Haryana Vet. (June, 2019) **58(1),** 111-113

Research Article

A GENETIC DISSECTION OF BREED COMPOSITION IN WELL-KNOWN TAURINE, INDICINE, AND CROSSBRED CATTLE BY USING SNP GENOTYPING

SHEIKH FIRDOUS AHMAD*, MANJIT PANIGRAHI, ROUF RASHID DAR[1], AJAZ ALI[1] and BHARAT BHUSHAN
Division of Animal Genetics, [1]Division of Animal Reproduction
Indian Veterinary Research Institute, Izatnagar, Bareilly-243122, India

Received: 20.10.2018; Accepted: 07.02.2019

ABSTRACT

In this study, the genotype data (50K SNP bead chip data) was used for a total of 156 representative animals belonging to seven indicine and four taurine breeds of cattle. A total of 43,041 SNPs common for all the breeds were taken into account for the present study. Using the Bayesian approach, STRUCTURE software was run on the data set. Under each value of K (2-12), the indicine and taurine breeds were distinctly separated, with poor differentiation of breeds within the two lineages. The results from principal component analysis were in accordance with separate identities of the two lineages. The study was extended under similar conditions to hybrid Santa Gertrudis individuals, with 10 individuals included from hybrid and purebred constituent populations and the tools used in the study efficiently differentiated the individuals from pure breed and crossbred populations. It may be concluded that these Bioinformatics and statistical softwares are efficient in determination of clustering, introgression and admixture levels in different purebred and crossbred populations of India.

Key words: Introgression, Indicine, K value, Lineage, STRUCTURE, Taurine

The origin and domestication of the bovine species are well understood with strong evidence for a major bifurcation between two lineages with distinct origins, i.e. Bos taurus and Bos indicus (Ajmone Marsan *et al.*, 2010). This divergence is reported to be of several hundred thousand years, through several genetic studies (Orozco-ter Wengel *et al.*, 2015).

Crossbreeding technique has been prevalent and meticulously finds a place in breeding policies of various countries, especially the developing ones. Santa Gertrudis is a milch type cattle breed developed in Southern Texas and was evolved using the genome of Brahman and Shorthorn cattle with eventual inheritance levels of 37.5 and 62.5%, respectively. This breed is hybrid cattle and possesses an introgressed genome of the two breeds and may act as good candidate to assess the efficiency of statistical models for admixture studies.

Microsatellite markers have long been used in population structure and divergence studies. Now-a-days, the single nucleotide polymorphism (SNP) markers have emerged rapidly in the modern era due to their beneficial properties. These properties include their robustness, cost effectiveness, automatic allele calling, low mutation rate, high incidence throughout the genome and bi-allelic nature; making them amenable to automated detection techniques (Singh *et al.*, 2014).

In this study, we have hypothesized the possibility of using the BovineSNP50BeadChip and maximum information on admixture levels of crossbred population, their population structure and genetic diversity can be elucidated by using Bioinformatics and statistical tools on indigenous and crossbred cattle populations. The main Bioinformatics' softwares used for this purpose include that of STRUCTURE (Pritchard *et al.*, 2000) and Principal Component Analysis (PCA).

Corresponding Author: firdousa61@gmail.com

MATERIAL AND METHODS

Data retrieval and number of animals

In order to study the introgression patterns applicable among indigenous and taurine cattle breeds in the Indian context, we retrieved the 50K SNP Bead Chip genotypic data from a Dryad digital repository (Decker *et al.*, 2014a; Decker *et al.*, 2014b) to compare the genetic makeup of the breeds involved in the study. The 50K SNP genotypic data were retrieved from the public repository for representative breeds of indicine and taurine cattle breeds. In this study, a total of 156 (n) animals were included which includes; 99 animals (n^1) belonging to the indicine group and 57 animals (n^2) were taken from the taurine cattle group. Seven and four representative breeds were taken from indicine and exotic breeds, respectively. The representative breeds from indicine breeds included Sahiwal (17), Gir (20), Tharparkar (12), Red Sindhi (10), Hariana (10), Kankrej (10) and Ongole (20). Whereas, the taurine group included the genotypic data from Holstein Friesian (20), Jersey (20), Brown Swiss (12), Guernsey (5) breeds. Genotypic data consisted of 156 animals in the data set with 43,041 loci of SNP variants of Illumina Bovine 50K SNP Bead Chip covered under it. This data was pruned by original authors (Decker *et al.*, 2014a; Decker *et al.*, 2014b) for parameters including call rate of 0.9 and an MAF (minor allele frequency) of 0.0005.

A separate data set was prepared for hybrid cattle i.e. Santa Gertrudis that possesses the inheritance of two lineages i.e. Brahman and Shorthorn breeds with the eventual inheritance of 37.5 and 62.5%, respectively. The dataset was formed by a total of 30 individuals from three populations i.e. one hybrid and two constituent breeds.

