Indian Journal of Animal Sciences **88** (11): 1275–1280, November 2018/Article

DNA profiling and paternity verification in Murrah buffaloes

RAMESH KUMAR VIJH[1], UPASNA SHARMA[2], UMA GAUR[3] and SURESH B GOKHALE[4]

ICAR-National Bureau of Animal Genetic Resources, Karnal, Haryana 132 001 India

Received: 6 October 2017; Accepted: 25 June 2018

ABSTRACT

The accuracy of selection for sire's merit depends upon the number of daughters born to the sire and also the correctness of performance records. The accuracy of records depends on confirmed paternity of the daughters. In this paper, we report paternity confirmation of a large progeny testing program in buffaloes. The work was undertaken in collaboration with BAIF, Pune under National Agricultural Innovation Project. Murrah bulls (12) were selected and around 120,000 buffaloes were inseminated at farmers' doorsteps. The area of operation was 52 districts of Uttar Pradesh**.** We targeted 1,000 female calves to be born to each of the 12 sires. The experiment was an effort to develop a reference population for identification of quantitative trait loci for various economic traits in buffaloes using halfsib design. Owing to financial constraints, we targeted 10,000 daughters born to these 12 sires. The DNA was isolated from these daughters. The DNA of sires was also isolated. Since there was no parentage verification kit commercially available for buffaloes, we developed a multiplex PCR reaction consisting of 11 microsatellite markers to confirm the paternity. The correctness of the paternity was established using microsatellite DNA markers and comparison of daughters' genotype with genotype of sires. The paper reports paternity testing of large reference family. We also developed a software 'Confirm Paternity' for comparing the genotypes of daughters with the respective sire. We report paternity correctness of 90.48% in buffaloes under field conditions.

Key words: Confirm paternity software, Multiplex PCR, Murrah, Parentage verification kit, Paternity

Progeny testing is a well established method of genetic improvement of trait for which the genes are present in both sexes but the expression is limited to one sex (sex-limited traits). The progeny testing program has not made much success in institutional/organized herds as large numbers of daughters per bull are required to be recorded for the traits of interest. Few attempts have been made in this direction in the form of Network Project on Buffaloes using Associated Herd Progeny Testing Program (Annual Reports Network Project on buffaloes, 2013–14, 2014–15 and 2015–16). Although a few institutions participated in this program, the number of daughters per bull that actually completed first lactation are abysmally low (5–10 daughters/sire) for the results obtained till date (Annual Report Network Project on Buffaloes, 2016–17). The only way to circumvent the situation is to have a large number of daughters per bull reared at farmers' door step and recorded for traits of interest. This is, however, both highly rewarding and challenging preposition. The most important aspect is correctness of paternity.

Present address: [1]Principal Scientist (rameshkvijh @gmail.com), [2]Research Associate (upasna30@gmail.com), Animal Genetics Division. [3]Postdoctoral Researcher (gaur.uma2906@gmail.com), University of Macau, Guangdong Sheng, China. [4]Director Research (suresh.gokhale45 @gmail.com), BAIF, Uruli Kanchan, Pune, Maharashtra.

The purpose of this study was to verify the paternity of large number of daughters born to 12 sires at the farmers' door step. The paternity check was a necessity since the records of the daughters of these sires were to be used for the identification of quantitative trait loci markers for milk yield, fat and protein percentage in buffaloes using a halfsib design. This is the first attempt for paternity testing at such a large scale at field level and perhaps the paternity has also not been tested for the organized herds assuming the manual records to be correct. In this paper, we report development of a single multiplex PCR reaction consisting of 11 microsatellite markers. We also report development of a software 'Confirm Paternity' to check the correctness of the paternity for large number of daughter-sire genotype combinations.

