Phylogeny and genetic diversity analysis in Chhattisgarhi, Chilika, and Kalahandi buffaloes of Central and Eastern India through the High-density SNP Array

Authors: *Swati* Rulyan[1,2], *Udayabanu* Malairaman[*1], *R.K.* Vijh[*2]

[1]Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Solan, 173234, Himachal Pradesh, India

[2]Department of Animal Genetics, ICAR NBAGR, Karnal, 132001, Haryana, India
Corresponding author(s): E-mail(s): udayabanu.m@juitsolan.in; rameshkvijh@gmail.com

Abstract. The genetic diversity and evolutionary pattern of the Indian buffalo groups can only be understood to come up with the effective conservation and breeding measures. This paper investigates the genomic composition and population structure of three regional buffalo populations, namely Chhattisgarhi (CG), Chilika (CH) and Kalahandi (KH) by high-density SNP genotyping. The analyses showed the general low-to-moderate population genetic variation. The pairwise F_{ST} estimates showed a moderate level of differentiation between Chilika and other two groups (0.044-0.049) but strong level of similarity between Chhattisgarhi and Kalahandi buffaloes (0.010). Principal component analysis (PCA) revealed that Chilika buffaloes had a higher number of internal diversity, whereas the Chhattisgarhi and Kalahandi buffaloes displayed a close clustering pattern, which corresponded to their proximity in the geographical location and the exchange of genes. Three genetically consistent groups were found to have a shared ancestry with phylogenetic relationships and identity-by-state (IBS) heatmaps Patterns of linkage disequilibrium suggest a moderate level of genetic diversity with indications of a recent decline in the effective population size. These findings would highlight the need to keep genetic uniqueness and adaptability to preserve diversity and make the buffalo genetic resources more productive, which would sustain the long-term viability of buffalo genetic resources in India.

1 Introduction

The Asian water buffalo (*Bubalus bubalis*) is one of the economically and culturally significant livestock species in South and Southeast Asia, the source of the milk, meat, draught power and organic manure contributing to the livelihood of millions of rural communities in the area [1, 2]. India, which hosts about 109 million buffaloes and nearly half of the total world buffalo population thanks to its position signifies the largest obtainable buffalo genetic assets at the globe, as per the 20th Livestock Census [3, 4]. Nevertheless, the pressure of such stakes on the sovereignty of most native Indian buffalo populations has been poorly recorded, especially the populations located in geographically remote or under-represented areas [5, 6].The Indian buffaloes are broadly divided into two major categories; riverine and swamp. Boons of ecological pressures, natural selection and mediated breeding choices, riverine buffaloes dominant on the Indian subcontinent have a history of over 5,000 years of domestication [6-8]. These actions have produced a tremendous amount of phenotypic diversity, which has led to ecologically segregated regional populations to fit a variety of agro-climatic regions [9, 10]. Nonetheless, the genome-wide foundation of this diversity is under-characterized regarding a number of domestic breeds on a local level.

Indian Buffaloes in the east and central India such as the Chhattisgarhi, Chilika and Kalahandi groups are some of such unexplored genetic resources. Chhattisgarhi buffalo is a descendant of mammals raised in the Chhattisgarh plains, which do not lose some of their morphic characteristics [11]. Chilika buffalo, a species of the marshy and lagoon ecosystem of the Chilika Lake in Odisha, is adapted to wetland environments, whereas the Kalahandi buffalo, a species of the semi-arid areas of Odisha, is famous because of its environmental adaptation and high resistance to diseases [12, 13]. Majority of previous studies on these populations have only characterized their phenotypes or of a genetic analysis using microsatellites providing otherwise cursory information on the structure of the genetics and evolutionary backgrounds of these populations [14, 15].

Although the microsatellite markers have also offered preliminarily based analysis of the genetic diversity in Indian buffaloes, they are not capable of observing the fine-scale structure, demographic histories, and adaptive genomic variation [10], [21]. High-density SNP arrays have transformed the genomics of livestock because they can now analyze high-resolution data on diversity and population stratification, admixture and signatures of selection on a massive scale [17], [18]. Specifically, the Axiom(r) Buffalo Genotyping Array and some of its region-specific offshoots have demonstrated usefulness as genome-explicit models of genetic organization as well as an instrument to map phylogeny, assess linkage disequilibrium (LD) and effective population size (Ne) in the buffalo populations of the world [19-20].

We have used in the study a high density Affymetrix SNP chip of 621,789 markers designed in swamp buffalo (Indian Patent Application 202111009710, obtained through the Indian Patent Office Public Search Portal: https://iprsearch.ipindia.gov.in/PublicSearch) to carry out an integrative genomic study of the Chhattisgarhi, Chilika, and Kalahandi buffalo populations. We aimed to (1) approximate the genome-wide parameters of diversity in each population (2) measure the genetic differentiation by comparing populations using FST values, (3) infer phylogenetic relationships and patterns of subdivision (4) daydream the population structure using the principal component analysis (PCA) (5) use the LD decay not only to estimate the effective population sizes (Ne) but also to assess historical demographic trends. The analyses were conducted using TASSEL as well as the supplementary statistical pipelines. This requires a thorough characterization of the complete genomics of these native buffalo populations in order to make informed conservation and sustainable use. Their

genomic structure, their diversity and evolutionary connections will inform the understanding of buffalo domestication as well as adaptive evolution in the Indian subcontinent and will supplement target breeding and conservation initiatives in the effort to preserve unique adaptive variants [21-22]. As Chhattisgarhi, Chilika, and Kalahandi buffaloes are accepted to be genetically resourceful and reasonably curious population, their genome-wide examination is crucial towards the secure production as well as supervision of the agro-climatic conditions in India with historically precious buffalo heritage in the near future.

2 Materials and Methods

2.1 Blood Sample Collection and DNA Extraction

A total of 82 buffaloes including Chhattisgarhi (n = 22) of the native tracts of Chhattisgarh, Chilika (n = 36) of Cuttack, Ganjam, Puri, and Khurda districts of Odisha, and Kalahandi (n = 24) of the Kalahandi district of Odisha, were sampled with blood. The unrelated animals that had healthy backgrounds and known pedigrees were used as samples; they were procured through local farmers and breed societies so as to have the purebred populations. The Roche DNA extraction kit was used to extract the genomic DNA, which was quantified in a Multiskan Sky spectrophotometer and made to a criterion concentration of 50 ng/mL and subjected to genotyping.

2.2 Genotyping and Quality Control

The genotyping of DNA samples was done using the high- density SNP chip which was developed in house. SNP calling and clustering of the raw intensity files (.CEL) were done using Affymetrix Power Tools. The quality control filtering was conducted using PLINK v1.90 which filtered out SNPs with a minor allele frequency (MAF) of less than 0.05, which had a call rate of less than 90 percent, Hardy-Weinberg deviation of less than 10 -6 as well as SNPs with missing genotype data greater than 5 percent. After carrying out the quality control process, 85,627 SNPs were retained for downstream analysis. The Variant call format (VCF) files of the three buffalo populations were combined to form a single dataset which was used in downstream analyses to determine principal component analysis (PCA), fixation index (Fst), and linkage disequilibrium (LD) analyses.

2.3 Principal Component Analysis (PCA)

EIGENSOFT v7.2.1 SmartPCA was used to do Principal Component Analysis (PCA) [23]. PLINK format files (bed/bim/fam) of input genotypes were used. The parameters were lsqproject= TRUE, numeric outlier = 2 and k=10 principal components (PCs). Besides, PCA and outliers were identified with PCAdapt (R package). The genotype data were transformed, and top five PCs were designated. The Mahalanobis distance was used to determine the SNPs that may be under selection, and the significance was set at q-value of < 0.05 through the Benjamini-Hochberg correction. Scree plots have been plotted to determine the percentage of variance accounted by the leading PCs.

2.4 Phylogenetic Analysis and F_{ST} Calculation

The phylogenetic tree was built with a phylogeny software called Fig Tree 1.4.4.exe [24] using two different layouts. On per-breed allele frequencies, global FST was obtained using the hierfstat package, and pairwise FST was obtained using the hierfstat package.

2.5 Linkage Disequilibrium and Distance Matrix

TASSEL v5.2.54 was used in analyzing the linkage disequilibrium (LD) decay [25]. All SNP pairs of the genome were calculated as pairwise r 2 and placed into physical distance bins (1-500 kb). A Reynolds genetic distance heatmap was also created of all individuals and breeds and visualized as a heatmap. TASSEL parameters were LD score calculation, LD window size = 50 and the estimation of r2 by the sliding-window technique. To create the distance matrix, we have used the SR plot [26].

3 Results

3.1 Pairwise F_{ST} differentiation among three buffalo breeds

There is a very close resemblance between Chhattisgarhi and Kalahandi buffaloes (0.010), implying that they have a common ancestor or mixed descent. Chilika demonstrates medium disjunction with both (0.044-0.049), and it is not very clear. The three are not very distant apart genetically and as expected their pattern of overlap in PCA matches.

