Evaluation of parentage testing in the Czech population of Holstein cattle

V. Řehout, E. Hradecká, J. Čítek

Agricultural Faculty, University of South Bohemia, České Budějovice, Czech Republic

ABSTRACT: A set of 233 Holstein calves, their 233 dams and 44 sires from 4 large-sized herds was genotyped for 10 microsatellites recommended by ISAG for paternity/parentage testing. Heterozygosity of microsatellites ranged from 0.607 (*SPS115*) to 0.835 (*TGLA227*), and PIC from 0.575 (*SPS115*) to 0.816 (*TGLA227*) confirming a high polymorphism of each analysed locus. Their combined exclusion probability reached 0.999, which made them suitable and sufficiently accurate for parentage testing. A conflict between putative parents and calf in at least 2 markers with combined exclusion probability > 0.9 was required to reject parentage. The pedigree was considered incorrect in 25 (10.73%) out of the evaluated progeny/parent trios, of which in 10 samples the genotype of both parents did not match their offspring, and in 2 samples the putative dam was in conflict with the calf genotype. This result shows that the interchange of calves on farms with large-sized herds plays the role as important as the errors in sire identification, or recording mistakes.

Keywords: dairy cattle; parentage testing; heterozygosity; exclusion probability

Correct pedigree information is the basic condition for a successful breeding program. Its importance even increased with the introduction of individual animal models for national genetic evaluation of cattle. Animal model evaluations are based on all known genetic relationships between the animals included in the calculation. The model assumes that all pedigrees and relationships are correctly recorded. This expectation is not always fulfilled as several studies showed. The proportion of misidentified progeny varies between 5 and 15% in Denmark (Christensen et al., 1982), 4 and 23% in Germany (Geldermann et al., 1986), 8 and 20% in Ireland (Beechinor and Kelly, 1987), 12% in The Netherlands (Bovenhuis and Van Arendonk, 1991), 2.9 and 5.2% in Israeli Holstein (Ron et al., 1996), 10% in UK dairy herds (Visscher et al., 2002) or 11.7% in Israeli Holstein again (Weller et al., 2004).

Pedigree errors are expected to bias the estimation of genetic parameters (Van Vleck, 1970), breeding values (Israel and Weller, 2000), and genetic progress (Geldermann et al., 1986). They may reflect in the structure of selection indexes because breeding values, their standard deviations, correlations and reliabilities of their estimates are together with economic weights, genetic standard deviations and genetic correlations the input data for the calculation of weighting coefficients (Přibyl et al., 2004). Geldermann et al. (1986) concluded that the loss in response should be similar to the proportion of progeny misidentified. Other studies found smaller losses. Van Vleck (1970) and Christensen et al. (1982) showed that the bias in heritability from half-sib groups was approximately $(1\text{-}p)^2$ if p was the error rate in the progeny of the bulls. Visscher et al. (2002) summarized a number of examples of reliabilities and losses in genetic gain as a function of the number of progeny n, h^2, and error rate p. For example, for an error rate of 10%, heritability of 25% and 50 progeny per sire, the reliability is reduced by 5%, and genetic progress is reduced by 3% (relative to 0% errors). Israel and Weller (2000) presented the same loss in selection response (3–4%) via the

Supported by Ministry of Education, Youth and Sports of the Czech Republic (Project No. MSM 6007665806).

stochastic simulation study of a large dairy cattle population with 10% incorrect paternity.

Christensen et al. (1982) summed up 7 reasons for errors in paternity recording: (1) mistakes by AI institutes in semen labelling; (2) AI technicians incorrectly identifying semen samples; (3) the insemination of cows already pregnant by a previous insemination; (4) errors when the bull's herdbook number or name is entered into the insemination record; (5) the use of natural-service bulls leading to pregnancies of previously inseminated cows which were assumed to be pregnant from the AI bull; (6) mistakes in sire identification when a cow enters the milking herd in schemes where pedigree information on milk recorded cow is obtained through the milk recording program; and (7) interchange of calves on the farm.