MATERIALS AND METHODS

Sires (12) were selected on the basis of their mothers' milk yield. These bulls were reared at Uruli Kanchan, BAIF, Pune. The semen was collected as per standard protocols and tested for various diseases and frozen in liquid nitrogen containers. The semen of these bulls were transported to the operational area of Uttar Pradesh. Inseminations (120,000) were carried out by BAIF field workers (inseminators) and a total of 58,000 progenies were obtained in the operational area. The experiment centered only on daughters of these sires. Thus only the female calves were

ear tagged as per the insemination records available in the field. More than 12,000 daughters were ear tagged which presumably were female calves of 12 experimental sires. The blood samples were collected from 10,075 of these identified daughters. The samples were collected in sterile vacutainer tubes which were EDTA coated. The area of blood sample collection was very large, hence we constituted 5 teams for collection of samples. The team invariably included 1 veterinary surgeon and 2 para-veterinarians and 2 helpers to handle the buffalo calves during blood collection. The samples were collected from operational area and brought to district headquarters. They were collected at regional centers, and transported to Delhi and then to Hyderabad for DNA isolation. The DNA isolation was carried out using the standard protocol of Proteinase K digestion and phenol-chloroform-isoamyl alcohol extraction. The quality and quantity of the DNA isolated was checked. Quality check was also carried out of 10% of the DNA samples at NBAGR, Karnal.

A multiplex PCR reaction was standardized by taking a set of 11 microsatellite markers. To begin with, 120 microsatellite markers were tested for developing a single PCR multiplex reaction. The considerations for selection for these 11 markers were the simultaneous amplifiability, size of PCR products, the fluorescent dye label, the annealing temperature and interaction among the primer pairs. Ultimately, one single multiplex PCR reaction was standardized for carrying out the paternity testing of such a large reference population.

The 12 sires were genotyped repeated so as to avoid any ambiguity about the genotypes. The PCR reaction mix for multiplex was: 50 ng template DNA, 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 1% Triton®-X, 1.5 mM $MgCl_2$, 200 µM dNTPs, 5 picomole of each primer pair of the 11 microsatellites, 1.0 U Taq DNA polymerase in a 5 µl reaction. Thermal cycle parameters were: Initial denaturation for 5 min at 95°C followed by 35 cycles of (45 sec for denaturation at 94°C, 45 sec for annealing at 55°C followed by extension for 1 min at 72°C) with a final extension step of 10 min at 72° C using a Eppendorf Pro-S thermocycler. The multiplex PCR reaction was diluted and the genotypes were called using ABI automated DNA sequencer 3130XL having 16 capillaries. The alleles were binned prior to comparison of daughters' genotype with sires' genotypes.

A very large number of daughters were to be assigned to daughters based on genotypes of the 11 microsatellite loci. The huge work cannot be done manually and thus necessitated the development of a software which we called as Confirm Paternity. The software compared the genotype of the daughters with the sires' genotypes. We kept a threshold of at least 10 marker genotypes (out of 11 markers) to match for the correct paternity. In case the daughters' genotype does not match for at least 10 loci, paternity was considered wrong. To take care of wrong recording of the sire in the field, we put in additional check in the software. In case the genotype did not match the recorded sire, we compared the genotype of daughter with other 11 sires before the daughter was removed from future recording of data. If it matched to any one of the other sires, we genotyped the sire and daughter with additional set of microsatellite markers to confirm the paternity.

RESULTS AND DISCUSSION

The number of markers required to check the paternity may vary and shall depend upon polymorphism level of microsatellite markers. In addition it shall also depend upon the relationship among the sires. Usha *et al.* (1995) and Heyen *et al.* (1997) suggested that at least 5 markers shall be sufficient to achieve the correct paternity with a reasonably well probability of exclusion of an incorrect sire if the sires are not related to one another. Paternity identification involves examining each calf's genotype at multiple different gene loci. The potential sires that do not share common alleles with the calf are excluded as the correct parent. Since paternity identification is a process during which we exclude the potential sires on the basis of their genotype, it is thus imperative that we should have the DNA from all the expected sires to be included in the paternity tests.

The amplification product of the multiplex PCR reaction was loaded on the ABI prism 3130XL automated DNA sequencer. The multiplex gave non-ambiguous peaks for parentage assignment with great confidence. A representative picture of the multiplex microsatellite run on automated DNA sequencer is represented in Fig. 1.

