STR microsatellite used in cattle genotyping is 12 loci (Baumung et al., 2004), and such panels would have lower P_E. The 23-loci STR panel helped ensure sufficient power to determine correct paternity for each calf in the sample, enabling a valid evaluation of the SNP assignment. Figure 2 shows the results of the first simulation study that examined the number of SNP loci that would be required to achieve a given P_E assuming 1 unknown parent, given equal or simulated (based on field study frequencies) MAF at each SNP locus, and independently assorting loci. It can be seen that the observed P_E (0.956) in this study with 28 SNP loci was lower than the theoretical maximal exclusion rate with equal MAF at the 28 SNP loci (0.976). This was due to some SNP loci having low MAF in the population examined (Table 1).

Figure 2 further illustrates that as the number of SNP loci increases, P_E and the probability of unique sire pedigree assignments increased. Increasing the number of SNP markers with optimal MAF, or the total number of SNP loci in the panel improved the P_E and hence the probability of single sire assignment. For the 28 SNP panel utilized in the field study, the probability of a calf having unique sire assignment among the 27 herd bulls was 0.315, and for an idealized panel with 0.5 MAF it was 0.535. The simulated SNP panel required markers at 64 loci to achieve the same exclusionary power (0.999) as the 23-loci STR panel and achieve a 0.974 probability of single sire assignment among 27 herd bulls. Inclusion of all 84 loci in the simulated SNP panel yielded a P_E of greater than 0.9999 and had a probability of progeny having a unique sire identified,

Table 1. Marker name, number of alleles, animals genotyped (n), and homozygotes (No. Homs), observed [H(O)] and expected [H(E)] heterozygosities, polymorphic information content (PIC), and loci exclusion probabilities, for the situation where genotypes were available for putative parents of one sex and the other parent was unknown, Excl(1), or both parental genotypes were available and one of the parents is known with certainty, Excl(2), for the 23 microsatellite (STR) and 28 SNP (GenBank accession number) genetic markers, and the minor allele frequency (MAF) for the 28 SNP genetic markers used in the field study

Panel	Locus	No. of alleles	No. of animals	No. of Homs	H(O)	H(E)	PIC	Excl(1)	Excl(2)	MAF
STR										
1	FCB19	2	649	508	0.22	0.21	0.18	0.02	0.09	
2	FCB	7	649	477	0.27	0.27	0.26	0.04	0.15	
3	BM410	8	649	420	0.35	0.36	0.35	0.07	0.21	
4	RM00	8	648	305	0.53	0.54	0.51	0.16	0.33	
5	TGLA12	7	649	207	0.68	0.64	0.58	0.23	0.38	
6	BM181	7	649	243	0.63	0.64	0.59	0.24	0.41	
7	TGLA9	8	649	226	0.65	0.66	0.60	0.24	0.41	
8	ETH1	8	649	201	0.69	0.69	0.63	0.26	0.43	
9	ETH15	7	647	201	0.69	0.70	0.64	0.26	0.43	
10	BM20	14	649	192	0.70	0.70	0.65	0.29	0.46	
11	ADCY	8	649	189	0.71	0.69	0.65	0.29	0.47	
12	SPS11	9	649	189	0.71	0.69	0.65	0.29	0.47	
13	BM182	6	648	177	0.73	0.70	0.66	0.29	0.47	
14	INRA2	8	649	168	0.74	0.71	0.66	0.30	0.47	
15	RM06	6	649	186	0.71	0.71	0.67	0.30	0.47	
16	TGLA12	13	648	158	0.76	0.72	0.68	0.32	0.50	
17	BR	12	649	191	0.71	0.72	0.69	0.33	0.51	
18	BM420	8	649	148	0.77	0.77	0.73	0.38	0.56	
19	ETH22	8	649	137	0.79	0.78	0.74	0.39	0.57	
20	TGLA22[1]	10	649	137	0.79	0.81	0.78	0.45	0.63	
21	BM88	10	649	129	0.80	0.82	0.79	0.46	0.64	
22	BM211	9	648	115	0.82	0.82	0.80	0.47	0.64	
23	CYP2	24	648	50	0.92	0.91	0.90	0.69	0.81	
SNP										
1	BTA_AY849380	2	617	432	0.30	0.30	0.25	0.04	0.13	0.18
2	BTA_AY863214	2	634	433	0.32	0.32	0.27	0.05	0.13	0.20
3	BTA_AY842475	2	631	424	0.33	0.32	0.27	0.05	0.14	0.20
4	BTA_AY939849	2	634	365	0.42	0.41	0.33	0.08	0.16	0.29
5	BTA_AY842473	2	651	368	0.44	0.41	0.33	0.09	0.16	0.30
6	BTA_AY860426	2	615	348	0.43	0.42	0.33	0.09	0.17	0.31
7	BTA_AY842474	2	646	352	0.46	0.45	0.35	0.10	0.17	0.34
8	BTA_AY851163	2	619	337	0.46	0.45	0.35	0.10	0.18	0.35
9	BTA_AY916666	2	617	335	0.46	0.46	0.35	0.11	0.18	0.36
10	BTA_AY773474	2	626	333	0.47	0.46	0.36	0.11	0.18	0.36
11	BTA_AY942198	2	607	333	0.45	0.47	0.36	0.11	0.18	0.37
12	BTA_AY842472	2	655	341	0.48	0.47	0.36	0.11	0.18	0.37
13	BTA_AY914316	2	640	335	0.48	0.47	0.36	0.11	0.18	0.38
14	BTA_AY853302	2	613	331	0.46	0.47	0.36	0.11	0.18	0.38
15	BTA_AY853303	2	647	340	0.47	0.48	0.36	0.11	0.18	0.39
16	BTA_AY761135	2	625	327	0.48	0.48	0.36	0.11	0.18	0.39
17	BTA_AY850194	2	639	316	0.51	0.48	0.37	0.12	0.18	0.41
18	BTA_AY857620	2	635	326	0.49	0.49	0.37	0.12	0.18	0.42
19	BTA_AY941204	2	644	316	0.51	0.49	0.37	0.12	0.18	0.42
20	BTA_AY943841	2	611	303	0.50	0.49	0.37	0.12	0.19	0.43
21	BTA_AY856094	2	648	324	0.50	0.50	0.37	0.12	0.19	0.45
22	BTA_AY841151	2	636	333	0.48	0.50	0.37	0.12	0.19	0.46
23	BTA_AY851162	2	613	310	0.49	0.50	0.37	0.12	0.19	0.46
24	BTA_AY919868	2	613	311	0.49	0.50	0.37	0.12	0.19	0.47
25	BTA_AY929334	2	637	309	0.52	0.50	0.37	0.12	0.19	0.47
26	BTA_AY858890	2	626	313	0.50	0.50	0.37	0.12	0.19	0.47
27	BTA_AY937242	2	653	315	0.52	0.50	0.38	0.13	0.19	0.49
28	BTA_AY776154	2	635	312	0.51	0.50	0.38	0.13	0.19	0.49

