Original Research Paper

BIS-CATTLE: A Web Server for Breed Identification using Microsatellite DNA Markers

[1]Sarika Jaiswal, [2]Sandeep Kumar Dhanda, [1]M.A. Iquebal, [1]Vasu Arora, [3]Tejas M. Shah, [1]U.B. Angadi, [3]Chaitanya G. Joshi, [2]Gajendra P.S. Raghava, [1]Anil Rai and [1]Dinesh Kumar

[1]*Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, PUSA, New Delhi 110012, India*
[2]*Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh 160036, India*
[3]*Department of Animal Biotechnology, College of Veterinary Science & Animal Husbandry, Anand Agricultural University, Anand 388 001, Gujarat, India*

Abstract: Domestic cow, *Bos taurus* is one of the important species selected by humans for various traits, viz. milk yield, meat quality, draft ability, resistance to disease and pests and social and religious reasons. Since cattle domestication from Neolithic (8,000-10,000 years ago) today the population has reached 1.5 billion and further it's likely to be 2.6 billion by 2050. High magnitude of numbers, breed management, market need of traceability of breed product, conservation prioritization and IPR issues due to germplasm flow/exchange, has created a critical need for accurate and rapid breed identification. Since ages the defined breed descriptors has been used in identification of breed but due to lack of phenotypic description especially in ova, semen, embryos and breed products molecular approach is indispensable. Further the degree of admixture and non-descript animals characterization, needs of molecular approach is imperative. Till date breed identification methods based on molecular data analysis has great limitations like lack of reference data availability and need of computational expertise. To overcome these challenges we developed a web server for maintaining reference data and facility for breed identification. The reference data used for developing prediction model were obtained from 8 cattle breeds and 18 microsatellite DNA markers yielding 18000 allele data. In this study various algorithms were used for reducing number of loci or for identification of important loci. Minimization up to 5 loci was achieved using memory-based learning algorithm without compromising with accuracy of 95%. This model approach and methodology can play immense role in all domestic animal species across globe in breed identification and conservation programme. This can also be modelled even for all flora and fauna to identify their respective variety or breed needed in germplasm management.

Keywords: Breed, Cattle, Microsatellite, Prediction, Web Server

Introduction

The domestic cow (*Bos taurus*) is economically and culturally important species of the globe facilitating nutrition to the entire human population. 800 different cattle breeds in the world have been selected by humans for various traits, viz. milk yield, meat quality, draft ability, resistance to disease and pests and social and religious reasons. Cattle domestication initiated sometime in the Neolithic (8,000-10,000 years ago) with subsequent spread of cattle throughout the world is intertwined with human migrations and trade (Willham, 1986). At present, more than 1.5 billion cattles are reported which is liable to expand to 2.6 billion by 2050, as per FAO (2013). This high magnitude of numbers and breed

management needs accurate identification tool to identify the breeds at molecular level.

Every breed is a unique combination of genes evolved in response to a geo-climate along with adaptation of gene pool in a given ecological niches. In livestock animals, correct individual identification is essential for breeding purposes since their abilities are directly passed down to the next generation. The previous methods applied in cattle were use of tattoos and ear tags which contained individual identification numbers, followed by more informative electronic chips (Seo *et al.*, 2000). With time, blood typing results based on the protein polymorphisms for parentage testing was also followed but due to experimental complexity of blood typing system, it was replaced with DNA testing as applicable in forensic sciences on human (Yoon, 2002).

Today, well defined breed descriptors declared by Breed Societies or Statutory Bodies are used to categorise the breed. These phenotypic descriptors have limitations as they cannot be used to identify semen, ova, embryo or a breed product. Moreover these phenotypic descriptors cannot predict breed in admixture or so called non-descriptive population.STR markers have been used to identify domestic animal breed in large number of studies.