Individuals from Brahman, Shorthorn and Santa Gertrudis breed formed three populations with an equal share from each population i.e. 10 individuals each. 43,041 SNP variant markers were retrieved from the same public platform, produced after same quality control conditions for this data set.

Application of Bioinformatics' tools

A total of 43,041 SNP markers common among the constituent breeds of the two lineages were used for the study. The data was fed to STRUCTURE software and assumptions of K values from 2-12 were made. The K value refers to the prior number of subpopulations for which the STRUCTURE program tries to divide the data set into. The Bioinformatics' software was made to run three iterative cycles for each value of K for 5000 and 25000 values of Burnin and MCMC runs, respectively. Bar plots (individual and population-wise), data plots, histograms, triangle plots were produced besides the parameters of alpha and FST for each value of K. Similar treatment was done to the second data set belonging to the hybrid Santa Gertrudis cattle population, but with K value of two to four (2-4).

The structure results from different iterative cycles for various runs of K values were zipped and fed to a harvester (Earl and vonHoldt, 2012). The harvester results were analysed mainly for deciding the best value of K for our data. The harvester produced the clump files for all animals under different values of K. A graphical plot of Ln Pr (K) and K was analysed for determining the best value of K fitting our study.

Principal Component Analysis (PCA)

PCA, when applied on the genotypic data, refers to a method that tries to infer individuals with new dimensions into different components explaining the variation between and among the individuals of the same and different species, respectively. R-pipeline was used for PCA analysis applying SNPRelate and gdsfmt packages for analysis. PCA, a statistical procedure, was applied on the genotypic data with the aim to check for stratification of two groups/lineages of cattle breeds. Principal component analysis in R-programming tries to stratify the populations in the data set by using different eigenvectors that explain the variation among different clusters of the data set. The PCA program under R-software was applied to both the datasets, with the aim to see the stratification patterns in the datasets.

RESULTS AND DISCUSSION

The four breeds used in the present study i.e. Holstein-Friesian, Jersey, Brown Swiss and Guernsey have been used extensively in India for crossbreeding. K values of 2-12 were used for analysis to check if the bioinformatics' software was able to detect the subpopulation and if the breed populations are separated efficiently. The STRUCTURE program forms different clusters based on the allele frequencies of different populations in the data set. Any two clusters can be plotted at two ends of the triangle against all other clusters at the third end. The animals belonging to taurine and indicine lineages were meticulously differentiated through the Structure analysis. Across all values of K in the bar plots, the indicine and taurine groups of breeds were distinguished on the basis of SNP genotypic data (Fig. 1). The value of K equal to 11 was analysed critically to check for any structure among the breeds of two lineages/ clusters. The individual lineages of taurine and indicine cattle were clustered up to the extent of 98.78% and 99.83%, respectively. Across all values of K, the barplot depicted some introgression of the genome for Ongole breed of indigenous cattle as depicted in Fig. 1. Sharma *et al.* (2015) made similar conclusions of Ongole breed being distinct from other Indian cattle breeds. Similarly, Ongole breed of cattle was reported to maintain abundant genetic variability in earlier studies based on other type of loci (Karthickeyan *et al.*, 2008; Devi *et al.*, 2017). Our results were thus in accordance with these studies.

To understand the breed admixture in crossbreds, STRUCTURE software was also successfully used to differentiate the pure breeds and to estimate the breed composition in Santa Gertrudis. We were able to cluster out the individuals of Brahman and Beef Shorthorn breeds up to the extent of 99.52% and 98.76%, respectively. Under the Structure software analysis, the Santa Gertrudis animals were found to be a blend of the two breeds i.e. Brahman and Beef Shorthorn up to the levels of 36.58 and 63.42%, respectively. The results of a blend of genotype up to these levels were in near-perfect accordance to the original inheritance of Santa Gertrudis breed by pedigree analysis. The bar plot was evidence for the same as depicted in Fig. 2. Here, only two clusters were formed in the triangle plot with the animals of Santa Gertrudis located in between, closer to the cluster of Shorthorn cattle population. The clustering results were thus in perfect accordance with STRUCTURE program results.

FST values from STRUCTURE program are effective depicters for the efficiency of structuring for subpopulations (Srivastava *et al.*, 2014). Generally, lower the FST values, efficient is the stratification. In data set with animals of two breed lineages, FST value of 0.1239 was produced during for K value of two that depicted the efficient sub-structuring of the constituent populations. Analysing the FST value of 0.1332 for K value of two in data set of hybrid cattle (Santa Gertrudis), it again depicted the effective sub-structuring of constituent populations. Thus, it was concluded that this approach of using Bioinformatics tool (STRUCTURE program) was equally

efficient in detecting the introgression of two cattle lineages, especially in the Indian context.

Using STRUCTURE harvester analysis, we tried to minimize the error in predicting the best value of K for our genotypic data from the software analysis. After the analysis of the scatter plot with Ln Pr (K) values presented as functions of the number of clusters, we were able to predict the cluster with a K value of two showed the minimum possible error and was perfect for populations under our study. In fact, the STRCUTURE harvester showed the minimum probabilistic error as shown for K values of two and three. These values imply that the data set was formed of two lineages and thus the K value of two showed the minimum probabilistic error. The minimum probabilistic error for K value of three reaffirmed the presence of somewhat unique genetic makeup of Ongole population in the data set.