[1]This locus was not in Hardy-Weinberg equilibrium ($P < 0.05$).

Table 2. Predicted and observed percentages of 625 calves assigned to 0, 1, 2, 3, 4, 5, or more of 35 candidate sires using calf and sire genotypes for a microsatellite (STR) or SNP genetic marker panel with differing cumulative exclusion probabilities (P_E)

No. of sires assigned per calf	STR (P_E = 0.999)		SNP (P_E = 0.956)	
	Predicted % of calves	Observed % (No. of calves)	Predicted % of calves	Observed % (No. of calves)
0	0	12.2 (76)	0	4.6 (29)
1	96.7	85.4 (533)	21.7	28.0 (175)
2	3.3	0.6 (4)	33.9	24.6 (154)
3	<0.1	0	25.7	13.9 (87)
4	0	0	12.6	10.9 (68)
5	0	0	4.5	5.4 (34)
≥6 or more	0	0	1.6	12.3 (77)
No result	0	1.8 (12)		0.0 (1)

among 27 sires, of greater than 0.9978. Coincidentally, the largest deviation between probability of identification of a single sire for a progeny using the P_E from the idealized panel or that of the panels with MAF less than 0.50 occurred with a SNP panel consisting of 28 loci as was used in this field study. Deviations between the maximal probability of single sire inclusions and those from the constructed panel decreased as more SNP loci were added to the system. However, as the number of loci was increased, the incremental improvements occurred at a decreasing rate. More than 20 additional SNP (~50 total) were required to achieve a 90%

