D² statistic technique used for analysis of genetic divergence among Gir crossbreds

Mahendra Gorakh Mote¹, Vishnu Eknath Narwade¹, Ulhas Shivaji Gaikwad² and Pramodkumar Shriram Sakhare¹

Abstract: D² analysis is used for selection of genetically divergent parents. It is used to measure degree of diversification and determines relative portion of each component trait to total divergence. The study aimed to estimate the genetic divergence among FG (50% HF+50% Gir), IFG (FG Interse), FJG (50% HF+25% Jersey+25% Gir), IFJG (FJG Interse) and R (50% HF+12.50% Jersey+37.50% Gir) crosses of Gir on the basis of age at first calving, first calving interval, lactation milk yield, lactation length and lactation milk yield per day of calving interval using Mahalanobis D² statistics. The genetic group differences were significant (P<0.01) for all traits separately and simultaneously (V test) based on five traits. The differences in the D² values among all the genetic groups combinations, were significant except IFG with R genetic group combination. The total D² values for AFC, 1st CI, LMY, LL and LMY/CI were 20.53, 0.71, 4.37, 4.16 and 0.26 respectively. The contribution of AFC to the total D² value was maximum as 68.37 per cent and lowest of LMY/CI as 0.87 per cent. Among the clusters formed on the basis of D² values, IFG, IFJG and R formed one cluster, whereas, FG and FJG formed second cluster. The magnitude of inter-cluster distance was greater than intra-cluster distance.

Keywords: D² analysis, Gir, canonical analysis and traits, Genetic divergence

Introduction

Genetic diversity has been defined as the variety of alleles and genotypes present in a population, and this is reflected in morphological, physiological and behavioral differences between individuals and populations. Recently, loss of genetic diversity within indigenous livestock breeds has been a major concern. Every year many species and breeds of animals become extinct thereby decreasing the biodiversity and genetic variation of populations. Hence, need for sustainable management and conservation strategies for these animal genetic resources are necessary. Previous efforts on the phenotypic characterization of breeds of livestock have been restricted to the use of analysis of variance, whereas the current trend in livestock classification involves the use of multivariate statistical tools. This is because univariate statistical analysis, analyze each variable separately and do not explain how the populations under investigations differ when all measured variables are considered jointly. Multifactorial discriminant analyses have been found to be more suitable in assessing variation within a population and can discriminate different population types when all variables are considered jointly Yakubu et al. (2010). The best way to understand the potential of the available germplasm is by analysing its genetic diversity. For an outstanding breeding programme in the cattle improvement, diversity analysis greatly helps the breeder in the identification and proper choice of parents for specific breeding objectives. The present study was conducted to assess the genetic diversity among different genetic groups of Gir for further breeding programme. The Mahalanobis distance is a very useful statistic in multivariate analysis. It plays a fundamental and important role in statistics and data analysis with multiple measurements. It is one of the most common measures in chemometrics, or indeed multivariate statistics. Extending to multivariate situation, Mahalanobis proposed a distance from the centre of the data (Brereton 2015). Hence, this study was done for genetic group differentiation using Mahalanobis D² statistics by combining information on important characters and to know the divergence through relative contribution of characters in different combinations.
Materials and Methods

Data collection

Data for the present investigation were collected from history sheet of Gir crossbred cattle at Research Cum Development Project on Cattle, MPKV, Rahuri. The data pertained to 158 FG (50% HF+50% Gir), 212 IFG (FG Interse), 107 FJG (50% HF+25% Jersey+25% Gir), 324 IFJG (FG Interse) and 71 R (50% HF+12.50% Jersey+37.50% Gir) crosses of Gir were distributed over a period of 44 years from 1972 to 2015. The traits considered were age at first calving, first calving interval, lactation milk yield, lactation length and lactation milk yield per day of calving interval. Cows with abnormal and incomplete records were excluded from the study. The lactation records of less than 200 days were considered as abnormal and were not included in the data.

Statistical analysis

As the data in the present study were non-orthogonal in nature with unequal subclass numbers, they were subjected to least squares analysis of variance without interactions using different models to examine the effect of genetic as well as non-genetic factors on various traits. Least-squares analysis of variance was conducted for the effect of period and season of birth/calving on age at first calving (AFC), first calving interval (1st CI), lactation milk yield (LMY), lactation length (LL) and lactation milk yield per day of calving interval (LMY/CI) by Harvey (1990). The data were corrected for the significant effects due to period and season of birth/calving as suggested by Gacula et al. (1968). Analysis of variance and covariance for the traits studied was done for testing the differences among the genetic groups for each character by F- test. Analysis of dispersion was done for the simultaneous test of differences between all the traits studied and for all the genetic groups. The significance of dispersion was observed using V-statistics test, Wilk’s criteria were used as described by Rao (1952). The original mean values of the character were transformed to uncorrelated variables by using pivotal condensation method (Rao, 1952). The diversities between different genetic groups based on various characters were estimated by using Mahalanobis (1936) D² Statistics as per following formula:

\[ D^2_p = \sum_{i}^{p} \sum_{j}^{p} W_{ij} (\bar{X}_{i1} - \bar{X}_{i2})(\bar{X}_{j1} - \bar{X}_{j2}) \]