To these reasons Weller et al. (2004) added factor (8) genotyping errors, mutation, or presence of "null alleles" followed by rejection of the proper pedigree. Weller et al. (2004), or Petersen and Bendixen (2000) described "null alleles" as those that failed to amplify: then the individual appears to be homozygous in the marker, even though it is in fact a heterozygote. The progeny of a sire heterozygous for a null allele that receive the null allele will appear to be homozygous for their maternal allele. If the maternal allele is different from the sire's observed allele, then no common allele will be found in the sire and his progeny, and paternity will be erroneously rejected.

Visscher et al. (2002) concluded that some of the above-mentioned causes could be prevented by a good recording and verification system at AI. The same authors also pointed out that only factor (7) would give rise to the equal identification error of sires and dams. Hence, a study in which trios of progeny and both of their putative parents are sampled should pinpoint whether the errors are because of poor recording or whether large-sized herds are involved. Weller et al. (2004) analyzed the main factors such as recorded sire, birth year, geographical region, herd and inseminator, which are responsible for pedigree errors when only the genotype of sire and daughter was detected. They found out that only the effect of inseminator and recorded sire were significant in all tested models (linear and nonlinear).

Traditionally, pedigree verification in dairy cattle has been carried out using blood group and protein polymorphism, but currently DNA microsatellites are the international standard system of identity verification in livestock (Bredbacka and Koskinen, 1999). The advantage of microsatellite-based tests is that any sample of the animal containing DNA can be used, which enables to apply easier, non-invasive sampling techniques. For example, Visscher et al. (2004) compared the efficiency of DNA extraction from milk, nasal, saliva, semen and hair samples. Ron et al. (2003) developed and tested a method to sample cows using vaginal swabs. The accuracy of the test using microsatellites is also much higher, as the probability of detecting mistaken paternity or maternity is a function of the polymorphism of genotyped loci (Ron et al., 1996). The microsatellites are generally polyallelic, and their number is almost unlimited. The informative value of each microsatellite is expressed by its polymorphic information content PIC (Botstein et al., 1980) and heterozygosity (Nei, 1978). The efficiency of any co-dominant allele in parentage testing is described by its exclusion probability (Jamieson and Taylor, 1997).

Weller et al. (2004) determined three categories with respect to the exclusion probability and paternity confirmation: confirmed, rejected or undetermined. If no discrepancies were found in at least 20 markers between the putative daughter and sire genotypes, and exclusion probability was > 0.9, then the paternity record was declared confirmed. If no discrepancies were found, but the exclusion probability was < 0.9, then paternity confirmation was declared undetermined. If 2 or more discrepancies were found among the first 20 valid genotypes, then paternity was rejected.

The objective of this study was to quantify the level of errors in pedigree identification, and to evaluate factors affecting the results of parentage testing in Czech Holstein cattle.

MATERIAL AND METHOD

DNA was extracted from 466 blood and 44 semen samples. A total number of 510 genotyped animals included 233 calves, their 233 dams and 44 sires. The calves were born in 4 different herds that were housed in high-capacity stables. All 10 microsatellites used in our study are recommended for bovine paternity tests by the International Society for Animal Genetics (ISAG). A list of analyzed microsatellites and their primer sequences are shown in Table 1. The genotyping was conducted by Laboratory of Applied Molecular Genetics in Brno.

Table 1. International panel of microsatellites recommended for paternity testing (ISAG)

Locus	Chromosome		Primer Sequence (5'–3')
BM1824	1	forward	GAG CAA GGT GTT TTT CCA ATC
		reverse	CAT TCT CCA ACT GCT TCC TTG
BM2113	2	forward	GCT GCC TTC TAC CAA ATA CCC
		reverse	CTT CCT GAG AGA AGC AAC ACC
INRA023	3	forward	GAG TAG AGC TAC AAG ATA AAC TTC
		reverse	TAA CTA CAG GGT GTT AGA TGA ACT C
SPS115	15	forward	AAA GTG ACA CAA CAG CTT CTC CAG
		reverse	AAC GAG TGT CCT AGT TTG GCT GTG
TGLA122	21	forward	CCC TCC TCC AGG TAA ATC AGC
		reverse (1)	AAT CAC ATG GCA AAT AAG TAC ATA C
		reverse (2)*	AAT CAC ATG GCA AAT AAG TAC ATA
TGLA126	20	forward	CTA ATT TAG AAT GAG AGA GGC TTC T
		reverse	TTG GTC TCT ATT CTC TGA ATA TTC C
TGLA227	18	forward	CGA ATT CCA AAT CTG TTA ATT TGC T
		reverse	ACA GAC AGA AAC TCA ATG AAA GCA
ETH3	19	forward	GAACCTGCCTCTCCTGCATTGG
		reverse	ACTCTGCCTGTGGCCAAGTAGG
ETH10	5	forward	GTT CAG GAC TGG CCC TGC TAA CA
		reverse	CCT CCA GCC CAC TTT CTC TTC TC
ETH225	9	forward	GAT CAC CTT GCC ACT ATT TCC T
		reverse	ACA TGA CAG CCA GCT GCT ACT