	Chilika buffalo	Chhattisgarhi buffalo	Kalahandi buffalo
Chilika buffalo	NA	0.048852	0.044275
Chhattisgarhi buffalo	0.048852	NA	0.010372
Kalahandi buffalo	0.044275	0.010372	NA

Table 1: Pairwise genetic difference F_{ST} among three buffalo populations (Chilika, Chhattisgarhi, and Kalahandi buffaloes).

3.2 Principal Component Analysis

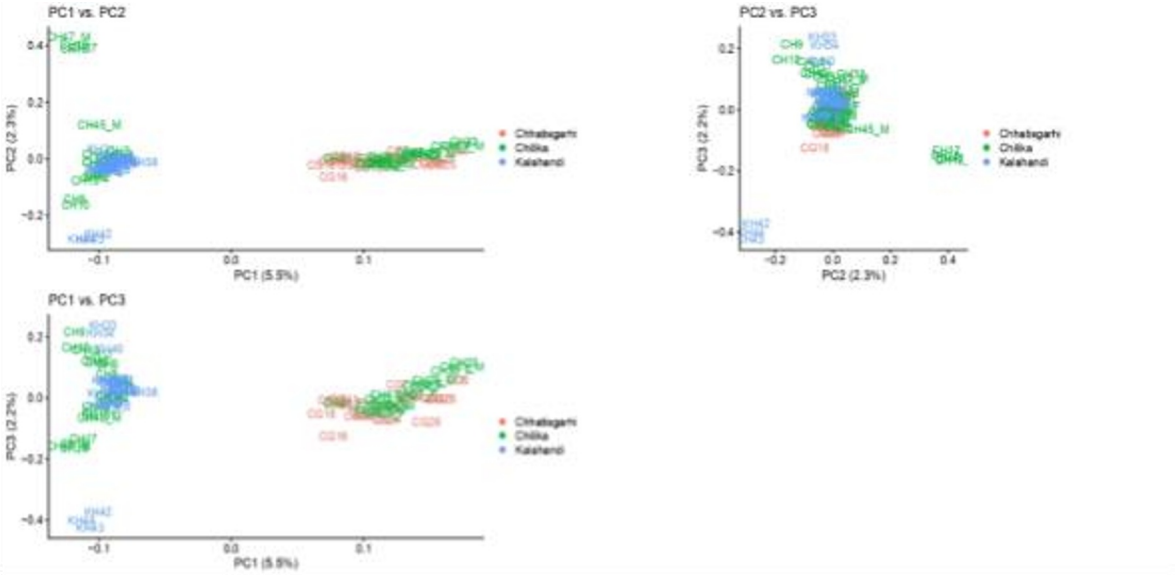

Fig.1 The grid plot displays pairwise combinations of the first three principal components; PC1 vs PC2 (top-left), PC2 vs PC3 (top-right), and PC1 vs PC3 (bottom-left) illustrating the genetic structure among Chhattisgarhi (red), Chilika (green), and Kalahandi (blue) buffaloes.

The Principal Component Analysis (PCA) in Fig. 1 shows the genetic relationship of Chhattisgarhi, Chilika, and Kalahandi buffaloes based on the first three principal components. Chhattisgarhi and Chilika buffaloes show partial proximity or overlap in their positions in the PC1 vs PC2 plot (top-left) showing moderate genetic relatedness between the two groups. On the other hand, Kalahandi buffaloes are strictly segregated on the

negative of PC1 indicating genetic diversity. This division is supported by the PC1 vs PC3 plot (bottom-left), in the sense that the Kalahandi buffaloes, in addition to being pointedly distinct with respect to the other groups, are also more variably clustered. Chhattisgarhi and Chilika buffaloes are found to be more overlapped in the PC2 vs PC3 plot (top-right), which indicates that more differentiation between them is lower, although Kalahandi buffaloes retain their own cluster. On the whole, the outcomes of the PCA demonstrate that Kalahandi and Chhattisgarhi buffalos are genetically diverse, and Chilika and Kalahandi are slightly close together along PC1 and PC2 respectively. The values PC1(5.5%), PC2(2.3%), and PC3(2.2) indicate that they are highly distinct genetically with the little or zero gene flow. . The low proportion of variance in the first three principal components suggest us considerable unexplained genetic variation and also the complex structure of genetic diversity in these populations. This could be also due to multiple factors such as historical gene flowadaptation to local environments and other population specific factors that may not be captured by the first few principal components.

The cumulative variance and variance was also done using R scripts. The circular heatmap in Fig. 2 represents the cumulative variance and variance among the principal components.

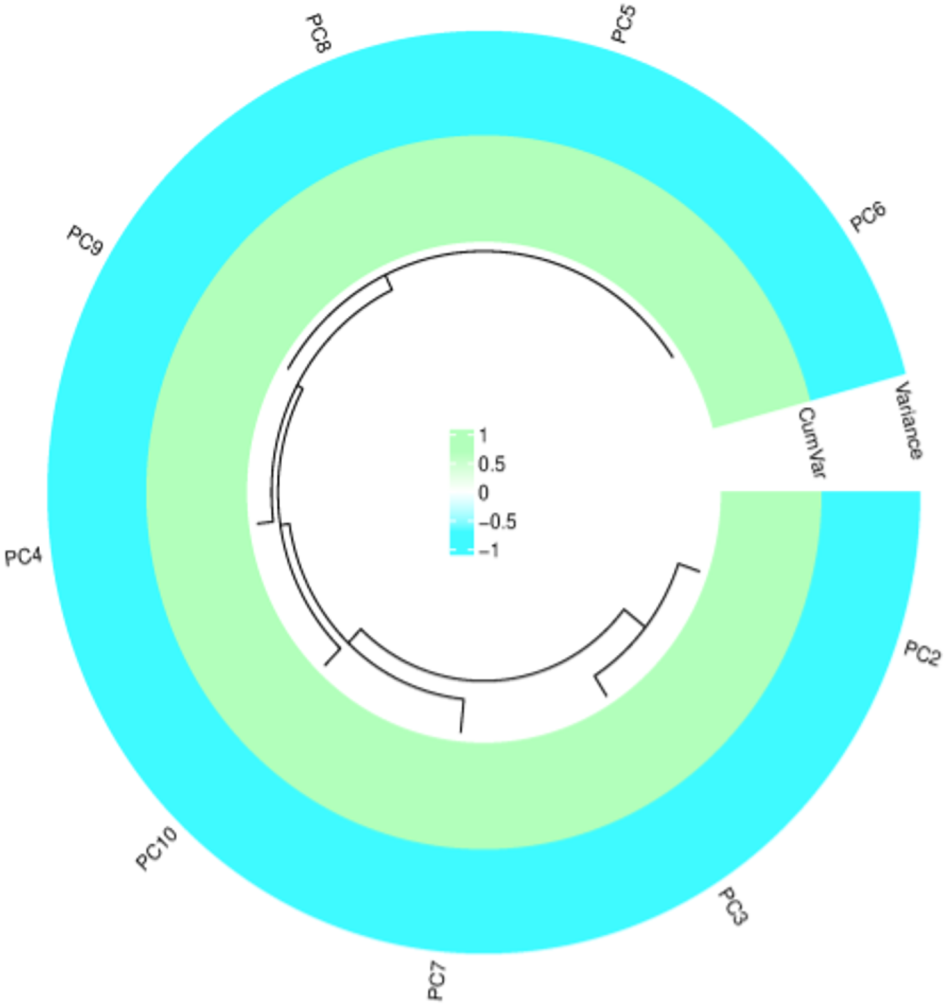

Fig. 2. Circular heatmap representing the cumulative Variance and variance among the ten principal components. The inner side of the circle (light green in colour) represents the Cumulative variance and outer ring represents the Variance of Principal Components.