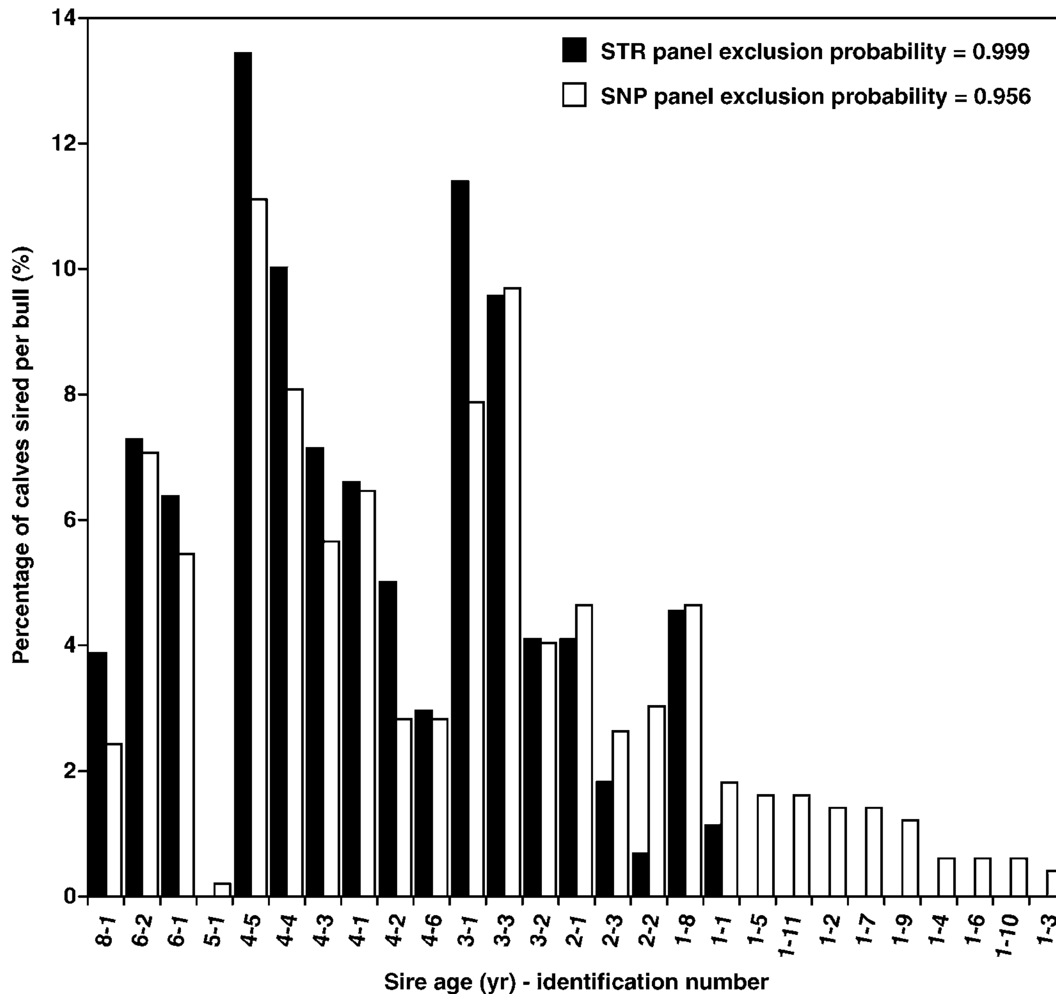

Figure 1. Calf output of 27 herd bulls of varying ages in a multiple-sire breeding pasture, as determined using genetic marker panels with differing probabilities of exclusion.