The heterozygosity and exclusion probability values for the different microsatellite markers are given in Table 1. The number of alleles varied from 5 (BMS2722) to 24 (CSSM47). The mean number of alleles were 12.36. The observed heterozygosity varied from 0.302 (BMS2063) to 0.633 (CSSM19). The calculated PIC values ranged from 0.49 to 0.889 (Table 1), with a mean of 0.6419. The exclusion probability was maximum for CSSM47 (0.80) while it was least for BMS2063 (0.19). The combined

Table 1. Number of alleles, observed heterozygosity and exclusion probability for 11 microsatellite loci used for paternity testing

Locus	Number of alleles observed	Observed heterozygosity	Polymorphic content	Probability of exclusion information for the marker
BMS2722	5	0.44	0.585	0.37
BMS2785	10	0.403	0.686	0.51
RM232	8	0.367	0.49	0.30
BMS2063	7	0.302	0.335	0.19
CSSM08	13	0.606	0.725	0.55
CSSM19	14	0.633	0.682	0.52
CSSM43	17	0.55	0.848	0.72
CSSM47	24	0.86	0.889	0.80
ILSTS05	12	0.551	0.622	0.43
ILSTS11	12	0.581	0.76	0.59
ILSTS49	14	0.34	0.437	0.25

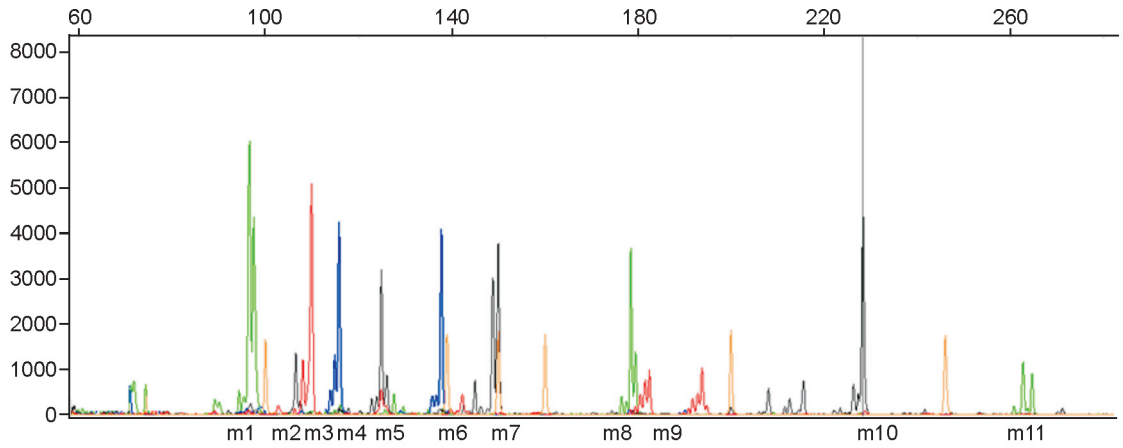

Fig. 1. Representative picture of the multiplex (11 microsatellite markers) run on automated DNA sequencer.

Table 2. Name of sires used, number of daughters of each sire identified in the field, number of samples collected and number of daughters with wrong paternity

Sire Code	Name of the sire	Total number of samples received from field	No of daughters with correct paternity	No. of daughters with incorrect paternity	Number of samples lost in transportation
S1	Sada	1,437	1,328	78	31
S2	Sahib	743	641	75	27
S3	Sahil	222	204	18	0
S4	Salman	850	749	76	25
S5	Sakha	1,269	1,067	173	29
S6	Sanam	286	254	20	12
S7	Satpal	1,229	1,109	86	34
S8	Srinath	1,047	954	71	22
S9	Sushil	1,510	1,270	181	59
S10	Shrirang	617	558	38	21
S11	Shagun	668	558	92	18
S12	Saiman	197	151	22	24
Total	12	10,075	8,845	930	300

exclusion probability for the set of 11 microsatellite markers was 99.97%. Thus these sets of markers were sufficient to assign paternity with almost 100% of exclusionary power.

