We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Repetitive DNA: A Tool to Explore Animal Genomes/Transcriptomes

Deepali Pathak and Sher Ali

Additional information is available at the end of the chapter

1. Introduction

The analysis of genetic diversity and relatedness within and between the different species and populations has been a major theme of research for many biologists. With the availability of whole-genome sequencing for an increasing number of species, focus has been shifted to the development of molecular markers based on DNA or protein polymorphism. DNA sequences originate and undergo evolutionary metamorphoses' and thus may be used as powerful genetic markers to characterize genomes of wide range of species. This type of analysis is called fingerprinting, profiling or genotyping. DNA profiling based on typing individuals using highly variable minisatellites in the human genome was first developed by Jeffreys et al (1985). He demonstrated short repeat sequences tandemly arranged within the gene(s) and each organism has a unique pattern of the arrangement of these minisatellites, the only exception being multiple individuals from a single zygote (e.g. identical twins). DNA fingerprinting technique was notably used to help solve crimes and determine paternity. In addition, with the advances in Molecular biology techniques, isolation of genes tagged with minisatellites has become the most powerful tool for genome analysis.

The term "repetitive sequences" (repeats, DNA repeats, repetitive DNA) refers to DNA fragments that are present in multiple copies in the genome. These sequences exhibit a high degree of polymorphism due to variation in the number of their repeat units caused by mutations involving several mechanisms (Tautz, 1989). This hypervariability among related and unrelated organisms makes them excellent markers for mapping, characterization of the genomes, genotype phenotype correlation, marker assisted selection of the crop plants, molecular ecology and diversity related studies. The nature of repeats provides ample working flexibilities over the other marker systems. This is because: (i) short tandem repetitive (STR) sequences are evenly distributed all over the genome (ii), are often

conserved between closely related species (iii) and are co-dominant. With these innate attributes, very small quantities of DNA can be used for simultaneous detection of the alleles tagged with STR employing minisatellite associated sequence amplification (MASA).

Current data base has the information on the genomes of various livestock species, like cattle, sheep, goat, pig, horse, chicken, Silkworm, Honey Bee, Rabbit, Dog, cat and duck (Georges and Andersson, 1996). However complete sequence analysis of several important species such as Yak, Banteng, Zebu, Donkey, Goose, Turkey, Camel and Water buffalo are still underway. Efforts are required to characterize genes controlling important traits in order to produce genetically healthy breeds and segregate superior germplasm wherever possible. Water buffalo, *Bubalus bubalis* is important domestic animal worldwide having immense potential in agriculture, dairy and meat industries. We have studied several repeat loci in buffalo genome using Restriction Fragment Length Polymorphism and characterized a number of important genes employing Minisatellite Associated Sequence Amplification (MASA).

MASA forms a rich basis of functional and comparative genomics contributing towards the understanding of genome organization, gene expression and development of molecular synteny. This approach also enables characterization of the same genes across the individuals within a species and amongst the individuals between the species. Thus, information about the organization of gene, its expressional, mutational and phylogenetic status, chromosomal location and genetic variations across the genomes maximize the chances of narrowing the search of possible genetic markers.

In this chapter, we discuss overall organization of the repetitive sequences, their origin, distribution, application in genome analysis and implications. In addition, use of repetitive sequences in bubaline genome mining is highlighted elucidating the potential of functional and comparative genomics. Thus, organizational variation and expressional profile of a single gene originating from a specific tissue may be studied in many ways to meet the varying requirements of biology.

2. Organization of repetitive sequences

The mammals have approximately 3 billion base pairs per haploid genome harboring about 20,000-25000 genes. A minor part of the genome (5-10%) is coding sequences (International Human Genome Sequencing Consortium, 2004; Hochgeschwender and Brennan, 1991) and the remaining part is non-coding representing repetitive DNA (Bromham, 2002). Comparison of the genome size of different eukaryotes shows that the amount of non-coding DNA is highly variable and constitutes 30% to about 99% of the total genome (Elgar and Vavouri, 2008; Cavalier-Smith, 1985). The non-coding repetitive sequences are dynamic elements, which reshape their host's genome by generating rearrangements, shuffling of genes and modulating pattern of expression. This dynamism of repeats leads to evolutionary divergence that can be used in species identification, phylogenetic inference and for studying process of sporadic mutations and natural selection. These repetitive sequences are mainly composed of interspersed and tandem repeats (Slamovits and Rossi,

