Indian J. Genet., 79(1) Suppl. 300-305 (2019)
DOI: 10.31742/IJGPB.79S.1.22

Genomic characterization of mithun (*Bos frontalis*) populations using high density SNP array

Anupama Mukherjee#, Sabyasachi Mukherjee*, Imsusosang Longkumer, Moonmoon Mech, Nazrul Haque, Kezhavituo Vupru, Kobu Khate and Chandan Rajkhowa

Animal Genetics and Breeding Lab., ICAR-National Research Centre on Mithun, Medziphema 797 106, Nagaland

(Received: February 2019; Revised: March 2019; Accepted: April 2019)

Abstract

Mithun (*Bos frontalis*) is a bovine species found in the four North-Eastern Hilly States of India *viz*., Nagaland, Arunachal, Manipur and Mizoram. The 770k BovineHDBeadChip (Illumina, CA), which is basically a high-density SNP array was utilized for genomic characterization of four mithun populations and to study genetic diversity present in the populations. SNP genotyping of a total 24 samples representing four mithun populations as per their geographic locations were carried out. The results indicated that the majority of polymorphic SNPs were found to be in HWE. Observed heterozygosity ranged between 0.263±0.040 (Nagaland) to 0.299±0.044 (Arunachal). Inbreeding coefficient ranged from 0.087±0.139 for Nagaland to 0.186±0.053 for Manipur mithun. The result of the Principal Component Analysis (PCA) indicated all the mithuns were distributed together in close proximity and no clear separation of the four populations was observed.The first and second principal components (PC1 and PC2) explained 9.59% and 88.0% of the total variation, respectively. The results of the STRUCTURE analysis also showed no population substructure as inferred from an increasing plot of cross validation error, similar to PCA. In our knowledge, this is the first study on genomic characterization of any Indian livestock species, through analysis of population structure and genetic diversity using high density SNP array. The baseline information generated in our study will help in developing proper strategies for implementing genomic selection and conservation of mithun.

Key words: Mithun population, genomic characterization, SNP array, genetic diversity

Introduction

Mithun is a unique bovine species indigenous to the South-Eastern parts of the Himalayas and is reared in the adjacent mountain ranges in four North Eastern Hill States (Arunachal Pradesh,Nagaland, Manipur and Mizoram) of India. Mithun, also described as 'Gayal' is believed to have its origin from wild Gaur, *Bos gaurus* even though its origin is still uncertain (Simoons and Simoons 1968; Phanchung and Roden 1996; Mukherjee et al. 2012, 2018). Phenotypically mithun looks like a transition between cattle and buffalo with a black coat and typical characteristic of white shanks or "stockings", while the cytogenetic analysis of the mithun chromosomes shows a diploid number 2n=58, compared to 2n=60 in cattle and 2n=48-50 in buffalo (Winter et al. 1984; Gupta et al. 1995; Mukherjee et al. 2012). Mithun is specifically reared by the tribal community as a sign of prosperity, gifts and sacrificial animal and for tribes that consume it as a source of meat (Shisode et al. 2009). Traditionally, mithun is reared on free range, grazing at altitudes of between 300-3000m AMSL on available pastures, jungle fodders, shrubs, herbs and other natural vegetation (Tenzin et al. 2017).

Genetic information on this animal is limited, even if India is having the highest mithun population (around 0.30 million heads) in the world (Anonymous 2012), distributed in small herds in remotely located hilly terrains. This overall small population necessitates an urgency to characterize the mithun populations and to estimate underlying genetic diversity of the species to develop better conservation and utilization efforts. Molecular markers such as single nucleotide polymorphisms are a valuable tool for determining genetic variation, mapping quantitative trait loci (QTL) and identifying genetic diversity because of their dense distribution in the genome, co-dominant inheritance and relative ease of genotyping (McKay 2008). The

