Buffalo genome research - a review

Muhammad Moaen-ud-Din*

Laboratories of Animal Breeding & Genetics, Faculty of Veterinary & Animal Sciences, PMAS-Arid Agriculture University, Rawalpindi (46300), Pakistan

(Accepted June 30, 2014)

Buffalo (Bubalus bubalis and Bubalus carabanesis) are important animals in livestock and agricultural economy of many countries, providing milk, meat and hides along with draught power. But, despite of immense advantages buffalo is a heavily neglected animal. The real potential of buffalo has never been realized and challenges such as reduced reproductive efficiency still needed to be addressed. At the onset of genomic era, after sequencing of entire cattle genome and recently sequencing of buffalo genome, there is a start of new chapter of buffalo research. This review is focused to highlight the recent advances in genome science of buffalo with respect to cytogenetic studies, whole genome mapping and application of next generation sequencing. These resources of genome science can be applied to provide knowledge and technologies to enhance the production potential to the optimum level, improve reproduction efficiency and increase disease resistance in buffalo. Buffalo and domestic cattle belong to same family and are closely related on phylogenetic tree. Therefore, available cattle genetic and genomic resources can be used as shortcuts by researchers to explore genome of buffalo and apply biotechnology tools for buffalo improvement.

KEYWORDS: buffalo / cytogenetics / genome mapping / genome sequencing

Buffalo plays a pivotal role in livestock and agriculture economy of many countries across the globe. Water buffalo produces milk, meat and is used as draught animal in developing countries [Kierstein et al. 2004, Yindee et al. 2010]. Domesticated species of buffalo are river and swamp buffalo (Bubalus carabanesis). The habitat of these animals is tropical, subtropical, wet grasslands, marshes and swamp regions of the world. These animals cool themselves by spending a substantial part of their time in water, mud holes or rivers.

*Corresponding author: drmoainawan@gmail.com
Genus *Bubalus* dispersed from Asia to Europe during the Pleistocene era. These animals become restricted to India, Indonesia, and parts of Southeast Asia when climate went dry. Water buffalo was introduced to Italy during the sixth century *via* Bay of Tunis. Moreover, buffalo was recently introduced in Australia, Africa, and both Americas [Kierstein *et al.* 2004, Yindee *et al.* 2010].

The current world buffalo population is 194 million [FAO 2010]. About 97.13% world buffalo is in Asia, with 76.92% in South Asia. India and Pakistan have 57% and 43% of the population, respectively. So, in Indo-Pak 149 million water buffalo comprise 97% of the world’s buffalo population, where more people are dependent on the water buffalo than on any other domesticated species. Buffalo world population is steadily increasing at the rate of 2% *per annum* during the last two decades.

Buffalo contribution in world milk supply is 12.8%. The milk is higher in nutrients (e.g. fat, lactose, protein, and minerals) and contains less water than cow’s milk [FAO 2010]. Buffalo milk is used to produce butter, butter oil, high quality cheeses, and other high quality dairy products and meat is reported to be leaner with less fat and cholesterol than beef. Buffalo hides are used to make high quality leather products. Finally, buffalo provides agriculture farm traction, specifically 20 to 30% of all farm power; it’s a superior draught animal in waterlogged conditions. Buffalo is a more efficient convertor of poor quality feeds and fiber than cattle. Most importantly, buffalo is used as cash during needy hours thus securing the economic status of many families.

Despite of immense advantages of raising buffalo as mentioned above, unfortunately, these animals remained underutilized. In particular, buffalo breeders and farmers are facing challenges in the term of poor reproductive efficiency, sub-optimal production potential, infertility, and lower calf survival rate.

Gene technologies are being promoted and utilized in many fields of livestock production. For instance, higher productivity is being realized through manipulating the genetic variation within and between breeds that multiplies genetic gain in traditional breeding value. Therefore, gene technologies have application in distinguishing molecular phenotypes and improving the use of genetic resources of domestic animals. Current review is focused on updates of buffalo available genome resources that can provide knowledge of technologies useful to optimize production potentials, reduce reproductive inefficiency, improve product quality and highlight immunity aspect in this important species.