2002). The later includes satellite, minisatellites and microsatellites. Satellite DNAs are predominantly associated with centromeric heterochromatin and the same is being increasingly utilized as a versatile tool for genome analysis, genetic mapping and for understanding chromosomal organization. On the other hand minisatellite and microsatellites are dispersed throughout the genome and are highly polymorphic in all populations studied. This arrangement has led to their extensive use as genetic markers for fingerprinting, genotyping, and for forensic analysis in human system. Based on their arrangements, repetitive DNA sequences are classed into two types (Figure 1).

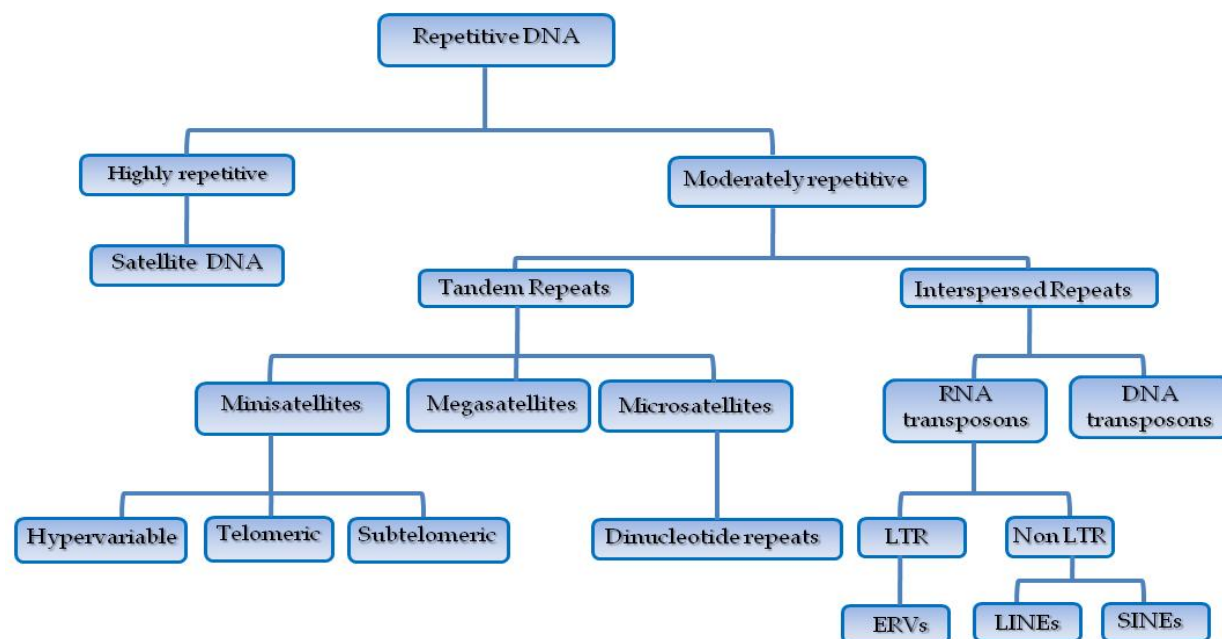## 2.1. Highly repetitive sequences

These are are short sequences (5 to10 bp) amounting 10% of the genome and repeated a number of times, usually occurring as tandem repeats (present in approximately 10^6 copies per haploid genome). However, they are not interspersed with different non-repetitive sequences. Usually, the sequence of each repeating unit is conserved. Most of the sequences in this class are located in the heterochromatin regions of the centromeres or telomeres of the chromosomes. Highly repetitive sequences interacting with specific proteins are involved in organizing chromosome pairing during meiosis and recombination.