MacHugh *et al.* (1998) used 20 STR for assessing genetic structure of seven European cattle breeds along with the locus minimization up to 10 where the correct breed designation can be inferred with accuracies approaching 100%. But the major limitations of this work are the standalone mode of analysis and lack of reference data. Genetic variability and relationships among six native French cattle breeds and one foreign breed were investigated using 23 microsatellite markers by Maudet *et al.* (2002) where French alpine breeds with smaller population sizes showed higher genetic variability than the larger Holstein breed. They used two different assignment tests for determining the breed of origin of individuals. The exclusion-simulation significance test correctly assigned fewer individuals than the direct approach but provided a confidence level (p<0.01) for each individual being assigned. The accuracy of assignment greatly decreased as the threshold level of confidence of assignment increased as well as when the level of population differentiation decreased below the level often found between related breeds (F_{ST}<0.1). Accurate assignment with high statistical confidence is required for animal traceability.

Breed signature using microsatellite DNA marker has been reported in various domestic animals for example, cattle (Blott *et al.*, 1999; Maudet *et al.*, 2002), sheep (Arranz *et al.*, 2001; Niu *et al.*, 2011), goat (Serrano *et al.*, 2009; Hoda *et al.*, 2011), pig

(Fan *et al.*, 2005), horse (Bjornstad and Roed, 2002), dog (Toskinen and Bredbadka, 1999), poultry and rabbit (Gotz and Thaller, 1998).

Present work aims at resolving challenging limitations in all earlier reported methods *viz.* reference data availability, relatively larger number of locus requirement and complexity of computation. Lack of reference data availability adds cost addition as each time one has to genotype all potential/suspected breeds in question. We present here a novel model approach for breed identification of using test data of 8 cattle breeds. This methodology can be used for more number of breeds/countries to have the advantage of less genotyping cost as reference data obviates the cost of genotyping each time. Web server based computation makes this approach much user friendly.

Materials and Methods

Data Generation/Availability

For reference allelic data generation, a total of 500 samples were collected using random stratified sampling in the respective native breeding tract. Blood samples were collected in vacutainer containing EDTA as anticoagulant. Due care was taken to collect genetically unrelated samples of each breed. The panel of 500 samples constituted Dangi (67) Khillar (66), Nimari (65), Malvi (68), Kankrej (56), Gir (33) Gaolao (75), Kenkatha (70) animals. These samples were considered as representative of gene pool for further DNA signature investigation. Eighteen microsatellite loci were chosen from the microsatellite loci panel recommended by the Food and Agricultural Organization (FAO) and the International Society for Animal Genetics and previous studies (Chaudhary *et al.*, 2009; Kale *et al.*, 2010). The 18 loci are CSRM60, ILSTS005, ILSTS011, ILST006, MM12, ILSTS030, BM1824, HAUT27, BM1818, ETH152, INRA035, ETH10, INRA005, ILSTS034, CSSM663, ETH3, INRA063 and ILSTS033. As microsatellite markers are co-dominant thus a combination of 18 co-dominant loci and 500 samples are expected to generate 18000 allelic data for the populations under study to develop reference data of breed signature. Allelic richness (R_t) from all over samples for each locus using the rarefaction method (El Mousadik and Petit, 1996) and pairwise genetic distance between breeds were estimated using Weir and Cockerham's estimate of Wright's F_{ST} (θ) (Weir and Cockerham, 1984) through FSTAT v2.9.3 (Goudet, 2002).

Statistical Approaches

Various statistical classifiers like Bayesian network, Memory-Based Learning with nearest neighbour (IB1) and Support Vector Machine (SVM) was applied to build accurate model for classification of cattle breeds.

After developing the model, the prediction quality of the model was examined through evaluation measures like accuracy, Mathew's Correlation Coefficient (MCC), sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV). The brief description of classifiers used for model development is as follows.