**Evolution and domestication of buffalo**

Buffalo belongs to family Bovidae. Asian buffalo (*Bubalus bubalus*) and the African buffalo (*Syncerus caffer*) are two main species of buffalo [Iannuzzi and Di Medeo 1995]. Asian buffalo originated in India where its domestication occurred during the third millennium BC [Cockrill 1981]. The second origin point of buffalo is China where it was domesticated earlier than in India (in fifth millennium BC) [Kumar *et al.* 2007b]. African buffalo originated distinctively than river buffalo and
swamp buffalo that came from Asia. Furthermore, a recent study by MacEachern et al. [2009] opposed to the previous mtDNA analysis results showing a distinct origin of water and swamp buffaloes [Kumar et al. 2007b].

There are two subspecies of African buffalo: the cape buffalo (*Syncerus caffer*) and the forest buffalo (*Syncerus caffer nanus*) [Van Hooft et al. 2002]. They have evolved independently to Asian buffalo [MacEachern et al. 2009]. Furthermore, recent studies of mtDNA sequence indicated on the divergence of central Africa and cape buffalo around 130000-100000 years ago, during the Pleistocene period [MacEachern et al. 2009, Smitz 2011, Van Hooft et al. 2002]. See Figure 1 [MacEachern et al. 2009].

![Phylogeny of the Bovini tribe using neighbor joining analysis](image)

**Fig. 1.** Phylogeny of the Bovini tribe using neighbor joining analysis (Anc: the Ancient, Ban: Banteng, Bis: Bison, BubB & BubC: Asian buffalo, water and swamp type, Ela: Eland, Gau: Gaur, Her: Hereford, Ind & Mur: Indian water buffalo, Mit: Mithan, Hol: Holstein, Syn: African buffalo, Tul: Tuli, and Yak) [MacEachern et al. 2009].

**Asian buffalo**

The number of chromosomes is 25 and 24 pairs in river and swamp buffalo, respectively. These subspecies differ by one chromosome because of a fusion between chromosomes 4 and 9 pair in the swamp buffalo genome. All chromosomes and
chromosome arms are preserved between these two subspecies [Gallagher and Womack 1992], crosses between them are fertile with 49 chromosomes in crossbreds. But, fertility and reproductive efficiency is reduced in subsequent matings [Harisah et al. 1989]. River buffalo has 5 biarmed chromosomes and the remaining chromosome 20 pairs are acrocentric including the sex chromosomes. Based on shared chromosome banding and gene order homology, it is evident that river buffalo and domestic cattle (two members of the Bovidae family) are closely related [Amaral et al. 2008]. Cattle and river buffalo chromosomes are cytogenetically matched arm-for-arm [Fig. 2, Amaral et al. 2007, Michelizzi et al. 2010]. The cattle genome has 29 acrocentric chromosome pairs, with XY sex chromosome pair, while the river buffalo has 5 biarmed and 19 acrocentric chromosome pairs along with sex chromosomes XY. Each two cattle acrocentric chromosomes, viz. BTA1 and BTA27, BTA2 and BTA23, BTA8 and BTA19, BTA5 and BTA28 fused into BBU1, BBU2, BBU3, BBU4 and BBU5 biarmed chromosome pairs of river buffalo, respectively. Thus, each of the acrocentric river buffalo chromosomes corresponds to one of the cattle chromosomes [Amaral et al. 2008].

Fig. 2. At the cytogenetic level, water buffalo chromosomes can be matched to bovine chromosomes arm for arm. Each biarmed water buffalo chromosome is derived from the fusion of two bovine acrocentrics. (a) shows similar banding patterns for bovine chromosomes 29 and 16 to water buffalo chromosome 5; (b) shows similar banding patterns for bovine chromosome 12 and water buffalo chromosome 13 [Michelizzi et al. 2010].

**African buffalo**

There are 26 chromosome pairs in Syncerus caffer and 27 chromosome pairs in Syncerus nanus. Likewise Asian buffalo, African buffalo can also be interbred and progeny has 53 chromosomes. The consequences are similar to those in Asian buffalo.
in terms of reduced fertility resulting from unbalanced gametes and zygotes [Cockett and Kole 2009]. There are four biarmed chromosomes in Syncerus caffer and three biarmed chromosomes in Syncerus nanus. All other chromosomes, including sex chromosomes, are acrocentric in this species. In Syncerus caffer biarmed pairs correspond to fused cattle chromosomes viz. 1, 13; 2, 3; 5, 20; and 11, 29 [Gallagher and Womack 1992].

