

# SCIENTIFIC REPORTS

OPEN

## High-density Genotyping reveals Genomic Characterization, Population Structure and Genetic Diversity of Indian Mithun (*Bos frontalis*)

Anupama Mukherjee<sup>1,3</sup>, Sabyasachi Mukherjee<sup>1</sup>, Rajan Dhakal<sup>2</sup>, Moonmoon Mech<sup>1</sup>, Imsusosang Longkumer<sup>1</sup>, Nazrul Haque<sup>1</sup>, Kezhavituo Vupru<sup>1</sup>, Kobu Khate<sup>1</sup>, I. Yanger Jamir<sup>1</sup>, Pursenla Pongen<sup>1</sup>, Chandan Rajkhowa<sup>1</sup>, Abhijit Mitra<sup>1</sup>, Bernt Gulbrandsen<sup>2</sup> & Goutam Sahana<sup>2</sup>

The current study aimed at genomic characterization and improved understanding of genetic diversity of two Indian mithun populations (both farm, 48 animals and field, 24 animals) using genome wide genotype data generated with Illumina BovineHD BeadChip. Eight additional populations of taurine cattle (Holstein and NDama), indicine cattle (Gir) and other evolutionarily closely related species (Bali cattle, Yak, Bison, Gaur and wild buffalo) were also included in this analysis (N = 137) for comparative purposes. Our results show that the genetic background of mithun populations was uniform with few possible signs of indicine admixture. In general, observed and expected heterozygosities were quite similar in these two populations. We also observed increased frequencies of small-sized runs of homozygosity (ROH) in the farm population compared to field mithuns. On the other hand, longer ROH were more frequent in field mithuns, which suggests recent founder effects and subsequent genetic drift due to close breeding in farmer herds. This represents the first study providing genetic evidence about the population structure and genomic diversity of Indian mithun. The information generated will be utilized for devising suitable breeding and conservation programme for mithun, an endangered bovine species in India.

Mithun (*Bos frontalis*), also known as gayal, a unique bovine species, has a limited geographical distribution primarily restricted to the North-Eastern Hilly (NEH) region of India, Myanmar, hilly provinces of Bangladesh, Bhutan and Yunan province of China<sup>1</sup>. Though it is difficult to ascertain the actual population of Mithun in the world, India possesses 0.30 million mithun which constitutes ~97.57% of the world population (19<sup>th</sup> Livestock Census, 2012)<sup>2</sup>. Mithun has a very special socio-cultural status among the indigenous tribal population of NEH region of India. Having the very high value in the barter system, owning mithun is considered as a sign of prosperity.

Mithun are traditionally raised as a meat animal in a free-range system in the sub-tropical rain forest with almost zero input. As a part of socio-cultural practices, mithuns are sacrificed during rituals and festivals to offer feast<sup>3</sup>. Mithun meat is one of the most preferred sources of animal protein among local tribes. Compared to the meat from local cattle and buffalo, mithun meat is preferred due to both quality (better marbling, finer texture, and tenderness) and quantity (higher dressing %)<sup>4-6</sup>. Mithun have higher feed conversion efficiency than local

<sup>1</sup>Animal Genetics and Breeding Lab., ICAR-National Research Centre on Mithun, Medziphema, Nagaland, 797106, India. <sup>2</sup>Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830, Tjele, Denmark. <sup>3</sup>Present address: Dairy Cattle Breeding Division, ICAR-National Dairy Research Institute, Karnal, Haryana, 132001, India. Anupama Mukherjee and Sabyasachi Mukherjee contributed equally to this work. Correspondence and requests for materials should be addressed to S.M. (email: [sabayasachimukherje@gmail.com](mailto:sabayasachimukherje@gmail.com))

Sl. No.	Name/Code used	Location of collected samples	Sample size	
			Admixture analysis	PCA analysis
1	Mithun – farm population ( <i>Bos frontalis</i> )/Farm	India	48	20
2	Mithun – field population ( <i>Bos frontalis</i> )/Field	India	24	20
3	NDama ( <i>Bos taurus</i> )/NDA	Africa	23	20
4	Bali cattle ( <i>Bos javanicus</i> )/BLI	Indonesia (Bali)	20	20
5	Gir ( <i>Bos indicus</i> )/ GIR	Brazil	50	20
6	Holstein ( <i>Bos taurus</i> )/HOL	Denmark	20	20
7	Wild buffalo ( <i>Bubalus depressicornis</i> )/OWB	Africa	10	10
8	Yak ( <i>Bos grunniens</i> )/OYK	China	4	4
9	Bison ( <i>Bison bison</i> )/OBB	USA	4	4
10	Gaur ( <i>Bos gaurus</i> )/OGR	USA	6	6

**Table 1.** Species/breeds included in the ADMIXTURE and PCA analyses genotyped with Illumina BovineHD BeadChip (HD) or BovineSNP50 (50 k).

cattle<sup>7</sup>. The current population size of mithun in India is not large. However, considering its importance in widening the biodiversity base and adaptation to a humid sub-tropical climate and a hilly topography, it could be exploited as an alternative means of livelihood as well as for improved food and nutrition security in its present habitat and in similar environments elsewhere.

Genetic improvement of mithun, however, is impeded by the way mithun is presently reared. Under traditional system of rearing, mithuns are let loose in the forest with minimal intervention. Recording of traits and pedigree information is almost non-existent. The size of a typical mithun herds ranges from 50 to 100. Herds are generally served by a few dominant bulls. The mithun population may therefore suffer from inbreeding. Under these circumstances, the genetic diversity of the mithun population needs to be described. This will assist ongoing mithun conservation and genetic improvement programme and for devising a suitable breeding policy.

Among a wide range of molecular markers developed, single nucleotide polymorphisms (SNPs) are the most abundant, widely dispersed throughout genomes, and have variable distribution among species<sup>8</sup>. The availability of high-throughput SNP genotyping platforms makes it feasible to undertake high-resolution scans by using large numbers of SNP markers distributed across the whole genome. SNPs are useful in studying livestock genetic diversity and population structure<sup>8,9</sup>. Although a large number of SNPs have been identified in bovine genome-sequencing projects, few of these have been validated outside *Bos taurus*, as for example in mithun<sup>10</sup>. The Illumina BovineHD BeadChip (Illumina, San Diego, CA) with 777k SNPs was introduced and utilized for genotyping studies in various breeds of cattle, but has also been used in other members of the bovidae, for example, yak, gaur, buffalo and wild anoa<sup>10–16</sup>.

