



## Genome-wide Single Nucleotide Polymorphism Analyses Reveal Genetic Diversity and Structure of Wild and Domestic Cattle in Bangladesh

Md. Rasel Uzzaman, Zewdu Edea, Md. Shamsul Alam Bhuiyan<sup>1</sup>, Jeremy Walker<sup>2</sup>,  
A. K. F. H. Bhuiyan<sup>1</sup>, and Kwan-Suk Kim\*

Brain Korea 21 Center for Bio-Resource Development, Department of Animal Science,  
Chungbuk National University, Cheongju 361-763, Korea

**ABSTRACT:** In spite of variation in coat color, size, and production traits among indigenous Bangladeshi cattle populations, genetic differences among most of the populations have not been investigated or exploited. In this study, we used a high-density bovine single nucleotide polymorphism (SNP) 80K Bead Chip derived from *Bos indicus* breeds to assess genetic diversity and population structure of 2 Bangladeshi zebu cattle populations (red Chittagong,  $n = 28$  and non-descript deshi,  $n = 28$ ) and a semi-domesticated population (gayal,  $n = 17$ ). Overall, 95% and 58% of the total SNPs (69,804) showed polymorphisms in the zebu and gayal populations, respectively. Similarly, the average minor allele frequency value was as high 0.29 in zebu and as low as 0.09 in gayal. The mean expected heterozygosity varied from  $0.42 \pm 0.14$  in zebu to  $0.148 \pm 0.14$  in gayal with significant heterozygosity deficiency of 0.06 ( $F_{IS}$ ) in the latter. Coancestry estimations revealed that the two zebu populations are weakly differentiated, with over 99% of the total genetic variation retained within populations and less than 1% accounted for between populations. Conversely, strong genetic differentiation ( $F_{ST} = 0.33$ ) was observed between zebu and gayal populations. Results of population structure and principal component analyses suggest that gayal is distinct from *Bos indicus* and that the two zebu populations were weakly structured. This study provides basic information about the genetic diversity and structure of Bangladeshi cattle and the semi-domesticated gayal population that can be used for future appraisal of breed utilization and management strategies. (**Key Words:** Bangladesh, Genome-wide Single Nucleotide Polymorphism Analyses, Gayal, Domestic Cattle)

### INTRODUCTION

Livestock are an integral part of the Bangladeshi agricultural economy. About 12% of the agricultural GDP comes from the livestock sector, and 10 million people are directly involved with the livestock sector for their livelihoods (Karim et al., 2010). In Bangladesh, cattle are the most versatile among livestock species in the existing integrated agricultural farming system. There are approximately 24.5 million heads of cattle in Bangladesh,

which comprise approximately 1.79% of the world cattle population and 5.47% of the Asian cattle population (Baker, 2004). Indigenous cattle breeds and types are better adapted to adverse climatic conditions, poor nutrition, and low management systems, and they are more resistant to local diseases and parasites. Bangladeshi cattle are zebu type, and they are identified by their local name or the place where they predominantly found, such as Pabna, red Chittagong (RC), Munshiganj, Madaripur, and north Bengal grey. The commonly available native types are called non-descript deshi (ND). Indigenous Bangladeshi cattle populations have lower productivity than the improved crossbred (e.g. Shahiwal, Sindhi and Holstein-Friesian) animals. There is also variation in coat color, size, live weight, and production and reproduction traits. However, genetic differences among these populations are not well understood. Gayal (*Bos frontalis*) is a large, semi-

\* Corresponding Author: Kwan-Suk Kim. Tel: +82-43-261-2547, Fax: +82-43-273-2240, E-mail: kwanskim@chungbuk.ac.kr

<sup>1</sup> Department of Animal Breeding and Genetics, Bangladesh Agricultural University, Mymensingh 2202, Bangladesh.

<sup>2</sup> GeneSeek, Neogen Corporation, Lincoln, NE 68521, USA.

Submitted Mar. 3, 2014; Revised May 23, 2014; Accepted Jun. 9, 2014

domesticated bovine species seen in the southeastern hilly regions of Bangladesh. Studies have shown species hybridization between gayal and the local cattle (*B. indicus*) (Giasuddin et al., 2003).