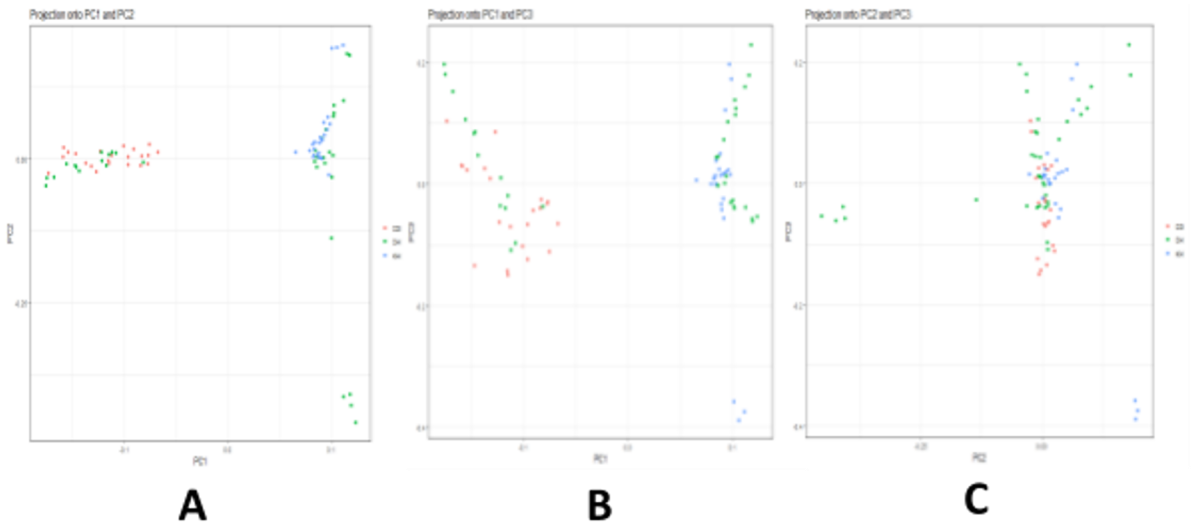

Fig. 3. shows the plots of PCA in A, B, and C

Principal Component Analysis (PCA) demonstrates that there is distinct genetic separation of the three populations of buffalo; Chhattisgarhi, Chilika, and Kalahandi. In PC1 vs PC2 plot (A), Chhattisgarhi and Chilika buffaloes exhibit a partially overlapping cluster, thus, revealing moderate genetic similarity, and Kalahandi buffaloes are segregated on PC1, thus demonstrating strong genetic divergence. The PC1 vs PC3 plot
(B) again indicates the clear genetic set up of the Kalahandi buffaloes which is also more dispersed indicating more internal genetic variation. Chhattisgarhi and Chilika showed greater inside overlapping which agrees with reduced genetic variation among them in PC2 vs PC3 curve (C), and the Kalahandi buffaloes still maintained a distinct cluster. All this implies that the genetics of the Kalahandi buffaloes and the Chhattisgarhi buffaloes are similar, as the genetic relatedness of Chilika buffaloes with the former is greater.

3.3 QQ plot, Statistics distribution and Chi square analysis

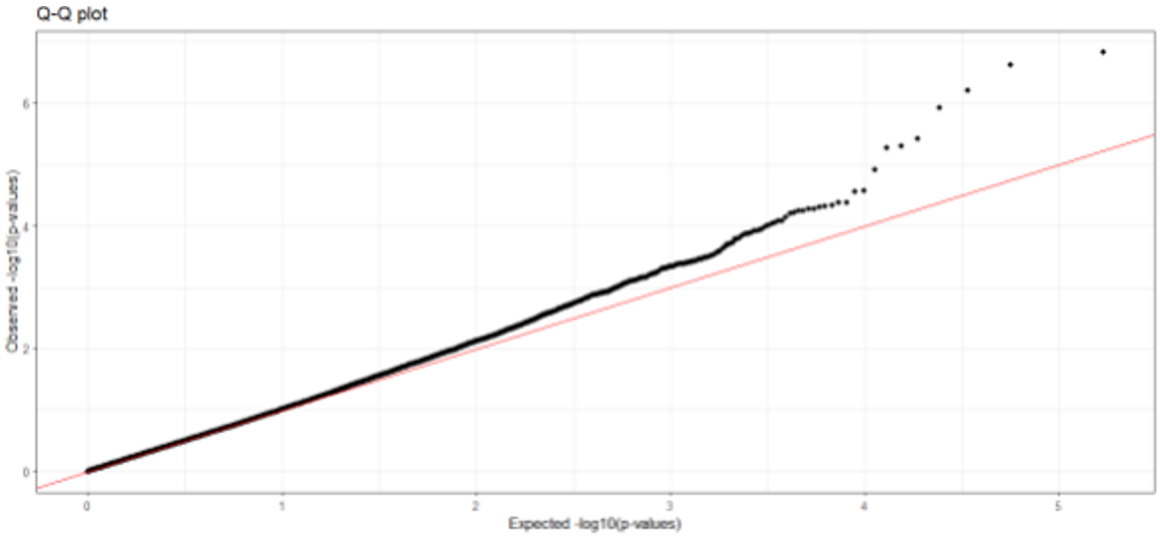

Fig. 4 shows the Q-Q plot in three buffalo breeds. The X-axis shows the expected -log10 (p values) and the Y-axis shows the Observed -log10 (p values).

The Q-Q plot presented in Fig. 4 represents the distribution between observed and expected -log10(p) values of the SNP-based association results in the three buffalo populations. The majority of SNPs are aligned with the

diagonal red reference line meaning that the test statistics observed are in agreement with the null expectation of analysis and this shows that there is little genomic inflation or systematic bias observed in the data. Along the distribution, we see that there is a deviation at the end most part of the tail, where there is a sub-set of SNPs with larger than anticipated -log10(p) values. Such positive deviations raise the possibility of likely true signals of genetic differentiation or loci selection among the three population. Generally, the plot validates good quality of the association results as well as indicating SNPs that can reflect meaningful biological variation.

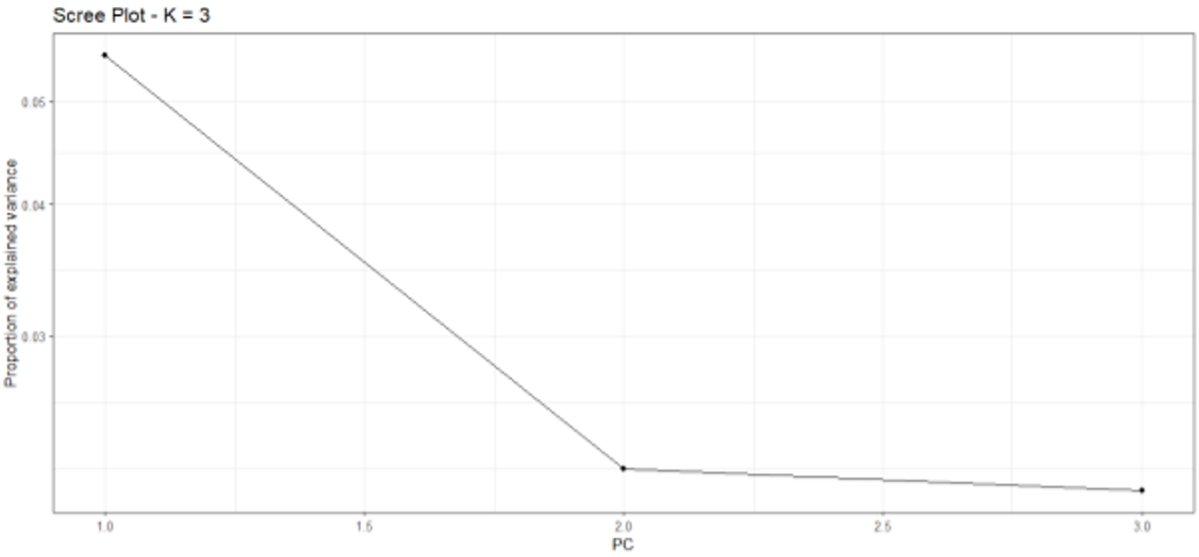

Fig. 5. shows the Scree plot at K=3. The X-axis shows the number of principal components, and the Y-axis shows the proportion of explained variance.

The scree plot shown in Fig. 5 shows the percentage of variance accounted by the top-three principal components (PCs) of SNP data on a genome scale. The contribution of the highest proportion of the explained variance is made by PC1 (~5.5%), and then by PC2 (~2.3) and PC3 (~2.1). The sharp decline between PC1 and PC2 shows that the first component is taking the most significant genetic difference among the 3 populations of buffalo, the second component is taking less and less future. This tendency indicates the existence of a prevailing genetic design, which is taken by PC1, and other smaller stratifications are indicated by PC2 and PC3. The post-PC1 flattening is an indication that the initial few of the components are informative in describing the population differentiation.

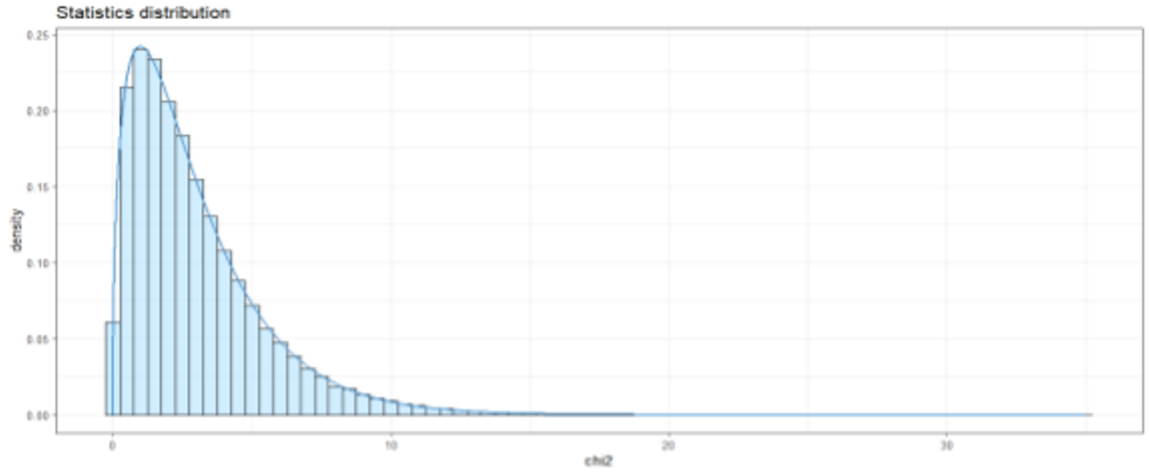

Fig. 6. shows the distribution of statistics for three buffalo breeds. The x-axis shows the chi2 values, and the Y-axis shows the density.