Other bovine species including indigenous cattle, yak and gaur are also found in the North Eastern parts of India along with mithun. Forest cover and natural habitats of mithuns in the North Eastern Hill States of India have shrunk over the years. This raises concerns about possible introgression of local cattle with mithun, sharing the same habitat. Sporadic instances of crossing mithuns with cattle bulls by mithun owners in the field to increase milk production have also been recorded<sup>17</sup>. Hence, we used the unsupervised clustering analysis (carried out with the ADMIXTURE software)<sup>18</sup> to estimate individuals' ancestries from SNP genotypes to assess the extent of admixture in mithun with other cattle species (Table 1).

Therefore, the present study aimed to generate high-resolution information on the genomic diversity and population structure of mithun using Illumina BovineHD BeadChip. To our knowledge, this is the first study using BovineHD BeadChip in mithun. We further studied the evolutionary relationship of mithun with closely related bovine species to understand mithun phylogeny and origin.

## Results

**Population diversity parameters - Observed  $H_o$ , expected  $H_e$ ,  $F_{IS}$  and ROH.** Observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities estimated in the Indian mithun population (farm and field population) ranged from 0.25 to 0.17, and 0.25 to 0.18, respectively (Table 2). The inbreeding coefficient estimates ( $F_{IS}$ ) based on the observed versus expected number of homozygous genotypes, were found to be  $0.06 \pm 0.02$  and  $0.02 \pm 0.01$  for the farm and field animals (Table 2). The mean  $F_{ST}$  estimate between farm and field mithun populations was  $0.03 \pm 0.01$ .

**Runs of homozygosity.** Runs of homozygosity (ROH) in the autosomes of 48 farm mithun and 24 field mithun animals were determined using PLINK1.9<sup>19</sup> and consisted of 139,350 SNPs in each of the farm and field data set after quality control.

The total length of ROH per animal was averaged within population in six sized windows: 250 kb–1 Mb, 1 Mb–2 Mb, 2 Mb–4 Mb, 4 Mb–8 Mb, 8 Mb–16 Mb and >16 Mb. The average total length of ROH per animal was 823.7, 242.1, 92.30, 47.03, 66.22 and 110.9 Mb in the farm and 735.8, 245.0, 121.3, 84.4, 86.0 and 87.4 Mb in the field population for these six length categories.

Summary statistics of ROH observed are outlined in Table 3 (farm) and 4 (field). Average numbers of ROH per animal for these six categories were 1793.0, 182.8, 35.3, 8.9, 6.0 and 4.25 in farm and 1596.0, 181.9, 45.5, 15.5,

Population	Location	No. of animals	Average $H_o$	Average $H_e$	Inbreeding Coefficient ( $F_{IS}$ )
Research- farm (Farm)	Nagaland	48 (28 males and 20 females)	$0.25 \pm 0.08$	$0.25 \pm 0.07$	$0.06 \pm 0.02$
Farmers' herds (Field)	Nagaland, Arunachal Pradesh, Manipur, Mizoram	24 (11 males and 13 females)	$0.17 \pm 0.03$	$0.18 \pm 0.03$	$0.02 \pm 0.01$

**Table 2.** Geographic location, sample size (N), composition of mithuns and diversity parameters in mithun population. Average observed and expected heterozygosities are indicated as  $H_o$  and  $H_e$ , respectively.

	Statistics	ROH length category (Mb)					
		250 kb-1 Mb	1-2 Mb	2-4 Mb	4-8 Mb	8-16 Mb	>16 Mb
Length of ROH per animal (Mb)	Mean	823.7	242.1	92.30	47.03	66.22	110.9
	SD	85.3	42.1	39.3	34.2	54.6	134.7
	Min	610.7	137.8	25.31	4.52	8.2	16.7
	Max	971.6	322.8	179.8	138.1	205.6	498.6
Number of ROH per animal	Mean	1793.0	182.8	35.3	8.9	6.04	4.25
	SD	199.8	30.1	14.5	6.2	4.7	5.02
	Min	1313.0	108.0	10.0	1.0	1.0	1.00
	Max	2125.0	240.0	69.0	26.0	18.0	19.00
$F_{ROH}$	Mean	0.335	0.098	0.038	0.019	0.027	0.045
	SD	0.035	0.017	0.016	0.014	0.022	0.007
	Min	0.248	0.056	0.010	0.002	0.003	0.007
	Max	0.395	0.131	0.073	0.056	0.084	0.202

**Table 3.** ROH and  $F_{ROH}$  in the farm mithun population.

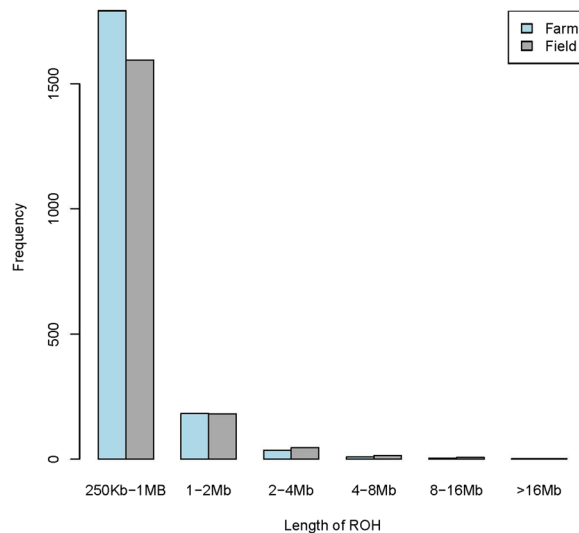
	Statistics	ROH length category (Mb)					
		250 kb-1 Mb	1-2 Mb	2-4 Mb	4-8 Mb	8-16 Mb	>16 Mb
Length of ROH per animal (Mb)	Mean	735.8	245.0	121.3	84.4	86.0	87.4
	SD	110.8	679.8	460.4	372.0	559.7	59.4
	Min	400.6	111.9	38.1	14.6	10.2	17.3
	Max	893.7	499.2	256.3	166.6	198.3	187.4
Number of ROH per animal	Mean	1596.0	181.9	45.5	15.5	7.7	3.6
	SD	235.7	48.9	17.1	6.7	5.0	2.4
	Min	894.0	82.0	15.0	3.0	1.0	1.0
	Max	1987.0	362.0	96.0	32.0	17.0	7.0
$F_{ROH}$	Mean	0.294	0.099	0.049	0.034	0.034	0.035
	SD	0.045	0.027	0.019	0.015	0.022	0.024
	Min	0.007	0.045	0.015	0.005	0.004	0.007
	Max	0.076	0.202	0.104	0.067	0.080	0.076