*Corresponding author's e-mail: sabayasachimukherje@gmail.com
#Present address: Dairy Cattle Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana 132 001
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

currently available Illumina®BovineHDGenotyping BeadChip panel provides extensive genome coverage and map resolution and has the potential to perform beyond other molecular markers such as microsatellites, resulting in much improved accuracy in finding the exact QTL locations (Matukumalli et al. 2009; Illumina 2012). Thus, the present study was aimed with the objectives to determine the diversity and variation existing in the mithun population using the Illumina®BovineHDGenotyping BeadChip panel so as to generate information which will be useful for ongoing conservation and genetic improvement programme of this livestock species. Use of highly sophisticated and modern genetic techniques can aid in speedy selection of economically important traits which will be useful for the benefits of all the stake-holders. An economically improved mithun will encourage ownership and better scope of utilization by the rural populace, leading to a sustainable *in situ* conservation programme for mithun in North East India.

Materials and methods

The present study included 24 mithun belonging to four mithun strains/populations (F_1, F_3 and F_4 populations) selected at random from a herd of animals maintained at the Institute research farm, Medziphema. The experiment was carried out as per guidelines and regulations dully approved by Institute Animal ethics Committee, ICAR-NRC on Mithun, Nagaland.

DNA isolation and genotyping

The blood samples from experimental animals were collected from jugular vein in EDTA vacutainer tube (BD Vector) by trained veterinarian and were immediately transported to the laboratory in chilled conditions with minimum jerks as soon as possible. DNA was isolated from the samples using a standard blood DNA isolation kit as per the manufacturer's protocol (Promega Cat. No. A1620). The Genomic DNA was quantified to have 50ng/ml concentration to fulfils the requirements for the Illumina® Infinium® SNP genotyping platform. A total of 24 animals were genotyped by Sandor Life sciences Pvt. Ltd., Banjara Hills, Hyderabad, India with the Illumina®Bovine HDGenotyping Bead Chip (777K) assay.

The genotype calls was defined based on the validated standard cluster file provided by the manufacturer. In the present study, only the SNPs on autosome and with known position in cattle genome were considered. The *bostaurus* genome assembly (UMD3.1) was used as the reference genome since reference genome assembly for mithun was not published till then. The quality control was performed using Illumina's Genome Studio software (https://support.illumina.com/array/arraysoftware/ genome studio /documentation.html). The date files (in .ped and .map format) generated from genome studio was used further analysis using Plink 1.9 software.

Pruning of data and genetic analysis

The Genotyping data was pruned with the GenABEL package (Aulchenko et al. 2007) for R (R Development Core Team, Anonymous 2009) using various attributes for filtering the data. The threshold for various attributes were MAF>5% genotypic call rate >95%IBS (= 0.95) for identifying the closely related individuals, HWE>0.001 for the SNP markers showing deviations from equilibrium. The pruning was carried out on the collective dataset prior to generation of Wright's fixation indices (based on fixation index F and weighted F_{ST}), STRUCTURE analysis and genetic distance. To study the genetic diversity of mithun populations, a number of population parameters *viz.*, observed and expected heterozygosity (H_o and H_e), inbreeding coefficients need to be estimated which were carried out using PLINK1.9 software.

Results

Descriptive statistics

Mithun genome (2n=58) differs from that of cattle (2n=60) in terms of diploid chromosome numbers. To compensate this, QC was run with 29 chromosomes, excluding the X chromosome. Initial number of individuals and SNPs before quality check were 24 and 777,962, respectively. After quality check 23 individuals (95.83%) and 127,432 SNPs (16.387%) were used in the analysis. 584705 (79.52%) markers excluded as having low (<5%) minor allele frequency (least common alleles) and 64296 (8.744269%) markers and 1 individual excluded because of low (<95%) call rate.

