International Journal of Animal and Veterinary Advances 4(4): 284-291, 2012
ISSN: 2041-2894

Submitted: June 19, 2012 Accepted: July 04, 2012 Published: August 20, 2012

Genetic Discontinuities and Sub structuring in buffaloes of Indo Gangetic Plains

[1]Upasna Sharma, [2]Priyanka Banerjee, [2]Jyoti Joshi and [2]Ramesh Kumar vijh
[1]Singhania University, Pacheri Bari, Jhunjhunu-313515, Rajasthan, India
[2]National Bureau of Animal Genetic Resources (ICAR), Karnal-132001 India

Abstract: The Indo-Gangetic plains of India are fertile basin and have more than 23% buffaloes of India. Bhadawari is the only one recognised breed of buffalo in the basin. The population structure of this vast buffalo population was not known. In this present study, we generated data on 11 microsatellite markers on 625 buffaloes from Indo-Gangetic plains. We utilised landscape genetic tools combining molecular genetic data and spatial coordinates to decipher the genetic structure. We utilised wombling analysis and Bayesian clustering using Markov models. We also utilised Monmonier's maximal difference algorithm for the identification of genetic barriers and tested their significance using 16 different genetic distances. The Womble analysis revealed three statistically significant barriers. The genetic bandwidth mapping revealed that buffaloes of Tarai area and Bhadawari buffaloes to have distinctive population structure. The buffaloes of Mau, Balia and Ghazipur also had a distinctive genetic structure. Most of the buffaloes of Central region of Indo-Gangetic plains were found to be heterogeneous. The variation in allele frequency in buffalo population in Indo Gangetic plains occurred at multiple scales. A sharp change in allele frequency in Eastern and Western Uttar Pradesh buffaloes was observed and detectable by all the techniques utilised and the buffalo population could be substructured into 5 clusters.

Keywords: Bayesian clustering, genetical bandwidth mapping, genetic barriers, landscape genetics, markov models, microsatellites, womble method

INTRODUCTION

The development of molecular markers in conjunction with statistical tools and powerful computers have led to the emergence of a new scientific field, landscape genetics, which is an amalgamation of population genetics and landscape ecology (Manel *et al*., 2003). This discipline aims to provide information on how landscape and environmental features influence gene flow, population structure and local adaptation. Landscape genetics tends to focus on the understanding of the micro-evolutionary processes that generate genetic structure across space. The 2 key steps of landscape genetics are the detection and location of genetic discontinuities and the correlation of these discontinuities with landscape and environmental features (e.g., mountains, rivers, roads, gradient of humidity and deforested areas) (Manel *et al*., 2003). Several new techniques of landscape genetics are being presently utilized to delineate the population genetic structure in the background of geographical parameters. The geographic visualization of patterns was originally published by Womble (1951) and rediscovered by Barbujani *et al*. (1989). This method focuses on a detailed visualization of the geographic areas associated with a considerable genetic change called as boundaries or barriers. The genetic structures can show correspondence with geography and hence general methods of geographic analysis can be successfully applied to population genetics. This implies the computation of neighbouring problems (computational geometry). Monmonier's maximum difference algorithm (Monmonier, 1973) has now been successfully used for the visualization on a geographic map with the dramatic rate of change in a given distance measure, which can be genetic and/or morphologic. The algorithm is applied to a geometric network that connects all the populations (sampled locations) using Delaunay triangulation (Brassel and Reif, 1979). The latest tool in the landscape genetics is usage of Genetic Bandwidth Map (GBM). The GBM is a new visual tool for investigating spatial variation of allele frequencies. The hierarchical Baye's clustering algorithm for studying population structure using individually geo-referenced multilocus data on the concept of hidden Markov random field can be utilised in this specific case of Indo-Gangetic plains. This is achieved by using Hidden Markov Random Fields (HMRFs) as prior distributions on cluster membership (Francois *et al*., 2006). The Indo-Gangetic plains have more than 23% of the total buffalo population but have not been defined for their genetic structures. A recent analysis of the microsatellite data

Corresponding Author: Ramesh Kumar vijh, National Bureau of Animal Genetic Resources (ICAR), Karnal-132001 India

(Upasna *et al*., 2011) revealed that the buffalo population of these plains reveal a continuity of allele frequencies and the estimate of isolation by distance is significant. In this study, we utilised the Womble's approach, Monmonier's maximum difference algorithm, Genetic Bandwidth Mapping and Baye's hierarchical clustering algorithm to identify subtle differences at the genetic level of buffaloes belonging to these geographical plains.