Bayesian Networks as Classifiers

A Bayesian network B may be induced and encodes a probability distribution $P_B(A_1, A_2,..., A_n, C)$ from a given training set. The resulting model can be used so that given a set of attributes $a_1, a_2,..., a_n$, the classifier based on B returns the label c which maximizes the posterior probability, i.e.:

$$P_B\left(c \mid a_1, a_2,..., a_n\right)$$

Let $D = \{u_1, u_2,..., u_N\}$ denotes the training data set. Here, each u_i is a tuple of the form $\left\langle a_1^i, a_2^i,..., a_n^i c^i \right\rangle$ which assigns values to the attributes $A_1, A_2,..., A_n$ and to the class variable C. The log likelihood function, which measures the quality of learned model can be written as:

$$LL(B \mid D) = \sum_{t=1}^{N} \log P_B\left(c^i \mid a_1^i, a_2^i,..., a_n^i\right)$$
$$+ \sum_{t=1}^{N} \log P_B\left(a_1^i, a_2^i,..., a_n^i\right)$$

The first term in this equation measures efficiency of B to estimates the probability of a class given set of attribute values. The second term measures how well B estimates the joint distribution of the attributes. Since the classification is determined based on $P_B(c \mid A_1, A_2,..., A_n)$, only the first term is related to the score of the network as a classifier i.e., its predictive accuracy. This term is dominated by the second term, when there are many observations. As n grows larger, the probability of each particular assignment to $A_1, A_2,..., A_n$ becomes smaller, since the number of possible assignments grows exponentially in n.

Memory-Based Learning (MBL) Algorithm

Memory-Based Learning (MBL) is based on the hypothesis that performance in cognitive tasks depends on reasoning on the basis of similarity of new situations to stored representations of earlier experiences, rather than on the application of mental rules abstracted from earlier experiences. This approach has surfaced in different contexts using a variety of alternative names such as similarity-based, example-based, exemplar-based, analogical, case-based, instance-based and lazy learning (Cost and Salzberg, 1993; Aha *et al.*, 1991). Historically, MBL algorithms are descendants of the k-Nearest Neighbour (henceforth *k*-NN) algorithm (Aha *et al.*, 1991).

An MBL system has two components, *viz.* a learning component which is memory-based since it involves adding training instances to memory and a performance component which is similarity-based. In the performance component, the product of the learning component is used as a basis for mapping input to output; this usually takes the form of performing classification.

The most basic metric that works for patterns with symbolic features is the Overlap metric (Hamming distance or Manhattan metric or city-block distance or L1 metric) given in following equations; where $\Delta(X,Y)$ is the distance between instances X and Y, represented by n features and δ is the distance per feature. The *k*-NN algorithm with this metric is called IB1 (Aha *et al.*, 1991):

$$\Delta(X,Y) = \sum_{i=1}^{n} \delta\left(x_i, y_i\right)$$

Where:

$$\delta\left(x_i, y_i\right) = \begin{cases} abc\left(\dfrac{x_i - y_i}{\max_i - \min_i}\right) & if\ number, else \\ 0 & if\ x_i = y_i \\ 1 & if\ x_i \neq y_i \end{cases}$$

The major difference of *k*-NN algorithm with IB1 algorithm, implemented in Tilburg Memory-Based Learner (TiMBL) software originally proposed by (Aha *et al.*, 1991), is that in TiMBL version, the value of k refers to k-nearest distances rather than k-nearest examples. With $k = 1$, for instance, TiMBL's nearest neighbour set can contain several instances that are equally distant to the test instance. TiMBL, which is used in the study is an open source software package implementing several Memory-Based Learning (MBL) algorithms.

Support Vector Machines

Support vector machines are relatively new type of supervised machine-learning techniques, proven to be particularly attractive to biological analysis due to their ability to handle noise and large input spaces (Vapnik, 1999). The choice of the proper kernel function is an important issue for SVM training because the power of SVM comes from the kernel representation that allows the nonlinear mapping of input space to a higher dimensional feature space. Linear, polynomial, RBF and sigmoid are some typical choices of kernel function (Cristianini and Shawe-Taylor, 2000).

SVM can handle large feature spaces, effectively avoid over fitting by controlling the margin and automatically identify a small subset made up of informative points, i.e., support vectors, etc. The use of appropriate decision function can give better classification.