At Minor Allele Frequency (MAF) pe"0.05 (common allele), SNPs in the Nagaland mithun population revealed lowest minor allele frequency of 85%, while all other populations had no MAF (Table 1). Identity by state (IBS) was least among Mizoram strain (0.813±0.020) and highest among Manipur mithun (0.933±0.057). Inbreeding coefficient ranged from 0.087±0.139 for Nagaland to 0.186±0.053 for Manipur mithun. Observed heterozygosity ranged

Table 1. Basic diversity indices across population based on 127,432 SNPS

	N	H_{ob} ±*SD*	H_{ex} ±*SD*	*Inbreeding f*	IBS±*SD*	% SNPS not in HWE (*P*d"0.05)	% Markers with MAFe"0.05
Arunachal		0.299±0.044	0.268±0.169	0.105±0.132	0.832±0.025	0.854±0.003	100
Manipur		0.267±0.017	0.245±0.183	0.186±0.053	0.933±0.057	0.715±0.004	100
Mizoram		0.287±0.036	0.257±0.171	0.161±0.104	0.813±0.020	0.718±0.004	100
Nagaland		0.263±0.040	0.259±0.263	0.087±0.139	0.857±0.029	1.169±0.001	85.4

Fixation indices and gene inflow

between 0.263±0.040 (Nagaland) to 0.299±0.044 (Arunachal). The majority of polymorphic SNPs were found to be in HWE, Percentage SNP deviation from HWE ($p<0.05$) was least among Manipur mithun (0.715±0.004) and highest among Nagaland mithun (1.169±0.001).

The estimate of global F_{ST} showed that deviations from HWE as a result of inbreeding coefficient (F_{IS}) and in total population were quite high (Table 2). The estimate of F_{IS} was 0.8705, while F_{IT} was 0.8778. The overall variation due to population substructure (F_{ST}) on the other hand was 5.63%. The

Table 2. Global and pair wise fixation indices for mithun strains using 127,432 SNPs

	Arunachal	Manipur	Mizoram	Nagaland
Arunachal	-	0.0658	0.0512	0.0566
Manipur	-	-	0.0808	0.0496
Mizoram	-	-	-	0.0561
Nagaland	-	-	-	-
F_{IS}	0.8705			
F_{IT}	0.8778			
F_{ST}	0.0563			

pair wise F_{ST} reveals that the highest values between Manipur and Mizoram (0.808), while least was recorded between Manipur and Nagaland (0.0496).

Principal component analysis (PCA)

PCA was utilized to understand the structure of mithun population. The result of the PCA generated with Adegenet (Jombart 2008) indicated all the mithun were distributed evenly in close proximity and no clear separation of the four mithun populations was found. The first and second principal components (PC1 and PC2) were explaining 9.59% and 88.0% of the total variation, respectively (Fig. 1).

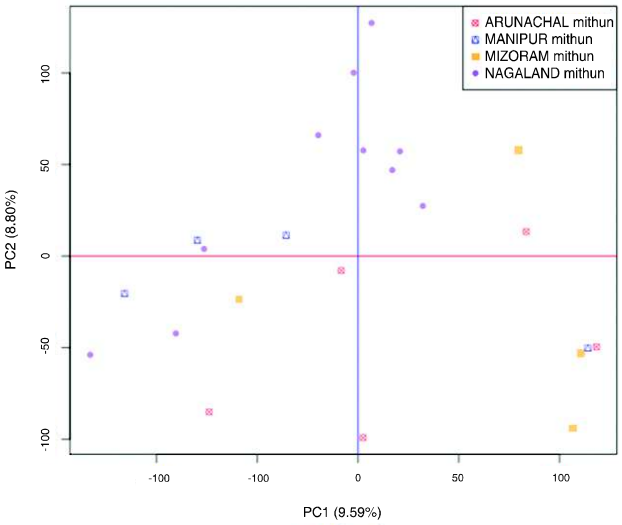

Fig. 1. Population stratification of mithun strains populations based on PCA across 127,432 SNPs. PC1 showing north-south orientation, while PC2 shows an east west orientation

The STRUCTURE analyses (Pritchard et al. 2000) were carried out on random subset of 127,432 SNPs with the admixture unlinked loci model set at 50000 burn-in followed by 20,000-100000 Markov Chain Monte Carlo (MCMC) repetitions, assuming K = 2-5, for all the mithun data sets (Fig. 2). This showed presence of ancestral populations.