**Table 4.** ROH and  $F_{ROH}$  in the field mithun population.

7.7 and 3.6, respectively in field population. Patterns of ROH are roughly similar up to about 2 Mb of length corresponding to more than about 25 generations ago. For ROH between 2 and 16 Mb (roughly 25–26 generations ago) the field population has more ROH. Proportions of very long ROH are higher in the farm populations. This suggest an older history that may be shared between the farm and field populations, a period of relative depression of the effective population size for the field population, followed by a period of relatively higher close breeding in the farm population. The largest total length of autosomal ROH observed corresponded to approximately 27% of the genome length for farm and 23% for field animals. The average number of ROH of different lengths in farm and field mithuns are plotted to show their frequency distribution (Fig. 1).

**Admixture Analysis.** Unsupervised clustering analysis using ADMIXTURE software program were carried out for 10 populations (total 209 samples; Table 1) including one indicine cattle breed, Gir and two taurine breeds (Holstein and NDama) to identify possible indicine and taurine introgression in the mithun population. The analysis was done using 50k markers.

The admixture analysis results of farm and field mithuns do not show any population structure (details not presented). The admixture analysis including other species grouped mithuns from farm and field together, and



**Figure 1.** Average number of ROH for different lengths in farm and field mithuns.

were fairly homogeneous (Fig. 2,  $K = 4-7$ ). With  $K = 4$ , we observed a strong genetic differentiation between mithun (colored in green) and the remaining bovine species included in this study, except gaur (OGR), yak (OYK) and bison (OBB), which shared a common genetic component. At  $K = 7$ , mithun was distinguished from all other bovine species except gaur. There were minute levels of indicine and taurine introgression in the mithun population. A plot using cross-validation errors (CVE) was drawn for the SNP dataset used for admixture analysis (Fig. 3), showing overall uniform values of CVE beyond  $K = 6$ .

**Principal Component Analysis.** The principal component analysis was done using 50k markers for eight *Bos* species and two mithun populations (farm and field). The first two principal components (PC1 and PC2) grouped farm and field mithuns in a cluster with gaur, yak and bison, separated from three other major clusters of two taurine cattle (NDama and Holstein), and indicine cattle (Gir; Fig. 4).

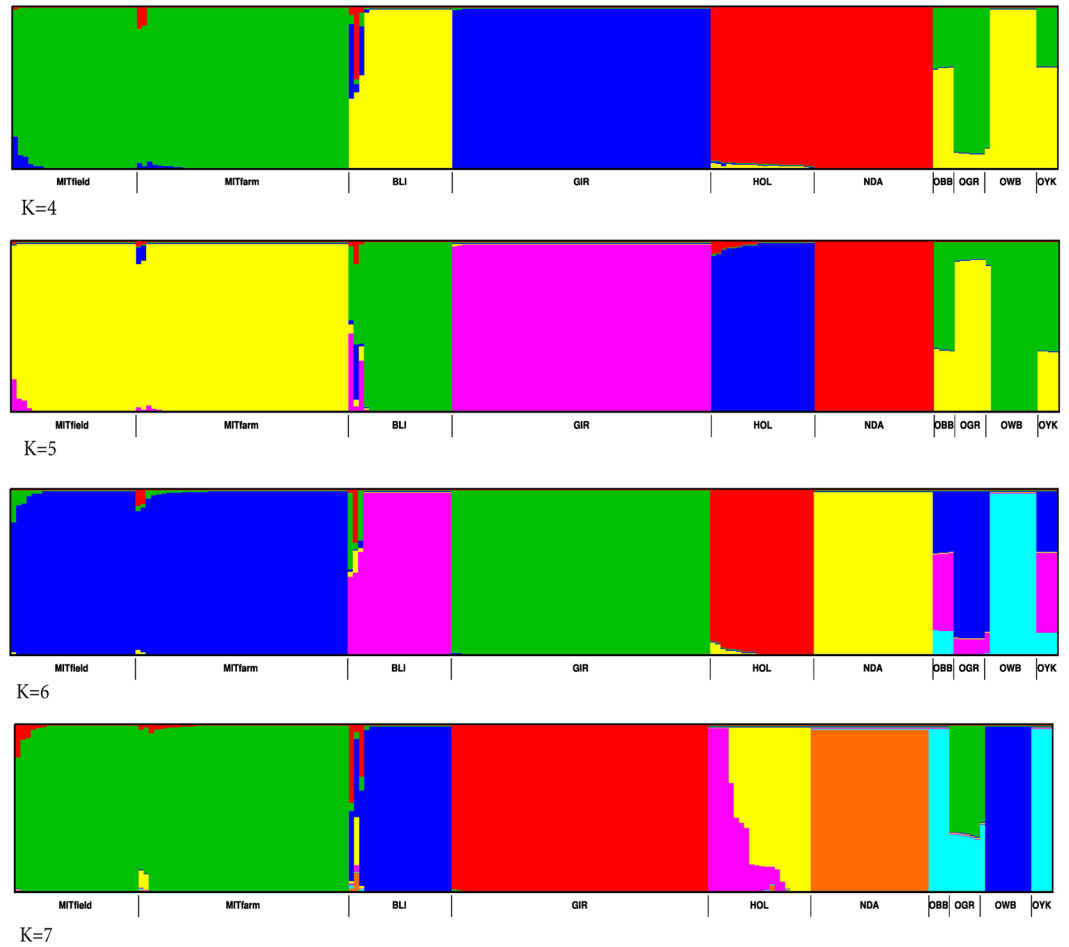
Principal component 1 (PC1) positioned mithuns along with gaur, yak, bison and Bali cattle in distinct clusters from indicine and taurine cattle populations. Principal component 2 (PC2) separated two taurine breeds into separate clusters. Principal component 3 (PC3) further separated Gir from other populations (Fig. 5).