Among the wide range of molecular markers developed, single nucleotides polymorphisms (SNPs) are the most abundant, are widely dispersed throughout genomes, and have variable distribution among species (Vignal et al., 2002). The availability of high-throughput SNP genotyping platforms makes it feasible to undertake high-density scans by using large numbers of SNP markers and are either distributed across the whole genome or focused in specific regions. The SNPs are useful in studying livestock genetic diversity and population structure (McKay et al., 2008; Lin et al., 2010). Although a large number of SNPs have been identified from the bovine genome-sequencing project, few of these have been validated, particularly in *B. indicus* breeds. Therefore, we used high density bovine SNP 80K to assess genetic diversity and population structure of Bangladeshi indigenous cattle populations and the semi-domesticated gayal.

## MATERIALS AND METHODS

### Populations and DNA sample collection

Among phenotypically categorized indigenous Bangladeshi zebu cattle breeds, we collected nasal samples from unrelated RC (n = 28) and ND (n = 28) animals. The two populations are notably different in their distinct coat color types (Bhuiyan et al., 2007a). The RC samples were collected from the Chittagong district, which is the only breeding area for this population. The ND samples were collected from Mymensingh region, but they had wide distribution (Bhuiyan et al., 2007b). For comparison, we also collected 17 samples from the gayal population of the Bandarban district, a hilly region of Bangladesh. Nasal samples were collected using Performagene Livestock's nasal swab DNA collection kit and DNA was extracted from nasal samples according to the manufacturer's recommendations (DNA Genotek Inc., 2012).

### Genotyping and marker selection

All samples were genotyped using the GeneSeek Genomic Profiler Indicine HD Beadchip, an Illumina

Infinium array consisting of nearly 80,000 SNPs derived mainly from *B. indicus* breeds (GeneSeek, Lincoln, NE, USA), according to Illumina's standard protocols (<http://www.illumina.com>). For minor allele frequency (MAF) estimation, we analyzed approximately 69,804 autosomal SNP markers. For diversity analysis, SNPs were screened based on the following criteria: call rate  $\geq 95$ , MAF  $\geq 5\%$ , and Hardy-Weinberg equilibrium, HWE  $\geq 0.001$ . The diversity analysis resulted in 51,366 SNPs. To save running time in the population structure analysis, 35,437 SNPs were screened based on the following criteria: call rate  $\geq 95$ , MAF  $\geq 20\%$ , and HWE  $\geq 0.001$ .

### Statistical analysis

Minor allele frequency and proportion of polymorphic SNPs were estimated using the Golden Helix SNP Variation Suite software version 7 (Golden Helix, 2013). Within breed genetic variability (observed and expected heterozygosity) and inbreeding (Weir, 1996) estimates were calculated using PowerMarker V3.25 software (Liu and Muse, 2005). Reynolds genetic distances (Reynolds et al., 1983) between pairs of populations were estimated using the same software. Population genetic structure was inferred by applying principal component analysis (PCA) using the Golden Helix SNP Variation Suite software version 7 (Golden Helix, 2013) for the whole data set and by model-based clustering using STRUCTURE 2.3.4 (Prichard et al., 2000). The PCA estimates were performed for the three studied populations by using the allele frequencies of 69,804 SNP markers. We ran STRUCTURE following the admixture ancestry model and correlated allele frequency for *K* values of 2 and 3 with a burn period of 20,000 generations and Markov chain Monte Carlo simulations of 100,000 iterations using the correlated allele model.

## RESULTS

### Minor allele frequency and genetic diversity

The mean MAFs and within population genetic variations are presented in Table 1. Common SNP variants with MAFs  $\geq 0.10$  and  $\leq 0.50$  ranged from 30.31% in gayal to 88.57% in ND populations. More than half of the SNPs (52.48%) in RC and ND displayed MAFs greater or equal to 0.30, whereas only 7.72% of SNPs in gayal were in this

**Table 1.** Minor allele frequency (MAF) distribution of three different Bangladeshi cattle populations

Population	No. of animals	Fixed (0)		Rare (>0 and <0.05)		Intermediate ( $\geq 0.05$ and <0.10)		Common ( $\geq 0.10$ and $\leq 0.50$ )		$\geq 0.30$ and $\leq 0.50$	
		SNP	Prop.	SNP	Prop.	SNP	Prop.	SNP	Prop.	SNP	Prop.
ND	28	689	0.99	2553	3.69	4654	6.73	61218	88.57	36384	52.64
RC	28	1104	1.58	2908	4.16	4799	6.87	60994	87.37	36522	52.32
Gayal	17	14287	20.46	15209	21.79	19141	27.42	21159	30.31	5390	7.72

SNP, single nucleotide polymorphism; Prop., proportion; ND, non-descript deshi; RC, red Chittagong.