2.1.1. Satellite DNA

These are represented by monomer sequences, usually less than 2000-bp long, tandemly reiterated up to 10^5 copies per haploid animals and located in the pericentromeric and or telomeric heterochromatic regions (Charlesworth et al 1994). Satellite DNA constitutes from 1 to 65% of the total DNA of numerous organisms, including that of animals, plants, and prokaryotes. The term "satellite" in the genetic sense was first coined by the Russian cytologist Sergius Navashin, in 1912, initially in Russian ("sputnik") and Latin (*satelle*), and was later translated to "satellite" (Battaglia, 1999). The more familiar usage of "satellite" relates to a small band of DNA with a density different (usually lower, because of a high AT-content) from the bulk of the genomic DNA, which are separated from the main band following CsCl centrifugation (Kit, 1961). Nucleotide changes and copy number variations fuel the process of their evolution within and across the species (Ugarkovic and Plohl, 2002). Satellite fraction(s), though not conserved evolutionarily (Bhatnagar et al 2004; Amor and Choo, 2002), are unique to a species and usually show similarity amongst related group of animals (Pathak et al 2011; Henikoff et al 2001; Ali and Gangadharan, 2000).

2.2. Moderately or dispersed repetitive sequences

These include short (150 to 300-bp) sequences or long ones (5-kbp) amounting about 40% and 1-2% of the total genome, respectively. These are dispersed throughout the euchromatin having 10^3-10^5 copies per haploid genome. These sequences are involved in the regulation of gene expression. In some cases, long dispersed repeats of 300 to 600-bp show homology with the retro viruses.

Figure 1. Schematic diagram showing biological categories of the different repetitive sequences.

On the basis of their mode of amplification, repetitive DNA sequences may be tandemly arranged or interspersed in the genome (Slamovits and Rossi, 2002).

2.2.1. Interspersed repetitive DNA

Interspersed repeat sequences scattered throughout the genome have arisen by transposition, having "ability to jump from one place to another in the genome" (Miller and Capy, 2004; Brown, 2002). Even though the individual units of interspersed repetitive noncoding DNA are not clustered, taken together they account for approximately 45% of the human genome. By the mechanism of their transposition, interspersed repeats are classified into two classes:

2.2.1.1. RNA transposons

RNA transposons also known as retroelements found in eukaryotic genome require reverse transcription for their activity. Based on their structural relationship, RNA transposons are divided into two general categories:

2.2.1.1.1. LTR elements

LTR includes retroviruses whose genomes are made up of RNA. They infect different types of vertebrates.

Endogenous retroviruses (ERVs)

These are retroviruses integrated into the vertebrate chromosomes and inherited from generation to generation as part of the host genome. Some are still active and might, at some stage in a cell's lifetime, direct synthesis of the exogenous viruses. However, majority of them are decayed relics and no longer have the capacity to form viruses (Patience et al 1997).

Retrotransposons are the biggest class of the transposons. An important characteristic of this type of transposable element is that they usually contain sequences with potential regulatory activity. They have features of non-vertebrate eukaryotic genomes (i.e. plants, fungi, invertebrates and microbial eukaryotes). These elements code for mRNA molecule which is processed and polyadenylated. Retrotransposons have very high copy numbers. In maize, these elements occupy half of the genome.

2.2.1.1.2. Non LTR elements

LINEs (Long Interspersed Nuclear Elements)

LINEs are several thousand base pairs in size and make up about 17% of the total human genome (Richard and Batzer, 2009). They contain reverse-transcriptase-like gene involved in retrotransposition process. Many LINEs also code for an endonuclease (e.g. RNase H). The most abundant LINE family is the 7-kbp, L1 repeat element having >500,000 copies and accounts for approximately 15% of the human genome (Lander et al 2001). Despite its abundance, no function of LINE 1 repeat is yet known. Initial studies on mouse have associated L1s in shaping the structure and expression of the transcriptomes (Han et al 2004; Han and Boeke, 2005).

SINEs (Short Interspersed Nuclear Elements)

SINEs are small elements, usually 100 to 500-bp in length, accounting for 11% of the human genome (Richard and Batzer 2009). SINEs do not have reverse transcriptase gene, instead they borrow reverse transcriptase enzymes from other retroelements. Well-known example of SINE in the human genome is *Alu* sequences (Capy et al 1998), which are 350 base pairs long, do not contain any coding sequences, and have over 1 million copies (Roy-Engel et al 2001)

2.2.1.2. DNA Transposons

DNA transposons do not require RNA intermediate and transpose in a direct DNA-to-DNA manner. In eukaryotes, DNA transposons are less common than retrotransposons, but they have a special place in genetics because a family of plant DNA transposons - the Ac/Ds elements of maize. There are two types of DNA transposons that both require enzymes coded by genes within the transposons.