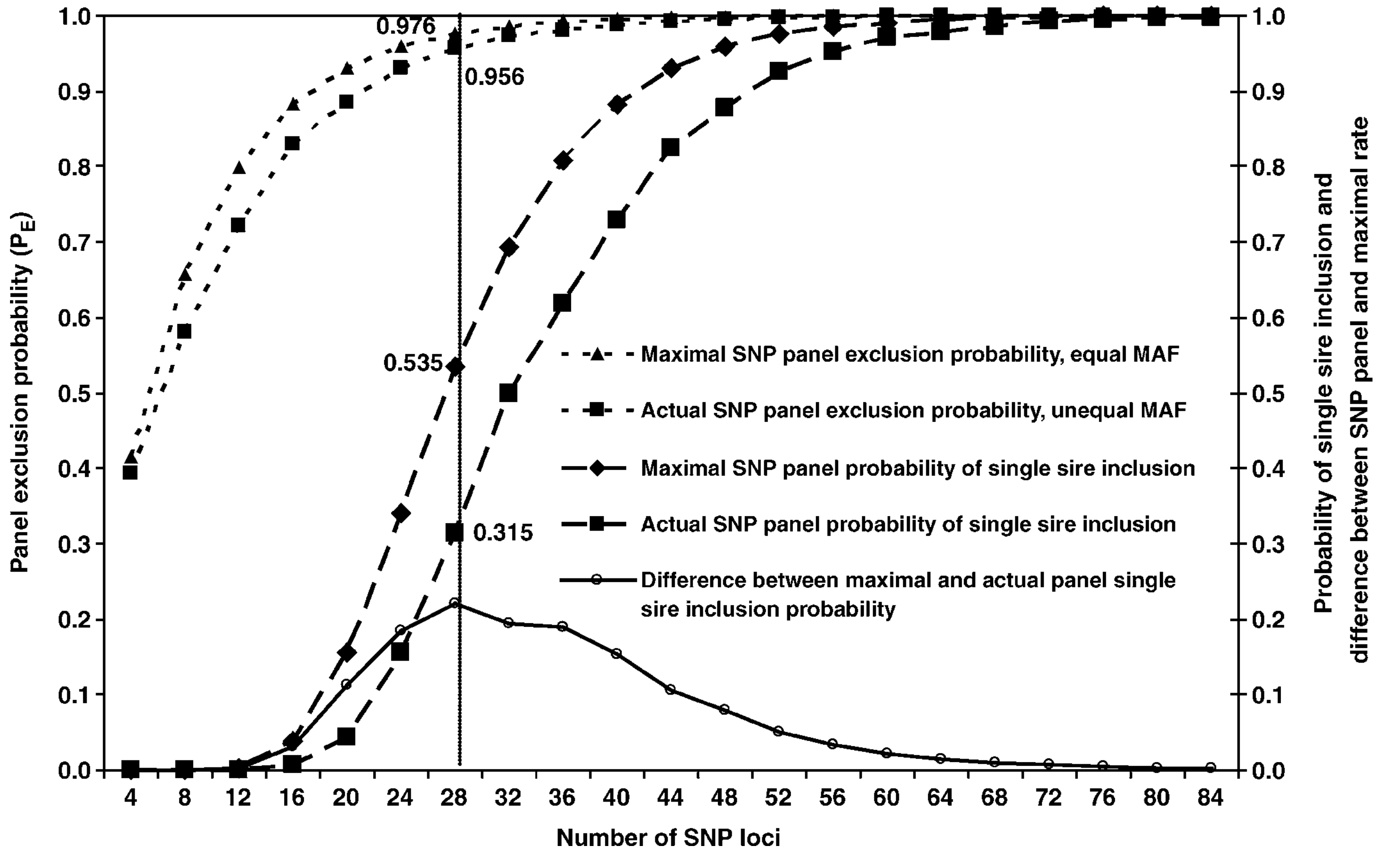

Figure 2. Theoretical maximal and actual SNP marker panel exclusion probabilities (P_E) with increasing numbers of SNP based on equal minor allele frequencies (MAF) and observed (unequal) MAF, respectively, and the probability of single sire inclusion for a multiple-sire breeding pasture containing 27 sires. The number of loci (28) included in the SNP panel used to analyze the field data set is indicated with a vertical line.

chance of obtaining a single sire assignment to a progeny when there were 27 putative sires.

The calculation of P_E based on allelic frequencies assumes all candidate bulls are unrelated. In the context of a commercial ranch setting where half-sib sires are common, it is important to recognize that as the number or relatedness of putative sires in a multiple-sire breeding group increases, additional numbers of marker loci would be required to maintain single sire assignments at a fixed rate (Pollak, 2005).

A comparison of the sire evaluations for weaning weight derived from pedigree assignments for the 2 panels used in the field study is shown in Table 3. In this analysis multiple bulls qualifying to a single calf were fractionally assigned to all qualifying males according to the likelihood score computed by Sire Match. The STR pedigree included 24 sires, 12 paternal grand sires, and 503 calf records, whereas the SNP pedigree included 35 herd sires, 14 paternal grand sires, and 558 calf records. The reason for the larger number of calf records in the genetic evaluation based on the SNP panel assignments was that the SNP panel analysis included records from some calves that were excluded from all known sires in the STR panel analysis. In the case of the SNP panel with a P_E of 0.956, bulls that actually sired no progeny according to the STR panel were incorrectly assigned fractional probabilities of calves summing to as many as 9 equivalent progeny (the sum of the fractional probabilities for each sire), which resulted in them receiving an erroneous EPD. Although the correlation between the STR and SNP-based EPD was 0.94 for the 24 bulls that had progeny according to the STR panel, the generation of EPD for bulls that actually sired no progeny would create obvious problems from the perspective of genetic improvement. In some cases, erroneous assignments based on the SNP-panel results placed bulls among the top 10 bulls for weaning weight EPD, although in this particular example the BIF accuracy of such EPD was never higher than 0.14, and the number of equivalent progeny was less than 10. Ideally, the resolution of the marker panel would be sufficient to minimize the assignment of equivalent progeny to such sires.