Cross Validation

In this study, five-fold cross validation technique was implemented to obtain estimate of prediction error, where the data sets were randomly divided into five equal sets. Among these, four sets were used for training and remaining one set for testing. The process was repeated five times such that each set has been used for testing. Average prediction on error estimation of five sets is calculated to estimate the prediction error.

Assessment of the Prediction Accuracy

The prediction quality of the model was examined through statistical estimation of the accuracy, Mathew's Correlation Coefficient (MCC), sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV). The evaluation measures are defined as follows:

$$Sensitivity = TP / (TP + FN)$$

$$Specificity = TN / (FP + TN)$$

$$PPV = TP / (TP + FP)$$

$$NPV = TN / (TN + FN)$$

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

$$MCC = \frac{(TP * TN - FP * FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Implementation of Server

The web server is developed using Hyper Text Markup Language and Java. It is a user-friendly web server launched using web server software Apache. The user has to submit microsatellite allelic data in base. This can also be uploaded using .csv format, .txt format or directly in the submission form. A number of flexibilities have been added in this server. The user may opt for breed identification with five loci, 10 loci and 18 loci. It has seven tabs *viz.* home, submission, algorithm, tutorial, team, links and FAQs. The server has tutorial for the users for easy understanding with a sample data.

Results and Discussion

The microsatellite DNA marker data was used in this study to develop a model web server minimizing the computational complexity and reduction in number of loci for breed identification. In order to achieve this, three classifiers *viz.* Bayesian network, Memory-based learning (IB1) algorithm and Support Vector Machine were applied over the 18000 allelic/ microsatellite data of cattle breeds. It was observed that models developed were with accuracies 95.45, 97 and 89.65%, respectively with five-fold cross validation. Hence, the model developed using Memory-based learning (IB1) algorithm was ultimately chosen to implement on the server using TiMBL: Tilburg Memory-Based Learner version 6.1 software (Daelemans *et al.*, 2010). The sensitivity, specificity and MCC were found to be 88, 98.29 and 0.86, respectively (Table 1). The breed wise accuracy and MCC of the implemented algorithm were rages from 91 to 100% and from 0.63 to 1.00, respectively (Table 2).

Though few *indicine* and *taurine* locus specific alleles are reported in very limited studies ARO23 (Metta *et al.*, 2004), OCAM (MacHugh *et al.*, 1967) and INRA124 (Giovambattista *et al.*, 2000) which shows STR alleles are rarely private. Thus they cannot be used as DNA signature directly. Instead of STR allele per say, the frequency and genetic distance can be used as signature. This was also observed in our studies. We did not find private or breed specific allele in any population.

Breed assignment tests can be performed by means of a number of statistical tools based on genetic distances and differences in allelic frequencies (Cornuet *et al.*, 1999). The later approach is widely reported in literature as it is usually obtained using a Bayesian approach (Ajmone-Marsan *et al.*, 2007; Talle *et al.*, 2005]. Ranking of loci were performed on the basis of contribution of each loci in the avergae prediction using best (TiMBL) algorithm. From the rank, here we applied an incremental feature addition approach and on each step performance was noted.

Similarly our results support the identification of eight breeds of cattle with 18 different loci with accuracy of 97%. In our investigation, the role of number of loci in breed identification is depicted in Fig. 1. It shows that the accuracy and MCC increases from 88.50 to 97.00% and 0.47 to 0.86 respectively with increase in the number of locus (Table 3). We can achieve accuracy upto 95.55% with MCC 0.80 by selecting just five loci (CSRM60, ILSTS005, BM1824, ILSTS034, ETH3). While minimising the number of locus, we found 95.55% accuracy using five STR locus and 96.1% accuracy with ten loci. Our finding shows that beyond ten loci, there is no need of