2.2.2. Tandem repeats

Tandem repeats consists of repeat arrays of two to several thousand-sequence units arranged in a head to tail fashion. Tandem repeats may be further classified according to the length and copy number of the basic repeat units as well as its genomic localization.

2.2.2.1. Mega satellite DNA

These are characterized by tandemly repeated DNA in which the repeat unit is approximately 50-400 times, producing blocks that can be hundreds of kilobases long. Some mega satellites are composed of coding repeats. For example: RNA genes, and the deubiquitinating enzyme gene USP17.

2.2.2.2. Minisatellite DNA

This comprises tandem copies of repeats that are 6-100 nucleotides in length (Tautz, D. 1993). Alec Jeffrey's first described minisatellites in 1985, from the non-coding (intron) regions of the human myoglobin gene. Since then similar DNA structures have been reported in many organisms including bacteria (Skuce et al 2002), avian (Reed et al 1996), higher plants (Sykorová et al 2006; Durward et al 1995), protozoan (Feng et al 2011; Bishop et al 1998), and yeast (Kelly et al 2011; Haber and Louis, 1998) genomes. Comparison of the repeat units in classical minisatellites led to early notion of consensus or core sequences, which exhibit some behavioral similarities with the *Chi* sequences of λ phage (GCTGTGG). Also called as variable number of tandem repeats (VNTR) (Brown 2002), majority of the minisatellites are GC rich, with a strong strand asymmetry. Often minisatellites form families of related sequences that occur at many hundred loci in the nuclear genome. In human genome, number of minisatellite loci is estimated to be approximately 3000 and each locus contains a distinctive repeat unit with respect to size and sequence content. The degree of repetition ranges from two to several hundreds. Repeat unit within a minisatellite usually display small variations in sequence. Minisatellite mutations usually consist of gains or losses of one or more repeat units. Such mutations at hypervariable minisatellite loci are up to 1000 times more common than mutations in protein coding genes (Debrauwere et al 1997).

In the humans, majority of minisatellites are clustered near sub-telomeric ends of the chromosomes limiting their usefulness for extensive gene mapping (Lopes et al 2006; Royle et al 1988), but there are examples of interstitial locations (alpha globin gene cluster (Proudfoot et al 1982) and type II collagen gene (Stoker et al 1985). Minisatellites of other species, such as mice or bovine (Georges et al 1991), are not always preferentially clustered at chromosomal termini as in the human genome, but are distributed along the entire length of chromosomes. Unlike microsatellites, which usually alter during the DNA synthesis stage of the mitotic cell cycle, minisatellites alter during meiosis, undergoing changes in overall length and repeat composition (Jarman and Wells, 1989; Jeffrey's et al 1998). Minisatellite tracts have proven very useful for genomic mapping (Legendre et al 2007; Jeffrey's et al 1985) and linkage studies (Nakamura et al 1987). Examples of human minisatellite used for fingerprinting include consensus sequence of 33.6, 33.15 repeat loci. List of other minisatellite sequences according to Ali and Wallace, (1988) are mentioned in Table 1.

2.2.2.2.1. Telomeric repeats

These are composed of multiple repeats of short sequence elements (typically 5 to 8-bp in length, with a GT-rich strand oriented 5' to 3' toward the end of the chromosome) and range in length from a few repeat units to >10-kbp. Long simple sequence tandem repeats of interstitial TTAGGG arrays form a three-dimensional nuclear network of poorly transcribed domains, which involve gene silencing by repositioning. This network, as well as clusters of retroelements properly positioned in the nucleus, form unique lineage-specific structures that affect gene expression (Tomilin, 2008). The repeated sequence $(TTAGGG)_n$ is found at telomeres in all vertebrates, certain slime molds, and trypanosomes; $(TTGGGG)_n$ and

(TTTGGGG)$_n$ are found in the ciliated protozoan *Tetrahymena* and *Oxytricha* species, respectively; and (TG$_{1-3}$)$_n$ is found in the yeast *Saccharomyces cerevisiae*. In organisms whose telomeres have been examined in detail, the GT strand extends 12 to 16 nucleotides (two repeats) beyond the complementary C-rich strand. The unique structure of telomere is involved in the maintenance of the integrity of the chromosome ends.