Figure 3 illustrates the results of the second simulation study and shows that for bulls with 30 progeny each, the accuracy of EPD was 0.81 when the true pedigree was used for estimation. Note that this accuracy was constant across all simulations examining varying number of loci and allele frequencies. Increasing the number of SNP loci improved the P_E for both panels and

Table 3. Weaning weight EPD and Beef Improvement Federation accuracies (Acc) calculated from adjusted weaning weight (kg) records representing 27 herd bulls and 8 AI sires from a single multiple-sire breeding pasture using parentage assignments derived from a microsatellite (STR; $P_E = 0.999$) or SNP panel ($P_E = 0.956$) with differing cumulative exclusion probabilities (P_E)

	STR panel						SNP panel				
Herd bulls	EPD	Acc	Equiv prog[1]	Act prog[2]	Rank	Herd bulls	EPD	Acc	Equiv prog[1]	Act prog[2]	Rank
1-8	13.5	0.32	19.0	19	1	1-8	9.0	0.31	20.7	37	1
4-1	5.6	0.36	25.0	25	2	4-1	4.0	0.34	24.6	33	6
8-1	4.6	0.28	15.3	18	3	8-1	2.0	0.21	12.5	28	11
6-1	2.5	0.34	22.0	22	4	6-1	3.1	0.29	18.7	31	8
4-2	1.3	0.33	21.0	21	5	4-2	−2.2	0.21	16.1	73	17
1-1	0.9	0.11	4.0	4	6	1-1	3.0	0.11	7.1	23	9
2-1	0.4	0.30	17.0	17	7	2-1	−0.5	0.29	21.3	59	14
4-3	−1.6	0.39	29.0	29	8	4-3	−0.5	0.31	25.9	91	13
3-1	−3.2	0.46	46.0	46	9	3-1	−2.7	0.40	38.2	87	19
4-5	−3.7	0.49	57.0	57	10	4-5	−1.9	0.41	43.3	69	16
2-2	−4.5	0.09	3.0	3	11	2-2	−6.6	0.24	21.7	171	23
6-2	−5.6	0.41	33.0	33	12	6-2	−2.5	0.36	30.9	58	18
4-4	−7.1	0.43	38.0	38	13	4-4	−6.5	0.37	33.1	58	22
3-2	−7.6	0.30	17.0	17	14	3-2	−13.9	0.28	17.6	29	25
4-6	−8.1	0.24	12.0	12	15	4-6	−4.8	0.15	12.8	63	21
3-3	−14.3	0.45	42.7	43	16	3-3	−17.5	0.44	43.4	55	26
2-3	−22.1	0.19	8.0	8	17	2-3	−20.3	0.20	14.2	66	27
1-6	.	.	.	0	.	1-6	6.6	0.07	3.8	21	2
1-11	.	.	.	0	.	1-11	6.2	0.10	8.4	44	3
1-7	.	.	.	0	.	1-9	4.8	0.10	6.3	27	4
1-9	.	.	.	0	.	1-7	4.6	0.11	8.8	59	5
1-2	.	.	.	0	.	1-2	3.2	0.14	9.8	87	7
1-4	.	.	.	0	.	1-4	2.5	0.02	2.4	25	10
1-5	.	.	.	0	.	1-5	1.8	0.09	4.5	14	12
5-1	.	.	.	0	.	5-1	−0.7	0.03	1.3	5	15
1-10	.	.	.	0	.	1-10	−4.6	0.07	6.6	64	20
1-3	.	.	.	0	.	1-3	−8.3	0.07	3.3	15	24
AI SIRES						AI SIRES					
AISIRE – 1	14.1	0.31	19.0	19	1	AISIRE - 1	7.6	0.22	15.3	40	3
AISIRE – 2	14.0	0.27	15.0	15	2	AISIRE - 2	15.1	0.22	15.9	70	1
AISIRE – 3	9.2	0.23	11.0	11	3	AISIRE - 3	8.0	0.21	13.1	43	2
AISIRE – 4	7.5	0.26	14.0	14	4	AISIRE - 4	6.4	0.18	12.2	46	4
AISIRE – 5	4.4	0.31	18.0	18	5	AISIRE - 5	5.0	0.28	19.5	32	6
AISIRE – 6	1.5	0.24	12.0	12	6	AISIRE - 6	−1.2	0.22	11.5	19	7
AISIRE – 7	−7.4	0.13	5.0	6	7	AISIRE - 7	−1.5	0.13	7.8	41	8
AISIRE – 8	.	.	.	0	.	AISIRE - 8	6.3	0.03	5.4	81	5

[1]Equiv prog is the sum of the fractional probabilities for each sire.
[2]Act prog is the actual number of progeny that included that sire with any nonzero probability.

resulted in increasingly reliable genetic evaluations, evidenced by the improved correlations between true breeding values and EPD. Larger SNP marker panels of 36 and 40 loci produced EPD with accuracies nearly identical to those EPD resulting from use of the true pedigree. It is not possible to ensure a minimum or equal number of progeny for each sire present in multiple-sire pastures and this presents a problem for genetic evaluations. Figure 3 shows that accuracies were reduced when only 5 instead of 30 progeny per sire were simulated and included in the evaluation. Likewise the inclusion of SNP loci with unequal allele frequencies or related bulls, as would be common in many applied situations, would also reduce the accuracy of breeding value estimates due to the inclusion of multiple bulls as qualifying sires.