*corrects the null allele problem

Electrophoresis was performed in an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). The fluorescent labelled products were analyzed with GeneScan v. 3.7. Software.

Expected heterozygosity (H_{exp}), polymorphism information content (PIC) and exclusion probability (EP) were calculated for each microsatellite on the basis of parents' allele frequencies. Expected heterozygosity was calculated as $H_{exp} = 1 - \Sigma p_{ij}^2$, where p_{ij} is the frequency of allele i on locus j, and polymorphism information content was

$$PIC = H_{\exp} - 2\sum_{i=1}^{n-1} p_i^2 \times \sum_{j=i+1}^{n} p_j^2$$

with the allelic frequencies p_i, p_j, and number of alleles n.

Exclusion probability, which expresses the probability that two random (unrelated) individuals do not share any allele, was defined for two cases:

(a) given two parents and one offspring; exclude a parent (Jamieson, 1994)

$$EP_A = 1 - 2\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} p_i^3 + 2\sum_{i=1}^{n} p_i^4 - 3\sum_{i=1}^{n} p_i^5 - 2(\sum_{i=1}^{n} p_i^2)^2 + 3\sum_{i=1}^{n} p_i^2 \times \sum_{i=1}^{n} p_i^3$$

(b) given two parents and one offspring; exclude both parents (Jamieson and Taylor, 1997)

$$EP_B = 1 + 4\sum_{i=1}^{n} p_i^4 - 4\sum_{i=1}^{n} p_i^5 - 3\sum_{i=1}^{n} p_i^6 - 8(\sum_{i=1}^{n} p_i^2) + 8(\sum_{i=1}^{n} p_i^2) \times (\sum_{i=1}^{n} p_i^3) + 2(\sum_{i=1}^{n} p_i^3)^2$$

Combining exclusion probability EP_c over n unlinked markers in both above formulae gives:

$$EP_c = 1 - (1 - EP_1)(1 - EP_2)...(1 - EP_n)$$

The shared alleles were checked for each parents-progeny trio. If no discrepancies were found between the putative parents and progeny genotypes, then parentage was declared confirmed. If single discrepancy was detected, it was assumed to be a

Table 2. Informativeness of analyzed markers

Marker	Number of alleles	Number of genotypes	FNA*	H_{ex}	H_{obs}	H_{obs-ex}	PIC
BM1824	6	17	0.287	0.764	0.788	0.024	0.723
BM2113	8	29	0.293	0.807	0.810	0.003	0.781
ETH3	9	24	0.475	0.705	0.744	0.069	0.670
ETH10	8	25	0.482	0.701	0.711	0.010	0.667
ETH225	9	22	0.355	0.752	0.762	0.010	0.714
INRA023	11	30	0.263	0.809	0.857	0.048	0.782
SPS115	8	20	0.591	0.607	0.625	0.018	0.575
TGLA122	14	50	0.324	0.816	0.815	–0.001	0.795
TGLA126	6	15	0.429	0.665	0.695	0.030	0.605
TGLA227	12	38	0.287	0.835	0.853	0.018	0.816
Mean	9.100	27.000	0.379	0.746	0.769	0.029	0.713

*frequency of the most frequent allele

mistake or possible mutation, and the parentage was not rejected, as the exclusion probability for such cases is not sufficient. If there were two or more discrepancies, the parentage was rejected. For each case of rejected parentage the combined exclusion probability was calculated by adding the probabilities for all genotype combinations, which results in exclusion.