The chi-square statistics distribution presented in Fig. 6 indicates distribution of the chi-square test statistics

across the three buffalo populations globally as a consequence of SNP-based differentiation analyses. The histogram shows that the distribution is skewed to the right with most SNPs tending to assume lower chi-square values which suggests that most loci show little deviation form neutrality or poor population differentiation. The shrinking curve on the chi-square as the density increases to large chi-square values represents the shrinking number of SNPs revealing strong signals. A low number of SNPs with large chi-square values are seen and these indicate that maybe there could be loci that are undergoing some form of selection or causing major genetic divergence of the populations. On the whole, the distribution validates the claim that the majority SNPs are neutral but a sub-population of SNPs could portray biologically significant variation as pertains to population structure.

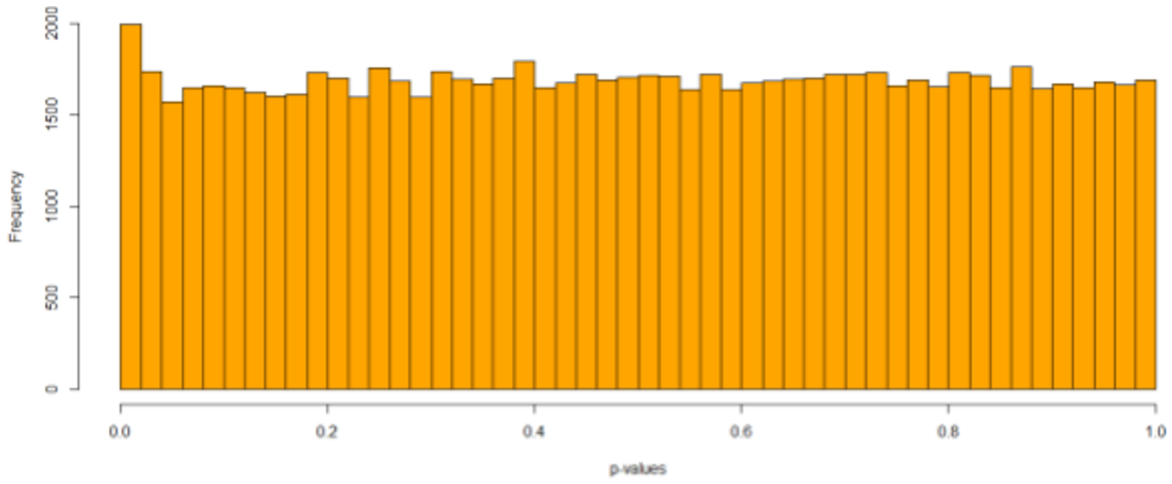

Fig. 7. shows the histogram plot for three buffalo breeds. The x-axis shows the p-values, and the Y-axis shows the Frequency.

The p-value distribution in Fig. 7 shows the occurrence of p-values throughout the genome as a result of SNP-association or SNP-identification tests in three buffalo populations. The histogram shows the relatively homogenous distribution of p-values along the 0-1-scale that would be expected under the null hypothesis of no correlation between most loci. Even distribution of bars indicates that no systematic bias, inflation or deflation of the statistical tests occurs, indicating us that the dataset used has been well-calibrated and has no significant confounding factors like population stratification or genotyping error. A very small fraction of p-values are found towards the lower end indicating that only a small number of SNPs can be taking the effect of differentiation or having the possibility of selection.

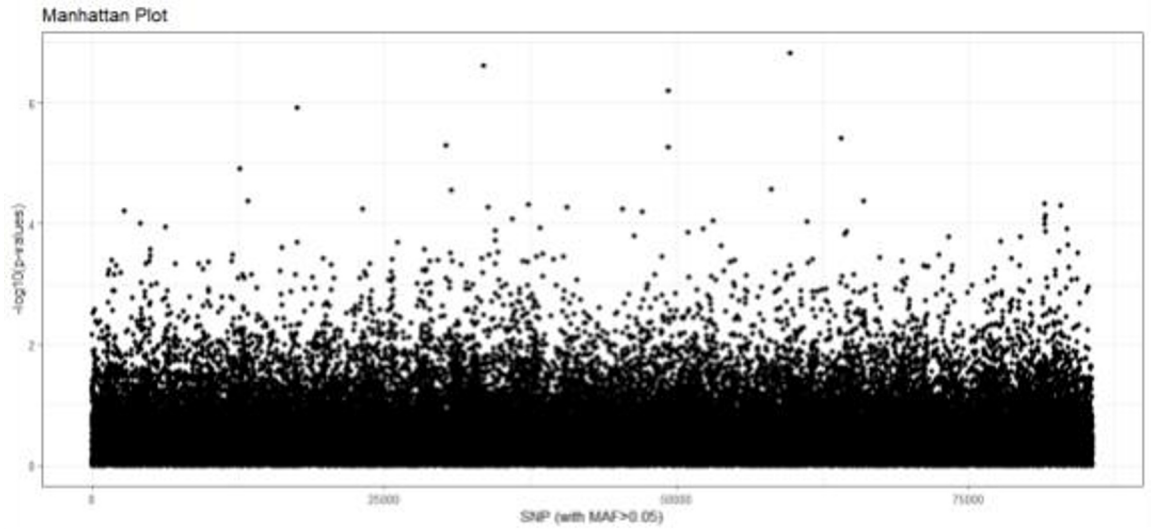

Fig. 8. shows the Manhattan plot for three buffalo breeds. The x-axis shows the SNP with MAF>0.05 values, and the Y-axis shows the –log10 (p-values).

The plot in Fig. 8 presents the Manhattan diagram of the distribution of all the genome-wide SNP association test statistics of all SNPs (MAF > 0.05) analyzed in the three buffalo populations. Every point shows an individual SNP and that is marked by genomic position (x-axis) and strength of the statistical significance/p-value in -log10 (y-axis). Most of the SNPs are around the baseline i.e. there are weak or not significant associations which is in line with a substantial neutral genomic background. The proportion of SNPs with -log10(p -value) values above a moderate level is small, and indicates that an extremely small number of loci have a strong signal of genetic differentiation or possible selection. The lack of strong peaks also confirms the fact that the populations are very genetically similar with the relatively few SNPs being the only cause of divergence.