The result further indicated there was no specific population substructure which can be inferred from plotting the cross-validation errors (Alexander et al. 2013). There was increasing trends (Fig. 3).

Genetic distances and phylogenetic tree

The results of the phylogenetic relationship and genetic distance between and within the four mithun populations based on Nei's genetic distance is shown in Table 3. Within and between breed genetic distances were quite similar. Average genetic distance across

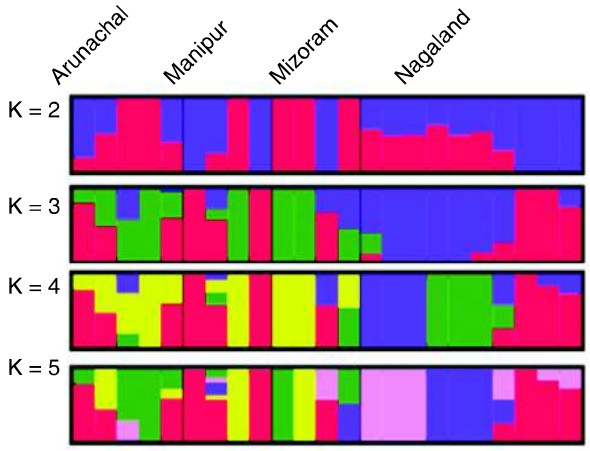

Fig. 2. Estimated Population Structure using 127,432 SNPs generated by ADMIX for *K*=2-5

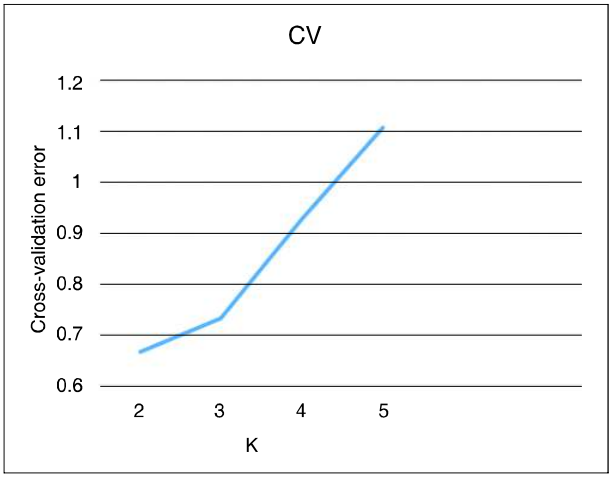

Fig. 3. Plot of cross validation error for *k* for the population structure analysis

all groups is 0.244. Within genetic distance does not vary much in each population. Nagaland had the least within group genetic distance (0.227), while Mizoram had the highest (0.248). Between groups genetic distance shows Manipur and Mizoram exhibiting highest distance (0.255) and least value of 0.238 between Arunachal and Mizoram. Net genetic distance was also highest with value 0.0013 between Mizoram–Manipur and Mizoram-Nagaland, while least was Mizoram-Arunachal (-0.008). A neighbour-joining tree was constructed to depict the genetic distance among the mithun populations using MEGA 5.2 (Fig. 4). This showed no specific clustering of the 4 populations which is also supported by earlier results through the STRUCTURE analysis.