PCA approach using genotypic data has recently emerged as one of significant methods of stratification of populations and dimension reduction of variances from correlated allelic frequencies. The few top principal components (PCs) explaining most of total trait variance are generally taken for analysis while other are left as such (Aschard *et al.*, 2014). The PCA analysis on whole 50K SNP genotypic data of sample population was able to differentiate the two lineages with the first two principal components explaining about 13.8% and 4.8% of variations, respectively, with the former component separating the two lineages while the later component stratifying different breeds among the taurine lineage cattle population (Fig. 3). It efficiently clustered the animal population from two lineages separately along various coordinates. The breeds from taurine lineage were clustered separately while as the indigenous breeds were clustered together, not showing any large variations. The results pointed towards distinct properties of the genome of these breeds that taurine breeds were efficiently clustered separately whereas the same could not be possible for indigenous breeds. The structure results for Ongole breed being varied from other indigenous breeds was reaffirmed from the results of PCA with most individuals of Ongole population presented separately from the other indicine breeds.

CONCLUSION

It may be concluded from the present study that the bioinformatics and statistical tools are efficient in detecting the admixture and evolutionary trend in hybrid composition of various crossbred strains present in India and all over the world. These tools may be applicable to estimate the breed composition and admixture in the Indian context where huge crossbred populations are maintained across different regions.

ACKNOWLEDGEMENT

This study was supported by the ICAR-Indian Veterinary Research Institute, Government of India for providing the necessary funding. The authors would also wish to acknowledge the help and support rendered by Director, IVRI, Bareilly, India, for providing necessary facilities to carry out this work.

REFERENCES

Aschard, H., Vilhjálmsson, B.J., Greliche, N., Morange, P.E., Trégouët, D.A. and Kraft, P. (2014). Maximizing the power of principal-component analysis of correlated phenotypes in genome-wide association studies. *Am. J. Hum. Genet.* **94(5)**: 662-676.

Ajmone Marsan, P., Garcia, J.F. and Lenstra, J.A. (2010). On the origin of cattle: how aurochs became cattle and colonized the world. *Evol. Anthropol.* **19(4)**: 148-157.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Alcalá, A.M., Sonstegard, T.S., Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L. and Babar, M.E. (2014a). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* **10(3)**: p.e1004254.

Decker, J.E., McKay, S.D., Rolf, M.M., Kim, J., Alcalá, A.M., Sonstegard, T.S., Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L., Babar, M.E., Regitano, L.A., Yildiz, M.A., Heaton, M.P., Liu, W., Lei, C., Reecy, J.M., Saif-Ur-Rehman, M., Schnabel, R.D. and Taylor J.F. (2014b). Data from: Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS Genet.* http://dx.doi.org/10.5061/dryad.th092.

Devi, K.S., Gupta, B.R., Vani, S., Asha, U., Kumar, U.R. and Krishna, C.H. (2017). Microsatellite analysis of Ongole cattle (*Bos indicus*) of AP. *Inter. J. Sci. Environ. Technol.* **6(1)**: 173-178.

Earl, D.A. and von Holdt, B.M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4(2)**: 359-361.

Karthickeyan, S.M.K., Kumarasamy, P., Sivaselvam, S.N., Saravanan, R. and Thangaraju, P. (2008). Analysis of microsatellite markers in Ongole breed of cattle. *Indian J. Biotechnol.* **7**: 113-116.

Orozco-ter Wengel, P., Barbato, M., Nicolazzi, E., Biscarini, F., Milanesi, M., Davies, W., Williams, D., Stella, A., Ajmone-Marsan, P. and Bruford, M.W. (2015). Revisiting demographic processes in cattle with genome-wide population genetic analysis. *Front. Genet.* **6**: 191.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genet.* **155**: 945–959.

Sharma, R., Kishore, A., Mukesh, M., Ahlawat, S., Maitra, A., Pandey, A.K. and Tantia, M.S. (2015). Genetic diversity and relationship of Indian cattle inferred from microsatellite and mitochondrial DNA markers. *BMC Genet.* **16(1)**: 73.

Singh, U., Deb, R., Alyethodi, R.R., Alex, R., Kumar, S., Chakraborty, S., Dhama, K. and Sharma, A. (2014). Molecular markers and their applications in cattle genetic research: A review. *Biom. Genomic Med.* **6(2)**: 49-58.

Srivastava, A.K., Chopra, R., Ali, S., Aggarwal, S., Vig, L. and Bamezai, R.N.K. (2014). Inferring population structure and relationship using minimal independent evolutionary markers in Y-chromosome: a hybrid approach of recursive feature selection for hierarchical clustering. *Nucleic Acids Res.* **42(15)**: e122-e122.