**Phylogenetic analysis using Treemix.** Treemix<sup>20</sup> analysis was used to study the population splits and subsequent gene flows. We constructed a phylogenetic tree of all the bovine species without adding any migrations and assuming wild buffalo (OWB) as outgroup using Treemix analysis (Fig. 6). Treemix reveals no evidence of admixture (the smallest test statistics was +2.76). We also analysed to fit 0–5 migrations in Treemix, but there was no evidence of migration (details not presented). While two mithun populations and gaur were in one single clade, the indicine breed, Gir, was in one group, two taurine breeds were in one group, Bali cattle was in a group by itself, and yak and bison formed another group. The taurine breeds Holstein and NDama animals formed a separate group. Gir, an indicine cattle breed, branched out from taurine as expected.

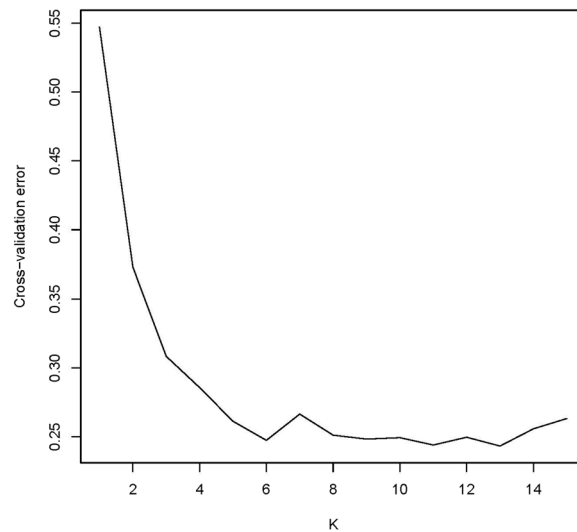
## Discussion

India has the largest mithun population (~97.57%) in the world (0.30 million, 19<sup>th</sup> Livestock Census, 2012). Myanmar (0.96%, approx. 3000), Bangladesh (0.32%, approx. 1000), China (0.96%, approx. 3000) and Bhutan (0.18%, approx. 570) also home small populations of mithun in Asia<sup>2</sup>. Information regarding genetic diversity and population structure of Indian mithun remains scanty. A conservation programme for mithun has been initiated at Khonoma and Thevopisu villages, Nagaland in their native breeding tracts in India. Genetic improvement in mithun through a systematic breeding program could not be initiated under field conditions due to the small number of animals, uncontrolled mating, and their scattered distribution across remote and inhospitable hilly terrain.

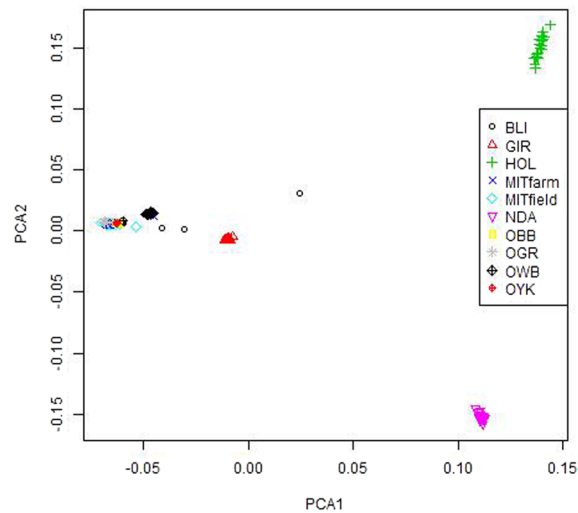
Genetic diversity of two groups of Indian mithun population (farm and field groups) were assessed by expected heterozygosity ( $H_e$ ), observed heterozygosity ( $H_o$ ) and estimates of inbreeding ( $F_{IS}$  and  $F_{ST}$ ).  $F_{ST}$  between two populations was low (0.03) indicating a close genetic connection between them<sup>21</sup>. Mithun is assumed to be a close evolutionary relative of gaur<sup>3,22,23</sup>. Hence, the BovineHD BeadChip was used to genotype mithuns and available literature on cattle was referred for comparison in the present study. Genetic diversity in terms of average heterozygosity in both mithun populations (0.17–0.25) was found to be similar to zebu and African taurine cattle<sup>14</sup>, and was in the same range as for taurine cattle from West Africa (from 0.18 for African Lagune to 0.22 for Somba cattle) or East Africa (0.24 for Sheko cattle)<sup>10</sup>. However, SNPs that are included in the analyses are polymorphic within Mithun, but originate from bovine and are thus expected to be old SNPs (polymorphic in the common ancestor of Mithun and *Bos taurus*) leading to biased allele frequency spectrum.



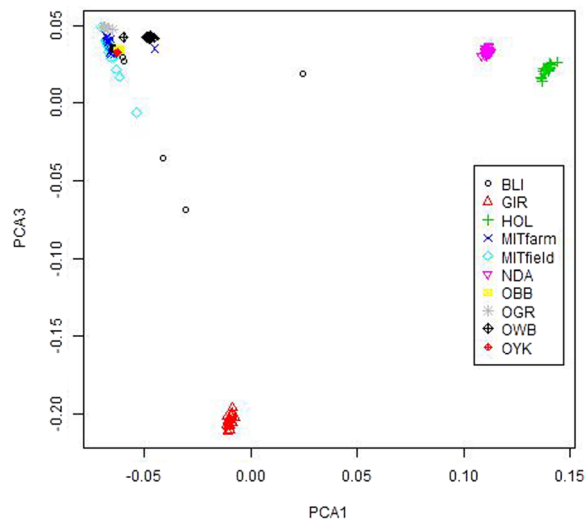
**Figure 2.** Plot of admixture analyses results for  $K = 4$  to  $7$  using 38,587 autosomal SNP.  $K = 6$  yielded the lowest cross-validation error. Abbreviations: MITfarm: Farm mithun, MITfield: Field mithun, BLI: Bali cattle, GIR: Gir cattle, HOL: Holstein, NDA: NDama; OBB: Bison; OGR: Gaur; OWB: Wild buffalo, and OYK: Yak.