Furthermore, Syncerus and Bubalus share no biarmed chromosome pairs. It suggests that the hybrid may be an unbalanced in the karyotyping results if these genera are crossed. Thus, chromosome morphology of these animals entitles them as two separate genera [Cockett and Kole 2009].

**Sex chromosomes**

All species of bovidae family show very high degree of homology among autosomal chromosomes. River buffalo, cattle, goat and sheep have similar gene order and banding patterns among the chromosomal arms [Cockett and Kole 2009, Ianuzzi et al. 2001,2009] contrary to sex chromosomes, which have a more complex rearrangement of sequences [Cockett and Kole 2009]. However, comparison of banding pattern of these chromosomes has revealed that larger parts of them are conserved. BBU-X has large blocks of constitutive heterochromatin that BTA-X lacks. Cytogenetic studies representing loci order on the sex chromosomes show complex rearrangements. This might have occurred during the karyotype evolution of river buffalo and cattle. There was a centromere translocation event with the loss of constitutive heterochromatin in cattle chromosome X, that is evident from the different centromere position. But, river buffalo and cattle X chromosomes share the same gene order [Cockett and Kole 2009]. Comparative FISH mapping indicates a similar arrangement of genes in river buffalo and cattle chromosomes Y. Cattle and buffalo chromosome Y differ in an inversion including the centromere and breakage points in both arms (pericentric inversion). Furthermore, buffalo chromosome Y gained heterochromatin regions and is larger than its counterpart in cattle [Cockett and Kole 2009].

**Nucleolar organizing regions (NORs)**

Five cattle, sheep, and goat chromosomes and six chromosomes of water buffalo have nucleolar organizing regions (NORs) in their telomeres [Cockett and Kole 2009]. These NORs coupled with conserved regions of bovine chromosomes, thus indicating the same nature of nucleolus organizer (NOCs). Some of the NORs are conserved to homologous chromosomes of different species [Gallagher et al. 1999]. Sheep, goat, cattle and river buffalo have one NOC in common. There are two NOCs shared between river buffalo and cattle showing a close evolutionary relationship [Cockett and Kole 2009].
Phylogenetics of water buffalo

A considerable, within and between population, differences have been observed in allelic variability in 21 microsatellite loci in 8 swamp and 3 river buffalo populations [Barker et al. 1997]. Bubalus bubalis and Syncerus caffer were reported as the most divergent species in the Bos clade when 20 bovine microsatellites among bovid species were studied [Ritz et al. 2000]. Kumar et al. [2006] studied genetic variability among Indian buffaloes using 27 microsatellite loci. The DNA fingerprinting analysis using amplified fragment length polymorphism (AFLP) revealed clustered African and water buffalo. This study included several phylogenetic trees with African buffalo and water buffalo [Cockett and Kole 2009]. Mitochondrial D-loop DNA sequence analysis of 19 swamp buffaloes and 61 river buffaloes suggested just one domestication event of water buffalo, (5,000 years ago in the subcontinent) [Kierstein et al. 2004]. In this study, researchers hypothesized that water buffalo was an interbred with wild buffalo or domestic buffalo from China, resulting in South-East Asian buffalo. Mitochondrial D-loop regions of Chinese swamp buffalo and other swamp and river buffaloes from Australia, India, Brazil, Italy, and Southeast posed an alternative notion of an independent domestication for river and swamp buffalo [Kumar et al. 2007a, Lei et al. 2007]. These results indicated an independent divergence of river buffalo and swamp buffalo, before the occurrence of domestication in the subcontinent and in China.

Water buffalo interleukin-12 (IL12) sequence showed significant identity with bovine IL12. There was also a functional cross-reaction with bovine immune cells [Cockett and Kole 2009]. Furthermore, Indian water buffalo interleukin-18 (IL18) showed 99% similarity in amino acid sequences with cattle IL18.

The whole genome mapping of water buffalo

Linkage, radiation hybrid (RH) and in situ hybridization mapping have been used for mapping of genomes. The linkage method was developed in the early 20th century. Principle of the linkage mapping is counting the number of offspring that receive either parental or recombinant allele combinations from a parent carrying two different alleles at two or more loci. This type of analyses tells whether loci are “linked” to each other, including their order and distance between them. Thus, this analysis needs polymorphic markers as well as reference populations.