3.4 Phylogenetic trees, Distance matrix and LD plots

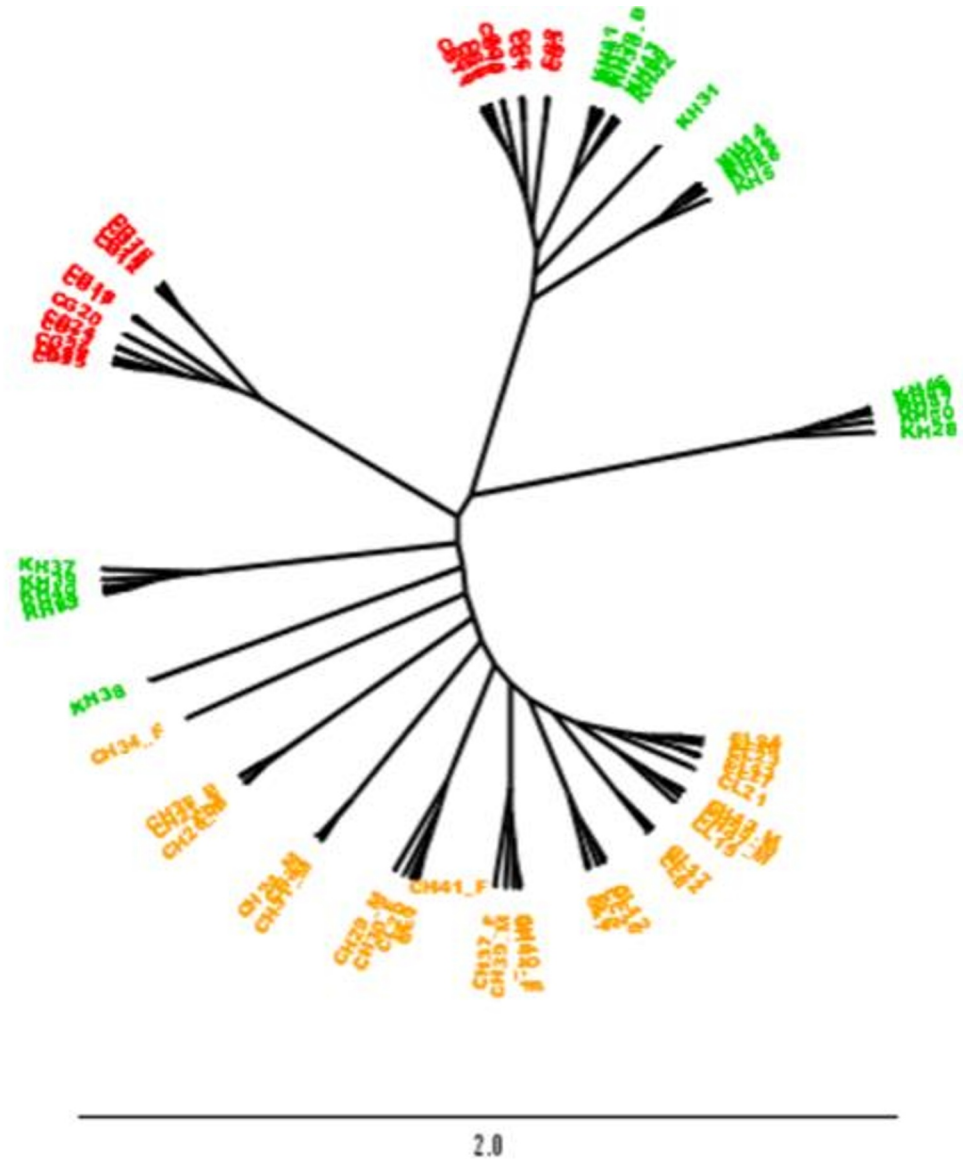

Fig. 9. shows the Radial tree Phylogenetic tree for three buffalo breeds.

The phylogenetic trees shown in Fig. 9 and Fig. 10 shows the genetic relationship between the three native buffalo tracts Chhattisgarhi (CG), Chilika (CL and CH), and Kalahandi (KH), using both the unrooted tree (top)

and circular dendrogram (bottom). Under the unrooted topology, individuals are clearly divided by their population labels as there is an obvious genetic structuring and little mixing between the two populations. The given separation is also justified with the help of the circular dendrogram, the branching structure of which allows seeing three distinct clades that are CG (red), KH (green) and CL (orange). The Chhattisgarhi animals are tightly clustered, which denotes the high level of within-population similarity, whereas the Kalahandi group is tightly clustered with the short length of the branch, which is an indication of low genetic departure among individuals. Chilika population also has a more diffused pattern yet it is a coherent clade pointing out to moderate intra population diversities. These phylogenetic reconstructions taken together indicate a steady pattern of population-specific clustering consistent with the genetic difference between the three populations of buffalos as well as a way to pin down fine variation within the groups.

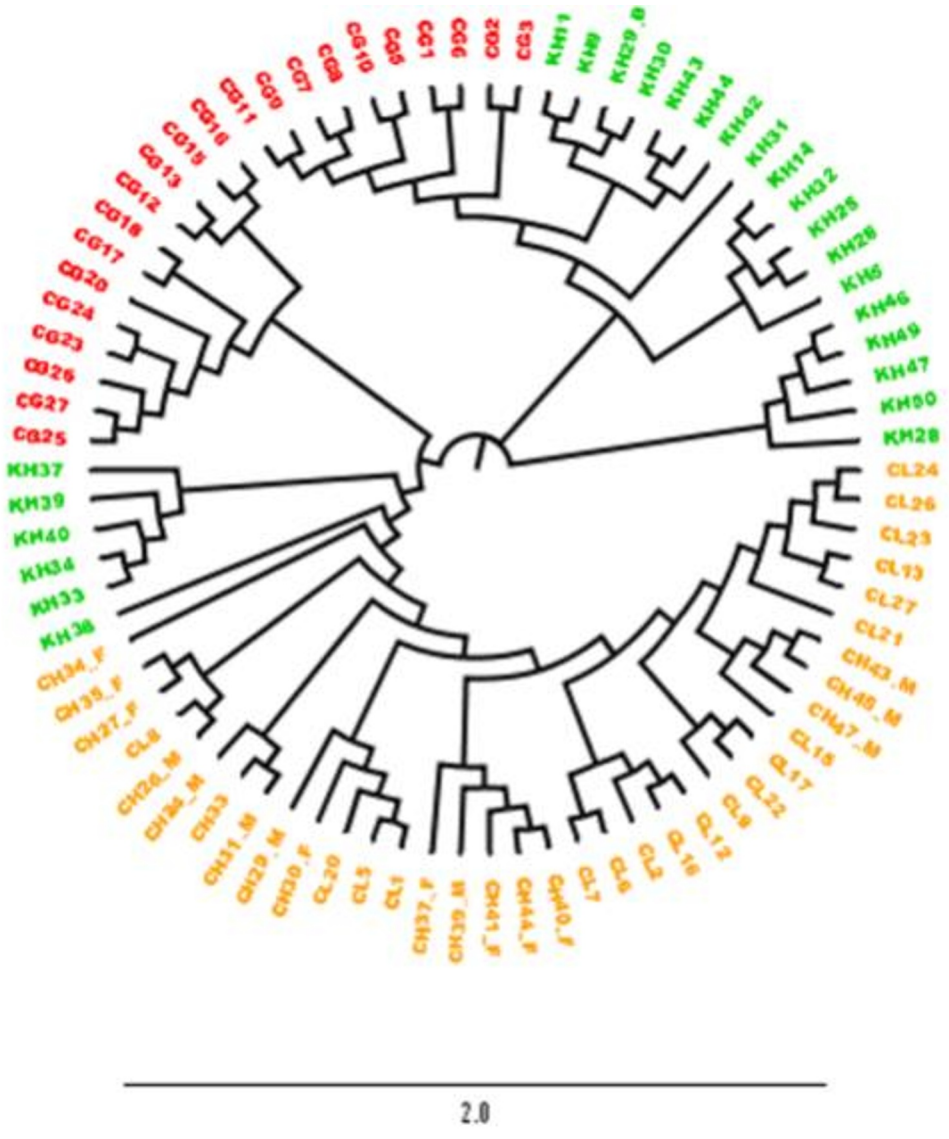

Fig. 10. shows the Circular tree Phylogenetic tree for three buffalo breeds.

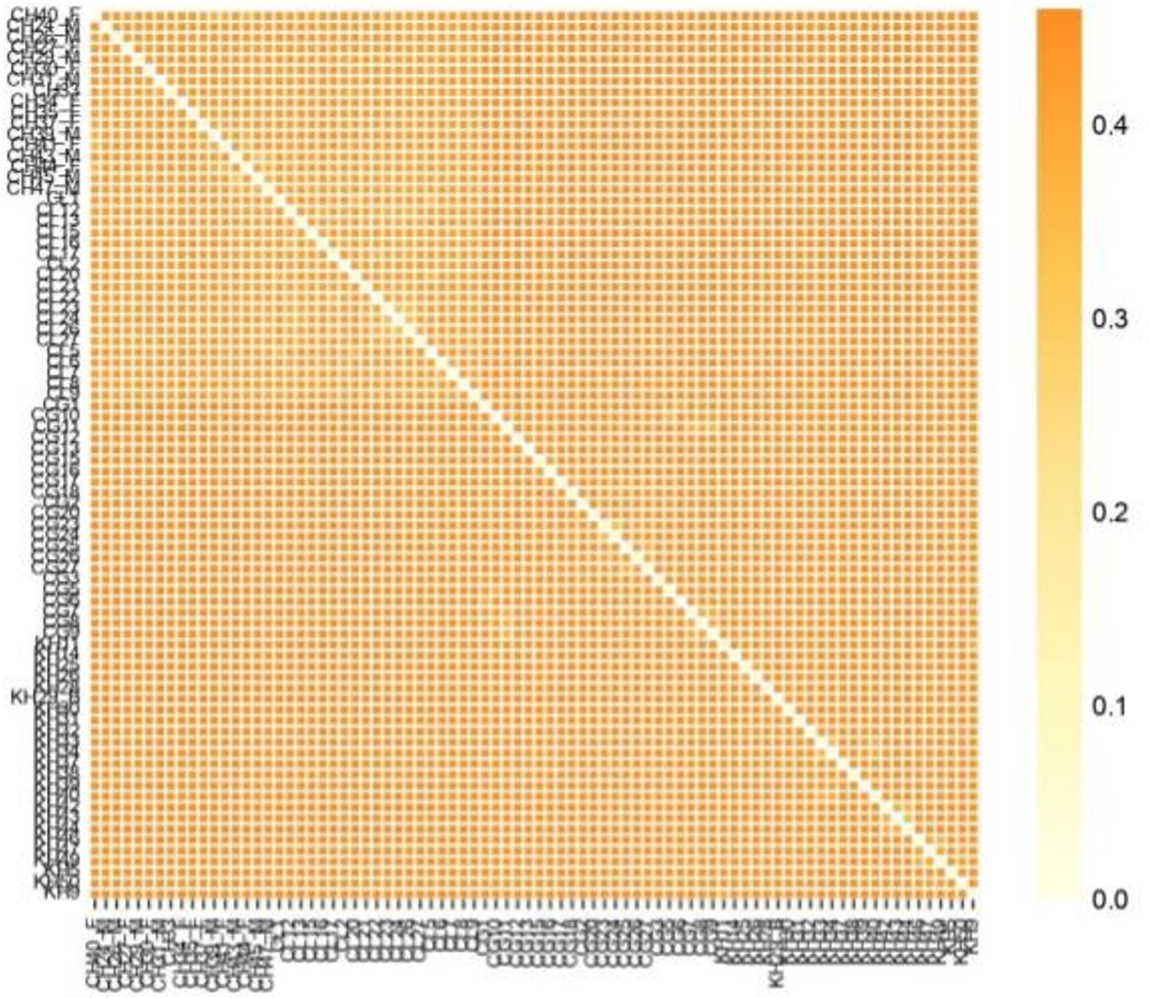

Fig. 11. shows the Distance matrix of three buffalo breeds.