The correlations between EPD produced from the true pedigree and those produced from inferred pedigrees derived from analysis using SNP marker panels of varying power are shown in Figure 4. It can be seen that as the number of SNP loci in the marker panel increases, the correlation between the EPD computed with the inferred pedigree and the true pedigree approached unity. This can be explained by the more complete and precise resolution of the relationships between sires and progeny that occurs with a more powerful panel. With low numbers of loci the correlations are poor, indicating low utility of the inferred pedigree. A panel of 36 or 40 SNP markers produced favorable correlations suggesting that a large portion of genetic variation was correctly attributed to true sires in the pedigree structure. However, this simulation study did not

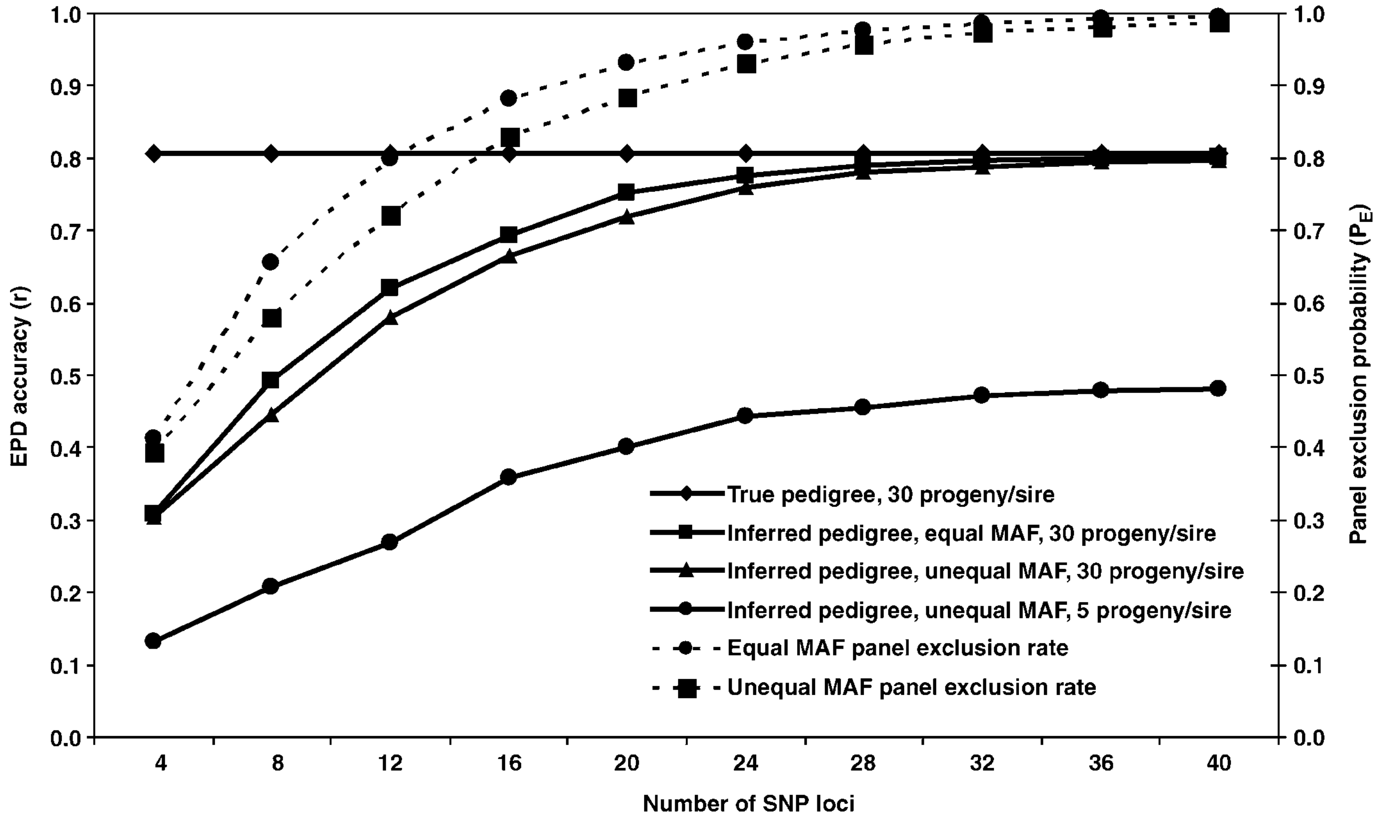

Figure 3. Correlation between simulated and predicted EPD from true or inferred pedigrees derived from genotype analysis using SNP marker panels with 4 to 40 loci and different minor allele frequencies (MAF) resulting in varying panel exclusion probabilities (P_E).

include the effect of misassignment of progeny that were sired by bulls missing from the set of putative sires, or progeny assigned to unproductive sires that in fact sired no progeny as was observed in the field study. Such situations are likely to be relatively common in field situations, and so the results of the simulation study may present an overly optimistic picture of the results that can be obtained with a 40 loci SNP panel under field conditions.