Fluorescent in situ hybridization (FISH)

In the in situ hybridization, specific RNA or DNA sequences are localized on a chromosome using labeled probe. FISH (for fluorescent in situ hybridization) is the modified in situ protocol that utilizes fluorescent tags. The effective length of a probe is 500 bp. In contemporary research, in situ hybridization and FISH is used as genome reference.
Di Meo et al. [2008] reported the first cytogenetic map for water buffalo with only 68 loci. They were mostly assigned using FISH. Minimum one bovine molecular marker was assigned in this map to each river buffalo chromosome. Later on, maps contained 171 known genes and 122 microsatellites. To date, 309 mapped loci have been reported for river buffalo covering all chromosomes and chromosome regions [Cockett and Kole 2009].

Recently, RsaI repeat fragments pDp1, pDp2, pDp3, pDp4 of 1331, 651, 603 and 339 base pairs, have been sequenced in buffalo. Furthermore, fluorescence in situ hybridization showed that repeats are distributed all across the chromosomes. These fragments represent retrotransposons and are part of some functional genes in cattle and buffalo [Pathak and Ali 2011].

Radiation hybrid mapping (RH)

RH mapping is based on cell lines generated by fusing irradiated donor cells from the species to be studied with a rodent cell line. Irradiation produces multiple fragments at random locations of each chromosome. The key advantage of this technique is that it doesn’t require breeding and polymorphic markers.

Medium to high resolution maps can be generated using radiation hybrid mapping. These are available for many mammalian species. Recently developed RH panel for river buffalo was used to construct RH maps for buffalo chromosomes [Cockett and Kole 2009, Goldammer et al. 2007, Ianella et al. 2008, Miziara et al. 2007, Stafuzza et al. 2007]. The preliminary RH maps were developed by using cattle markers and thus proved that cattle genome is a pretty informative source of markers for mapping of the buffalo genome. This information may be used as a tool for a quick and efficient transfer of genetic information from cattle to buffalo. Thus, efforts have been made to construct a first generation genome RH map of the river buffalo using BBURH5000 covering all chromosomes [Amaral et al. 2008]. The map demonstrates improved coverage with considerable increase in the number of mapped markers, when compared to preliminary maps previously constructed.

Furthermore, based on the first generation whole genome RH map for river buffalo, a comparison was made for marker order of bovine genome sequence assembly. Buffalo marker order within the linkage groups for the chromosomes was the same as bovine genome assembly [Amaral et al. 2008, FAO 2000]. Buffalo Y chromosome RH maps contained 28 markers distributed within one linkage group [Stafuzza et al. 2009].

In recent studies six loci involving genes engaged in the dioxin metabolism pathways (ARNT, AHR, CYP1A1, CYP1A2, CYP1B1 and AHRR) were assigned to chromosomes by comparative FISH-mapping and RH mapping [Genualdo et al. 2011]. Five genes, MYOD1, MYF5, MYF6 and MYOG of basic helix-loop helix protein family have been mapped in buffalo using FISH [Strazzullo et al. 2010]. Using minisatellite-associated sequence amplification and FISH, SARS2 gene has
been localized on chromosomal mapping of Dp10 [Pathak et al. 2010]. Recently an improved cytogenetic map of buffalo has been published for chromosome X, that added up 25 loci [Perucatti et al. 2012].

The whole genome sequencing of water buffalo

Sequence data are available in the public domain for many livestock species (cattle, pigs, chicken, sheep, horse, cat and dog) in the GenBank nucleotide databases (http://www.ncbi.nlm.nih.gov). Furthermore, the whole genome sequencing has been completed for cattle, horse, chicken and dog. For buffalo, a total of 1027150 nucleotide sequences have been deposited in the GenBank database. There is a total 1868 EST, 4050 GSS, 4983 protein, 502 SNPs, 8648 UniSTSand 29 OMIA records. Out of total 1027150 sequences, 1028950 are for water buffalo, 971 for swamp buffalo and 1416 for African buffalo.