S.No.	Probe name	Sequence 5'-3'	Total length	Length repeat unit	No. Repeat units
1	0-33.6-22	(CCTCCAGCCCT)$_2$	22	11	2
2	0-33.6-37	GCCCTTCCTCCGGAGCCCTCCTCCAGCC CTTCCTCCA	37	37	1
3	0-33.15-32	(CACCTCTCCACCTGCC)$_2$	32	16	2
4	0-33.15-80	(CACCTCTCCACCTGCC)$_5$	80	16	5
5	0-AY-29	GAGGARYAGAAAGGYGRGYRVTGTGGG CGC	29	37	1
6	0-YN-124	TCCTGAACAACCCCACTGTACTTCCCA	27	31	1
7	0-33.1	GTGCCTGCTTCCCTTCCCTCTCTTGTC	27	62	1
8	0-34BHI	CCTGCTCCGCTCACGTGGCCCACGCAC	27	?	1
9	0-CCR-26	(CCR)$_8$CC	26	3	8.67
10	0-CCA-26	(CCA)$_8$CC	26	3	8.67
11	0-H-Ras	CACTCCCCCTTCTCTCCAGGGGACGCCA	28	28	1
12	0-GACA-16	(GACA)$_4$	16	4	4
13	0-GACA-24	(GACA)$_6$	24	4	6

Table 1. Sequences and hybridization characteristics of the oligonucleotides probes (Ali and Wallace, 1988)

2.2.2.2.2. Subtelomeric repeats

Sub-telomeric Repeats are the classes of repetitive sequences that are interspersed within the last 500,000 bases of non-repetitive DNA located adjacent to the telomere. Some sequences are chromosome specific whereas others seem to be present near the ends of all the human chromosomes (Norman, 2001).

2.2.2.3. Microsatellite /Short Sequence Repeats (SSRs)

Tandem repeats are made up of usually, di-, tri-, or tetranucleotide units (1-6 bps), were earlier called simple sequences (Tautz and Renz, 1984). Later, this class of DNA was coined as microsatellites by Tautz 1989. Microsatellites or simple sequence repeats (SSRs) are ubiquitously interspersed in coding and non-coding regions of the eukaryotic and prokaryotic genomes (Gur-Arie et al 2000; Toth et al 2000). All the SSRs taken together occupy about 3% of the human genome in which they are widely dispersed and associated

with many genes (Subramanian et al 2003). The significance of specific microsatellite in different regions has not been completely understood. However, some microsatellites occurring in flanking regions of coding sequences are believed to play significant roles in regulation of gene expression by forming various DNA secondary structures and offering a mechanism of unwinding (Catasti et al 1999). The variation of length and unit type of simple repeats in upstream activation sequences might influence transcriptional activity (Kim and Mullet, 1995; Epplen et al 1996; Martienssen and Colot, 2001; Zhang et al 2004), and affect interaction with different regulatory proteins during translation (Lue et al 1989).

Microsatellites are usually characterized by low degree of repetition at a particular locus. However, these elements containing identical motifs may be found at many thousand genomic loci. When the occurrence of SSRs in different functional genome regions is considered, it turned out that most of them show much higher density in non-coding regions. Exceptions to the rule are trimers and hexamers that are nearly two times more prevalent in exons compared to introns and intergenic regions. Their high frequency in coding regions may be explained by the fact that they do not change the reading frames and gene coding properties, thus, are much better tolerated than other SSRs. Their positive selection in exons suggests some functions for these repeats.

The high mutation rate of these repeats and their frequent length polymorphism suggest that they may be involved in the regulation of gene expression thus leaving quantitative effects on the phenotype. Few examples of repeat units used for fingerprinting and transcriptome analysis includes (GATA/GACA)n, CA, (AT)n, (GAA)n, (TCC)n, (GGAT)n, (GGCA)n, and (TTAGGG)n.

3. Evolution and inter-species variation of repeat sequences

Several mechanisms have been proposed for their evolution, such as stand slippage during replication, base misalignment and unequal cross over between homologous chromosomes during meiosis, sister chromatid exchanges or even insertion of the viral genome (Barros, 2008; Jeffrey's et al 1985; Tautz, 1989).