Out of 10,075 daughters belonging to 12 sires, only 9,775 daughters could be tested for the paternity. A total of 300 samples were lost owing to clotting of blood samples, loss of identity of the sample and due to other technical reasons. Out of 9,775 daughters tested, 8,845 (90.48%) could be assigned to the recorded sires (Vijh 2014). The most obvious reason for daughters classified to have wrong paternity was attributable to mutation in the primer binding sites which lead to non amplification of one or more microsatellite loci during the PCR reaction. A daughter was considered as with correct paternity if at least 10 out of 11 microsatellite genotypes matched sires' genotype. Thus even if 2 markers did not amplify in any daughter it was classified as a case of wrong paternity.

In case the genotype did not match to the recorded sire, we matched the genotypes of daughters to the genotypes of other sires assuming there may be wrong recording of the semen dose or wrong insemination records. Such provision has also been made in the software developed termed as Confirm Paternity. In few of the instances, the daughters were assigned to more than one sires, in such cases more number of markers were utilised to test the paternity with high level of accuracy (in some cases even 25 markers) for correct assignment to the sires as the genotype of the sires were quite similar. To overcome this, animals were further genotyped with additional number of microsatellite markers and comparisons made at additional loci.

Software for comparison of genotypes

With large number of sire-daughter combinations and 11 loci, comparison of genotypes was not feasible using manual procedure. To carry out the work, a software called 'confirm paternity' was developed. The data utilized by the software is the genotypes (2 values at each loci) for sires and daughters. The software has been developed in MS Excel using its DBMS and macro features. The data are stored in MS Excel in various sheets. The microsatellite genotypic data for sires and their probable daughters are maintained in separate sheets. Commands of Visual Basic for Applications from Microsoft Office, have been used for resolving the complexities of the system. Macros were written to filter the data for the correct daughters and the result is produced on the separate result sheet which can be

printed or saved as desired. There is no limitation on the number of sires, their daughters and marker loci as far as MS Excel supports it. The software is available for download from the website (14.139.252.118:8084/BuffaloQTL/index.html) and from a direct link: http://www.mediafire.com/file/919u3qhnqwmaa26/Confirm_Paternity_1.0.rar).

The basic purpose of this study was to check the paternity of the 10,075 daughters born to 12 sires. The creation of 12 halfsib families was done at field level. There were greater chances of wrong parentage owing to several factors like wrong recording by the inseminator, pregnancy from earlier natural service or insemination, natural service after the insemination as the farmers tend to be more interested in getting their buffaloes pregnant so that they earn money rather than caring for long term economic gains through genetic improvement. Similar reasons have also been recorded under field conditions in Mehsana buffalo by Jakhsera *et al*. (2012) and for farm conditions by Christensen *et al.* (1982).

These 12 halfsib families were created for the identification of the QTLs for various traits of economic importance in buffaloes. In the absence of any commercial parentage verification kit for buffaloes, it was important to create a multiplex PCR reaction and also one software for comparison of the genotypes. We initiated the development of a multiplex PCR reaction from a set of approximately 120 microsatellite markers and ultimately developed a multiplex PCR reaction of 11 microsatellite markers which are amplifiable in a single PCR. The criterion for selection of microsatellite loci were—similar annealing temperature, different sizes of alleles, large number of alleles for the microsatellite loci and thereby a high PIC value, and heterozygosity. Further the microsatellite loci were selected from the cattle database; the microsatellites were tested for their unambiguous peaks and amplifiability of loci in buffaloes. The 11 microsatellite were differently labeled with four dyes, viz. Fam, Ned, Pet and Vic. The overlapping of the PCR product sizes were resolved using different dye labels while Liz was used as a size standard.