MATERIALS AND METHODS

Blood samples of 625 buffaloes were collected from throughout the Indo-Gangetic plains. The districts from which there were very few samples were clubbed with the adjoining districts totalling of 34 conglomerates. The DNA was isolated following normal protocols (Sambrook *et al*., 2001) and 11 heterogonous microsatellites were selected from cattle database. The microsatellite data was generated using fluorescently labelled primers and using ABI Automated DNA sequencer. The data generated was extracted using GeneMapper software v4.1 (Applied Bio systems).

We utilized a software package Womb soft (Crida and Manel, 2007) written in R for carrying out the Wombling analysis. The individually geo-referenced multi-locus genotypes were used for the inferences of genetic boundaries between sampling locations. The significance of boundaries using the binomial test was carried out to see if the number of candidate boundary elements at each point of the grid was significant. The Binomial test was carried out on the data with p-value taken as 0.05 and cbe value calculated earlier. The wombling was done using the parameters bandwidth as 50, 70, 100, 200, 300 and 700. The calculation of “pseudo slopes” from the genetic and geographical distance matrix was derived by the software Alleles in Space (AIS) (Miller, 2005) as the quotient of congruent elements from the genetic and geographical distance matrices. The genetic barriers were calculated using the software Barrier v 2.0 (Manni *et al*., 2004). The significance of the barriers was tested using 16 different genetic distance matrices. The various genetic distances utilized for the purpose were Bhattacharyya (1946), Cavalli-Sforza (1967, 1969), Goldstein distance (1995), Latter (1972), Nei’s distance (1972, 1973, 1987), Prevosti (1975), Reynolds *et al*. (1983), Rogers Eucledean distance (1972), Sanghvi (1953), Shared allele distance *D SA* (Chakraborty and Jin, 1993), Shriver (1995) and Slatkin (1995).

For genetic bandwidth mapping and for evaluating the buffalo cryptic structure of Indo Gangetic plains, the software GenBMap (Cercueil *et al*., 2007) was used. The resolution of X and Y was kept at 1000 for getting a genetic bandwidth map. For hierarchical Baye's analysis we used the software TESS version 2.3.1. The parameters set were burning of 10,000 sweeps and running period of 50,000 sweeps. Preliminary (short) runs were used to calibrate length. We utilised the model with admixture, starting with Kmax = 2. The analysis was run until the bar plot stabilized at Kmax = 5. Several runs with the above parameters were carried out and we kept the 20% lowest Deviance Information Content runs and used CLUMPP software version 1.1.2 (Jakobsson and Rosenberg, 2007) to process the output to perform averages over these runs. We utilised Destruct software version 1.1 (Rosenberg, 2004) to display the estimated membership coefficients.

RESULTS AND DISCUSSION

The allele frequency data was subjected to statistical analysis for characterizing the spatial variation of allele frequencies. Wombling methods aim at detecting regions of abrupt changes in allele frequency. The Womble approach located boundaries across allele frequency surface by searching for regions in which the absolute value of the surface slope is large. The systemic function obtained from the present study of 11 microsatellite markers on buffaloes of Indo Gangetic plains has been depicted below in Fig. 1. The boundaries which were found to be significant on the basis of Binomial test have been depicted in Fig. 2. This method focused on a detailed visualization of the geographic areas associated with a considerable genetic change, called boundaries or barriers. The method permitted different variables within the same landscape to be considered together. First, individual surfaces were differentiated such that steep slopes become peaks and flat plains fall to zero. Second, the magnitudes of the derivatives of surfaces from different variables can be added to get a composite picture of barriers derived from all variables. By introducing a percentile consideration of significance (i.e., by considering values in the top 20% to represent barriers as carried out in this present analysis), values within the landscape could be compared, thus controlling for the effect of Isolation By Distance (IBD) which was significant for the present data (Upasna *et al*., 2011). The sampling locations which are in the shaded grey and show heterogeneity on the right hand side are Allahabad, Mirzapur, Pratapgarh, Sultanpur, Jaunpur, Ambedkar Nagar, Azamgarh, Deoria, Gorakhpur, Sant Kabir Nagar, Siddharth Nagar, Kushi Nagar, Chandauli, Ghazipur, Mau and Balia. The districts on the left side and shaded with grey are Meerut, Baghpat and Ghaziabad. The only district that shows heterogeneity at the lowest end of the landscape is Lalitpur sampling area (Fig. 2). Monmonier’s maximum difference