This study, which included a field trial using commercial ranch records and conditions, highlights some important factors to consider with respect to parentage testing for sire evaluations of bulls on large commercial ranches. There were obviously bulls missing from the field data set. This may have occurred for a variety of reasons including neighboring bulls jumping the fence, precocious bull calves, or inadvertent omission of sire(s) from sample collection. Missing samples of sires when using DNA marker-based parentage for genetic evaluation decreases the rate of genetic gain (Dodds et al., 2005). Sampling bulls for DNA testing as they are put out with cows, rather than at a later time, would help to minimize the missing sire samples. A paternity testing study in 10 large commercial herds in Australia similarly reported an average of 9.9% of calves with missing sires, with a range from 0 to 33.6% (Holroyd et al., 2002). Missing sires were especially problematic for the panel with a low P_E because it was not sufficiently powerful to exclude nonparent sires from calves when DNA from the true sire was not in the sire sample pool. This problem, which is not generally included in simulation studies or controlled research herds examining the effectiveness of paternity assay systems, is likely to be relatively frequently encountered in commercial situations.

There was a large variability in calf output reflecting variation in mating success to the extent that a large proportion of young bulls did not sire any offspring. Other studies have reported similar variability in calf output among herd sires, and further found success is moderately repeatable providing the composition of the bull mating groups remains similar (DeNise, 1999; Holroyd et al., 2002). The unexpectedly large number of young bulls that did not produce any progeny in this particular trial would present an obvious problem for sire evaluation programs based on progeny tests, especially when considering that the years of service subsequent to genetic evaluation have a significant effect on the return on investment for progeny testing (Weaber, 2005). A separate multiple-sire breeding pasture for yearling bulls would be advantageous.

Using DNA testing to generate on-farm EPD for sires in multiple-sire breeding groups represents a promising application of biotechnology. The simulation and field

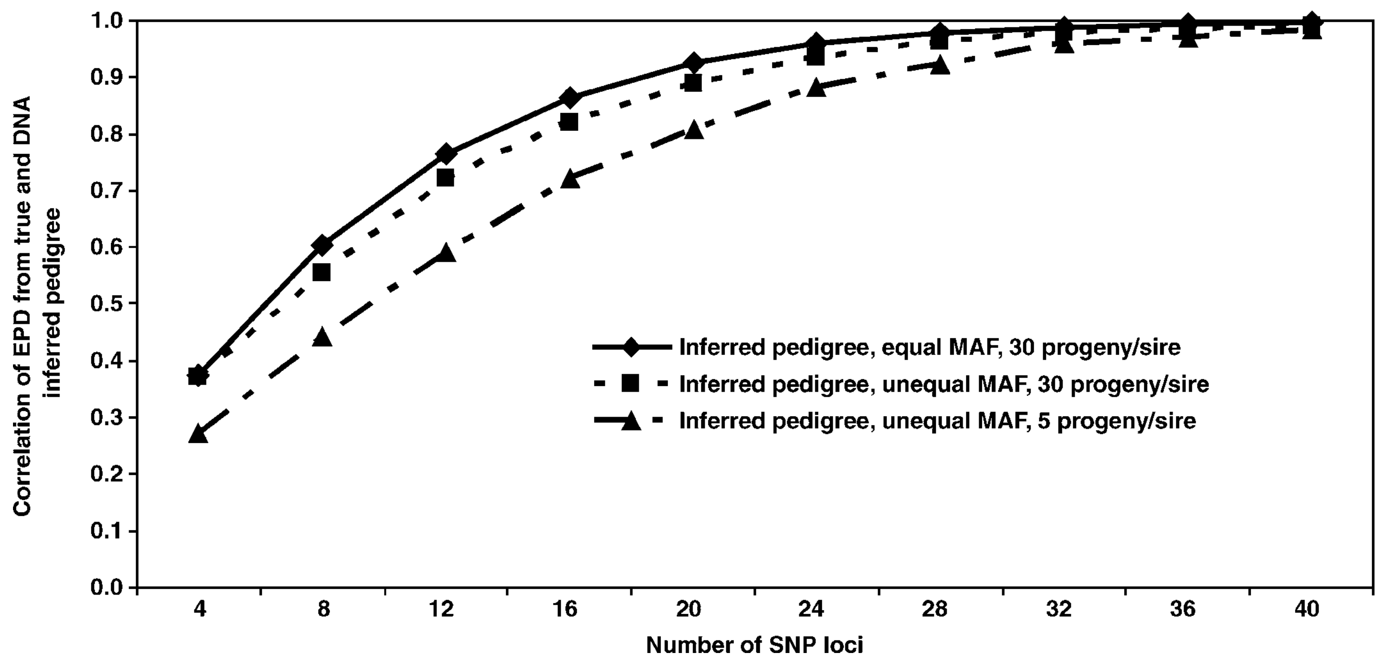

Figure 4. Correlation between EPD computed on sires with differing numbers of progeny based on the true and DNA-inferred pedigrees derived from SNP marker panels with 4 to 40 loci with equal or unequal minor allele frequencies (MAF).