The accurate estimation of the genetic parameters is of utmost importance for finding the genetic relationship among the individuals in the halfsib families that were created. Misidentification or in this present case wrong paternity can seriously impair the finding of the identification of QTLs for economic traits in buffaloes. For progeny testing, wrong paternity can seriously hamper the estimation of breeding value and thereby reduce the genetic gain with sire models (Geldermann *et al.* 1986) and may have even greater effect with the models that account for all assumed genetic relationships (Wiggans *et al.* 1989). Thus the correct paternity is central to all animal breeding experiments. There are not many reports of paternity testing at organised farms or at farmers' door step. Christensen *et al.* (1982) reported the misidentification in the range of 5 to 15% in Danish cattle. The value was reported to be 13% in German dairy cattle (Geldermann *et al.* 1986) while the values were 5% for Israeli Holstein cows based on 173 cows and their 4 putative sires (Ron *et al.* 1996). There has been only one report in Mehsana buffaloes of Gujarat where the parentage was tested under field conditions based on 100 dam-daughter pairs and 12 sires (using semen samples of 5 sires for extraction of DNA). A total of 19 daughters out of 100 tested did not match with the recorded sire (Jakhesara *et al.* 2012). In this present study, the recording was at farmers' household at the field level in 52 districts of Uttar Pradesh. The mother of the daughters was never in question and there was no ambiguity. The correct paternity was very important but could be doubted owing to several reasons. The error could have crept in data recording, errors in identification, insemination and recording of semen straws. The wrong paternity could also be attributed to fertilization during previous inseminations or natural service. The farmers could simultaneously subject the animals to natural service after artificial insemination as conception of the buffalo is considered to be more important compared to genetic improvement because of immediate economic concerns. Taking into consideration the technical reasons of amplification of microsatellite loci, we matched a minimum of 10 markers out of 11 markers to record it as correct paternity. The matching of the genotypes of daughters and sires did not pose much problem in the experiment due to different genetic background of sire and dams that resulted into the halfsib families. The microsatellite alleles in local buffaloes were different from the alleles of the experimental sires. In case the alleles were common at some of the microsatellite loci, the allele frequencies were starkly different. The alleles and the frequencies of the 12 sires were however very similar to one another as all of them belonged to the Murrah breeding tract in Haryana. The marker name, allele in the twelve sires, their allelic frequencies, frequency of sire alleles in the reference population and the number of alleles observed in the reference population are given in Table 3. In case of wrong paternity, the genotype of the recorded sire did not match the daughter's genotype. Assuming erroneous recording the genotype of daughters was matched to the genotype of other sires. There were few instances in which the genotype matched to another sire from the set but in few instances the genotypes matched more than one sires, such daughters were further genotyped at 18–20 microsatellite to resolve the ambiguity in their paternity.

Declaring the wrong paternity when more than one marker does not match is likely to increase the pedigree error detection rate but would also increase the false positive rate (Jakhesara *et al.* 2012). In the present experiment, most of the wrong paternity was attributable to non amplification of 2 or more markers due to mutation in primer binding sites. Such problems were easily resolved by using another multiplex PCR reaction. The other set was used on all the daughters where the genotype of the daughters did not fulfill the correct paternity criterion of matching 10 microsatellite genotypes.

Thus we could check the paternity of large number of

Table 3. Markers used in parentage verification.

Marker name	Allele and their allele sizes observed in 12 sires	Allele frequency in sires used	Frequency of sire alleles in reference population created	No. of alleles observed in reference population created
BMS2684	98	0.0417	0.256	8
	100	0.625	0.4952	
	102	0.3333	0.248	
BMS2722	109	0.0833	0.205	4
	111	0.5000	0.4327	
	113	0.4167	0.3618	
BMS 2785	108	0.4583	0.0607	11
	110	0.0417	0.0299	
	114	0.5000	0.2031	
CSSM08	181	0.2083	0.1672	11
	187	0.375	0.3856	
	193	0.375	0.3088	
	195	0.0417	0.048	
CSSM19	131	0.125	0.0169	17
	135	0.4167	0.095	
	139	0.1667	0.3768	
	141	0.125	0.1095	
	145	0.1667	0.1079	
CSSM43	240	0.2500	0.0448	23
	250	0.4583	0.1307	
	252	0.0833	0.2863	
	262	0.0417	0.0172	
	264	0.0417	0.0153	
	268	0.1250	0.0095	
CSSM47	129	0.3333	0.1488	27
	135	0.2500	0.164	
	141	0.1250	0.144	
	143	0.0833	0.0552	
	157	0.2083	0.1256	
ILSTS05	178	0.2917	0.4056	14
	180	0.4167	0.3832	
	198	0.2917	0.0968	
ILSTS11	262	0.5417	0.1168	19
	264	0.0833	0.36	
	268	0.3750	0.148	
ILSTS49	136	0.7083	0.668	10
	138	0.2917	0.2056	
RM232	113	0.0417	0.2128	8
	115	0.7500	0.5928	
	117	0.2083	0.1552	