Discussion

Bos frontalis is a unique livestock species, historically

Table 3. Nei's genetic distance (lower diagonal) and average net genetic distance (upper diagonal) between groups from 127,432 SNPs among the four mithun populations

	Aruna-chal	Mani-pur	Mizoram	Naga-land	Average genetic distance
Arunachal	-	0.008	-0.008	0.019	
Manipur	0.248	-	0.013	0.009	
Mizoram	0.238	0.255	-	0.013	
Nagaland	0.254	0.241	0.251	-	
Within group	0.244	0.237	0.248	0.227	0.244

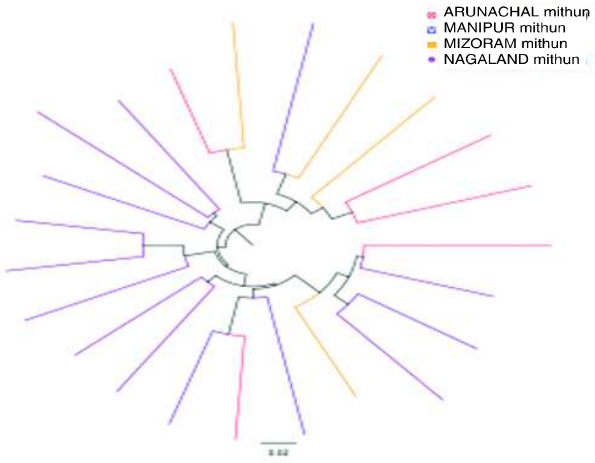

Fig. 4. Neighbour-joining tree reconstructed using MEGA 5.2 software from 127,432 SNPs among mithun populations

reared by tribals in a few remote hilly terrains of North Eastern Region of India. The preservation of such a species will preserve the culture and livelihood of the people in this region. Conservation and utilization programs need *a priori* information on diversity and variation within the population assessed using molecular markers. Among such markers SNPs are the most densely populated across the genome and easily assayed with reproducible results (Gibbs and Singleton 2006). A large number of SNPs have been identified from the bovine genome-sequencing project. These were utilized for the development of Illumina BovineHDBeadChip assay by taking samples different bovine species including gaur (Illumina Data Sheet, 2015).However, use of such a sophisticated SNP genotyping assay in mithun (*Bos frontalis*) was scanty so far (Uzzaman et al. 2014; Mukherjee et al. 2018). Prior to this study, no SNPs have been described for these populations individually. In this study, we reported the first preliminary findings on genomic

diversity of the four populations of mithun, and estimated both within and between populations variability in North-East Region of India using Illumina SNPs assay.

The mithun population diversity was tested based on Wright's global fixation indices, population sub-structure. Genetic distances within and between populations were also evaluated. The farther populations are from each other geographically the less they mate or breed with other, resulting in less alleles shared among the population. With time the population differentiates from each other and becomes fixed to their separate allele frequency. As a result of inbreeding within the population it soon becomes homogenous with time and fixed to a theoretical single allele frequency with high observed homozygosity and low heterozygosity. The Wrights overall fixation index (F_{ST}) recorded here showed low value (0.0563). F_{ST} is reported to generate low value among high diverse, unselected populations viewed as substantially distinct and among bi-allelic single nucleotide polymorphic markers when compared to microsatellite or others markers (Jakobsson et al. 2013). The paired F_{ST} values signify level of shared alleles as a result of present or past within paired populations breeding. Values show more relatedness among the mithun populations which should more or less be expected most likely due to spatial distance. The variation reported is also supported by the values of observed (H_O) and expected (H_E) heterozygosity found in this study (H_O ranged between 0.263±0.040 in Nagaland to 0.299±0.044 in Arunachal) showing the panmitic nature of the populations. The populations also exhibit some level of inbreeding (ranging between 0.087±0.139 in Nagaland to 0.186±0.053 in Manipur populations, Table 1) which might have been responsible for slight deviations from HWE (Mukherjee et al. 2018). The results indicate that the mithun population were having high within population heterozygousity. Between and net Nei's genetic distance showed low divergence between the mithun populations. Both within and between divergences were similar, this might be due to the fact that there was no population sub-structure in spite of geographical separation, and no separate selection pressures on individual population was available.