locus genotyping saving cost of breed identification upto 50%. Addition of further eight more loci will increase accuracy by just 0.9% unnecessarily making the genotyping cost in magnitude of almost double which is not desirable. We found interesting values of top five loci which are highly differentiated (F_{ST}>0.15 and Rt >8.0). In case of top 10 loci, we found similar range except two loci *viz.*HAUT27 and INRA035 (Table 4). We also found in these loci where genetic differentiation (F_{ST} value) is relatively less, the allelic richness (R_t value) was high enough to compensate the informativeness of locus for potential breed identification. This finding is supported from literature too. There are cases of domestic animal breed predictions with as low as three loci in horse (Bjornstad and Roed, 2002). Minimum number of locus with high accuracy is always desirable and such success comes when loci are highly differentiable i.e. high F_{ST} values for example in case of horse, F_{ST} is 0.2-0.25. The maximum individual assignment success with F_{ST} of 0.18 across 10 loci has been reported in dog (Koskinen, 2003). The results of genetic differentiation and analysis supported differentiation of the Murciana and Granadina populations with 25 microsatellites loci even with a low F_{ST} value (0.0432) and with assignment of individuals to their populations with a success rate of more than 80% (Martinez *et al.*, 2010).

Table 1. Performance of different classifiers

Classifiers	Sensitivity	Specificity	Accuracy	MCC
Bayesian network	81.80	97.40	95.45	0.79
Support vector machine	58.60	94.09	89.65	0.53
TiMBL-IB1 algorithm	88.00	98.29	97.00	0.86

Table 2. Prediction accuracies obtained for eight breeds of cattle

Breeds	Sensitivity	Specificity	Accuracy	MCC
DAG	100.00	100.00	100.00	1.00
GAO	100.00	100.00	100.00	1.00
GIR	60.61	99.57	97.00	0.73
KAN	69.64	96.62	93.60	0.67
KEN	100.00	100.00	100.00	1.00
KHL	100.00	100.00	100.00	1.00
MAL	80.88	96.53	94.40	0.76
NIM	73.85	93.56	91.00	0.63
Wt. Avg	88.00	98.29	97.00	0.86

DAG-Dangi; GAO- Gaolao; GIR- Gir; KAN- Kankrej; KEN- Kenkatha; KHL-Khillar; MAL- Malvi; NIM- Nimari

Table 3. Trend for average of sensitivity, specificity, accuracy and MCC with incrementing locus according to their rank

No. of Loci	Sensitivity	Specificity	Accuracy	MCC
1	54.00	93.43	88.50	0.47
2	70.00	95.71	92.50	0.66
3	73.80	96.26	93.45	0.70
4	77.60	96.80	94.40	0.74
5	82.20	97.46	95.55	0.80
6	81.80	97.40	95.45	0.79
7	83.00	97.57	95.75	0.81
8	83.00	97.57	95.75	0.81
9	83.00	97.57	95.75	0.81
10	84.40	97.77	96.10	0.82
11	85.40	97.91	96.35	0.83
12	85.00	97.86	96.25	0.83
13	85.20	97.89	96.30	0.83
14	86.00	98.00	96.50	0.84
15	86.40	98.06	96.60	0.84
16	87.20	98.11	96.95	0.85
17	87.80	98.26	96.95	0.86
18	88.00	98.29	97.00	0.86