Microsatellites tend to be highly polymorphic, suggesting a 'stepwise mutation' model in which most variations are introduced by replication slippage, changing the array length by only one or two repeats at a time, but also with occasional larger 'jumps' in size at much lower frequency. Minisatellites, evolve more readily by larger-scale mechanisms such as unequal exchanges. For all classes, there appears to be a general bias towards increase in array length through evolutionary time. Highly repetitive DNA tends to accumulate only in regions of low recombination such as centromeres and telomeres, where recombination is suppressed, while repeats occurring in euchromatin are much more susceptible to crossing-over and tend to be more variable in copy number relative to their array length.

As mentioned above, mechanism of loss or gain of repeat by unequal cross over and gene conversation can lead to molecular drive of any given variant in a sexually dimorphic population. During the evolution of repetitive elements by unequal cross over, some

variants will be lost whereas others will increase in frequency, eventually replacing all others. These evolutionary changes leads to homogeneity in the repeats of an array within a species and heterogeneity in the units of the corresponding array in different species, giving rise to inter- species variations (Harris and Wright, 1995). This phenomenon however is affected by overall male female ratios, population size and possibility of infusion of newer genetic materials in a given gene pool and allele fixation involving evolutionary incubation time.

4. Functional significance of repetitive sequences

With respect to functional roles of these sequences, uncertainty persisted for a long time and it was largely believed that they represent detritus part of the genome (Ohno, 1972). However, recent studies have shown repeat elements influencing the structure, function, and evolution of the chromosomes in the host genomes (Sinden, 1999; Dey and Rath, 2005; Tang, 2011). Their association with the promoters and coding regions of the genes has made them very attractive objects of the study. Transcription, mRNA processing, translation, folding, stability and aggregation rates, as well as gross morphology have been found to be incrementally affected by the alterations in the tracts of tandem repeats (Fondon and Garner, 2004; Vinces, 2009). The human genome provides many instances of regulatory regions embedded in the remnants of repeat elements (Jordan et al. 2003) and studies have documented participation of repeat sequences in regulation of gene expression (Boeva et al., 2006). This suggests that the repeat elements play a major architectonic role in higher order of physical structuring of the genome (Shapiro and Sternberg, 2005; Vermaak et al 2009). More studies on repeat sequences will lead to an increased understanding on the functions and dysfunctions of the genomes.

5. Significance of repetitive sequences as marker

Primers based on VNTR provide an unprecedented opportunity to develop potential molecular markers for a particular species. Where a complete genome sequence is available for an organism, repeats may be annotated with their physical position on the genome. Markers may then be selected either for their location within a specific region of interest or for their even distribution across the regions. Where a full genome sequence is unavailable, location may be predicted through synteny using a sequenced genome or through previous mapping exercises. Alternatively, for a genome whose sequences are not known can still be analyzed employing primers from other species for gene amplification. A gene so amplified may then be localized onto the chromosomes employing FISH. Similar set of primer may be used to amplify cDNA of the species. This approach circumvents the need for screening the genomic library.

Furthermore, for species which exhibit low levels of polymorphism at repeat loci, candidate polymorphic loci may be predicted through mining large sequence datasets. The presence of short sequence repeat (SSR) polymorphisms within aligned sequences of different origin

would be indicative of the level of polymorphism at that locus. These selection strategies could greatly reduce the time and cost associated with the development of repeat markers. Integration of this repetitive sequence data with genome databases would provide further benefits to genome researchers.

6. *Bubalus bubalis* genome

The water buffalo (*Bubalus bubalis*) population in the world is actually about 168 million head, of which 161 million can be found in Asia (95.83 percent); 3717 million are in Africa and Egypt (2.24 percent); 3.3 million (1.96 percent) in South America, 40 000 in Australia (0.02 percent); 500 000 in Europe (0.30 percent). Asian buffalo or Water buffalo is classified under the Genus: *Bubalus*, Species: *bubalis*. Asian buffalo includes two subspecies known as the River and Swamp types, the morphology and purposes of which are different so are the genetics. The River buffalo has 50 chromosomes of which five pairs are sub-metacentric, while 20 are acrocentric: the Swamp buffalo has 48 chromosomes, of which 19 pairs are metacentric. Swamp buffaloes are stocky animals with marshy land habitats. They are primarily used for draught power in paddy fields and haulage but are also used for meat and milk production. They produce a valuable milk yield of up to 600 kg milk per year, Swamp buffaloes are mostly found in South East Asian countries. A few animals can also be found in the northeastern states of India (Sethi, 2003). River buffaloes are generally large in size, with curled horns and are mainly found in India, Pakistan and in some countries of western Asia. They prefer to enter clear water, and are primarily used for milk meat and draught purposes. Each subspecies includes several breeds. Buffaloes are known to be better at converting poor-quality roughage into milk and meat. They are reported to have a 5 percent higher digestibility of crude fiber than high-yielding cows; and a 4-5 percent higher efficiency of utilization of metabolic energy for milk production (Mudgal, 1988).