**Table 2.** Genetic diversity indicators calculated for three different cattle populations

Breed	No. of animals	MAF	H <sub>E</sub>	H <sub>O</sub>	F <sub>IS</sub>
ND	28	0.29±0.14	0.401±0.107	0.420±0.143	-0.027±0.206
RC	28	0.28±0.14	0.396±0.112	0.415±0.143	-0.029±0.192
Gayal	17	0.09±0.11	0.153±0.139	0.148±0.143	0.061±0.229

MAF, minor allele frequency; H<sub>O</sub>, observed heterozygosity; H<sub>E</sub>, expected heterozygosity; F<sub>IS</sub>, inbreeding coefficient; ND, non-descript deshi; RC, red Chittagong.

range. On average, 95.18% of SNPs (69,804) displayed polymorphisms (MAF≥0.05) within the *B. indicus* populations (RC and ND). However, only 57.73% of SNPs showed polymorphisms in gayal. The observed heterozygosity was 0.415±0.143 and 0.420±0.143 in RC and ND, respectively (Table 2). Gayal had the lowest observed and expected heterozygosities (0.148±0.143 and 0.153±0.139, respectively) and a significant heterozygosity deficiency (0.061±0.229 [F<sub>IS</sub>]). The estimated F<sub>IS</sub> values in RC and ND were -0.029±0.192 and -0.027±0.206, respectively (Table 2). Gayals had a larger proportion of monomorphic SNPs, which accounted for 79.31% of monomorphic SNPs shared with *B. indicus* (RC and ND) (Table 3). However, approximately 60% of polymorphic SNPs were shared by gayal and the *B. indicus* population. High differences in the number of monomorphic SNPs were found on chromosome 1, wherein 152 SNPs were common and accounted for 85.88% of the monomorphic SNPs on the

**Table 3.** Numbers of polymorphic (MAF ≥0.05) and monomorphic (MAF <0.05) SNP markers in domestic cattle (ND and RC) and gayal populations across chromosomes

Chromosome No.	Monomorphic SNPs			Polymorphic SNPs		
	Cattle	Gayal	Common (% share of cattle)	Cattle	Gayal (% share of cattle)	Common
1	177	2,315	152 (85.88)	4,082	1,943 (47.60)	1,918
2	196	1,567	143 (72.96)	3,487	2,115 (60.65)	2,062
3	177	1,142	134 (75.71)	3,171	2,206 (69.57)	2,163
4	187	1,192	145 (77.54)	3,069	2,064 (67.25)	2,022
5	151	1,461	125 (82.78)	3,139	1,829 (58.27)	1,803
6	90	704	54 (60)	3,110	2,496 (80.26)	2,460
7	184	1,317	119 (64.67)	2,870	1,737 (60.52)	1,672
8	155	1,445	121 (78.06)	2,882	1,592 (55.24)	1,558
9	74	1,000	62 (83.78)	2,824	1,898 (67.21)	1,886
10	164	1,230	130 (79.27)	2,667	1,601 (60.03)	1,567
11	108	1,137	92 (85.19)	2,774	1,744 (62.87)	1,728
12	115	1,120	92 (80)	2,354	1,348 (57.26)	1,325
13	127	1,286	103 (81.10)	2,183	1,022 (46.82)	998
14	106	1,037	93 (87.74)	2,209	1,278 (57.85)	1,265
15	115	989	98 (85.22)	2,329	1,454 (62.43)	1,437
16	99	1,133	91 (91.92)	2,194	1,159 (52.83)	1,151
17	80	1,273	77 (96.25)	2,006	812 (40.48)	809
18	105	651	82 (78.10)	1,788	1,242 (69.46)	1,219
19	119	1,165	108 (90.76)	1,754	708 (40.36)	697
20	114	1,071	98 (85.96)	1,965	1,008 (51.30)	992
21	127	989	102 (80.31)	1,942	1,080 (55.61)	1,055
22	82	532	53 (64.63)	1,680	1,230 (73.21)	1,201
23	55	491	40 (72.73)	1,551	1,115 (71.89)	1,100
24	74	692	65 (87.84)	1,724	1,106 (64.15)	1,097
25	64	534	50 (78.13)	1,230	760 (61.79)	746
26	105	730	95 (90.48)	1,401	776 (55.39)	766
27	56	316	38 (67.86)	1,295	1,035 (79.92)	1,017
28	65	517	47 (72.31)	1,340	888 (66.27)	870
29	93	460	59 (63.44)	1,420	1,053 (74.15)	1,019
Total	3,364	29,496	2,668 (79.31)	66,440	40,299 (60.65)	39,603