The identity-by-State (IBS) heatmap in Fig. 11 shows the genomic similarity of the overall population of the Chhattisgarhi (CG), Chilika (CL), and Kalahandi (KH) buffalo. The data is symmetrically placed in the matrix where the diagonal is self-comparisons (IBS = 1.0), which is indicated by the lightest color. Continuous orange color scale depicts that the rest of the cells have pairwise IBS values of about 0.20 to 0.45. It is possible to identify a steady block-like gradient of the system with the individuals within the same population having rather bigger IBS, as they are more genetically similar to one another. The three populations are not visually segregated by specific color blocks with moderate inter-population relatedness yet sub-optimistic trend patterns of clusters depict themselves Chhattisgarhi individuals have a moderately stronger concentration of increased IBS values, whereas Chilika and Kalahandi animals have a more extensive assortment of genomic comparability. Allowing the overall homogeneity of the heatmap indicates that the three populations are significantly genome-wide allied in spite of the evident population structure in phylogenetic and PCA analyses. This tendency points to their similar origin and the potential movement of the genetic material in the past though still it presupposes the existence of the population-specific genetic patterns.

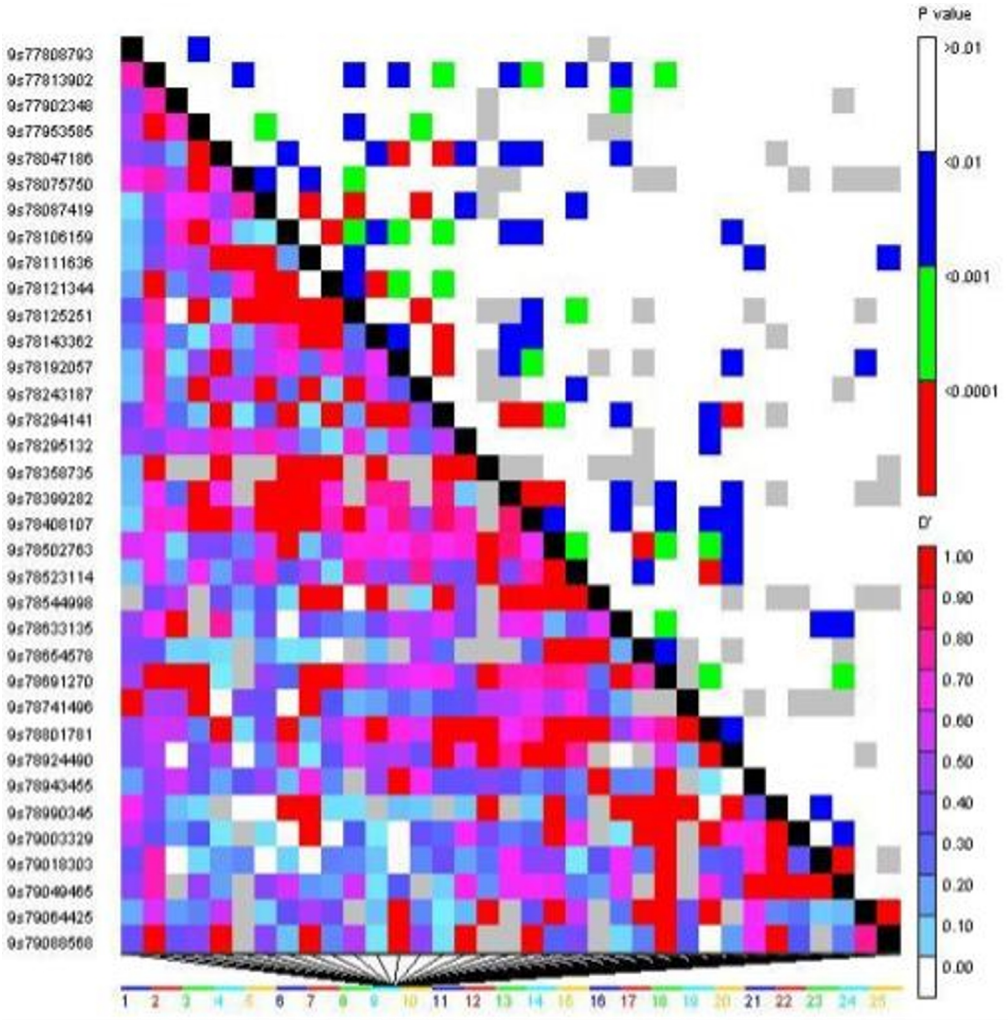

Fig. 12A. shows the LD plot of three buffalo breeds. The upper triangle shows the P-values, and the lower triangle shows the D‘ values.

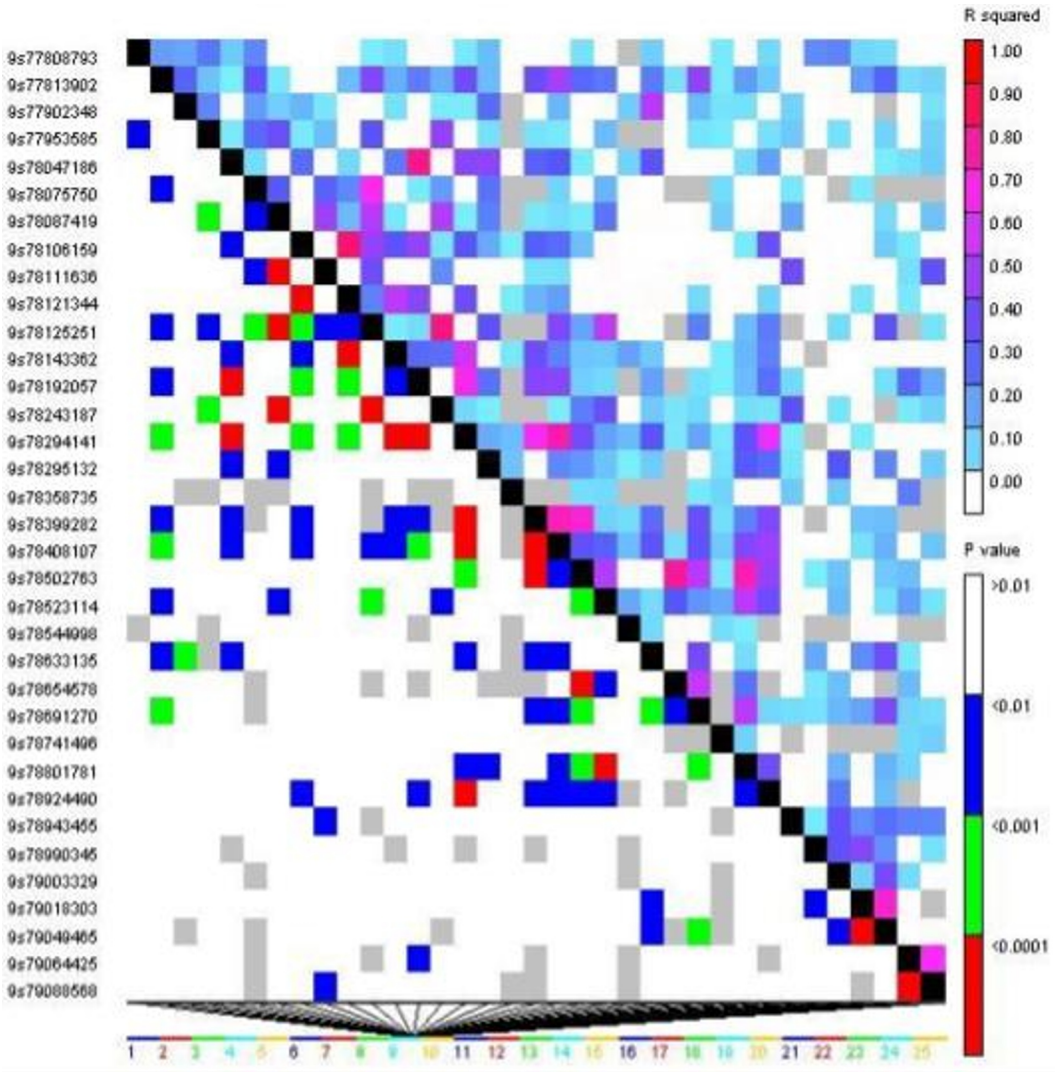

Fig. 12B. shows the LD plot of three buffalo breeds. The upper triangle shows the R^2 values, and the lower triangle shows the P values.