This preliminary study being the first of its kind on four different populations of mithun has been able to show that 80% of the SNPs on the array have MAF < 0.05, so approximately 20% were polymorphic (1/5). The present study generated valuable genomic information on mithun population found in North Eastern Hill Region of India regarding diversity and population structure. The mithun population show little divergence from each other based on the polymorphic loci and can be concluded that there is no sub-structure present in the mithun population based on geographical locations and there is fairly uniformity in the genetic make-up of these populations. The results obtained in the present study will be helpful in devising suitable breeding policy of mithun for conservation and genetic improvement programme.

Authors' contribution

Conceptualization of research (AM,SM, CR); Designing of the experiments (AM,SM); Contribution of experimental materials (NH,KK,KV); Execution of field/lab experiments and data collection (SM,IL,AM, MM); Analysis of data and interpretation (AM, SM);Preparation of manuscript (AM,SM)

Declaration

The Authors declare no conflict of interests.

Acknowledgement

The authors gratefully acknowledge the Director of the Institute for providing all the facilities to carry out this work under one Institute funded research project of the first author. Lab. facility rendered by the Institute Biotech Hub, Dept. of Biotechnology (DBT), Govt. of India to carry out some of the experimental work is thankfully acknowledged.

References

Alexander D. H., November J. and Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res., **19**: 1655-1664.

Anoymous. 2009. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. http://www.R-project.org (accessed on 12.12.2017).

Anonymous. 2012. Annual report of the National Research Centre on Mithun, 2010-20. ICAR-National Research Centre on Mithun. Jharnapani, Nagaland, India. Pp.**84**.

Aulchenko Y. S., Ripke S., Isaacs A. and Van Duijn C. M. 2007. GenABEL: an R library for genome-wide association analysis. Bioinformatics, 2007, **23**(10): 1294-1296.

Benjamini Yoav and Hochberg and Yosef. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Royal

Statistical Society, Series B, **57**(1): 289-300.

Georges M. 2007. Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. Annu. Rev. Genom. Hum., G., **8**: 131.

Gibbs J. R. and Singleton A. 2006. Application of Genome-Wide Single Nucleotide Polymorphism Typing: Simple Association and Beyond. PLoS Genet., **2**: e150. doi: 10.1371/journal.pgen.0020150.

Goddard M. E. and Hayes B. J. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet., **10**: 381-391.

Gupta N., Verma N. D. Gupta S. C., Kumar P. and Sahail R. 1995. Chromosomes of Mithun (*Gaveaus frontalis*). Indian J. Anim. Sci., **65**(6): 688-690.

Hu Z. L., Park C. A., Wu X. L. and Reecy J. M. 2012. Animal QTLdb: An improved database tool for livestock animal QTL/association data dissemination in the post-genome era. Nucleic Acids Res., 2012.

Illumina. 2012. BovineHD genotyping bead chip data sheet. http://res.illumine.com/documents/products/datasheets/datasheet bovinehd.pdf. Pg: 4. (accessed on 12.12.2014).

Jakobsson M., Edge M. D. and Rosenberg N. A. 2013. The relationship between Est and the frequency of the most frequent allele. Genetics, **193**: 515-528.

Jombart T. 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, **24**: 1403-1405.

Kinsella R. J., Kahari J., Staines D., Derwent P., Kerhornou A., Kersey P. and Flicek P. 2011. Ensembl BioMarts: A hub for data retrieval across taxonomic space. Database (Oxford): Bar030.

Ma L., Runesha H. B., Dvorkin D., Garbe J. R. and Da Y. 2008. Parallel and serial computing tools for testing single-locus and epistatic SNP effects of quantitative traits in genome-wide association studies. BMC Bioinformatics, **9**: 315.