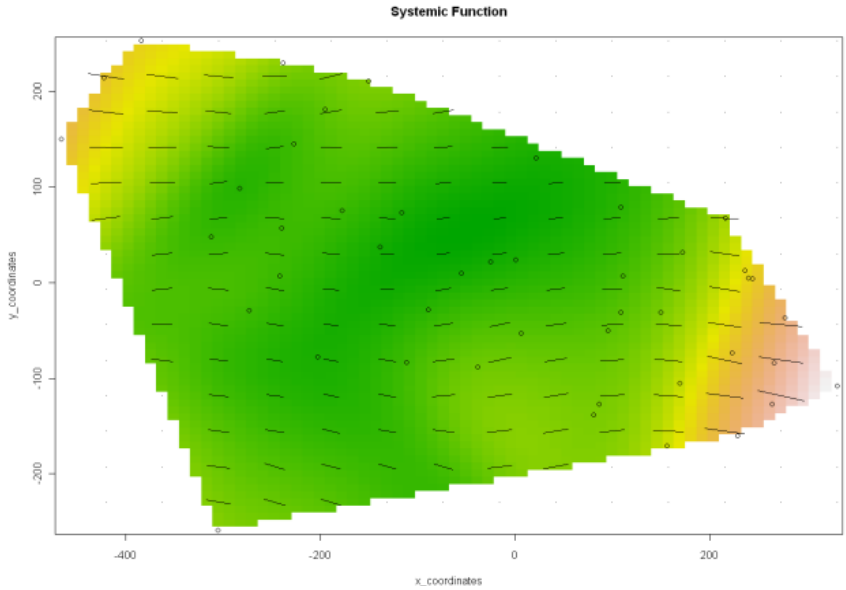


Fig. 1: The systemic function obtained using wombling method. The homogeneity (green) and heterogeneity (pink) is depicted with yellow is between two extremes

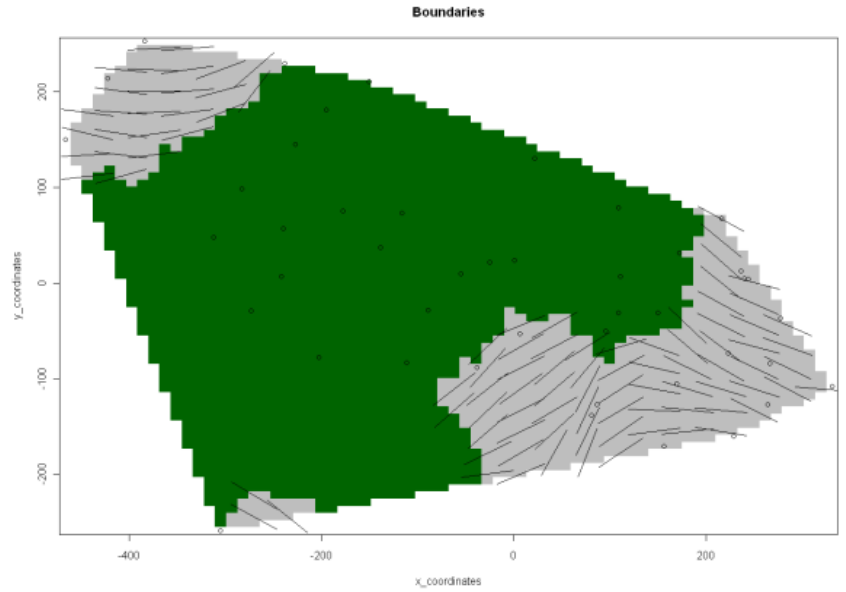


Fig. 2: The boundaries obtained using womb soft with bandwidth of 200. The lines represent the direction of allele frequency changes which are random in this present analysis. The green colour depicts the homogeneity and grey colour heterogeneity

algorithm enables a better interpretation of micro-evolutionary processes, such as gene flow, genetic drift and selection. It also helps to identify hidden boundaries resulting from secondary gene flow among previously isolated populations. The application of Monmonier's algorithm leads to an understanding of the processes that caused the patterns. The software Barrier version 2 (Manni *et al*., 2004) permitted a significance test using bootstrap matrices analysis. The bootstrapping reduced the noise and genetic barriers were more robust and could be identified. The genetic barrier highlights the geographic area where a discontinuity exists in terms of allele frequencies. In the present study, we identified 5 barriers (Fig. 3)

