

Discriminatory Analysis of Principal Components (DAPC) for Ancestry Estimation in Indian Sheep Populations

Priyanka Swami^{1*}, Jaswant Singh², Mahender Miland Lakeshar³, Radha Rani Swami⁴, M.K. Verma⁵,
Nidhi Verma⁶ and Sunil Kumar Meena⁷

¹M.V.Sc, Department of Animal Genetics and Breeding, College of Veterinary Science & Animal Husbandry, A.N.D.U.A.T., Kumarganj, Ayodhya (Uttar Pradesh), India.

²Professor, Department of Animal Genetics and Breeding, College of Veterinary Science & Animal Husbandry, A.N.D.U.A.T., Kumarganj, Ayodhya (Uttar Pradesh), India.

³Assistant Professor, Department of Veterinary Microbiology, Sri Ganganagar Veterinary College, Tantia University Sri Ganganagar (Rajasthan), India.

⁴Ph.D. Scholar, Department of Animal Genetics and Breeding,

College of Veterinary Science & Animal Husbandry, Bikaner, RAJUVAS (Rajasthan), India.

⁵Assistant Professor, Department of Animal Genetics and Breeding, College of Veterinary Science & Animal Husbandry, A.N.D.U.A.T., Kumarganj, Ayodhya (Uttar Pradesh), India.

⁶Ph.D. Scholar, Department of Animal Nutrition, College of Veterinary Science & Animal Husbandry, A.N.D.U.A.T., Kumarganj, Ayodhya (Uttar Pradesh), India.

⁷Ph.D. Scholar, Department of Animal Genetics and Breeding, College of Veterinary Science & Animal Husbandry, R.A.J.U.V.A., Udaipur (Rajasthan), India.

(Corresponding author: Priyanka Swami*)

(Received: 20 November 2023; Revised: 30 November 2023; Accepted: 17 December 2023; Published: 05 January 2024)
(Published by Research Trend)

ABSTRACT: Discriminatory Analysis of Principal Components (DAPC) is a powerful method used in population genetics to estimate the ancestry and genetic structure of various populations. In this research paper, we applied DAPC to analyze Indian sheep populations using different SNP densities and four distinct methods: Combine, Delta, Information, and FST. The central focus was to evaluate the variation in "a" score values, a crucial parameter in DAPC, across different SNP densities and methods. Our results reveal significant variations in the "a" score values across different SNP densities and methods, indicating the efficiency of DAPC in estimating ancestry and discerning genetic structure in Indian sheep populations. The 20K marker panel consistently demonstrated superior performance, closely resembling the original dataset. Comparative insights from previous studies on sheep populations are also discussed to put our findings in context.

Keywords: SNP, FST, Marker, PCA, Population, Ancestry.

INTRODUCTION

Population genetics, a field that unravels the complexities of genetic diversity within and among populations, relies on sophisticated methodologies to estimate ancestry and decipher genetic structures. Discriminatory Analysis of Principal Components (DAPC) has emerged as a powerful analytical tool, offering insights into the intricate genetic makeup of diverse populations (Haidar *et al.*, 2020).

The use of DAPC in population genetics has proven instrumental in unraveling the evolutionary history and genetic intricacies of different populations. By capturing the principal components of genetic variation and employing discriminant functions, DAPC provides a nuanced understanding of population structure and ancestry (Jombart & Collins 2015).

Understanding the genetic diversity and ancestry of populations is a fundamental aspect of population

genetics. Discriminatory Analysis of Principal Components (DAPC) is a widely used method for assessing population structure and estimating ancestry based on genetic markers (Thia, 2023). Jombart and Ahmed (2011) describe Adegenet as a specialized R-package designed for conducting exploratory analyses on genetic data. This software incorporates a range of tools, covering diverse techniques such as multivariate methods, spatial genetics, and the analysis of single nucleotide polymorphism (SNP) data on a genome-wide scale. In essence, Adegenet serves as a comprehensive toolkit for researchers and analysts working with genetic information, offering capabilities that span various aspects of genetic data exploration and interpretation (Jombart *et al.*, 2010). In this study, we employed DAPC to investigate the genetic structure and ancestry of Indian sheep populations using different SNP densities and methods. In Discriminant Analysis of Principal Components (DAPC), the "a score" value

is a statistic used to assess and measure the quality of the discriminant analysis and the effectiveness of the method in differentiating between populations or clusters within a genetic dataset (Deperi *et al.*, 2018). The "a score" is computed by comparing the variance explained by the discriminant functions (between-group variance) to the total variance in the dataset. It's expressed as a proportion, typically ranging from 0 to 1. A high "a score" (close to 1) indicates that the discriminant functions effectively separate the groups, meaning that there is substantial genetic differentiation between the populations or clusters. In contrast, a low "a score" (closer to 0) suggests that the discriminant functions are less effective at distinguishing the groups, indicating less genetic differentiation (Dal Pra *et al.*, 2023). We this value to assess the level of genetic differentiation and population structure within the data. Higher "a scores" are indicative of strong population differentiation, while lower values suggest weaker differentiation or greater genetic admixture between groups (Jin *et al.*, 2021). The aim of this research was to assess the performance of DAPC in estimating ancestry in Indian sheep populations and identify the optimal marker panel for this purpose.