Table 4. R_t and F_{ST} of top ten loci

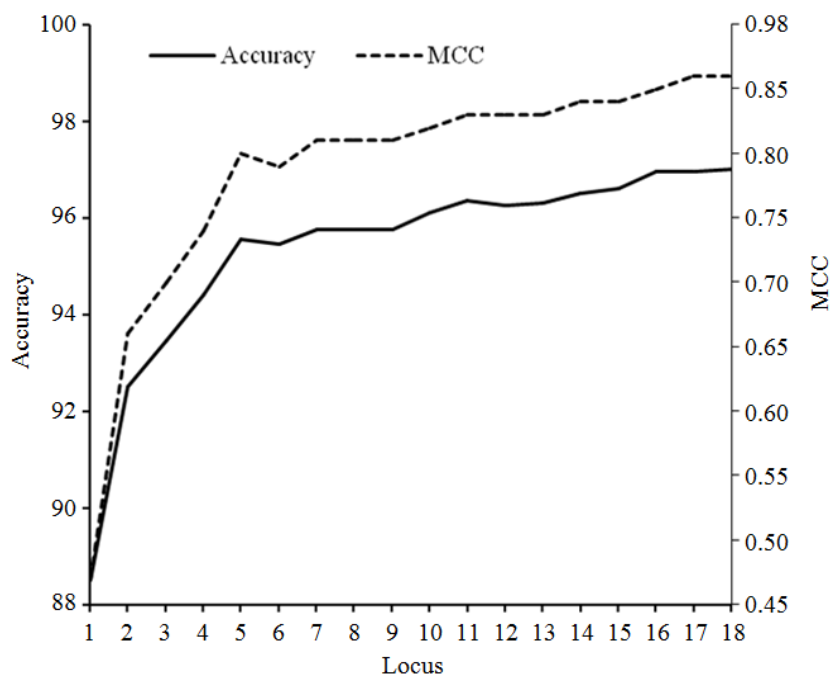| Locus	R_t (Allelic richness)	F_{ST} (Genetic differentiation)
CSRM60	13.69	0.150
ILSTS005	8.20	0.270
BM1824	8.17	0.272
ILSTS034	14.83	0.151
ETH3	8.09	0.359
ILSTS030	5.36	0.150
HAUT27	8.32	0.088
ETH152	9.33	0.321
INRA035	11.66	0.093
INRA005	8.23	0.113

Fig. 1. Trend of Accuracy and MCC with locus

Although available bovine High-Density (HD) SNP Chip (778K) (Bai *et al.*, 2012) and Low-Density (LD) SNP Chip (54 K) (Kuehn *et al.*, 2010) can also be used for breed identification but at the moment they are not cost effective for most part of the globe. Moreover, for breed differentiation, using SNP 50 K data can be done by limited number of software for example Mendel (Lange *et al.*, 2013) which is again not in server mode further compounding the issue of user-friendliness.

Conclusion

The present study reports world's first model web server for domestic animal breed prediction. We report accuracy of 95.5, 96.10 and 97% of 8 cattle breeds with 5, 10 and 18 loci respectively. Selecting less number of loci will not only reduce the cost drastically but also provide greater computational ease to identify the breed at molecular level with degree of admixture too. This can be an indispensable tool for existing breed and new synthetic commercial breeds with their IP protection in case of sovereignty and bio-piracy dispute. This web server can be used as a model for other domestic species as well as all flora and fauna across globe in germplasm management. Though we develop this model on microsatellite DNA markers but similar server based approach with reference data is going to be warranted for high thorough put SNP chip based data to reap the benefit of genomics and computational tools.

Acknowledgement

Author contributions

Sarika Jaiswal: Created the work-flow, web application and performed data analyses. Drafted the manuscript, read and approved the manuscript.

Sandeep Kumar Dhanda, Vasu Arora and U.B. Angadi: Created the work-flow, web application and performed data analyses, read and approved the manuscript

M.A. Iquebal: Created the work-flow, web application and performed data analyses, drafted the manuscript, read and approved the manuscript

Tejas M. Shah: Participated in sample collection and data generation, read and proved the manuscript.

Chaitanya G. Joshi: Drafted the manuscript, read and approved the manuscript.

Gajendra P.S. Raghava and Anil Rai: Conceived this study, read and approved the manuscript.

Dinesh Kumar: Conceived this study, drafted the manuscript, read and approved the manuscript.

Conflict of Interest

The authors declare that they have no conflict of interest.