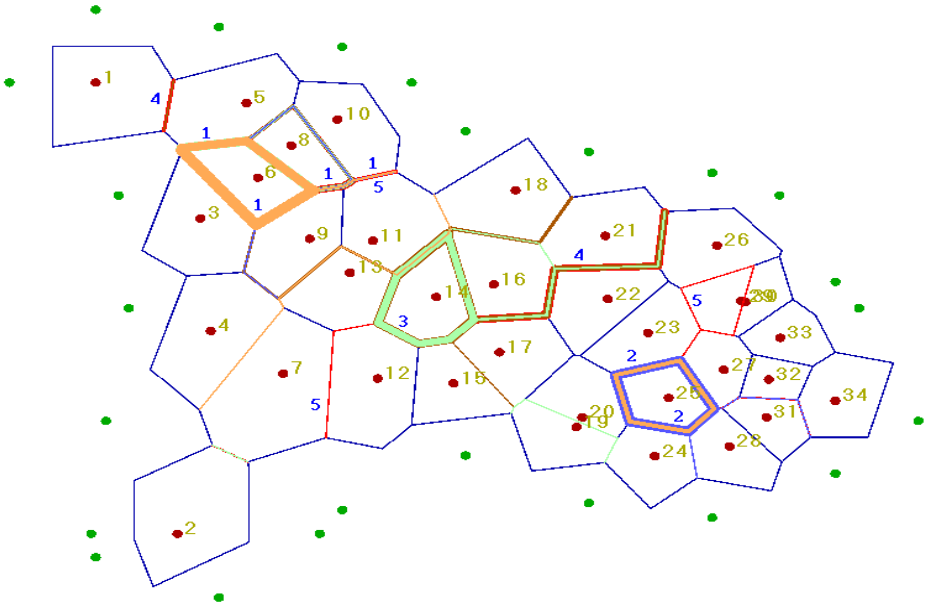


Fig. 3: Figure represents 5 Genetic barriers. All the barriers have been given distinctive numbers (1-5) and the sampling locations have been given the district number. The thickness of the barriers is based on the bootstrap values. The barrier numbers are marked in blue colour (1-5)

which were highly significant based on the bootstrap matrices (using 16 different genetic distances to check the significance of the barrier.

The first barrier (Orange Colour marked 1) encapsulates sampling location 6 (Badaun district). A weak barrier also exist separating Meerut, Baghpat, Bareilly, Rampur and Pilibhit from rest of the Indo-Gangetic plains. This prominently is the Tarai area. The 2nd barrier (Blue Colour) is around district of Jaunpur and a weak barrier between the districts of Mau and Balia. The 3rd Barrier (Green Colour) covers Unnav and extends between districts of Lucknow-Raibareily, Gonda- Faizabad and Gonda-Siddharth Nagar. It also seperates the districts of Lucknow-Sitapur and Lucknow-Gonda. The 4th barrier (Brown colour) seperates the districts of Lucknow-Raibareilly, Gonda-Faizabad and Gonda-Siddharth Nagar and also Sitapur and Gonda and Baghpat-Meerut and Rampur. The 5th barrier depicted in red colour is between districts of Bareilly-Farrukhabad and Pilibhit-Hardoi and also Kanpur-Hamirpur-Jalaun. A barrier also separates Gorakhpur from Siddharth Nagar, Ambedkar Nagar and Azamgarh. The Monmonier's algorithm was also utilised in the software Allele in Space (Fig. 4) using the Pseudo-slope method. The five genetic barriers obtained using these procedures are depicted in different colours. The 5 barriers obtained using the pseudoslope methods were: Barrier 1 divides the districts Ghazipur, Mau and Balia from rest of UP. The Barrier 2 encapsulates Mirzapur, Allahabad, Pratapgarh, Raibareilly, Fatehpur. The Barrier 3 covers Kanpur, Etawah, Kannauj, Firozabad, Badaun, Farrukhabad districts; Barrier 4 touches Lucknow, Barabanki, Sitapur and Gonda districts. The Barrier 5 encapsulates Siddharth Nagar and Maharajganj area. The Genetic Bandwidth Mapping (GBM) is a visual tool for investigating spatial variation of allele frequencies. The information is displayed through a two-dimensional graphical representation of a local structural parameter. This parameter could be interpreted as the shortest distance to areas of significant changes as well as the radius of the largest zone for which the genetic structure can be thought of as being spatially homogeneous. The definition of genetic bandwidths fit in the framework of Wombling methods as the systemic map provides a natural measure of spatial homogeneity. Genetic Bandwidth Mapping and Wombling are nevertheless fundamentally distinct. The main difference resides in the fact that Wombling estimates the systemic function by using a fixed local parameter (e.g. a window size). Due to the high sensitivity of this parameter, systemic maps might hardly be estimated unambiguously, because each value of the local parameter might lead to a new map. Deciding which value minimizes the statistical error was difficult. In contrast, the GBM avoids these issues by adopting a reverse perspective, focusing on bandwidths rather than on the systemic map itself. The bandwidths were estimated on the basis of local homogeneity