MAF, minor allele frequency; SNP, single nucleotide polymorphism; ND, non-descript deshi; RC, red Chittagong.

same chromosome. Less difference were found on chromosome 27, wherein only 38 monomorphic SNPs were common and accounted for 67.86% of the monomorphic SNPs. Almost all of the polymorphic SNPs on gayal chromosomes were also found to be polymorphic in the zebu population, which may reflect a high deficiency in heterozygosity observed in the gayal population.

Genetic differentiations between the two Bangladeshi zebu populations were very low, and approximately 99% of the genetic variation was retained within the breed. Similarly, genetic distance (Reynolds distance) showed a close relationship between RC and ND (0.020). With an  $F_{ST}$  value of 0.33 and Reynolds distances of 0.31, gayal showed strong differentiation from *B. indicus* subspecies.

### Population structure

The first and the second principal components (PC1 and PC2) explained 90.24% of the total variation and evidently distinguish the two zebu populations from gayal. The results coincided well with the STRUCTURE output at  $K = 2$  and  $K = 3$  (Figure 2). The output at  $K = 2$  seems plausible; it clearly distinguished the gayal from *B. indicus* populations (RC and ND) with some level of gene flow between the breeds. The output at  $K = 3$  suggests higher admixture than expected in ND and RC, which does not agree with the PCA result in this study. This might be because of the influence of exotic blood due to indiscriminate crossbreeding. Approximately 11% of the gayal population is considered to share common ancestry with the ND and RC populations.

## DISCUSSION

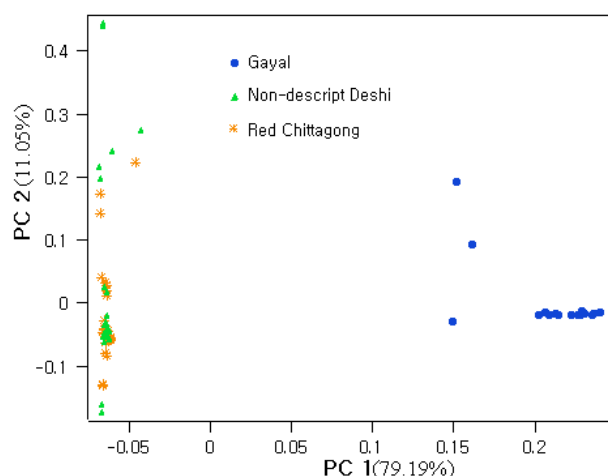
Evaluation of genetic diversity and assessment of population structure is necessary to appraise the utilization and management of farm animal genetic resources. In this study, we estimated genetic diversity and population structure of Bangladeshi zebu cattle populations and the semi-domesticated gayal breed by using a high-density SNP genotyping chip recently developed from indicine cattle. The indigenous zebu cattle of Bangladesh have been reared using traditional husbandry practices, and intensive selective breeding, for milk or meat, is rare. The practices result in an effective population size that is much larger than the recognized breed. The attributes contributed to the high diversity measures in Bangladeshi zebu cattle. The low genetic differentiation observed between RC and ND could be because of their common origin and gene flow. In addition, it is possible that a spontaneous mutation or aggregation of genes for coat-color differentiated RC occurred from ND and has been maintained by the farmers of the greater Chittagong district. However, apart from this

phenotypic feature, there is no remarkable difference between the two zebu populations in production traits.