For each sample the rejected (0) or confirmed (1) pedigree was modelled as a function of the putative sire (44 cases) or putative sire line (4 cases; NX, NBY, NGA, NEB). The pedigree confirmation score was analyzed using simple linear and logistic models with Statistica v. 7.

RESULTS AND DISCUSSION

Paternity or parentage testing effectiveness closely depends on the level of informativeness provided by the markers. The results of the microsatellite marker potential, expressed by expected heterozygosity (H_{exp}) and polymorphic information content (PIC), are shown in Table 2. Expected heterozygosity ranged from 0.607 (*SPS115*, 8 alleles) to 0.835 (*TGLA227*, 12 alleles), with average value 0.746. Polymorphism information content ranged from 0.575 (*SPS115*, 8 alleles) to 0.816 (*TGLA227*, 12 alleles) with average value 0.713, which confirms a high polymorphism of each analyzed microsatellite. Our results are comparable to those of Holstein cattle found by Heyen et al. (1997) in the USA, Visscher et al. (2002) in the UK, or Czerneková et al. (2006) in the Czech Republic, and higher than those published by Radko et al. (2005) in Poland. Generally, the higher the heterozygosity, the higher the genetic variation of the population and its genetic polymorphism, and the more suitable is the marker for individual identification. In all but two (*SPS115*, *TGLA126*) of the analyzed microsatellites heterozygosity was higher than 0.7. In *SPS115* the expected heterozygosity reached 0.607, but the frequency of its most frequent allele overcame 0.5, which makes the usefulness of this microsatellite somewhat questionable.

A difference between the observed and expected heterozygosity, based on Hardy-Weinberg equilibrium, was computed for each marker, and is also listed in Table 2. The average frequency of observed heterozygotes over all markers was 0.769 and the difference between the observed and expected heterozygosity ranged from –0.001 (*TGLA122*) to 0.069 (*ETH3*) with the mean 0.029.

Exclusion probabilities for all loci corresponded with their heterozygosity with the highest value in *TGLA227* (EP_A = 0.677, EP_B = 0.853), *TGLA122* (EP_A = 0.648, EP_B = 0.832) and *INRA023* (EP_A = 0.623, EP_B = 0.803). The genotype parent-offspring discrepancies were also most frequent in loci *TGLA122* (22 cases), *ETH225* (20 cases), *TGLA227* (19 cases) and *INRA023* (19 cases), which corresponded with the highest values of their exclusion probabilities. The lowest probabilities were calculated for loci *SPS115*, *TGLA126*, *ETH3* and *ETH10*, when their values did not exceed 50% for one parent exclusion (see Table 3). Curi and Lopes

Table 3. Probabilities of one (EP_A) or both (EP_B) parents exclusion

Marker	EP_A	EP_B
BM1824	0.535	0.713
BM2113	0.623	0.804
ETH3	0.488	0.685
ETH10	0.488	0.689
ETH225	0.534	0.721
INRA023	0.623	0.803
SPS115	0.397	0.596
TGLA122	0.648	0.832
TGLA126	0.407	0.582
TGLA227	0.677	0.853
Combined EP_C	0.999	0.999

(2002) found out very low informativeness and exclusion probabilities for those loci in Gyr breed, though they are recommended for parentage testing. The same authors deduced that this fact accentuated the need of characterization for different populations or lineages within a breed in which one wants to perform a paternity testing, since the number of allelic frequencies can be different in different populations of the same breed.

The genotype discrepancies were detected between the sire and progeny in 2 samples and between the dam and progeny in 3 samples. The genotype of both parents did not match their offspring in 10 samples. Twelve calves shared one allele with both parents, but it was not possible to determine which one (or both of them) was incorrectly recorded. This result indicates that the interchange of calves on farms (factor (7) according to Christensen et al., 1982) with large-sized herds may play the role as important as the errors in sire identification, or poor recording, which was reminded by Visscher et al. (2002). Also Ron et al. (1996) stated that the switching of two calves born on the same day was a possible cause of misidentification in large-sized herds.