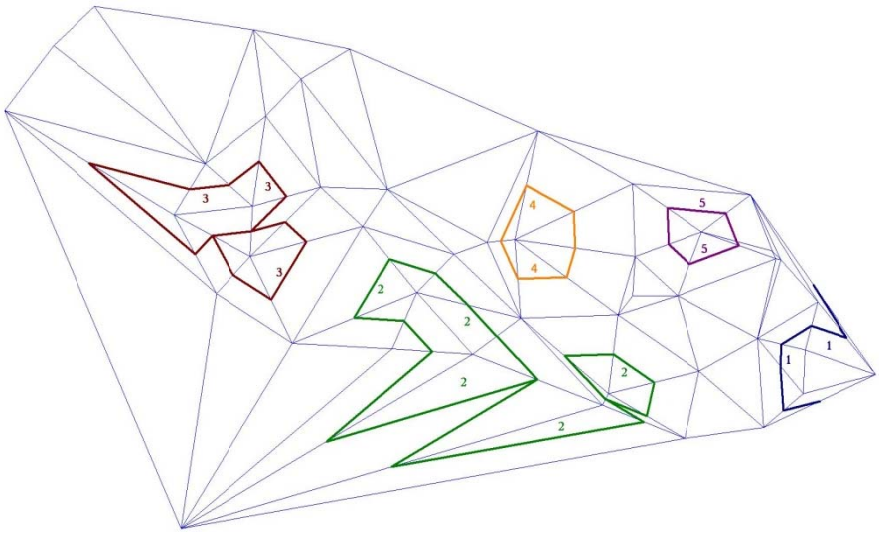


Fig. 4: Exhibiting five barriers obtained using Monmonier's algorithm and 'Pseudoslope' method

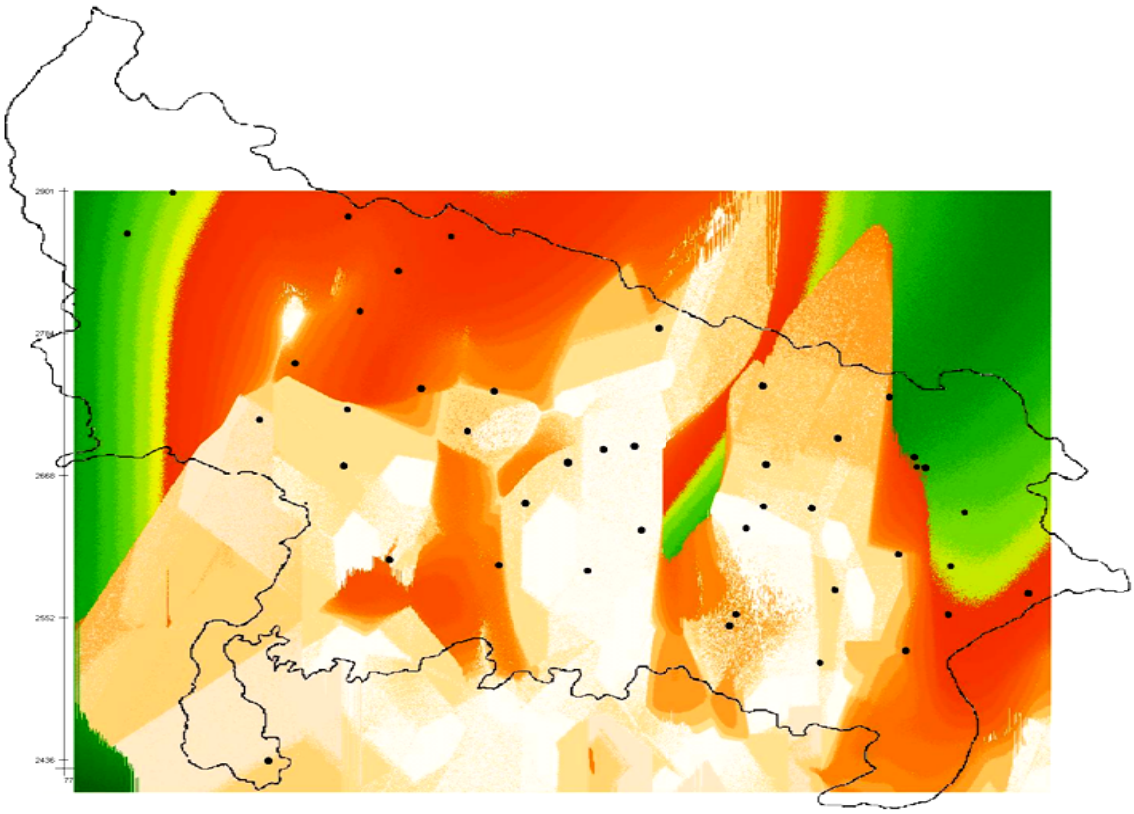
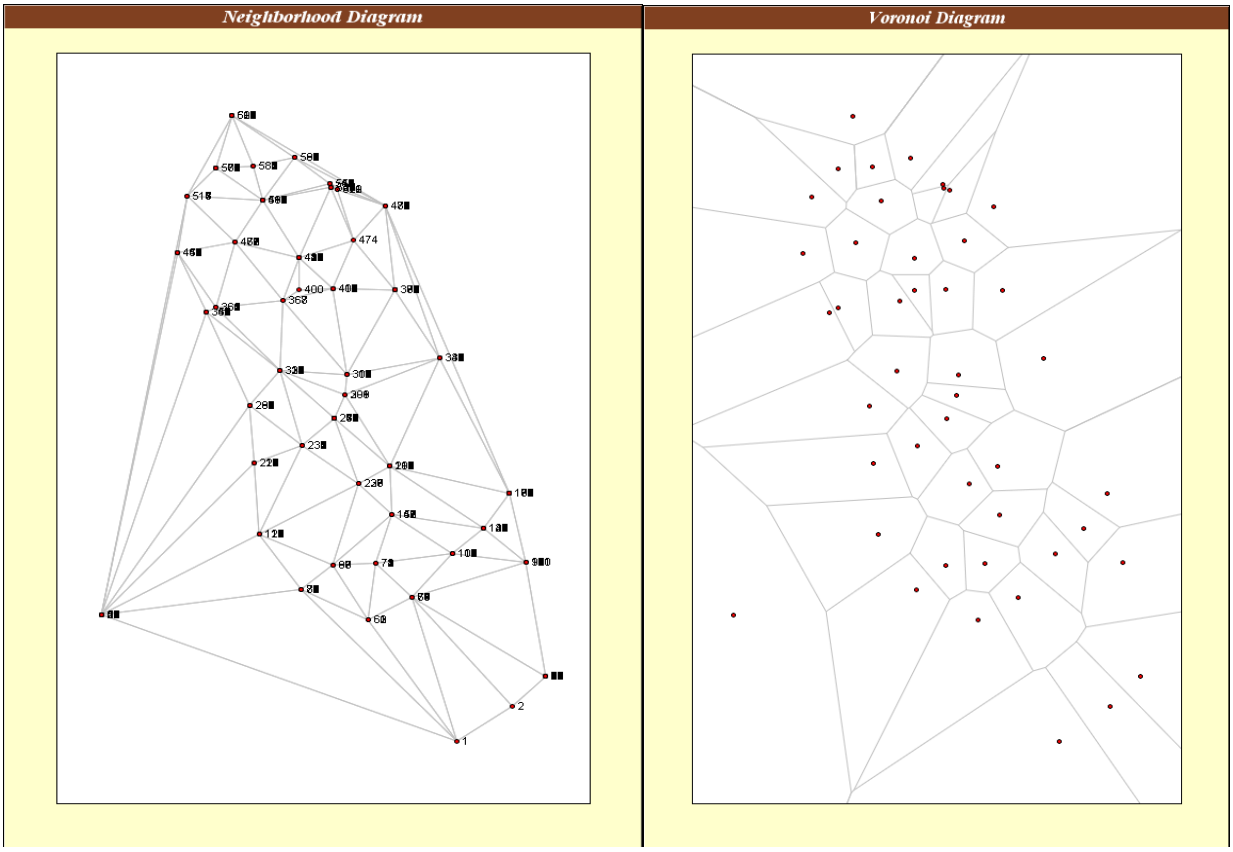