McKay Stephanie D., Schnabel Robert D., Murdoch Brenda M., Matukumalli Lakshmi K., Aerts Jan, Coppieters W., Crews Denny, Neto Emmanuel Dias, Gill Clare A., Gao Chuan, Mannen Hideyuki, Wang Zhiquan, Tassell Curt P. Van, Williams John L., Taylor Jeremy F. and Moore Stephen S. 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genetics, **9**, 37, https://doi.org/10.1186/1471-2156-9-37.

Mukherkee G., Jha A., Basu S. Kumar, Rajkhowa C., Biswas Palab Kumar, Goswami P., Chowdhurry B. D. and Roy Debananda. 2012. Mithun: A potential resource for leather industry – A humble exploration. In: Strategies for propagation and augmenting productivity of mithun in north eastern hill region (Eds. Nazrul Haque and K. K. Baruah), 2012. ICAR-National Research Centre on Mithun, Jharnapani, Medziphema, Nagaland, Pp. 65-83.

Mukherjee Anupama, Mukherjee Sabyasachi, Rajan Dhakal, Moonmoon Mech, Imsusosang Longkumer, Haque N., Kobu Khate, Kezhavituo, Yanger Jamir, Pursenla Pongen, Chandan Rajkhowa, Abhijit Mitra, Bernt Guldbrandtsen and Goutam Sahana. 2018. High Density Genotyping reveals Genomic Characterization Population Structure Genetic Diversity of Indian Mithun (*Bos frontalis*). Scientific Reports, **8**: 10316. DOI: 10.1038/s41598-018-28718-x.

Mukherjee Sabyasachi, Mukherjee Anupama and Longkumer Imsusosang. 2012. Breeding for genetic improvement in Mithun. In: Strategies for propagation and augmenting productivity of Mithun in north eastern hill region (Eds. S. Mukherjee, A.Mukherjee, M. Mondal, N. Haque and K.K Baruah). ICAR-National Research Centre on Mithun, Jharnapani, Medziphema, Nagaland, India. pp.11-19.

Matukumalli L. K., Lawley C. T., Schnabel R. D., Taylor J. F., Allan M. F., Heaton M. P., Jeff O'Connell, Stephen S. Moore, Timothy P. L. Smith, Tad S. Sonstegard, Curtis P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS One, **4**: doi: 10.1371/journal.pone.0005350.

Phanchung and Roden J. A. 1996. Characterisation of the Siri breed and the Mithun cross Siri in Bhutan. AGRI 1996, **20**: 27-34.

Shisode M. G., Khanvilkar A. V., Kulkarni M. D., Samant S. R., Yadav G. B. and Bawaskar M. S. 2009. Mithun: The Pride animal of North-eastern hilly region of India. Vet. World, **2**: 480-48.

Simoons F. J. and Simoons E. S. 1968. A ceremonial ox of India. The mithun in nature, culture and history. 3-13 (The University of Wisconsin Press, Madison, WI, USA).

Tenzin S., Dorji J., Dorji T. and Kawamoto Y. 2017. Assessment of genetic diversity of Mithun (*Bos frontalis*) population in Bhutan using microsatellite DNA markers. Animal Genetic Resources/ *Ressourcesgénétiquesanimales/Recursosgenéticos animales* **59**: 1-6, https://doi.org/10.1017/s2078633616000072.

Uzzaman Md. Rasel, Edea Zewdu, Bhuiyan Md. Shamsul Alam, Walker Jeremy, Bhuiyan A. K. F. H. and Kim Kwan-Suk Kim. 2014. Genome-wide Single Nucleotide Polymorphism Analyses Reveal Genetic Diversity and Structure of Wild and Domestic Cattle in Bangladesh. Asian Australas. J. Anim. Sci., **27**: 1381-1386

Winter H., Mayr B., Schleger W., Dworak E., Krutzler J. and Burger B. 1984. Karyotyping, red blood cell and haemoglobin typing of the Mithun (*Bos frontalis*), its wild ancestor and its hybrids. Res. Vet. Sci., **36**: 276-283.