MATERIALS AND METHODS

The genotyping information was sourced from publicly accessible databases, consortia, and/or datasets provided in existing scientific literature. The data must be specific to the Ovine50KSNP Beadchip density and obtained from Indian Sheep breeds (Changthangi, Tibetan, Deccani, Garole), Asian Sheep breeds (Bangladeshi Garole, Bangladeshi East), and Exotic Sheep breeds (Rambouillet and Australian Merino). We conducted our study on four datasets (A, B, C, and D), each representing Indian sheep populations with different SNP densities. For each dataset, we applied DAPC using four distinct methods: Combine, Delta, Information, and FST. The "a" score value, a measure of the effectiveness of DAPC in estimating ancestry, was used to evaluate the results. The discriminatory power of different lower density SNP panels generated from earlier steps was assessed along with their

suitability and stability on other similar data. The assessment was done using discriminatory analysis of principal components (DAPC) methodology under "adegenet" package in R-programming environment.

RESULTS AND DISCUSSION

Our analysis yielded a range of "a" score values across different SNP densities and methods. The 20K marker panel consistently showed superior performance, closely resembling the original dataset. In dataset A, using the Delta method at LDP 20K, the highest "a" score value of 0.8197 was achieved, while in dataset B, the FST method at LDP 1K yielded the highest "a" score value of 0.6878. In datasets C and D, the Delta method at LDP 20K produced the highest "a" score values of 0.8074 and 0.8197, respectively. The results are summarized in Tables 1, 2, 3, 4. Our findings indicate that the 20K marker panel obtained through DAPC efficiently estimated the ancestry level in Indian sheep populations, consistently outperforming other SNP densities. These results are in line with previous studies in sheep population genetics. Moradi *et al.* (2012) identified a set of 201 SNPs for distinguishing Baluchi sheep breeds in Italy. Similarly, According to Chhotaray *et al.* (2020) the markers selected through DAPC, it was observed that the inheritance level of Sahiwal in Frieswal population was 38%, 38.26%, and 39.1% when estimated with 500, 1000 and 2000 markers, respectively. While, inheritance from Holstein-Friesian breed was 62%, 61.74%, and 60.89% estimated with 500, 1000, and 2000 markers, respectively.

Dimauro *et al.* (2015) used statistical techniques to select informative markers for discriminating Italian sheep populations, finding that a panel of 108 markers could distinguish 21 different sheep populations and their geographic areas of origin. Tortereau *et al.* (2017) designed 249 SNPs for population assignment in thirty French sheep breeds. Our DAPC analysis reveals that Indian sheep breeds cluster separately from other Asian and exotic sheep breeds, indicating a distinct genetic identity (Jin *et al.*, 2021).

Table 1: The "a" score value calculated by DAPC using various methods and the densities of SNPs through data set A.

| Densities | Combine Methods | Delta Methods | Info. Method | F _{ST} Method |
|-----------|-----------------|---------------|--------------|------------------------|
| 1K | 0.6328 | 0.6322 | 0.6079 | 0.6612 |
| 3K | 0.6324 | 0.6845 | 0.6630 | 0.6070 |
| 5K | 0.6756 | 0.7170 | 0.6961 | 0.6668 |
| 10K | 0.6338 | 0.6944 | 0.6922 | 0.6742 |
| 20K | 0.6999 | 0.6937 | 0.7137 | 0.6843 |

Table 2: The "a" score value calculated by DAPC using various methods and the densities of SNPs through data set B.

| Densities | Combine Methods | Delta Methods | Info. Method | F _{ST} Method |
|-----------|-----------------|---------------|--------------|------------------------|
| 1K | 0.6835 | 0.6217 | 0.6372 | 0.6878 |
| 3K | 0.5967 | 0.5794 | 0.6279 | 0.6427 |
| 5K | 0.6270 | 0.6365 | 0.6498 | 0.6672 |
| 10K | 0.6450 | 0.6310 | 0.6132 | 0.6955 |
| 20K | 0.6791 | 0.6652 | 0.6808 | 0.6162 |

Table 3: The "a" score value calculated by DAPC using various methods and the Densities of SNPs through data set C.