Fig. 5: Map produced by GenBMap Software for the buffalo population of Indo-Gangetic plains.

tests using all values of the local parameter. The GBM therefore produced a unique map for buffaloes of Indo Gangetic plains. The GBM proved successful at identifying genetic discontinuities and sharp clinal variation. The variation of allele frequencies in buffalo populations in the Indo-Gangetic plains occurred at multiple scales. The GBM seemed particularly relevant for this study of population of buffaloes of Indo Gangetic plains in which the fine-scale population structure was stronger



(a) (b)

Fig. 6: Showing (a) neighbourhood diagram and (b) Vornoi Diagram of sampling areas of buffaloes

than in the long range (Kayser *et al*., 2005; Marjanovic *et al.,* 2005 and Klopfstein *et al*., 2006). The colour coding representation used by the graphical interface (GENBMAP) may however be partly responsible for the fact that large-scale structure may confound the program when local structure is present because the program uses few colours. Nevertheless secondary structures associated with large-scale variation could be unambiguously identified. The green colour depicts the maximum genetic bandwidth and also homogeneity while the orange colour gives the intermediate homogeneity and white colour depicts the heterogeneity. The green colour may also show the absence of sampling locations. The green colour is exhibited for the districts of Meerut, Baghpat and Ghaziabad in the left side and Maharajganj, KushiNagar, Deoria, SantKabirNagar, Mau and Gorakhpur in the right side. The districts of Bareilly, Rampur, Pilibhit, Budaun, Etah, Hardoi get separated with sharp change in gene frequency as depicted in the Fig. 5. The districts of Etawah, Firozabad and Mainpuri show brown colour and coincide with the breeding tract of Bhadawari buffaloes. The districts of Lucknow, Barabanki, Gonda and Rai-Bareilly show large heterogeneity as depicted by white colour. The districts of Faizabad, Sultanpur and Mirzapur also show a large heterogeneity. The districts of Chandauli, Ghazipur and Balia show homogeneity. The genetic bandwidth thus shows cryptic genetic structures in great details. The differences are in congruence with the genetic barriers obtained using Monmonier's Maximum difference algorithm. In the present study of the buffalo populations of Indo Gangetic plains, it was presumed that there was a continuous variation in allele frequencies and hence HMRFs model were expected to be powerful when detecting geographical discontinuities in allele frequencies (Upasna *et al.*, 2011) and regulating the number of clusters. The TESS software which uses these models revealed the Neighborhood diagram and Voronoi diagram which are shown below as Fig. 6a and b. Utilizing a dataset of 11