Mitochondrial genome sequencing of water buffalo

There are 26 entries for Mitochondria in gene bank. Mitochondrial genome is a circular DNA molecule. There are many copies of the mitochondrial molecule in every cell. But, the copy number is generally quite variable, e.g. oocyte contains 100,000 to 1,000,000 mitochondrial DNA molecules, but in sperm it is limited to 100-1000 molecules.

Hainan Medical College, Haikou, China (Genbank accession number: AY702168), Centre for Cellular and Molecular Biology, Hyderabad, India (Genbank accession number: AF547270) and Istituto Spallanzani, Milano, Italy (Genbank accession number: AY488491) sequenced complete water buffalo mitochondrial genome. Each mtDNA molecule in water buffalo encodes 37 genes. Out of these 37 genes, 13 are for...
proteins (polypeptides), 22 for transfer RNA (tRNA) and one each for the small and large subunits of ribosomal RNA (rRNA) [FAO 2000] (Fig. 3).

The length of the thirteen mtDNA buffalo sequences coding for four genes (COX1, ND4L, ND4 and ND6A) is same as that of Bison bison (NC_012346), Bos grunniens (NC_006380), Bos indicus (AF492350), Bos taurus (NC_006853), Capra hircus (NC_005044), Equus asinus (NC_001788), Equus caballus (NC_001640), Lama glama (NC_012102), Oryctolagus cuniculus (NC_001913), Ovis aries (NC_001941) and Sus scrofa (NC_000845). D-loop region of mtDNA (size 926 bp) is shorter in water buffalo than in cattle. Furthermore, in the water buffalo mitochondrial genome, 16 out of 37 genes overlap. This also includes two three-gene-overlaps (ND1/tRNA-Ile/tRNA-Gln and ATP8/ATP6/COX3), and five two-gene-overlaps (ND2/tRNA-Trp, tRNA-Tyr/COX1, ND4L/ND4, ND5/ND6 and tRNA-Thr/tRNA-Pro) (GeneBank acc. no. AY488491). The size of overlap is between 1 and 40 bp [Michelizzi et al. 2010]. The D-loop and CYTB regions of water buffalo have also been studied. To date, there are 911 and 497 entries for the D-loop and CYTB regions, respectively, in the GenBank databases.

**Nuclear gene/genomic sequencing of water buffalo**

There are 290 genes added since year 2010 [Michelizzi et al. 2010], and a total of 1020 known gene sequences have been contributed to the GenBank river buffalo to date. Current annotation showed 567 functional genes based on the BLAST searches against orthologous genes in mammals. There are 825 entries for sequences of genes related to growth and milk production. These include oxidized low density lipoprotein (lectin-like) receptor 1 (57 entries), leptin and leptin receptor (90), growth hormone and growth hormone receptor (131), growth hormone 1 (31), alpha lactalbumin, (19), caseins (159), insulin-like growth factor 1 (59), stearoyl-CoA desaturase (9), myostatin (15), diacylglycerol O-acyltransferase homolog 1 (19) and butyrophilin, subfamily 1, member A1 (14), lactoferrin (31), solute carrier family 11, member 1 (87), integrin beta 2 (33).

Immunity-related genes have 307 entries that include major histocompatibility complex (107 entries), CD14 (15), toll-like receptor (116), lysozyme (32), cathelicidin (17) and interleukin 2 (20).

Variations in reproduction-associated genes have been studied in buffalo. There are total 194 of them that include studies on FSH and follicle stimulating hormone receptor (25 entries), sex determining region Y (27), luteinizing hormone and receptor (18), cytochrome P450 family (19), subfamily A polypeptide 1 (9), estrogen and estrogen receptor (33 entries), progesterone (8) and prostaglandin (74).
Annotation of known gene sequences in water buffalo

