Theriogenology 83 (2015) 1445–1450

Contents lists available at ScienceDirect

Theriogenology

journal homepage: www.theriojournal.com

Designing of an artificial neural network model to evaluate the association of three combined Y-specific microsatellite loci on the actual and predicted postthaw motility in crossbred bull semen

Rajib Deb [a,*], Umesh Singh [a], Thirvvothur Venkatesan Raja [a], Sushil Kumar [a], Shrikant Tyagi [b], Rafeeque R. Alyethodi [a], Rani Alex [a], Gyanendra Sengar [a], Sheetal Sharma [a]

[a] Division of Animal Genetics and Breeding, Molecular Genetics Laboratory, ICAR-Central Institute for Research on Cattle, Meerut, Uttar Pradesh, India
[b] Animal Reproduction Section, Semen Freezing Laboratory, ICAR-Central Institute for Research on Cattle, Meerut, Uttar Pradesh, India

ARTICLE INFO

Article history:
Received 9 September 2014
Received in revised form 9 January 2015
Accepted 13 January 2015

Keywords:
Artificial neural network
Crossbred bull
Multiple regression analysis
Postthaw motility
Semen

ABSTRACT

The freezing of bull semen significantly hamper the motility of sperm which reduces the conception rate in dairy cattle. The prediction of postthaw motility (PTM) before freezing will be useful to take the decision on discarding or freezing of the germplasm. The artificial neural network (ANN) methodology found to be useful in prediction and classification problems related to animal science, and hence, the present study was undertaken to compare the efficiency of ANN in prediction of PTM on the basis of the number of ejaculates, volume, and concentration of sperms. The combined effect of Y-specific microsatellite alleles on the actual and predicted PTM was also studied. The results revealed that the prediction accuracy of PTM based on the semen quality parameters was comparatively lower because of higher variability in the data set. The ANN gave better prediction accuracy (34.88%) than the multiple regression analysis models (32.04%). The root mean square error was lower for ANN (8.4353) than that in the multiple regression analysis (8.6168). The haplotype or combined effect of microsatellite alleles on actual and predicted PTM was found to be highly significant ($P < 0.01$). On the basis of results, it was concluded that the ANN methodology can be used for prediction of PTM in crossbred bulls.

1. Introduction

The freezing of the bull semen significantly hamper the motility of sperm which leads to reduced conception rate in dairy cattle. The volume of semen, concentration of sperms, and number of ejaculations will have a direct effect on the motility and postthaw motility (PTM) of the sperm [1–3].

The prediction of PTM based on the other sperm quality parameters would be useful to discard the inferior germplasm before freezing which will reduce the time, money, and wastage of skilled man power. Generally, a multiple regression analysis (MRA) is used for prediction that suffers with some disadvantages *viz.*, multicolinearity, number of independent variables included in the model, number of observations, and henceforth. The artificial neural network (ANN), also known as a connectionist model, is the recent technology which can overcome the disadvantages of MRA technique and found to be suitable for most of the studies

* Corresponding author. Tel.: +91 121 2645598; fax: +91 121 2657134.
E-mail address: drrajibdeb@gmail.com (R. Deb).

0093-691X/$ – see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.theriogenology.2015.01.014

in animal sciences [4–9]. So far, report on the application of ANN on reproductive data is very scanty, and hence, the present study was undertaken to design an ANN model for prediction of PTM in crossbred bull semen and assess the efficacy of the ANN model over the conventional MRA.

Deb et al. [10] earlier reported that Y-specific microsatellite markers can be used as an alternative biomarker for screening bull semen quality. In addition, the present study also aimed to identify the effect of haplotype or three combined Y-chromosome–specific microsatellite loci on the actual and predicted PTM of crossbred bull semen.

2. Materials and methods

All experimental procedures were approved by the Institutional Animal Ethics Committee of the Indian Council of Agricultural Research-Central Institute for Research on Cattle, Meerut, Uttar Pradesh, India.