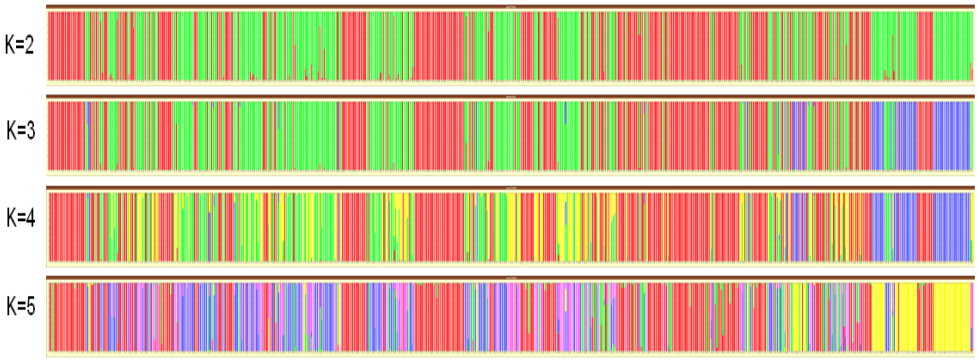


Fig. 7: Showing Bar plot with same color represents similar clusters along the Indo Gangetic plains

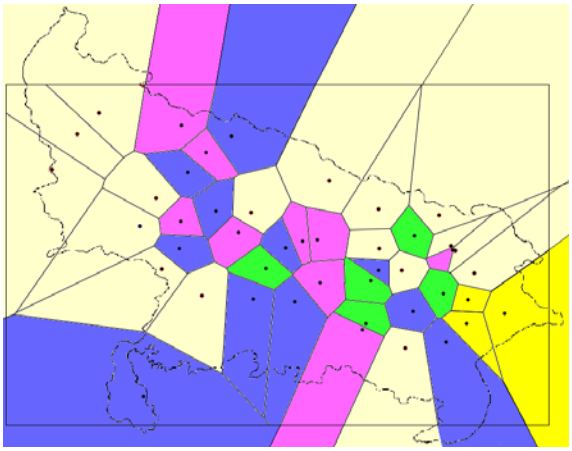


Fig. 8: Estimated cluster configuration for the buffalo data set in Indo Gangetic plains of UP with HMRF model (5 clusters)

microsatellite loci, we could delineate the substructure of buffalo population of UP into 5 distinctive clusters. The problem of local differentiation was studied in terms of change in correlation with distance where individuals living nearby tend to be more alike than those living far apart. Thus HMRF study in the present case was basically same concept except that statistical correlation was hidden at the cluster membership level. The five clusters identified in the buffalo population are not contiguous but exist in a fragmented manner (Fig. 8). This is also depicted in bar plots for Kmax from 2 to Kmax of 5 (Fig. 7).

CONCLUSION

The analysis revealed that a linear significant correlation between the genetic distance and geographical region may not be sufficient for deriving inferences about the populations of livestock species distributed over a large geographical area. The robustness of the present analysis was greatly enhanced by inclusion of continuous variation in allele frequencies through space. A smaller number of genetic markers shall be required for making inferences about the sub-structuring of any livestock species especially in Indian context where strong population structures do not exist owing to lack of breed associations/societies (selection pressure is quite low). The inferences drawn from this study are useful in developing breeding plans for buffaloes of Indo-Gangetic plains which contribute one quarter of the total buffalo population of India.

ACKNOWLEDGMENT

The financial support provided from NAIP ICAR, component 4 (C-1050) is gratefully acknowledged.

REFERENCES

Barbujani, G., N.L. Oden and R.R. Sokal, 1989. Detecting areas of abrupt change in map of biological variables. Syst. Zool., 38: 376-389.

Bhattacharyya, A., 1946. On a measure of divergence between two multinomial populations. Indian J. Stat., 7: 401-407.

Brassel, K.E. and D. Reif, 1979. A procedure to generate thiessen polygons. Geogr. Anal., 325: 31-36.