**Figure 3.** Cross-validation errors for the SNP dataset used for admixture analysis for varying numbers of population components ( $K$ ).



**Figure 4.** Principal components analysis of Indian mithun along with other bovine species based on autosomal SNPs. PCA1 and PCA2 explained 36.6% and 11.8% of the variance respectively. MITfarm: Farm mithun, MITfield: Field mithun, BLI: Bali cattle, GIR: Gir cattle, HOL: Holstein, NDA: NDama; OBB: Bison; OGR: Gaur; OWB: Wild buffalo, and OYK: Yak.

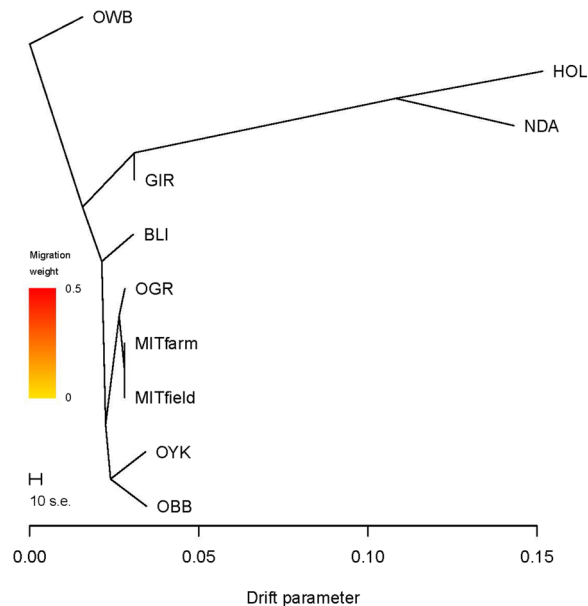


**Figure 5.** Plot of principal components 1 and 3 positioned Indian mithuns along with OGR: gaur, OYK: yak, and OBB: bison in distinct cluster from major taurine and indicine cattle populations. PCA1 and PCA3 explained 36.6% and 7.4% of the variance, respectively.

We used  $F_{IS}$  as a measure to study inbreeding within farm and field mithun populations. Estimated  $F_{IS}$  values indicated low levels of inbreeding.  $F_{IS}$  was lower in the field (0.02) than in the farm population (0.06). We see some evidence of mating between relatives in the farm population, while there is little evidence of mating among close relative in the field population. The  $F_{IS}$  values in our study were similar to that reported<sup>24</sup> in mithuns from Bangladesh using the Illumina BovineSNP50 BeadChip, mean expected heterozygosity of  $0.148 \pm 0.14$  with a heterozygote deficiency of 0.06 ( $F_{IS}$ ).

Lower heterozygosity values  $H_o$  and  $H_e$  of mithun population in our study compared to taurine and indicine cattle breeds as reported by the Bovine HapMap Consortium<sup>12,13</sup> and Korean cattle populations<sup>16</sup> may also be attributed to ascertainment bias in SNP discovery of the BovineHD BeadChip. Previously low heterozygosity estimates of Tunisian cattle populations were likewise attributed to ascertainment bias in the design of the BovineSNP50 BeadChip as polymorphic sites of African origin present in the genome of Tunisian cattle were not included in the chip<sup>14</sup>. The Bovine HapMap Consortium reported that nucleotide diversity in (Brahman) cattle, a crossbred of indicine breeds, is more than twice that observed within Holstein and Angus breeds<sup>25,26</sup> as expected.

Our study revealed a higher number of short and medium ROH (250 kb–1 Mb, 1–2 Mb, 2–4 Mb, and 4–8 Mb) than longer categories (8–16 Mb and >16 Mb) in both the farm and field mithun population. This reflects ancestors shared between the parents long ago, can be indicative of selective sweep, ancient inbreeding or bottleneck.



**Figure 6.** Phylogenetic tree to infer the position of mithun among eight bovine species. The bootstrap support for all the branching ranged between 99 and 100%.

The farm population showed greater ROH length and numbers compared to field population, which was as per expectation due to higher level of inbreeding and small population size parental population in the farm. Similar observations in various cattle breeds were reported as short ROH are generally due to older haplotype relatedness, while longer ROH result from more recent inbreeding<sup>27–29</sup>.

Length and frequency of ROH provide information about demographic history as well as recent inbreeding in individuals<sup>27,30,31</sup>. Long ROH indicates consanguinity between an individual's parents<sup>27</sup>. Shorter ROH indicate more distant demographic and selective events after repeated fragmentation of chromosome segments by recombination<sup>27</sup>. In particular, recent inbreeding resulting from the mating of closely-related ancestors leads to a high occurrence of long ROH. On the other hand, very long ROH sometimes occur in outbred populations<sup>32</sup>.

In human, short ROH (<1.5 Mb) is due to ancient linkage disequilibrium through inheriting parental common haplotypes, whereas, bigger ROH is due the history of related parents in recent times<sup>33</sup>. Level of inbreeding can be estimated accurately in cattle population without knowledge of the pedigree using Illumina BovineHD SNP genotyping assay<sup>34</sup>.  $F_{ROH}$ , the genomic inbreeding coefficient, was found to be highly correlated with pedigree inbreeding in a small group of populations<sup>30,34</sup>.

Estimates of mean  $F_{ROH}$  for the various categories (250 kb–1 Mb, 1 Mb–2 Mb, 2 Mb–4 Mb, 4 Mb–8 Mb, 8 Mb–16 Mb and >16 Mb) in the farm and field mithun population were generally low, probably indicating larger effective population sizes in the past. We observed relatively higher mean  $F_{ROH}$  for the 250 kb–1 Mb category in the farm mithun (0.335). Such short ROH reflect events at least 25 generations ago. With generation interval of approximately 5–10 years in mithun, this translates to selection 100 or 200 years in the past. It seems that probably mithun population has experienced a population bottleneck in the past.