Where,

- \( D^2 \) = D² based on p number of characters
- \( W_{ij} \) = Inverse of estimated variance covariance matrix
- \( \bar{X}_{i1} \) = Mean for iᵗʰ character of genetic group 1
- \( \bar{X}_{i2} \) = Mean for iᵗʰ character of genetic group 2
- \( \bar{X}_{j1} \) = Mean for jᵗʰ character of genetic group 1 and
- \( \bar{X}_{j2} \) = Mean for jᵗʰ character of genetic group 2

The test of significance of D² values was done for any two out of five genetic groups by using statistics suggested by Rao (1952). The percentage contribution of each character to overall diversity was calculated on the basis of D² values and rank basis. For grouping of the genetic groups into various clusters, Tocher’s method (Singh and Choudhary 1985) was used. To verify the clustering pattern canonical analysis (Rao 1948) was done. The data were analysed with the help of two programmes Harvey (1990) and SAS version 9.3 (2013). The CANDISC (SAS 2013) procedure was used to perform multivariate analysis that calculated the Mahalanobis distance between the five genetic groups and to compute canonical discriminant analysis to derive canonical functions. Based on the Mahalanobis distance dendogram was created using PROC CLUSTER (SAS 2013).

Results and Discussion

The analysis of variance indicated that the genetic group difference was significant (P<0.01) for each trait separately. The V-statistics revealed that the influence of genetic group on traits was significant (P<0.01) when all the five traits were considered together. Boujenane (2015) studied on multivariate characterization of Oulmes Zaer and Tidili cattle and reported that the multivariate test for differences between the breeds (Wilks’ Lambda) was significant (P<0.001).

Relative diversities

The differences in D² values among all genetic group combinations were significant except between IFG with R group (Table 1). Mahalanobis squared distance values were maximum between FG and R group as 6.58. It’s due to higher difference in performance between two groups for the traits under study. The maximum D² values among FG and R genetic groups might be due to the fact that cows of FG group calved in early age than cows of R group and also due to higher lactation milk yield in FG than R group. The higher performance of FG (Fₜ) might be due to heterotic effect. The inter-genetic D² value (0.16) between IFG and IFJG group was least. The minimum D² value among these groups is due to similarity in performance of these interse groups.

D² values and character wise D² values with ranks

The genetic diversity among five genetic groups was measured by employing D² statistics and was grouped into two clusters using Tocher’s method. The total D² value contributed by all five traits was 30.03 (Table 2). The per cent contribution of AFC to the total D² value was maximum as 68.37, whereas, LMY/CI contributed minimum as 0.87 in total diversity. These results resembled with
Singh and Bhat (1986) reported in HF × Sahiwal crosses who noted contribution of CI as 2.15 per cent. Lower contribution of FCI than the present results was reported by Singh and Parekh (1989) in Gir crossbreds as zero per cent. The contribution of age at first calving in FJG with R and FG with R genetic group combinations were maximum among all combinations. The highest $D^2$ values in AFC were between FJG and R (4.58) groups which indicate maximum difference in AFC in both groups. This might be due to higher Gir inheritance in R crossbred (37.50%) than FJG crossbred (25.00%). The minimum $D^2$ value for AFC was noticed between FG with FJG genetic group combination. The rank total indicated that LMY/CI had higher rank, whereas, AFC had the lowest rank. This indicates minimum per cent contribution of LMY/CI and maximum per cent contribution of AFC in overall divergence. The present results were in accordance with Sangwan and Singh (1995) observed in Hariana crosses who noted contribution of LMY/CI as zero per cent. The discriminant quantitative variable and Mahalanobis distance between two

**Fig. 1** Dendogram showing cluster formation

**Fig. 2** Cluster diagram depicting intra and inter-cluster distances based on Tocher’s method

**Fig. 3** Clustering pattern plotted with 2-dimensional diagram (canonical method)
N’Dama populations of agro-ecological zones were assessed using discriminant analysis by N’goran et al. (2018). They reported that the Mahalanobis pair wise distance between the two populations of N’Dama was equal to 3.69, which was significant (P<0.05). They further noticed that discriminant analysis showed N’Dama populations were separated following the two agro-ecological zones.

Cluster formation

Clustering of genetic groups under study are presented in Table 2 and Figures 1 to 3. Based on the D^2 values all the genetic groups were grouped into two clusters, signaling the presence of diversity for different traits. The cluster I had the three genetic groups (IFG, IFJG and R) and cluster II had two genetic groups (FG and FJG). Similar cluster pattern was observed by Jagtap et al. (1989) in Gir crossbreds showing FG, FJG and JFG in one cluster. Singh and Parekh (1989) in Gir crossbreds noticed that FG and FJG formed common cluster due to minimum genetic divergence between them.