Cavalli-Sforza, L.L., 1969. Human diversity. Proceeding of the 12th Intl Cong. Genet. Tokyo, 3: 405-416.

Cavalli-Sforza, L.L. and A.W.F Edwards, 1967. Phylogenetic analysis: Models and estimation procedures. Am. J. Hum. Genet., 19: 233-257.

Chakraborty, R. and L. Jin, 1993. A unified approach to study hypervariable polymorphisms: Statistical considerations of determining relatedness and population distances. EXS, 67: 153-175.

Cercueil, A., O. François and S. Manel, 2007. GenBMap: The Genetical Bandwidth Mapping: A Spatial and Graphical Representation of Population Genetic Structure Based on the Wombling Method. Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, BP 53, F-38041 Grenoble cedex 9, France.

Crida, A. and S. Manel, 2007. WOMBSOFT: R Package for Wombling Analysis. Retrieved from: http://www-leca.ujf-grenoble.fr/logiciels.htm.

François, O., S. Ancelet and G. Guillot, 2006. Bayesian clustering using hidden markov random fields in spatial population genetics. Genetics, 174: 805-816.

Goldstein, D.B., L.A. Ruiz, L.L. Cavalli-Sforza and M.M. Feldman, 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA, 92: 6723-6727.

Jakobsson, M. and N.A. Rosenberg, 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics (Oxford) 23: 1801-1806.

Kayser, M., O. Lao, K. Anslinger, C. Augustin, G. Bargel, J. Edelmann, S. Elias, M. Heinrich, J. Henke and L. Henke, 2005. Significant genetic differentiation between Poland and Germany follows presentday political borders, as revealed by Y-chromosome analysis. Hum. Genet., 117: 428-443.

Klopfstein, S., M. Currat and L. Excoffier, 2006. The fate of mutations surfing on the wave of a range expansion. Mol. Biol. Evol., 23: 482-490.

Latter, B.D.H, 1972. Selection in finite populations with multiple alleles III: Genetic divergence with centripetal selection and mutation. Genetics, 70: 475-490.

Manel, S., K. Michael Schwartz, G. Luikart and P. Taberlet, 2003. Landscape genetics: Combining landscape ecology and population genetics. Trends Ecol. Evol., 18: 189-197.

Manni, F., E. Guérard and E. Heyer, 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by "Monmonier's algorithm". Hum. Biol., 76(2): 173-190.

Marjanovic, D., S. Fornarino, S. Montagna, D. Primorac, R. Hadziselimovic, S. Vidovic, N. Pojskic, V. Battaglia, A. Achilli and K. Drobnic, 2005. The peopling of modern Bosnia-Herzegovina: Y chromosome haplogroups in the three main ethnic groups. Ann. Hum. Genet., 69: 757-763.

Miller, M.P., 2005. Alleles in Space: Computer software for the joint analysis of interindividual spatial and genetic information. J. Heredity, 96: 722-724.

Monmonier, M., 1973. Maximum-difference barriers: An alternative numerical regionalization method. Geogr. Anal., 3: 245-61.

Nei, M., 1972. Genetic distance between populations. Am. Naturalist., 106: 283-292.

Nei, M., 1973. The Theory and Estimation of Genetic Distance. In: Morton, N.E. (Ed.), Genetic Structure of Populations. University Press of Hawaii, Honolulu, pp: 45-54.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.

Prevosti, A., J. Ocana and G. Alonzo, 1975. Distances between populations for Drosophila Subobscura based on chromosome arrangement frequencies. Theo. Appl. Genet., 45: 231-241.

Reynolds, J., B.S. Weir and C.C. Cockerham, 1983. Estimation of the Coancestry coefficient: Basic for a short-term genetic distance. Genetics, 105: 767-779.

Rogers, J.S., 1972. Measures of Genetic Similarity and Genetic Distance. University of Texas Publication Austin, TX, Studies in Genetics VII, 7213: 145-153.

Rosenberg, N.A., 2004. DISTRUCT: A program for the graphical display of population structure. Mol. Ecol. Notes., 4: 137-138.

Sambrook, J. and D.W. Russell, 2001. Molecular Cloning: A Laboratory Manual. 3rd Edn., Cold Spring Harbor, New York, 1(6): 6.4-6.62.

Sanghvi L.D., 1953. Comparison of genetical and morphological methods for a study of biological differences. Amer. J. Phys. Anthropol., 11: 385-404.

Slatkin M., 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics., 139: 457-462.

Upasna S., J. Joshi, P. Banerjee and R.K. Vijh, 2011. Genetic landscape and demography of buffaloes in Indo- Gangetic plains. Indian J. Anim. Sci., (In Press).

Womble W.H., 1951. Differential systematics. Science, 114: 315-322.