Reference

Aha, D.W., D. Kibler and M. Albert, 1991. Instance-based learning algorithms. Mach. Learn., 6: 37-66. DOI: 10.1007/BF00153759

Ajmone-Marsan, P., R. Negrini, E. Milanesi, L. Colli and M. Pellecchia, 2007. Breed assignment of Italian cattle using biallelic AFLP® markers. Anim. Genet., 38: 147-53. DOI: 10.1111/j.1365-2052.2007.01573.x

Arranz, J., Y. Bayon and F.S. Primitivo, 2001. Differentiation among Spanish sheep breeds using microsatellites. Genet. Sel. Evolut., 33: 529-42. DOI: 10.1186/1297-9686-33-5-529

Bai, Y., M. Sartor and J. Cavalcoli, 2012. Current status and future perspectives for sequencing livestock genomes. J. Anim. Sci. Biotechnol., 3: 8-8. DOI: 10.1186/2049-1891-3-8

Bjornstad, G. and K.H. Roed, 2002. Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses. Anim. Genet., 33: 264-70. DOI: 10.1046/j.1365-2052.2002.00868.x

Blott, S.C., J.L. Williams and C.S. Haley, 1999. Discriminating among cattle breeds using genetic markers. Heredity, 82: 613-619. DOI: 10.1046/j.1365-2540.1999.00521.x

Chaudhary, M.V., S.N.S. Parmar, C.G. Joshi, C.D. Bhong and S. Fatima *et al.* 2009. Molecular characterization of Kenkatha and Gaolao (*Bos indicus*) cattle breeds using microsatellite markers. Anim. Biodivers Conserv., 32: 71-6.

Cornuet, J.M., S. Piry, G. Luikart, A. Estoup and M. Solignac, 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics, 153: 1989-2000. PMID: 10581301

Cost, S. and S. Salzberg, 1993. A weighted nearest neighbor algorithm for learning with symbolic features. Mach. Learn., 10: 57-78. DOI: 10.1007/BF00993481

Cristianini, N. and J. Shawe-Taylor, 2000. An Introduction to Support Vector Machines and other Kernel-based Learning Methods. 1st Edn., Cambridge University Press, Cambridge, ISBN-10: 0521780195, pp: 189.

Daelemans, W., J. Zavrel, K. Van der Sloot and A. Van den Bosch, 2010. TiMBL: Tilburg memory based learner, version 6.3, reference guide. ILK Research Group Technical Report.

El Mousadik, A. and R.J. Petit, 1996. High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa (L.) Skeels)* endemic to Morocco. Theor. Applied Genet., 92: 832-9. DOI: 10.1007/BF00221895

Fan, B., Y.Z. Chen, C. Moran, S.H. Zhao and B. Liu *et al.*, 2005. Individual-breed assignment analysis in swine populations by using microsatellite markers. Asian-Aust. J. Anim. Sci., 11: 1529-34. DOI: 10.5713/ajas.2005.1529

FAO, 2013. The state of the world's animal genetics resources for food and agriculture. FAO.

Giovambattista, G., M.V. Ripoli, J.C. De Luca, P.M. Mirol and J.P. Lirón *et al.*, 2000. Male-mediated introgression of *Bos indicus* genes into Argentine and Bolivian Creole cattle breeds. Anim. Genet., 31: 302-5. DOI: 10.1046/j.1365-2052.2000.00658.x

Gotz, K. and G. Thaller, 1998. Assignment of individuals to populations using microsatellites. J. Anim. Breed Genet., 115: 53-61. DOI: 10.1111/j.1439-0388.1998.tb00327.x

Goudet, J., 2002. FSTAT, a program to estimate and test gene diversities and fixation indices Version 2.9.3.2.

Hoda, A., G.A. Hyka, S. Dunner and G. Obexer-Ruff, 2011. Genetic diversity of Albanian goat breeds based on microsatellite markers. Arch. Zootec., 60: 607-15.

Kale, D.S., D.N. Rank, C.G. Joshi, B.R. Yadav and P.G. Koringa *et al.*, 2010. Genetic diversity among Indian Gir, Deoni and Kankrej cattle breeds based on microsatellite markers. Ind. J. Biotechnol., 9: 126-30.