There were only single discrepancies detected in samples 19 and 112. In sample 19 the putative sire, dam and calf shared allele 210 in *INRA023*, but the other calf's allele was in conflict with both parents. In sample 112 the genotype of calf did not match the genotype of dam in *TGLA122*. As stated by Weller et al. (2004), if the cow and putative sire are genotyped for many markers, and paternity recording is incorrect, then it is very unlikely that only a single discrepancy should be observed. According to Visscher et al. (2002) the probability that only one locus of ten shows parent exclusion is lower than 10%. These single discrepancies may be due to either mutations or genotyping mistakes. Ron et al. (1996) reported that the observed frequency of mutation of short tandem repeats was 0.01 per locus per gamete per generation, so to reject the possibility of mutation, the exclusion should be confirmed by two independent loci. As for genotyping mistakes, they are on the order of 1%. As reminded by Weller et al. (2004), when the results of 33 laboratories on 40 DNA samples were compared, identical results were not obtained for any of the 9 basic ISAG markers, and 2 laboratories differed in more than one allele. Thus, with the exclusion probabilities lower than 0.9 (0.803 in sample 19 and 0.832 in sample 112), taking the probability of genotyping mistake or mutation into account, the parentage in these two cases (sample 19 and 112) was not rejected.

With respect to combined exclusion probabilities and number of loci in the mismatch we can reject 25 of 233 evaluated pedigrees, which represents 10.73% of incorrect pedigrees. For example Visscher et al. (2002) and Weller et al. (2004) found a similar rate of pedigree errors (10–11%) in their studies in Holstein cattle.

Most of the studies (Heyen et al., 1997; Curi and Lopes, 2002; Visscher et al., 2002; Weller et al., 2004) were focused on paternity testing, as the genotyping of both parents and progeny would be too expensive for routine application. Weller et al. (2004) tested the effects of inseminator, region, herd, sire and birth year on incorrect paternity by logistic and linear analysis. The effects of sire and inseminator were found out to be highly significant, while the region and birth year were marginal or non-significant.

Pedigree errors were found in the putative progeny of 16 sires out of 44 evaluated. Each sire has on

Table 4. Genotypisation results according to sire lines

Line	Progeny (*n*)	Sires (*n*)	Genotype errors	Relative frequency
NX	77	13	11	0.14
NXA	3	1	0	0.00
NBY	46	8	4	0.09
NGA	23	4	0	0.00
NEB	84	18	10	0.12
Total	233	44	–	–

Table 5. Characteristics of linear and logistic model evaluation

Model type	Factor	Levels	*F* (χ2)-values*	Significance
Linear	sire	44	0.841	0.763
	sire line	5	1.113	0.351
Logistic	sire	44	0.029	0.865
	sire line	5	0.288	0.591

*for the logistic model χ2 values for the maximum likelihood estimates are given instead of *F*-values

average 5 genotyped calves, with the highest representation in NX 961. This sire had 15 calves assigned, but 4 of them showed genotype discrepancies. The relatively highest occurrence of misidentified progeny was found in sires from frequent lines, such as NX – 14% or NEB – 12% (Table 4), as with a higher number of cases grows the probability of incorrectly identified semen sample or mistakes in entering the name or number of bull in the insemination record, which are considered to be the main factors of paternity mistakes (Weller et al., 2004). Evaluating both effects by a linear and logistic model, we did not find any significant differences in the presence of pedigree errors among sires and lines (Table 5).

CONCLUSION

DNA microsatellites can be efficiently used to determine incorrect parentage attribution. All analyzed loci showed high polymorphism and sufficient informativeness, though in *SPS115* the frequency of one of its alleles exceeded 0.50. The overall frequency of rejected parentage in the sample of Czech Holstein population was 10.73%. The comparison of calf genotype with both putative parents showed that interchanges of calves on farms with large-sized herds were among the major factors of incorrect pedigree assignment, when in a half of the samples with detected discrepancies the genotype of calf did not match to the genotype of either dam or both parents. The effect of sire or sire line on the frequency of pedigree errors was not significant.

REFERENCES

Beechinor J.G., Kelly E.P. (1987): Errors of identification amongst cattle presented as progeny of some bulls used in the artificial insemination service in Ireland. Ir. Vet. J., *41*, 348–352.