The zebu cattle MAFs that we found were higher than that reported for Ethiopian cattle populations (Edea et al., 2013) and for *B. indicus* and *B. taurus* breeds (Lin et al., 2010). Variation across breeds may be because of the difference in the chip employed. The higher MAF reported here might be because the SNP panels that we used were derived from *B. indicus* breeds, and as expected, the minor allele was higher in the zebu breeds. The level of polymorphism we found was higher than that shown in previous studies of taurine (95.21%) and African cattle populations (83.96%) (Edea et al., 2013). As Bangladeshi zebu cattle are non-selected random bred indigenous population, expectedly higher heterozygosity remains over the generations, and therefore, the observed and expected heterozygosities were higher than those reported for African cattle populations (Edea et al., 2013) and taurine breeds (Lin et al., 2010).

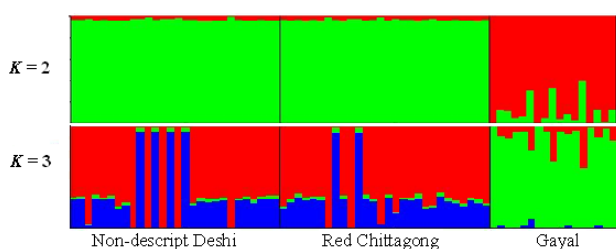
A total of 29,496 SNP markers were monomorphic (MAF<0.05) in the gayal population, whereas the total number of monomorphic SNP makers was 3364 in the zebu population (Table 3). Interestingly, approximately 79.31% of monomorphic markers in the zebu population were also monomorphic in the gayal population, which might indicate an influence of zebu on gayal. This result supports the hypothesis that gayal are hybrid descendants of wild gaurs and domestic cattle of either *B. indicus* or *B. taurus* (Payne, 1970). Species hybridization between gayal and local cattle (*B. indicus*) has been reported (Giasuddin et al., 2003). In Bangladesh, some tribal family rear gayal with native cattle and their hybrids are sometimes found in local markets. The total number of polymorphic (MAF $\geq$ 0.05) SNPs screened were 40,299 and 66,440 in gayal and zebu populations, respectively. Regardless of the effect of origin of the chip, the high genome-wide number of monomorphic SNPs in the gayal population could be attributed to the  $F_{IS}$  of 0.06. The Gayal population has been declining over time, which has led to the reduced effective population size in the region. The observed variation of SNPs between the two subspecies should be studied further to increase the understanding of the genetics of the phenotypic differences. The candidate gene approach could be used to determine the role of specific genes in these bovine species.

The PCA results clustered the RC and ND as one population, whereas the gayal population was observed as a separate cluster (Figure 1). Within the gayal cluster, a few were outgroups. This is likely because a few hybrid individuals were noted within the studied gayal population. The structure output also revealed the same genetic architecture in the two zebu populations, with some introgressions of genes from domestic cattle (i.e., zebu) to



**Figure 1.** Individual animal clustering-based principal component analysis.

the gayal population and vice versa. This might be due to inter-specific hybridization among the *Bos* species (zebu, taurine, gayal, and yak) which has been documented in South and Southeast Asian cultures (Tanaka et al., 2011). The admixed zebu and gayal individuals in the studied population might be due to sampling errors despite they have the phenotypic resemblance and every precautions were taken during sampling. It is notable that the remnant of inheritance in admixed individuals is very difficult to trace phenotypically in some cases. In addition, the pedigree information is almost unknown to farmers in Bangladesh perspective. The bottleneck of genetic features in the gayal population might be because of inbreeding within a small population. In fact the geographic distribution of gayal habitats is very specific where people of several tribal families used to keep their animals together in a form of small group in the particular territory. Random mating occurs in that small sub-population over the years despite hierarchy of breeding bulls is a question. Moreover, the chance of interbreeding among the sub-population is negligible. This faulty breeding practices and lack of



**Figure 2.** Genetic structure of three cattle populations inferred using structure 2.3.4 (Pritchard et al., 2000). Each individual animal is represented using a single vertical line divided by k colors, where k is the number of clusters assumed and the length of the colored segment represents the individual's estimated proportion of membership to a particular cluster.

scientific management leads to increase homogeneity within the small group. However, our attempts unveiled the genetic architecture of gayal for the first time and above all, gave us hints to take immediate attention from different scientific institutes of Bangladesh to overcome the bottlenecks.