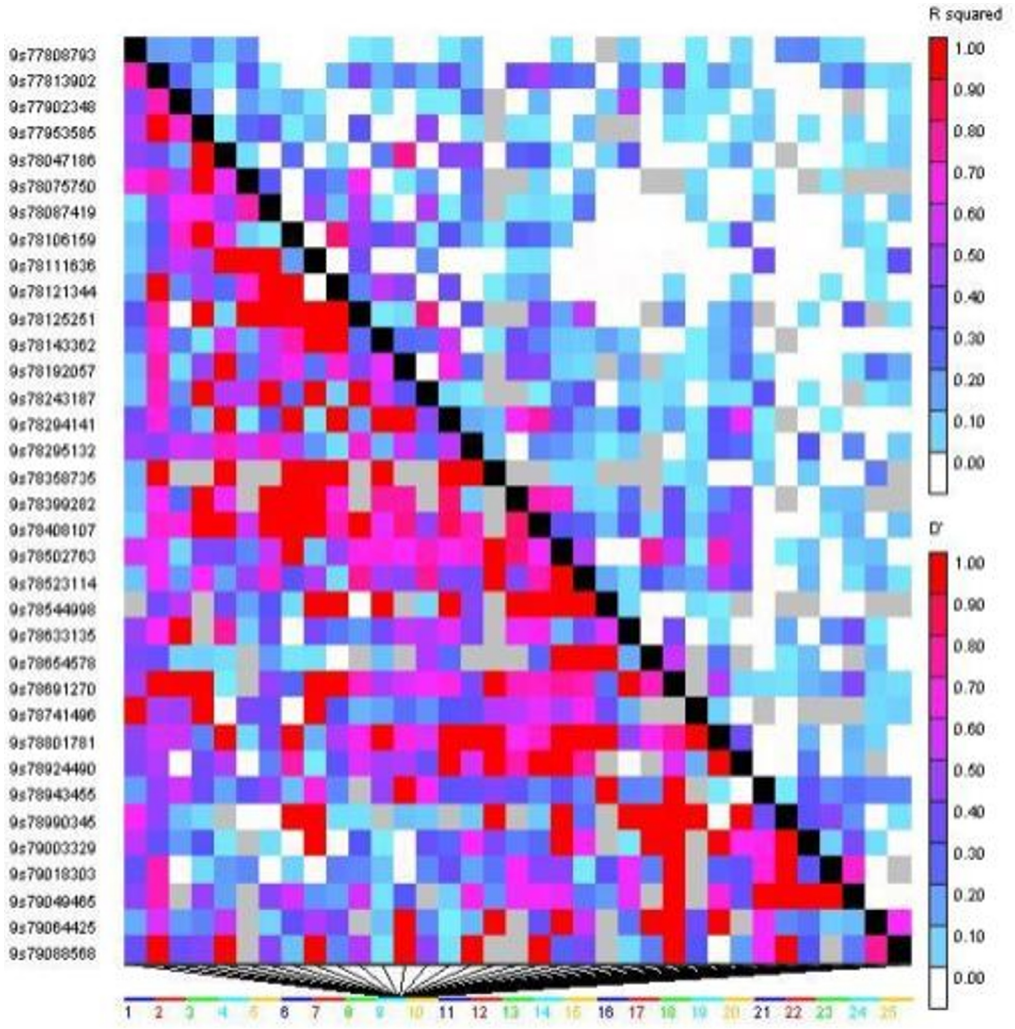

Fig. 12C. shows the LD plot of three buffalo breeds. The upper triangle shows the R^2 values, and the lower triangle shows the D‘ values.

The upper triangle of the LD (linkage disequilibrium) heatmap in Fig. 12A shows the P values which represents the statistical significance of pairwise LD between SNPs. The P-values shown in red colour have the high significance with $p < 0.0001$, while the p values shown in green are less significant p values >0.01. The lower p-values suggests us a stronger and more statistically significant non-random association of alleles. Fig. 12B has central diagonal squares (black in colour) because R2 =1.0 indicates a perfect self-correlation of the same SNP pairs at each locus that is a baseline to verify data integrity in LD analysis. The D' values in Fig.12C represents the normal correlation between SNP pairs which ranges from 0 (no LD shown in blue colour) to 1 (which means perfect LD, shown in red colour). The heatmap shows lowerD' values which indicates stronger recombination and weaker association, while higher D' values shows the regions with less recombination and stronger LD values.

4 Discussion

The thorough genomic studies of the three buffalo groups Chilika (CH), Chhattisgarhi (CG) and Kalahandi (KH) have shown that there is a certain degree of genetic structuring and differentiation, which implies unique evolutionary direction and higher or lower rates of gene flows. The PCA and Fst results demonstrate a clear substructure of the population between Chhattisgarhi, Chilika, and Kalahandi buffaloes as found in relation to geographic isolation patterns of the breed of Indian buffaloes. Chhattisgarhi-Kalahandi coincidence (Fst=0.010) indicates commonalities in the world due to the separation, and the moderate divergence of Chilika (Fst=0.045-0.049) and intermediate positioning at the PCA indicate local adaptation in Chilika, due to unique lacustrine environment. The QQ plot showed that there was a good fit between distribution and the observed and expected distribution and genomic inflation and data reliability were low. Minor fluctuations at the tail end indicate that there is a small group of SNPs that may be exposed to selection pressure which are essentially biological relevant genomic regions. Likewise, the Chi-square and p-value distributions indicate that majority of the SNPs do not vary in any particular way, but some of them have significant variations as there are localized selection events that cause the differentiation among the populations. The scree plot also substantiates the idea that most genetic variation is explained by PC1, which supports the idea that most population differentiation can be found on the first genetic axis, with the remaining variations of PC2 and PC3 being more fine-tuning, of population-specific variations. The IBS and the phylogenetic analyses provide a macro level overview of the population associations. The dendrogram and the circular tree explicitly separate the three groups of buffalo in different clades, which confirms the genetic differentiation of FST and PCA. The clusters between Chhattisgarhi and Kalahandi buffaloes were small with a short distance between the branches which tended to show low within-group variability. The length of the branches was longer and more dispersed in agreement with high levels of heterogeneity among the Chilika population.

These findings are also supported by the IBS heatmap, where genetic similarity varies between groups and is moderate and high but there is overall similarity in populations in terms of shared background, which indicates the existence of a common pool of gene and historical introgression. The studies referenced in [16] and [21] proivde us the valuable insights that relate our findings on the genetic diversity of buffaloes. Lastly, the LD analysis showed moderate-to-great linkage disequilibrium blocks in each and every population particularly in the genomic regions of 0.75-0.88 Mb indicating genomic regions with low recombination or hotspots of selection. The consequent pattern of LD decay also implies differences in recombination rates and history of population with Chhattisgarhi and Kalahandi buffaloes possessing tighter LD owing to the fact that they are lower in number or occur more recently in their lineages and Chilika buffaloes having a more diffused LD pattern owing to more varied ancestries. The overall findings present a consistent image of Chilika buffaloes as more genetically intermediate and diverse population (possibly, as a result of ecological and geographical isolation issues along the coastal region). Chhattisgarhi and Kalahandi buffaloes do not seem as genetically distant, on the contrary, they are tightly clustered into a sub-cluster that probably represents recent divergence rather than further migration aided by the closeness of geographical location. The small total differentiation and specific clustering is an indication that although the three populations of buffalos may have the same origin, they have adapted locally, are breeding under isolation, and are under environmental selection pressure that has led to the present genetic variations. This type of observation is crucial in designing successful breed conservation and genetic enhancement programs to maintain the special adaptive merit and also distribute the common genetic heritage of these native buffaloes.