| Densities | Combine Methods | Delta Methods | Info. Method | F _{ST} Method |
|-----------|-----------------|---------------|--------------|------------------------|
| 1K | 0.6274 | 0.6118 | 0.6141 | 0.6579 |
| 3K | 0.6141 | 0.7041 | 0.6842 | 0.6617 |
| 5K | 0.6315 | 0.7540 | 0.7286 | 0.6929 |
| 10K | 0.7470 | 0.7855 | 0.7919 | 0.7273 |
| 20K | 0.8074 | 0.8197 | 0.8030 | 0.7669 |

Table 4: The "a" score value calculated by DAPC using various methods and the densities of SNPs through data set D.

| Densities | Combine Methods | Delta Methods | Info. Method | F _{ST} Method |
|-----------|-----------------|---------------|--------------|------------------------|
| 1K | 0.5883 | 0.6570 | 0.6494 | 0.6203 |
| 3K | 0.6829 | 0.7440 | 0.6938 | 0.6113 |
| 5K | 0.7164 | 0.7381 | 0.7842 | 0.6535 |
| 10K | 0.7641 | 0.7732 | 0.7663 | 0.7427 |
| 20K | 0.7964 | 0.8018 | 0.7962 | 0.7818 |

CONCLUSIONS

Discriminatory Analysis of Principal Components (DAPC) is a valuable tool for estimating ancestry and assessing population structure in Indian sheep populations. The 20K marker panel consistently produced the best results, demonstrating its efficiency in capturing the genetic diversity of Indian sheep breeds. Our findings provide important insights into the genetic structure of Indian sheep populations and contribute to the broader field of population genetics. Future research can further explore the genetic diversity of Indian sheep populations using advanced genomic techniques and larger datasets. Additionally, investigating the functional implications of genetic diversity and ancestry in these populations can offer valuable insights for sheep breeding and conservation efforts.

Acknowledgment. The author would like to express gratitude to the Dean, College of Veterinary Science and Animal Husbandry, A.N.D.U.A. & T., Kumarganj, Ayodhya (Uttar Pradesh) for providing funds and necessary support for the research.

Conflict of Interest. None.

REFERENCES

- Chhotaray, S., Panigrahi, M., Pal, D., Ahmad, S. F., Bhushan, B., Gaur, G. K., & Singh, R. K. (2020). Ancestry informative markers derived from discriminant analysis of principal components provide important insights into the composition of crossbred cattle. *Genomics*, *112*(2), 1726-1733.
- Dal Pra, A., Bozzi, R., Parrini, S., Immovilli, A., Davolio, R., Ruozzi, F., & Fabbri, M. C. (2023). Discriminant analysis as a tool to classify farm hay in dairy farms. *Plos one*, *18*(11), e0294468.
- Deperi, S. I., Tagliotti, M. E., Bedogni, M. C., Manrique-Carpintero, N. C., Coombs, J., Zhang, R., & Huarte, M. A. (2018). Discriminant analysis of principal

components and pedigree assessment of genetic diversity and population structure in a tetraploid potato panel using SNPs. *Plos one*, *13*(3), e0194398.

- Haidar, O., Ball, S., & Barrett-Jolley, R. (2020). Discriminant Analysis of Principle Component analyses of Physiological Data. *bioRxiv*, 2020-01.
- Jin, D., Henry, P., Shan, J., & Chen, J. (2021). Classification of cannabis strains in the Canadian market with discriminant analysis of principal components using genome-wide single nucleotide polymorphisms. *Plos one*, *16*(6), e0253387.
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070-3071.
- Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, *11*(1), 1-15.
- Jombart, T., & Collins, C. (2015). A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0. 0. London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling.
- Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G., & McEwan, J. C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC genetics*, *13*(1), 1-15.
- Thia, J. A. (2023). Guidelines for standardizing the application of discriminant analysis of principal components to genotype data. *Molecular Ecology Resources*, *23*(3), 523-538.
- Tortoreau, F., Moreno, C. R., Tosser-Klopp, G., Servin, B., & Raoul, J. (2017). Development of a SNP panel dedicated to parentage assignment in French sheep populations. *BMC genetics*, *18*(1), 1-11.
- Wilkinson, S., Wiener, P., Archibald, A. L., Law, A., Schnabel, R. D., McKay, S. D., & Ogden, R. (2011). Evaluation of approaches for identifying population informative markers from high density SNP chips. *BMC genetics*, *12*, 1-14.

How to cite this article: Priyanka Swami, Jaswant Singh, Mahender Miland Lakeshar, Radha Rani Swami, M.K. Verma, Nidhi Verma and Sunil Kumar Meena (2024). Discriminatory Analysis of Principal Components (DAPC) for Ancestry Estimation in Indian Sheep Populations. *Biological Forum – An International Journal*, *16*(1): 41-43.