Our results fit well with other genotyping studies in various breeds of cattle using Illumina BovineHD BeadChip and validated through Illumina BovineSNP50 BeadChip, viz. average autozygosity calculated from ROH ( $F_{ROH}$ ) with lengths above 1 Mb was in intensely selected Brown Swiss breed (0.151–0.156)<sup>33,35</sup> and in Holstein cattle (0.116)<sup>29</sup>, while lower  $F_{ROH}$  for unselected cattle breeds Pinzgauer and Tyrolean Grey domestic cattle (0.062 and 0.066, respectively)<sup>35</sup>, and the lowest in unselected, preserved Polish Red breed (0.057)<sup>29</sup>.

We found Indian mithun studied here constitute a genetically uniform group. Influence from taurine and indicine cattle was comparatively minor. Admixture analysis detected a small proportion of admixture of Indian mithun with indicine or taurine cattle. However, based on Treemix analysis there was no evidence of direct gene flow to Indian mithun from cattle, indicating the local tribal practice of crossing mithun with cattle under the field conditions. This crossing probably was with a cattle population not closely related to the breeds represented in our study. Treemix results indicated there was considerable genetic similarity between Indian mithun and gaur, and consistently placing mithun and gaur in the same clade.

## Conclusion

To our knowledge, this is first study aiming to assess the genetic structure of Indian mithun and population diversity using the BovineHD BeadChip SNP array. Our results provide information about the genomic diversity, population structure and origin of Indian mithun inhabiting North Eastern Hilly region. We found evidence of admixture of taurine and indicine in some mithuns, signifying crossing with cattle under field conditions. We also showed that the Indian mithuns are having distinct genetic characteristics and common ancestry with gaur. There was a substantial amount of inbreeding detected as ROH, which has to be considered in future sustainable

breeding and conservation programs for the species. Overall introgression from other bovine species into Indian mithun was limited. Our results are consistent with the mithun being domesticated from a population related to the extant gaur. Our results do not support the hypothesis that mithun originated from crossing gaur bulls with indigenous cattle.

Our study provides a comprehensive picture of the genetic structure and population diversity of Indian mithuns and their phylogenetic relationship with other bovine species.

## Methods

**Study Populations and sampling.** The mithuns were selected randomly from a randomly mating population maintained at the ICAR-NRC on Mithun research farm, Medziphema, Nagaland ( $n = 48$ ) and a diverse field population ( $n = 24$ ) from four locations of North Eastern Hill Region (Nagaland, Arunachal Pradesh, Manipur and Mizoram) falling in the native mithun breeding tracts of India. All the experiments were performed in accordance with relevant guidelines and regulations approved by the Institutional Animal Ethics Committee, ICAR-NRC on Mithun, Nagaland.

**DNA processing and genotyping with BovineHD BeadChip (777 k).** *DNA extraction.* Blood samples were collected from jugular vein in a vacutainer tube (BD) containing EDTA by a qualified veterinarian and were transported to the laboratory in cool pack as soon as possible. The DNA was isolated using a standard blood DNA isolation kit (Promega #A1620) as per the manufacturer's instructions.

*Genotyping and Quality Control.* Genomic DNA from each mithun was quantified to assure a concentration of at least 50 ng/ $\mu$ l genomic DNA required by the Illumina<sup>®</sup> Infinium<sup>®</sup> SNP genotyping platform. A total of 72 mithun (39 males and 33 females) were genotyped by Sandor Lifesciences Pvt. Ltd., Banjara Hills, Hyderabad, India with the Illumina<sup>®</sup> BovineHD BeadChip assay, following the manufacturer's protocol. Genotypes were called using the validated standard cluster file provided by the manufacturer. Only the autosomal SNPs were considered in this study.

Samples and marker based quality control was performed using the Illumina's Genome Studio software (<https://support.illumina.com/array/arraysoftware/genomestudio/documentation.html>). Genome Studio was also used to generate PLINK1.9 data files (in *.ped* and *.map* format)<sup>19</sup> for further analyses. The SNPs located on sex chromosomes, not polymorphic or without known position in the cattle genome were excluded from further analysis. The *Bos taurus* genome assembly (UMD3.1; [http://www.ensembl.org/Bos\\_taurus/Info/Index](http://www.ensembl.org/Bos_taurus/Info/Index)) was used as a reference genome due to the absence of a published mithun genome.

To identify closely related individual, a pair-wise identity-by-state (IBS) distance analysis was performed using PLINK1.9<sup>19</sup>. No closely related individuals were detected, based on a significance test criterion of whether two individuals belong to the same population (i.e. do not merge clusters that contain significantly different individuals). SNP markers showing deviation from Hardy-Weinberg proportions (HWP) based on parental genotype data ( $p < 0.001$ ) were removed. The attributes considered for quality control finally include filtering of SNPs with call rate  $\geq 95\%$ , MAF  $\geq 5\%$  and HWP  $\geq 0.001$  using PLINK1.9 software. After this filtration 139,350 polymorphic SNPs remained for analysis.

*Population genetics analysis.* The parameters estimated to study the genomic diversity in mithun populations included estimates of heterozygosity, detection of runs of homozygosity (ROH) and estimates of inbreeding coefficients.

The observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosities, as well as the estimates of inbreeding for mithun populations were estimated using PLINK1.9 software<sup>19</sup>. The  $F_{ST}$  was estimated using PLINK software, which uses method introduced by Weir and Cockerham (1984)<sup>36</sup>.

*Detection of autozygosity in genomic region.* Runs of homozygosity (ROH) were detected to determine the extent of autozygosity in the mithun populations. A ROH is defined as a contiguous length of homozygous genotypes. A ROH of sufficient length indicates the two copies of the chromosome in this region are identical-by-descent (IBD)<sup>30</sup>. The length and frequency of ROH is useful to get insights into the history of inbreeding of an individual and the population.

PLINK1.9<sup>19</sup> was used to identify ROH by running a sliding window that scans the genomic distribution of SNP data to identify stretches of homozygous SNPs. A minimum number of 10 consecutive homozygous SNP and zero heterozygotes were allowed in each window. A maximum gap between SNP of 1,000 kb was allowed. With high-density SNP data this approach mostly detects truly autozygous segments<sup>37</sup>. Moreover, this strategy is particularly suitable for livestock populations because they have much higher levels of autozygosity than model organisms making identification of longer ROH easy<sup>35,37</sup>.