5 Conclusion

The present study used high density SNP genotyping in analyzing the genetic composition of three buffalo populations as found in Chhattisgarhi (CG), Chilika (CH), and Kalahandi (KH). The findings indicate low to mid genetic variation in general. The pair-wise FST data indicate that there is a moderate separation between Chilika and other two (FST 0.044-0.049) whereas Chhattisgarhi and Kalahandi buffalo are very similar (FST 0.010). This trend correlates with the results of PCA scatter plots where Kalahandi buffaloes has a very narrow range on one side of PC1, Chilika has a scattered range on PC2 which means it has internal diversity and Chhattisgarhi buffaloes slightly overlaps with Chilika buffaloes. Three distinct groups of population sharing genetic backgrounds as demonstrated by phylogenetic trees and IBS heatmaps have diagonal blocks of high similarity within each group and moderate orange tones between groups, which is indicative of past genetic flow and shared origin. The quality of data is checked by QQ-plots, chi-square distributions, p-value histograms, and Manhattan plots, where most SNPs are when the tail is insignificant. Scree plots emphasize the role PC1 plays that captures the key genetic divide (5.5% variance), with PC2 (2.3%) and PC3 (2.2%) having smaller roles. LD heatmaps indicate moderate-to-high blocks (R2 0.75-0.88) and decay patterns, which indicate big historical values of effective population size but recent decreases that require intervention. With the isolation and adaption of the coast, Chilika buffalo is more diverse. Chhattisgarhi and Kalahandi buffalo exist in a tight cluster, which could be the result of the inland close location of the two and thus the aspect of exchange of genes, as

Chhattisgarhi buffaloes are more ancestral or rather wild-like as evidenced in its phylogenetic location. These buffaloes are also evolutionary relatives but locally adapted to local environments Chilika buffalo to wetlands, Chhattisgarhi and Kalahandi buffaloes to various inland environments. To preserve a more integrated approach is appropriate since they are closely related, and specific strategies are critical: preserve the unique spread of Chilika, keep a check on the ancestral character of Chhattisgarhi buffaloes, and keep a track of the homogeneity of Kalahandi buffaloes. These breeds are the source of milk and sustainability to the dairy industry in India. It is important to create a balance between increasing productivity and maintaining diversity during times of climate change, disease, and market changes. These genomic tools and knowledge provide a pathway to genomic selection, breeding initiatives, and more studies on the genes of adaptation. We can preserve these worthy resources to enable buffalo farming to be sustainable in future by the combination of modern science and the preservation of traditional farming knowledge and community activities.

Acknowledgement

The work was supported by National Bovine Genomic Centre Indigenous Breeds (NBGC-IB) project. The authors also thank ICAR-NBAGR, Karnal and JUIT Solan, for providing infrastructure and facilities to complete the research work.

References

[1] Nanda, A. S., & Nakao, T. *Role of buffalo in the socioeconomic development of rural Asia: Current status and future prospectus.* Anim. Sci. J., **74(6)**, 443-455 (2003). DOI **10.1046/j.1344-3941.2003.00138.x)**

[2] Borghese, A. *Buffalo Production and Research*. FAO Regional Office for Europe, Rome (2005). ISSN**:** 1020-3737

[3] FAOSTAT, F. (2023). Livestock primary. *Food and Agriculture Organization of the United Nations*.

[4] DAHD. *20th Livestock Census Report*, Government of India (2022).

[5] Sangwan, M. L. *Analysis of genetic diversity of Indian buffalo breeds by DNA markers*, J. Buffalo Sci., 1(1), 91 (2012). DOI 10.6000/1927-520X.2012.01.01.16

[6] Kumar, S., Nagarajan, M., Sandhu, J. S., Kumar, N., & Behl, V. B. *Phylogeography and domestication of Indian river buffalo*. BMC Evol. Biol., **7(1)**, 186 (2007). DOI 0.1186/1471-2148-7-186.

[7] Nagarajan, M., Nimisha, K., & Kumar, S. *Mitochondrial DNA variability of domestic river buffalo (Bubalus bubalis) populations: genetic evidence for domestication of river buffalo in Indian subcontinent*. Genome Biol. Evol., **7(5)**, 1252-1259 (2015). DOI 10.1093/gbe/evv067

[8] Luo, X., Zhou, Y., Zhang, B., Zhang, Y., Wang, X., Feng, T., ... & Liu, Q. (2020). *Understanding divergent domestication traits from the whole-genome sequencing of swamp-and river-buffalo populations.* Natl. Sci. Rev., **7(3)**, 686-701 (2020). DOI 10.1093/nsr/nwaa024

[9] Michelizzi, V. N., Dodson, M. V., Pan, Z., Amaral, M. E. J., Michal, J. J., McLean, D. J., ... & Jiang, Z. *Water buffalo genome science comes of age***.** Int. J. Biol. Sci., **6(4)**, 333 (2010). DOI https://doi.org/10.7150/ijbs.6.333

[10] Colli, L., Milanesi, M., Talenti, A., Bertolini, F., Chen, M., Crisà, A., ... & AdaptMap Consortium. *Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes.* Genet. Sel. Evol., **50(1)**, 58 (2018). DOI 10.1186/s12711-018-0422-x

[11] Vohra, V., Singh, M., Mukherjee, K., & Kataria, R. S. *Identification and characterization of Chhattisgarhi buffalo population in India* (2017). DOI 10.56093/ijans.v87i2.67711

[12] Dash, S. K., Sethi, B. P., Rao, P. K., & Prakash, B. *Characterization and prospective of Chilika buffalo - A unique germplasm of Eastern India*. Livest. Biodivers. J., Society for Conservation of Domestic Animal Biodiversity: Karnal, India (2010).

[13] Gahlyan, R. K., Vohra, V., Chhotaray, S., & Kataria, R. S. *Diversity assessment of a lesser known buffalo population from Central India and its comparative evaluation reveals presence of sufficient genetic variation and absence of selection.* Anim. Biotechnol., ***35*(1)**, 2305550 (2024). DOI https://doi.org/10.1080/10495398.2024.2305550

[14] Vijh, R. K., Tantia, M. S., Mishra, B., & Bharani Kumar, S. T. *Genetic relationship and diversity analysis of Indian water buffalo (Bubalus bubalis).* J. Anim. Sci., ***86*(7),** 1495-1502 (2008). DOI https://doi.org/10.2527/jas.2007-0321

[15] Singh, S. G., Kour, A., Taggar, R. K., Kumar, D., & Chakarborty, D. *Genetic Diversity Studies using Microsatellite Markers to Analyse Genetic Variation among the Buffalo Breed Populations in Jammu Region.* J. Anim. Res., **11(6)**, 1001-1006 (2021). DOI 10.30954/2277-940X.06.2021.8

[16] Bhave, K., Joshi, S., Shirsath, T., Potdar, V., Joshi, A., Punde, N., ... & Gaundare, Y. *Adoption of Genomic Tools for Genetic Improvement in Smallholder Dairy Cattle and Buffalo.* Anim. Sci. Cases, ascs20250027 (2025). DOI 10.1079/animalsciencecases.2025.0027

[17] Fan, B., Du, Z. Q., Gorbach, D. M., & Rothschild, M. F. *Development and application of high-density SNP arrays in genomic studies of domestic animals.* Asian-Australas. J. Anim. Sci.**, *23*(7)**, 833-847 (2010). DOI 10.5713/ajas.2010.r.03

[18] Deng, T. X., Ma, X. Y., Lu, X. R., Duan, A. Q., Shokrollahi, B., & Shang, J. H. *Signatures of selection reveal candidate genes involved in production traits in Chinese crossbred buffaloes.* J. Dairy Sci., **105(2)**, 1327-1337 (2022). DOI 10.3168/jds.2021-21102

[19] Mokhber, M., Shahrbabak, M. M., Sadeghi, M., Shahrbabak, H. M., Stella, A., Nicolzzi, E., & Williams, J. L. *Study of whole genome linkage disequilibrium patterns of Iranian water buffalo breeds using the Axiom Buffalo Genotyping 90K Array*. PLoS One, **14(5)**, e0217687 (2019). DOI 10.1371/journal.pone.0217687

[20] Nascimento, A. V., Cardoso, D. F., Santos, D. J. A., Romero, A. R. S., Scalez, D. C. B., Borquis, R. R. A., ... & Tonhati, H. *Inbreeding coefficients and runs of homozygosity islands in Brazilian water buffalo.* J. Dairy Sci., **104(2)**, 1917-1927 (2021). DOI 10.3168/jds.2020-18397

[21] Mousbah, A. M., Abdullah, H. M., Mohammed, W. S., El-Refy, A. M., & Helmy, M. *Buffalo Genome Projects: Current Situation and Future Perspective in Improving Breeding Programs*. *arXiv preprint arXiv:2304.04977* (2023). DOI 10.48550/arXiv.2304.04977

[22] Zhang, Y., Colli, L., & Barker, J. S. F. *Asian water buffalo: domestication, history and genetics.* Anim. Genet., **51(2)**, 177-191 (2020). DOI 10.1111/age.12911

[23] Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. *Principal components analysis corrects for stratification in genome-wide association studies*. Nat. Genet., ***38*(8)**, 904-909 (2006).

[24] Rambaut, A., & Drummond, A. FigTree v1. 3.1 Institute of Evolutionary Biology. *University of Edinburgh* (2010).

[25] Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. *TASSEL: software for association mapping of complex traits in diverse samples.* Bioinformatics, **23(19)**, 2633-2635 (2007). DOI 10.1093/bioinformatics/btm308

[26] Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., ... & Wang, Y*. SRplot: A free online platform for data visualization and graphing*. *PloS one*, ***18*(11)**, e0294236 (2023). DOI 10.1371/journal.pone.0294236

[27] Pauciullo, A., Gaspa, G., Versace, C., Cosenza, G., Piscopo, N., Gu, M., ... & Zhang, Y. (2025). *New insights into genetic diversity and differentiation of 11 buffalo populations using validated SNPs for dairy improvement.* Genes, **16(4)**, 400 (2025). DOI 10.3390/genes16040400