daughters of these 12 sires and found the correctness of the paternity to be more than 89%. This high success could be attributed to the trained manpower of BAIF Pune and also the cooperation of the farmers who were sensitized about the objectives of the project. This value is higher from those obtained from the field (Jakhesara *et al.* 2012) for Mehsana buffaloes as well as in the organized herds abroad (Geldermann *et al.* 1986). In most of the districts belonging to the area of operation of the project paternity was found to be correct while the districts, which were recently adopted by BAIF for AI revealed more number of daughters with incorrect paternity.

The information on the daughters of the 12 sires with respect to date of birth of the buffalo calf, name of sire, paternity check, their tag numbers, name of district, center, village and the owners name and the phenotype generated on first lactation was compiled and published by authors under National Agricultural Innovation Project of Indian Council of Agricultural Research entitled 'Identification of Quantitative Trait Loci for Milk Yield, Fat and Protein Percentage in Buffaloes'- Buffalo Reference Family Germplasm Catalogue. The catalogue has entries of 8,027 buffaloes (out of a total of 8,845 daughters with correct paternity) reared in the project.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support of National Agriculture Innovation Project of Indian Council

of Agricultural Research (Grant no. 415401-02) under Component IV (Basic and Strategic Research in Agriculture). The authors also acknowledge large number of field functionaries, block level and district level officers of BIRD, Allahabad who whole heartedly supported this scientific endeavour. The authors acknowledge the directions and monitoring of such a large experiment by members of Consortium Advisory Committee of the project namely Drs A E Nivsarkar, S L Goswami, B K Joshi and Dinakar Raj.

REFERENCES

Annual Reports of Network Project on Buffalo Improvement 2013–14, 2014–15 and 2015–16. Available at *http://www.cirb.res.in/attachments/289_nwp.pdf*

Christensen L G, Madsen P and Petersen J. 1982. The influence of incorrect sire identification on the estimates of genetic parameters and breeding values. *Proceedings of the 2nd World Congress on Genetics Applied to Livestock Production,* Madrid, Spain. pp 200–208.

Geldermann H, Pieper U and Weber W E. 1986. Effect of misidentification on the estimation of breeding value and heritability in cattle. *Journal of Animal Science* **63**: 1759–68.

Heyen D W, Beever J E, Da Y, Evert R E, Green C, Bates S R E, Ziegle J S and Lewin H A. 1997. Exclusion probabilities of 22 bovine microsatellite markers in fluorescent multiplexes for semiautomated parentage testing. *Animal Genetics* **28**: 21–27.

Jakhesara S J, Rank D N, Kansara J D, Parikh R C, Patel V M, Vataliya P H and Solanki J V. 2012. Parentage verification in field progeny testing program of Mehsana buffalo. *Journal of Animal Science* **90**: 4723–28.

Ron M, Blane Y, Band M, Ezra E and Weller J I. 1996. Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. *Journal of Dairy Science* **79**: 676–81.

Usha A P, Simpson S P and Williams J L. 1995. Probability of random sire exclusion using microsatellite markers for parentage verification. *Animal Genetics* **26**: 155–61.

Vijh R K. 2014. Identification of Quantitative Trait Loci for Milk Yield, Fat and Protein percentage in Buffaloes. *Buffalo Reference Family Germplasm Catalogue*. Published by Indian Council of Agricultural Research. Pp 1–671.

Wiggans G R and VanRaden P M. 1989. USDA-DHIA animal model genetic evaluation. *Nationa Cooperative Dairy Herd Improvement Progamme Handbook*, Fact-sheet H-2, Washington, DC.