There have been extensive studies in water buffalo that involved markers i.e. satellite, minisatellite and microsatellite sequencing. Nagoya University, Chikusaku, Japan [Tanaka et al. 1999] and the National Institute of Immunology, New Delhi, India have extensively contributed to satellite studies [Pathak et al. 2006]. Digestion of genomic DNA with BamHI and StuI, yielded one ~1,400 bp and ~700 bp tandem repeat units in water buffalo [Tanaka et al. 1999]. They have 79% and 81% similarity to the bovine satellite I DNA, respectively. Furthermore both satellite DNAs are located to the centromeric regions of all chromosomes. Hybridization of satellite I DNA has been identified on the acrocentric autosomes and chromosome X. It was stronger than that on biarmed autosomes and chromosome Y. The hybridization signals with satellite II DNA was equally distributed over all the buffalo chromosomes, including chromosome Y. Pathak and her colleagues [Pathak et al. 2006] identified 1378-bp and 673-bp repeat satellite sequences in the buffalo genome. Furthermore, real-time PCR analysis revealed 1234 and 3420 copies of 1378-bp and 673-bp fragments per haploid genome, related to 30 and 68 copies per chromosome, respectively. Both 1378-bp and 673-bp repeat fragments are abundantly expressed in the spleen and liver.

To date there are 398 minisatellite sequences deposited in the GenBank which came mostly from the National Institute of Immunology, New Delhi, India. Minisatellite-associated sequence amplification with an oligo (5’ CACCTCTCCACCTGCC 3’) has been performed. The oligo was designed using 33% repeat loci from cDNA of various tissues of water buffalo [Srivastava et al. 2006]. The minisatellites with the size of 1263, 846/847, 602, 576, 487, and 324 bp were identified. Furthermore, BLAST searches showed that the 846/847-bp fragment has homology with the adenylate kinase gene and the 1,263, 324, and 487-bp fragments showed homology with the secreted modular calcium binding protein (SMOC-1), leucine-rich repeat neuronal 6A (LRRN6A) mRNA, and human TTTY5 mRNA, respectively. There are currently a total of 1001 microsatellite records in GenBank, mainly submitted by the Centre for Cellular and Molecular Biology, Hyderabad, India (312; with accession numbers AY775830 - AY775944, AY779565 - AY779623, AY787147 - AY787166, AY805331 - AY805389 and AY912133 - AY912182).

Next generation sequencing of water buffalo genome

BAC sequencing and shotgun sequencing has been mainly used for whole genome sequencing. These methods use Sanger method that has drawbacks of being costly and time consuming [Metzker 2005]. These pitfalls of traditional methods have given birth to low-cost next-generation sequencing.

At Anand Agricultural University, Anand, India’s researchers completed <0.53x genome sequencing of water buffalo in year 2009 using Roche 454 GS-FLX Titanium technology and updated the entry in 2012. Total 1,498,523 buffalo sequences resulting
from NGS were submitted to the GenBank (ACZF020000001-ACZF021498523). The submitted sequences range from 92 bp to 4,726 bp in size. Furthermore, entries with 201–500 bp account for 83.3% [Cockett and Kole 2009]. Overall, the 454 GS-FLX sequencing technology has contributed 23,173,770 bp in total to the species of *Bubalus bubalis*.

The whole water buffalo genome was sequenced in year 2011 [Tantia 2011]. In collaboration with the University of Florida, Washington State University. Illumina GAIIx technology was used to produce approximately 40 Gb of sequences for water buffalo.

Zimin *et al.* [2013] reported on high quality *de novo* assembly of the water buffalo genome generated from mix of Illumina and 454 data from MSR-CA assembler. That work was collaborated by the University of Maryland, USA, USDA-ARS, USA and CASPUR, Italy. The assembly contained 2.76 Gb in scaffolds. This was followed by collaborative work between Bangladesh (Lal Teer Livestock Limited) and China (Beijing Genomics Institute) in year 2014 to sequence buffalo genome with the size of about 2.77 Gb.

Concluding, the vast amount of genetic/genomic resources for cattle research will serve as shortcuts for the water buffalo community to study functional genomics. Buffalo genome resources are going to contribute to develop genome technologies for researchers and policy makers and may serve to improve reproductive performance and production potential of water buffalo.

REFERENCES


27. PATHAK D., ALI S., 2011 – RsaI repetitive DNA in Buffalo Bubalus bubalis representing retrotransposons, conserved in bovids, are part of the functional genes. **BMC Genomics** 12, 338.
29. PATHAK D., SRIVASTAVA J., SAMAD R., PARWEZ I., KUMAR S., 2010 – Genome-wide search of the genes tagged with the consensus of 33.6 repeat loci in buffalo Bubalus bubalis employing minisatellite-associated sequence amplification. **Chromosomes Research** 18, 441-458.