Detected ROH were classified in windows: ROH in the range of 250 kb–1 Mb, 1 Mb–2 Mb, 2 Mb–4 Mb, 4 Mb–8 Mb, 8 Mb–16 Mb and  $> 16$  Mb.

In the present study, genomic autozygosity,  $F_{ROH}$  of each individual was estimated as the sum of length of autosomal ROH divided by the total length of the autosomes covered by markers<sup>34</sup>.

*Population structure and origin of Mithun.* The main goal of the present study was genomic characterization of two mithun populations (farm and field). Another eight Illumina<sup>®</sup> BovineHD BeadChip or BovineSNP50 BeadChip genotype data sets were collected from previously published work, available at [<http://widde.toulouse.inra.fr/widde/>]<sup>38</sup> and were included along with two mithun populations to explore the evolutionary relationship



of mithun with other bovine species (Table 1). The wild gaur (*Bos gaurus*, OGR,  $n = 6$ ) was included assuming contribution of gaur to mithun genomes as its presumed wild relative; while yak (*Bos grunniens*, OYK,  $n = 4$ ), bison (*Bison bison*, OBB,  $n = 4$ ) and Bali cattle (*Bos javanicus*, BLI,  $n = 20$ ) were taken into account to find out possible genetic contribution towards mithun genome, if any. As it is assumed that there might be some introgression due to crossing between mithun and cattle under the field conditions, the indicine cattle, Gir (*Bos indicus*) genotype data was included. Taurine cattle (NDama, NDA,  $n = 23$ ; Holstein, HOL,  $n = 20$ ) were included to identify any taurine introgression and anoa, the wild buffalo (*Bubalus depressicornis*, OWB,  $n = 10$ ) was included as an outgroup in the present study.

Genotype data from a total of 10 populations were included in subsequent analysis (Table 1). Mithun genotype data were merged with the other data into a single dataset by retaining markers shared between the HD and the 50k panels. The final data included total 209 samples and 38,587 SNPs. Markers and individuals were removed from the dataset using PLINK1.9 software<sup>19</sup> if they did not have call rate of at least 95%. No filtering was done at this step for MAF, except SNPs monomorphic across all breeds. Summary statistics such as heterozygosity and inbreeding coefficient were computed and results are presented (Table 2). For principal component analysis, the dataset included total 144 samples keeping the maximum number of individuals set to 20 (Table 1). Population structure was inferred in two ways: principal component analyses (PCA) by *smartpca* from the Eigenstrat package<sup>39</sup> and ADMIXTURE v. 1.23<sup>18</sup>. PCA uses an orthogonal transformation to convert a set of correlated variables into a set of linearly uncorrelated variables called principal components and maps individuals onto these major axes of variation. In contrast, ADMIXTURE provides maximum likelihood estimates of individual ancestries.

Phylogenetic analysis was carried out using Treemix<sup>20</sup> and wild buffalo as an outgroup with between 0 and 5 migrations allowed. For Treemix analysis we used 21,939 SNPs polymorphic in mithun out of 38,587 SNPs used in PCA and ADMIXTURE analysis. 18 samples (3 Bali cattle, 9 mithuns from farm and 6 mithuns from field) showing admixture were removed from the Treemix analysis. Nodes robustness was estimated with 500 bootstrap replicates and plotted using the Treemix bootstrap function of BITE<sup>40</sup>. Admixture detection was done using the F3 statistics computed using threepop program from the Treemix software package.

**Ethics approval and consent of mithun owners.** Approval of Institutional Animal Ethics Committee was obtained for this study including collection of mithun blood samples and DNA extraction from the Institute research farm. Blood samples of the mithun were collected from the farm and with the permission of the mithun owners from the field by qualified veterinarians.

**Data availability.** Data supporting this paper were generated by ICAR-NRC on Mithun. The phenotype and genotype data are available with the Data Cell of the Institute and should be requested directly from the corresponding authors or the ICAR-NRC on Mithun.

## References

1. Tenzin, S., Dorji, J., Dorji, T. & Kawamoto, Y. Assessment of genetic diversity of Mithun (*Bos frontalis*) population in Bhutan using microsatellite DNA markers. *Animal Genetic Resources/Ressources génétiques animales/Recursos genéticos animales* **59**, 1–6, <https://doi.org/10.1017/s2078633616000072> (2017).
2. Faruque, M. O. *et al.* Present status of gayal (*Bos frontalis*) in the home tract of Bangladesh. *Bangladesh Journal of Animal Science* **44**, 75–84 (2015).
3. Simoons, F. J. & Simoons, E. S. *A ceremonial ox of India. The mithun in nature, culture, and history*. 3–13 (The University of Wisconsin Press, Madison, WI, USA, 1968).
4. Ge, C. R., Tian, Y. B., Chen, T. & Wu, Y. Studies on the meat feature of gayal (*Bos frontalis*). *Sci. Agric. Sin.* **29**, 75–78 (1996).
5. Giasuddin, M., Huque, K. S. & Alam, J. Reproductive potentials of gayal (*Bos frontalis*) under semi-intensive management. *Asian Austral J Anim* **16**, 331–334 (2003).
6. NRCM Annual Report (National Research Centre for Mithun, Nagaland, India, 2014–15).
7. Das, K. C., Baruah, K. K., Haque, N., Sharma, R. & Rajkhowa, C. Effect of Different Dietary Protein Levels on Growth, Nutrient Utilization and Rumen Fermentation in Mithun (*Bos frontalis*) Fed on Rice Straw based Diet. *Anim Nutr Feed Techn* **12**, 305–313 (2012).
8. McKay, S. D. *et al.* 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> (2008).
9. Lin, B. Z., Sasazaki, S. & Mannen, H. Genetic diversity and structure in *Bos taurus* and *Bos indicus* populations analyzed by SNP markers. *Anim Sci J* **81**, 281–289, <https://doi.org/10.1111/j.1740-0929.2010.00744.x> (2010).
10. Gautier, M., Laloe, D. & Moazami-Goudarzi, K. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS one* **5**, e13038, <https://doi.org/10.1371/journal.pone.0013038> (2010).
11. Gautier, M. *et al.* A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC genomics* **10**, 550, <https://doi.org/10.1186/1471-2164-10-550> (2009).
12. Matukumalli, L. K. *et al.* Development and characterization of a high density SNP genotyping assay for cattle. *PLoS one* **4**, e5350, <https://doi.org/10.1371/journal.pone.0005350> (2009).
13. Decker, J. E. *et al.* Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *PLoS genetics* **10**, e1004254, <https://doi.org/10.1371/journal.pgen.1004254> (2014).
14. Ben Jemaa, S., Boussaha, M., Ben Mehdi, M., Lee, J. H. & Lee, S. H. Genome-wide insights into population structure and genetic history of tunisian local cattle using the illumina bovinesnp50 beadchip. *BMC genomics* **16** (2015).
15. Orozco-terWengel, P. *et al.* Revisiting demographic processes in cattle with genome-wide population genetic analysis. *Frontiers in genetics* **6**, 191, <https://doi.org/10.3389/fgene.2015.00191> (2015).
16. Sharma, A. *et al.* A genome-wide assessment of genetic diversity and population structure of Korean native cattle breeds. *BMC genetics* **17**, 139, <https://doi.org/10.1186/s12863-016-0444-8> (2016).
17. Phanchung & Roden, J. A. Characterisation of the Siri breed and the Mithun cross Siri in Bhutan. *Animal Genetic Resource Information* **20**, 27–34 (1996).
18. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* **19**, 1655–1664, <https://doi.org/10.1101/gr.094052.109> (2009).
19. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559–575, <https://doi.org/10.1086/519795> (2007).

20. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics* **8**, e1002967, <https://doi.org/10.1371/journal.pgen.1002967> (2012).
21. Hartl, D. L. & Clark, A. G. *Principles of population genetics*. 3rd edn, (Sinauer Associates, 1997).
22. Felius, M. *Encyclopedia of Cattle Breeds*. (Misset, 1995).
23. Lenstra, J. A. & Bradley, D. G. *Systematics and Phylogeny of Cattle*. 1–14 (CAB International, 1999).
24. Uzzaman, M. R. *et al.* 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, <https://doi.org/10.5713/ajas.2014.14160> (2014).
25. Bovine HapMap, C. *et al.* Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532, <https://doi.org/10.1126/science.1167936> (2009).
26. Lewin, H. A. Genetics. It's a bull's market. *Science* **324**, 478–479, <https://doi.org/10.1126/science.1173880> (2009).
27. Kirin, M. *et al.* Genomic runs of homozygosity record population history and consanguinity. *PLoS one* **5**, e13996, <https://doi.org/10.1371/journal.pone.0013996> (2010).
28. Zhang, Q., Guldbrandtsen, B., Bosse, M., Lund, M. S. & Sahana, G. Runs of homozygosity and distribution of functional variants in the cattle genome. *BMC genomics* **16**, 542, <https://doi.org/10.1186/s12864-015-1715-x> (2015).
29. Szmatala, T. *et al.* Characteristics of runs of homozygosity in selected cattle breeds maintained in Poland. *Livest Sci* **188**, 72–80, <https://doi.org/10.1016/j.livsci.2016.04.006> (2016).
30. Purfield, D. C., Berry, D. P., McParland, S. & Bradley, D. G. Runs of homozygosity and population history in cattle. *BMC genetics* **13**, 70, <https://doi.org/10.1186/1471-2156-13-70> (2012).
31. Curik, I., Ferencakovic, M. & Solkner, J. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livest Sci* **166**, 26–34, <https://doi.org/10.1016/j.livsci.2014.05.034> (2014).
32. Gibson, J., Morton, N. E. & Collins, A. Extended tracts of homozygosity in outbred human populations. *Human molecular genetics* **15**, 789–795, <https://doi.org/10.1093/hmg/ddi493> (2006).
33. Ferencakovic, M. *et al.* Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of animal breeding and genetics = Zeitschrift fur Tierzucht und Zuchtungsbiologie* **130**, 286–293, <https://doi.org/10.1111/jbg.12012> (2013).
34. McQuillan, R. *et al.* Runs of homozygosity in European populations. *American journal of human genetics* **83**, 359–372, <https://doi.org/10.1016/j.ajhg.2008.08.007> (2008).
35. Ferencakovic, M., Solkner, J. & Curik, I. Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics, selection, evolution: GSE* **45**, 42, <https://doi.org/10.1186/1297-9686-45-42> (2013).
36. Weir, B. S. & Cockerham, C. C. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* **38**, 1358–1370, <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x> (1984).
37. Howrigan, D. P., Simonson, M. A. & Keller, M. C. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC genomics* **12**, 460, <https://doi.org/10.1186/1471-2164-12-460> (2011).
38. Sempere, G. *et al.* WIDDE: a Web-Interfaced next generation database for genetic diversity exploration, with a first application in cattle. *BMC genomics* **16**, 940, <https://doi.org/10.1186/s12864-015-2181-1> (2015).
39. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904–909, <https://doi.org/10.1038/ng1847> (2006).
40. Milanese, M. *et al.* BITE: an R package for biodiversity analyses. *bioRxiv*, <https://doi.org/10.1101/181610> (2017).

## Acknowledgements

The authors would like to thank the field veterinarians and mithun owners/farmers for their participation and cooperation in the collection of mithun blood sampling from the field. Financial assistance in the form of DBT Overseas Associateship is greatly acknowledged. This study was supported by an ICAR funded research project of ICAR-NRC on Mithun, Nagaland, India and Department of Biotechnology, under DBT Overseas Associateship, Govt. of India.

## Author Contributions

A.M., S.M. and C.R. conceived and designed the study. A.M., S.M. and R.D. carried out the analyses and drafted the manuscript. B.G. and G.S. critically monitored the analyses and revised the manuscript. AMI revised Introduction part and checked punctuation of the manuscript. M.M., I.L., K.V., K.K., N.H., I.J. and P.P. were involved in sampling and running lab work. All authors read and approved the final manuscript.

## Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018