Koskinen, M.T., 2003. Individual assignment using microsatellite DNA reveals unambiguous breed identification in the domestic dog. Anim. Genet., 34: 297-301. DOI: 10.1046/j.1365-2052.2003.01005.x

Kuehn, L.A., J.W. Keele, G.L. Bennett, T.G. McDaneld and T.P.L. Smith *et al.*, 2010. Predicting breed composition using breed frequencies of 50,000 markers from the US Meat Animal Research Center 2,000 Bull Project. J. Anim. Sci., 89: 1742-50. DOI: 10.2527/jas.2010-3530

Lange, K., J.C. Papp, J.S. Sinsheimer, R. Sripracha and H. Zhou *et al.*, 2013. Mendel: The Swiss army knife of genetic analysis programs. Bioinformatics, 29: 1568-70. DOI: 10.1093/bioinformatics/btt187

MacHugh, D.E., M.D. Shriver, R.T. Loftus, P. Cunningham and D.G. Bradley, 1967. Microsatellite DNA variation and the evolution, domestication and phylogeography of Taurine and Zebu Cattle (*Bos taurus and Bos indicus*). Genetics, 146: 1071-86. PMID: 9215909

MacHugh, D.E., R.T. Loftus, P. Cunningham and D.G. Bradley, 1998. Genetic structure of seven European cattle breeds assessed using 20 microsatellite markers. Anim. Genet., 29: 333-40.
DOI: 10.1046/j.1365-2052.1998.295330.x

Martinez, A.M., J.L. Vega-Pla, J.M. Leon, M.E. Camacho and J.V. Delgado *et al.*, 2010. Is the Murciano-Granadina a single goat breed? A molecular genetics approach. Arq. Bras. Med. Vet. Zootec., 62: 1191-8.
DOI: 10.1590/S0102-09352010000500023

Maudet, C., G. Luikart and P. Taberlet, 2002. Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. J. Anim. Sci., 80: 942-50. PMID: 12002331

Metta, M., S. Kanginakudru, N. Gudiseva and J. Nagaraju, 2004. Genetic characterization of the Indian cattle breeds, Ongole and Deoni (*Bos indicus*), using microsatellite markers-a preliminary study. BMC Genet., 5: 16-16.
DOI: 10.1186/1471-2156-5-16

Niu, L.L., H.B. Li, Y.H. Ma and L.X. Du, 2011. Genetic variability and individual assignment of Chinese indigenous sheep populations (*Ovis aries*) using microsatellites. Anim. Genet., 43: 108-11.
DOI: 10.1186/1471-2156-10-61

Seo, K.S., Y.M. Cho and H.K. Lee, 2000. Development of network system for the application of HACCP in livestock production stage. AgroInformat. J., 1: 1-4.

Serrano, M., J.H. Calvo, M. Martinez, A. Marcos-Carcavilla and J Cuevas *et al.*, 2009. Microsatellite based genetic diversity and population structure of the endangered Spanish Guadarrama goat breed. BMC Genet., 10: 61-61.
DOI: 10.4321/S0004-05922011000300049

Talle, S.B., E. Fimland, O. Syrstad, T. Meuwissen and H. Klungland, 2005. Comparison of individual assignment methods and factors affecting assignment success in cattle breeds using microsatellites. Acta Agric. Scandinavae, 55: 74-9.
DOI: 10.1080/09064700500435416

Toskinen, M.T. and P. Bredbadka, 1999. A convenient and efficient microsatellite-based assay for resolving parentages in dogs. Anim. Genet., 30: 148-9.
DOI: 10.1046/j.1365-2052.1999.00438.x

Vapnik, V., 1999. The Nature of Statistical Learning Theory. 2nd Edn., Springer-Verlag Press, New York, ISBN-10: 0387987800, pp: 314.

Weir, B.S. and C.C. Cockerham, 1984. Estimating F-statistics for the analysis of population structure. Evolution, 38: 1358-70. DOI: 10.2307/2408641

Willham, R., 1986. From husbandry to science: A highly significant facet of our livestock heritage. J. Anim. Sci., 62: 1742-58.

Yoon, D.H., 2002. Molecular genetic diversity and development of genetic markers in association with meat quality for Hanwoo (Korean cattle). Ph.D. Thesis.