2.1. Experimental animals, data collection, and DNA extraction from bull semen

Semen samples were collected from 82 mature Frieswal (Holstein Friesian × Sahiwal) bulls maintained at the Bull Rearing Unit of Central Institute for Research on Cattle, Meerut, during the period from December 2010 to November 2011, were used for the study. Immediately after collection, the ejaculates were stored at 34 °C in a water bath to evaluate the fresh semen quality traits including semen volume (mL) and sperm concentration (million/mL). The fresh semen was then diluted with glycerol–egg yolk–citrate–Tris buffer and then cryopreserved. After storage in liquid nitrogen for 1 to 2 days, two straws were randomly obtained from each ejaculate, thawed at 37 °C for 60 seconds, and immediately evaluated for PTM (%) with phase-contrast microscope. Genomic DNA was extracted from the sperm using the GenElute Blood Genomic DNA Kit (Sigma–Aldrich, USA). The DNA samples were dissolved in elution buffer (supplied with kit) and stored at −20 °C for further use.

2.2. Primer designing and PCR amplification, genotyping, and their combinations

For the amplification of each Y-specific microsatellite loci from genomic DNA, primer sequence data were obtained from scientific literatures and gene bank (Table 1). All forward primers are 6-FAM (6-carboxyfluorescein) labeled at the 5′ end (Golbia, Bioserver Pvt Ltd., India). The polymerase chain reactions (PCR) were carried out in a total volume of 25-μL solution containing 50 ng/μL of template DNA, 1X buffer (Tris-HCl, 100 mmol/L, pH 8.3; KCl, 500 mmol/L), 0.25 μmol/L primers, 2.0 mmol/L $MgCl_2$, 0.25 mmol/L dNTPs, and 0.5-U Taq DNA polymerase (Sigma–Aldrich, USA). The PCR protocol was 94 °C for 5 minutes, followed by 35 cycles of 94 °C for 30 seconds, annealing at 55 °C to 58 °C for 30 seconds and 72 °C for 30 seconds, and a final extension at 72 °C for 8 minutes. The PCR products were separated on 1.0% agarose gel (Sigma–Aldrich, USA) including 0.5 μg/mL of ethidium bromide, photographed under the Gel Documentation system (Alphaimager EP). Amplified PCR products were gel purified and sent for genotyping from outsourcing (Science Genome Pvt Ltd., India). The major alleles of the three microsatellite loci were combined as haplotypes to study their effect on the actual and predicted PTM of Frieswal bull semen.

2.3. Prediction of PTM by ANN and MRA models

2.3.1. Development of ANN models

The detailed description and principle of ANN methodology has been well explained by [14], and the readers are suggested to refer the article for better understanding. A multilayer feedforward network with back propagation of error-learning mechanism was developed for predicting the PTM. The Neural Network Toolbox of MATLAB 7.8.0 was used to develop the ANN model. A general schematic diagram of the multilayer feedforward network with an input layer, two hidden layers, and an output layer is depicted in Supplemental Materials. The input layer consisted of three nodes/neurons representing three independent variables *viz.*, the number of ejaculations (X_1), volume of semen (X_2), and spermatozoal concentration (X_3), whereas the output layer consisted of one node for the PTM (Y). The number of hidden layers and the number of neurons in each hidden layer were randomly modified to find the optimum model for prediction of PTM from semen quality parameters. The number of hidden layers ranged from one to two, and the number of neurons in each hidden layer ranged from one to 20. Initial weights and bias matrix were randomly initialized between −1 and 1. A nonlinear transformation (or activation) function tangent sigmoid was used to compute the output from the summation of weighted inputs of neurons in each hidden layer. A pure linear transformation function was used at the output layer for getting network response. The designed network was trained in supervisory mode with the Bayesian regularization algorithm, and the network was trained using previously mentioned learning algorithms for up to 2000 epochs or till the algorithms truly converged. The input and target data were preprocessed, so

Table 1
List of Y-specific microsatellite markers used for the present study.

Sl. No.	Marker	Primer sequences (5′–3′)	Repeat motif	Number of alleles	Range of alleles (bp)	Reference
1	INRA126	F: GTTGTTGCCTCTGCAGAGTAGG R: GACACTCTTTCTATTTTCAAGG	$(TG)_{11}$	3	182–186	Vaiman et al., 1994 [11]
2	INRA 189	F: TACACGCATGTCCTTGTTTCGG R: CTCTGCATCTGTCCTGGACTGG	$(TG)_{22}$	9	68–124	Kappes et al., 1997 [12]
3	BM861	F: TTGAGCCACCTGGAAAGC R: CAAGCGGTTGGTTCAGATG	$(GT)_6$ (C) $(TG)_9$	6	144–158	Bishop et al., 1994 [13]

Abbreviations: CV, coefficient of variation; SD, standard deviation.