## ACKNOWLEDGMENTS

This study was designed by a Memorandum of Understanding agreement between Chungbuk National University and Bangladesh Agricultural University, and was supported by a grant from Next Generation Bio-green21 Program under Rural Development Administration.

## REFERENCES

- Baker, F. 2004. Selected indicators of food and agriculture development in asia-pacific region 1993-2003. FAO Regional Office for Asia and the Pacific, Bangkok, Thailand.
- Bhuiyan, A., M. Hossain, and G. Deb. 2007a. Indigenous cattle genetic resources of Bangladesh and a way forward to their development. *Bangladesh J. Prog. Sci. Tech.* 5:105-112.
- Bhuiyan, M. S. A., A. K. F. H. Bhuiyan, D. H. Yoon, J. T. Jeon, C. S. Park, and J. H. Lee. 2007b. Mitochondrial DNA diversity and origin of Red Chittagong cattle. *Asian Australas. J. Anim. Sci.* 20:1478-1484.
- Dadi, H., J. J. Kim, D. Yoon, and K. S. Kim. 2012. Evaluation of single nucleotide polymorphisms (SNPs) genotyped by the illumina bovine SNP50K in cattle focusing on Hanwoo breed. *Asian Australas. J. Anim. Sci.* 25:28-32.
- Dalvit, C., M. De Marchi, R. Dal Zotto, E. Zanetti, T. Meuwissen, and M. Cassandro. 2008. Genetic characterization of the Burlina cattle breed using microsatellites markers. *J. Anim. Breed. Genet.* 125:137-144.
- DNA Genotek Inc. 2012. Performa-gene-Livestock (PG-100). Available at: <http://www.dnagenotek.com/US/products/PG100.html> Accessed April 8, 2012.
- Edea, Z., H. Dadi, S. W. Kim, T. Dessie, and K.-S. Kim. 2012. Comparison of SNP variation and distribution in indigenous Ethiopian and Korean Cattle (Hanwoo) populations. *Genomics Inform.* 10:200-205.
- Edea, Z., H. Dadi, S.-W. Kim, T. Dessie, T. Lee, H. Kim, J.-J. Kim, and K.-S. Kim. 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. *Front. Genet.* 4:35.
- Giasuddin, M., K. S. Huque, and J. Alam. 2003. Reproductive potentials of gayal (*Bos frontalis*) under semi-intensive management. *Asian Australas. J. Anim. Sci.* 16:331-334.
- Golden Helix, I. 2012. SNP & Variation Suite Manual, Version 7.6.9. Available at: <http://www.goldenhelix.com> Accessed April 10, 2012.
- Karim, Z., K. S. Huque, G. Hussain, Z. Ali, and M. Hossain. 2010. Growth and development potential of livestock and fisheries in Bangladesh. Bangladesh Food Security Investment Forum. May 26-27, 2010; Dhaka, Bangladesh.
- Lin, B. Z., S. Sasazaki, and H. Mannen. 2010. Genetic diversity and structure in *Bos taurus* and *Bos indicus* populations

- analyzed by SNP markers. *Anim. Sci. J.* 81:281-289.
- Liu, K. and S. V. Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128-2129.
- McKay, S., R. Schnabel, B. Murdoch, L. Matukumalli, J. Aerts, W. Coppieters, D. Crews, E. Neto, C. Gill, and C. Gao. 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. *BMC Genetics* 9:37.
- Payne, W. J. A. 1970. Cattle production in the tropics. Vol. I. Breeds and Breeding. Longman Group Ltd., London, UK.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Reynolds, J., B. S. Weir, and C. C. Cockerham. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767-779.
- Tanaka, K., T. Takizawa, T. Dorji, T. Amano, H. Mannen, Y. Maeda, Y. Yamamoto, and T. Namikawa. 2011. Polymorphisms in the bovine hemoglobin-beta gene provide evidence for gene-flow between wild species of *Bos* (*Bibos*) and domestic cattle in Southeast Asia. *Anim. Sci. J.* 82:36-45.
- Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genet. Sel. Evol.* 34:275-306.
- Weir, B. 1996. *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA, USA.