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

Bovenhuis H., Van Arendonk A.M. (1991): Estimation of milk protein gene frequencies in crossbred cattle with maximum likelihood. J. Dairy Sci., *74*, 2728–2736.

Bredbacka P., Koskinen M.T. (1999): Microsatellite panels suggested for parentage testing in cattle: Informativeness revealed in Finnish Ayrshire and Holstein Friesian populations. Agric. Food Sci. Finl., *8*, 233–237.

Christensen L.G., Madsen P., Petersen J. (1982): The influence of incorrect sire identification on the estimates of genetic parameters and breeding values. In: Proc. 2nd World Congr. Genet. Appl. Livest. Prod., Madrid, Spain, *7*, 200–208.

Curi R.A., Lopes C.R. (2002): Evaluation of nine microsatellite loci and misidentification paternity frequency in a population of Gyr breed bovines. Brazil. J. Vet. Res. Anim. Sci., *39*, 129–135.

Czerneková V., Kott T., Dudková G., Sztankóová Z., Soldát J. (2006): Genetic diversity between seven Central European cattle breeds as revealed by microsatellite analysis. Czech J. Anim. Sci., *51*, 1–7.

Geldermann H., Piper U., Weber W.E. (1986): Effect of misidentification on the estimation of breeding value and heritability in cattle. J. Anim. Sci., *63*, 1759–1768.

Heyen D.W., Beever J.E., Da Y., Evert R.E., Green C., Bates S.R.E., Ziegle J.S., Lewin H.A. (1997): Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplex for semi-automated parentage testing. Anim. Genet., *28*, 21–27.

Israel C., Weller J.I. (2000): Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. J. Dairy Sci., *83*, 181–187.

Jamieson A. (1994): The effectiveness of using co-dominant polymorphic allelic series for (1) checking pedigrees and (2) distinguishing full-sib pair members. Anim. Genet., *25*, 37–44.

Jamieson A., Taylor StC.S. (1997): Comparison of three probability formulae for parentage exclusion. Anim. Genet., *28*, 397–400.

Nei M. (1978): Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, *89*, 583–590.

Petersen A.H., Bendixen C. (2000): Null-alleles in the standard set of loci for cattle parentage control. In: Proc. Int. Conf. Anim. Genet., St. Paul, MN., 89.

Přibyl J., Šafus P., Štípková M., Stádník L., Čermák V. (2004): Selection index for bulls of Holstein cattle in the Czech Republic. Czech J. Anim. Sci., *49*, 244–256.

Radko A., Żyga A., Ząbek T., Słota E. (2005): Genetic variability among Polish Red, Hereford and Holstein-Friesian cattle raised in Poland based on analysis of microsatellite DNA sequences. J. Appl. Genet., *46*, 89–91.

Ron M., Blanc Y., Band M., Ezra E., Weller J.I. (1996): Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. J. Dairy Sci., *79*, 676–681.

Ron M., Domochovsky R., Golik M., Seroussi E., Ezra E., Shturman C., Weller J.I. (2003): Analysis of vaginal swabs for paternity testing and marker-assisted selection in cattle. J. Dairy Sci., *86*, 1789–1796.

Van Vleck L.D. (1970): Misidentification in estimating the paternal sib correlation. J. Dairy Sci., *53*, 1469–1474.

Visscher P.M., Woolliams J.A., Smith D., Williams J.L. (2002): Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. J. Dairy Sci., *85*, 2368–2375.

Weller J.I., Feldmesser E., Golik M., Tager-Cohen I., Domochovsky R., Alus O., Ezra E., Ron M. (2004): Factors affecting incorrect paternity assignment in the Israeli Holstein population. J. Dairy Sci., *87*, 2627–2640.

Received: 2006–04–20
Accepted after corrections: 2006–10–15

Corresponding Author

Prof. Ing. Václav Řehout, CSc., Agricultural Faculty, University of South Bohemia in České Budějovice, Studentská 13, 370 05 České Budějovice, Czech Republic
Tel. +420 387 772 590, fax +420 387 772 593, e-mail: rehout@zf.jcu.cz