data presented in this study suggest that SNP panels for some commercial applications may need to have a higher P_E than can be achieved with the 32 - 37 SNP loci panels that have been proposed for use in bovine parentage analysis (Heaton et al., 2002; Werner et al., 2004). Variable calf output and MAF, large sire cohorts, relatedness among sires, and missing data can all negatively impact the accuracy of on-farm EPD. In field situations where several of these variables occur concurrently, the use of marker panels with high P_E values will be required to obtain accurate EPD.

Single nucleotide polymorphism discovery is ongoing, and already bovine SNP panels for parentage determination that use over 100 loci are commercially available. The use of progeny testing to develop within-herd EPD for herd sires on economically-relevant traits has the potential to generate value by improving the response to selection for targeted traits. The return on investment that results from such progeny testing was found to be greatly influenced by the cost of parentage determination (Weaber, 2005). New SNP genotyping platforms continue to drive down the cost to generate SNP genotypes, and the future will undoubtedly see the introduction of inexpensive genotyping assays using high resolution SNP parentage panels. This will improve the accuracy of sire assignments and on-farm genetic evaluations, and may result in progeny testing becoming an economically viable option for commercial ranchers.

This case study illustrated some problems that may be encountered in paternity testing in large commercial herds. Field data are likely to include both missing sires and sires that did not produce any progeny. Low resolution marker panels and large cohorts of potential herd sires are particularly problematic and may result in sire-assignment errors and imprecise genetic evaluations. The frequency of sire misassignment can be minimized by using a high resolution panel or by simple management practices that include dividing large herds into smaller multiple-sire breeding groups with fewer sires while maintaining the same bull:female ratio, genotyping all potential bulls before breeding, sorting bulls into sire groups with divergent genotypes, keeping young bulls in separate breeding groups, and minimizing relatedness among bulls.

LITERATURE CITED

Anderson, E. C., and J. C. Garza. 2006. The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics 172:2567–2582.

Baumung, R., H. Simianer, and I. Hoffmann. 2004. Genetic diversity studies in farm animals—A survey. J. Anim. Breed. Genet. 121:361–373.

BIF. 2002. Guidelines for uniform beef improvement programs. 8th ed. Beef Improv. Fed., Athens, GA.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314–331.

DeNise, S. K. 1999. Using parentage analysis in commercial beef operations. Pages 183–190 in Proc. 31st Annu. Res. Symp. Beef Improv. Fed., Roanoke, VA.

Dodds, K. G., M. L. Tate, and J. A. Sise. 2005. Genetic evaluation using parentage information from genetic markers. J. Anim. Sci. 83:2271–2279.

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.

Holroyd, R. G., V. J. Doogan, J. De Faveri, G. Fordyce, M. R. McGowan, J. D. Bertram, D. M. Vankan, L. A. Fitzpatrick, G. A. Jayawardhana, and R. G. Miller. 2002. Bull selection and use in northern Australia. 4. Calf output and predictors of fertility of bulls in multiple-sire herds. Anim. Reprod. Sci. 71:67–79.

Jamieson, A., and S. C. Taylor. 1997. Comparisons of three probability formulae for parentage exclusion. Anim. Genet. 28:397–400.

Marshall, T. C., J. Slate, L. E. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7:639–655.

Pollak, E. J. 2005. Application and impact of new genetic technologies on beef cattle breeding: A 'real world' perspective. Aust. J. Exp. Ag. 45:739–748.

Sherman, G. B., S. D. Kachman, L. L. Hungerford, G. P. Rupp, C. P. Fox, M. D. Brown, B. M. Feuz, and T. R. Holm. 2004. Impact of candidate sire number and sire relatedness on DNA polymorphism-based measures of exclusion probability and probability of unambiguous parentage. Anim. Genet. 35:220–226.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34:275–305.

Weaber, R. L. 2005. A simulation study of replacement sire selection and genetic evaluation strategies for large commercial ranches. PhD Diss. Cornell Univ., Ithaca, NY.

Werner, F. A. O., G. Durstewitz, F. A. Habermann, G. Thaller, W. Kramer, S. Kollers, J. Buitkamp, M. Georges, G. Brem, J. Mosner, and R. Fries. 2004. Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. Anim. Genet. 35:44–49.