Intra and inter-cluster distance

Intra and inter cluster divergence values (D^2) by Tocher’s method between and within two clusters are presented in the Table 3 and Figures 1 to 3. The magnitude of inter-cluster distance was greater than intra-cluster distance. This showed distinct difference between grades belonging to different clusters. Lower inter cluster
values between clusters are indication of close relationship and similarity for most traits in the genetic groups hence selection of parents from these clusters are to be avoided. The inter-cluster distance value between cluster I and cluster II was 2.17.

**Canonical analysis**

This method was used to plot the various grades into two dimensional picture by considering the mean values of the canonical variates ($Z_1$ and $Z_2$). The nearness observed between these grades on the basis of $Z_1$ and $Z_2$ values may be considered as more appropriate than distance observed on the basis of $D^2$ values. It’s due to the fact that the first two roots explained 98.91 per cent of variability (Table 4). Canonical discriminant analysis was used by Boujenane (2015) to obtain the function of all body measurements necessary for the separation of Oulmes Zaer and Tidili cattle breeds. He revealed that the squared Mahalanobis distance between breed means was 11.77 ($P<0.001$).

**Conclusions**

$D^2$ analysis offers a reliable technique to estimate the divergence present in the population. This technique measures the forces of differentiation at intra cluster, inter cluster level and further helps in selection of genetically divergent parents for exploitation in hybridization programme based on their superior mean performance. The $D^2$ statistics indicated that reproductive trait age at first calving played an important role than all other traits studied in discriminatory analysis. The contribution of LMY/CI was least among all the traits studied. The IFG, IFJG and R crosses formed one cluster, whereas, FG and FJG formed second cluster with higher intra cluster distance.

**Acknowledgement**

The authors extend their sincere thanks to MPKV, Rahuri for providing the facilities for conducting the present investigation.

**References**

Boujenane I (2015) Multivariate characterization of Oulmes zaer and Tidili cattle using the morphological traits. Iranian J Appl Anim Sci 5: 293-299

Brereton RG (2015) The Mahalanobis distance and its relationship to principal component scores. J Chemometrics 29: 143-145. DOI: 10.1002/cem.2692.

Gacula MC (Jr.), Gaunt SN, Demon RA (Jr.) (1968) Genetic and environmental parameters of milk constituents for five breeds. 1. Effect of herd, year, season and age of cow. J Dairy S 51: 428-437

Harvey WR (1990) Least-squares analysis of data with unequal subclass numbers. ARS H-4, U. S. D.A., Washington

Jagtap DZ, Jaiswal UC, Khanna AS, Bhagat SS (1989) Genetic divergence among different genetic groups of crossbreed dairy cattle. Indian Vet J 66: 1022-1026

Mahalanobis, PC (1936) On the generalized distance in statistics. Proceedings National Institute of Science, India 2: 49-55

N’goran KE, Kouassi NGC, Loukou NGE, Sangare M (2018) Multivariate analysis for morphological characteristics of N’Dama cattle breed in two agro-ecological zones of Côte d’Ivoire. Eur Sci J 14: 602-621

Rao CR (1948) On some problems arising out of discrimination with multiple characters. Sankhya 9: 343-366

Rao CR (1952). Advanced statistical methods in biometric research. John Willey and sons. Inc. New York. P-390

Sangwan ML, Singh B (1995). Genetic divergence on production traits among Hariana and its crosses. Indian J Anim Sci 65: 801-803

Singh B, Bhat PN (1986). Divergence among genetic grades of Sahiwal × Holstein crossbred. Indian Vet J 63: 566-572

Singh RK, Chaudhary BD (1985) Biometric methods in quantitative genetic analysis. Kalyani publishers, Ludhiana, New Delhi, India. P- 303

Singh M and Parekh HKB (1989). Genetic divergence among Gir crosses with exotic dairy breeds. Indian J Anim Sci 59:1411-1415

Statistical Analysis System (SAS) (2013) Statistical Analysis System User’s guide: Release 9.2. SAS Institute, Inc., Cary, NC, US

Yakubu A, Idahor KO, Haruna HS, Wheto M and Amusan S (2010) Multivariate analysis of phenotypic differentiation in Bunaji and Sokoto Gudali cattle. Acta Agriculcarae Slovenica 96: 75-80

---

**Table 4** Mean values of the first 2 canonical variates ($Z_1$ and $Z_2$) for five genetic groups

| Genetic group | $Z_1$ | $Z_2$ |
|---------------|-------|-------|
| FG            | 4.89  | 4.39  |
| IFG           | 7.27  | 3.84  |
| FG            | 5.34  | 3.38  |
| IFJG          | 6.93  | 3.66  |
| R             | 7.45  | 4.24  |

Per cent contribution of first two roots

- Canonical root 1: 92.08
- Canonical root 2: 6.83
- Total: 98.91