India has about 97 million animals, which represents 92% of the world buffalo population. India possesses the best River milk breeds in Asia e.g. Murrah, Nili-Ravi, Surti Jaffarabadi, Mehsana, Kundi, Bhadavari and Nagpuri which originated from the north-western states of India (Sethi, 2003). However, despite the importance of buffalo to the economic and social fabric of the region, its population has been declining. There are many reasons for the decline of buffalo populations, foremost of which are: increased agricultural mechanization; increased urbanization, industrialization, and reforestation limiting paddy areas for buffaloes; growing buffalo slaughter rate to satisfy meat demands of a fast-growing population; poor reproductive performance; and lack of proper attention by policy makers and researchers. The low reproductive efficiency in female buffalo can be attributed to delayed puberty, higher age at calving, long postpartum anoestrus period, long calving interval, lack of overt sign of heat, and low conception rate. In addition, female buffaloes have few primordial follicles and a high rate of follicular atresia. Understanding potential quantitative trait loci associated with economically important traits will help in segregating genetically superior breeds.

7. Repetitive sequences as molecular markers in bovid genome

Based on repeat sequences, a number of probes with varying length and sequence complexities have been successfully used as genetic markers (Kapur et al 2003; Jobling and Tyler-Smith, 2003; Bashamboo and Ali, 2001; Amos et al 1991; Tourmente et al 1994; Ali et al 1986). Earlier conventional protein and biochemical markers were used for breeding program of bubaline species (Wilson and Strobeck, 1999). Subsequently, diallelic Restriction Fragment Length Polymorphism (RFLP) for the loci homologous to cattle (Blott et al 1999) were used. However due to low levels of polymorphism detected with these markers, their application remained limited. RFLP technology was followed by Random amplification of Polymorphic DNA (RAPD), followed by Amplified fragment length Polymorphism (AFLP) besides minisatellite markers. A series of synthetic oligonucleotide probes were developed as markers for genetic analysis and molecular systematics of Bubaline and related genomes. While probes based on repeat sequences are available there is no clear cut experimental approach that could assist identification and segregation of elite animals with superior QTL loci. This is because most of the physical and physiological attributes recognized to be the part of the elite animals, are controlled by several genes and it is extremely challenging to uncover all such genes implicated with superior germplasm. However marker based analysis would possibly bridge the gap and facilitate much-needed advance research to segregate genetically superior germplasm in the context of animal genetics in general and animal biotechnology in particular.

8. Restriction Fragment Length Polymorphism (RFLP)

The basic technique for detecting RFLPs involves the fragmentation of genomic DNA by a restriction enzyme. The resulting DNA fragments are then separated by length through a process known as agarose gel electrophoresis, and transferred to a membrane via the Southern blot procedure. Hybridization of the membrane to a labeled DNA probe then determines the size of the fragments, which are complementary to the probe. An RFLP occurs when the size of a detected fragment varies between individuals. Each fragment size is considered an allele, and can be used in genetic analysis. RFLP's are quick, simple and inexpensive ways to assay DNA sequence differences. It is the first DNA polymorphism to be widely used for genomic characterization, which detects variations ranging from gross rearrangements to single base changes. The polymorphisms are found by their effects on sites for restriction enzyme mediated cleavage of preparations of high molecular weight DNA. In buffalo, RFLP approach has been used to gain insight into organization and allele length variation of satellite fractions (Chattopadhyay et al 2001; Bhatnagar et al 2004). From our laboratory, *BamH*1 derived pDS5 and pDS4 and *Rsa*I derived pDp1-pDp4 were found to be conserved only in buffalo, cattle, goat, and sheep (Pathak et al 2006; 2011).

9. Minisatellite Associated Sequence Amplification (MASA)

MASA involves random amplification of genomic or cDNA with primers specific to minisatellites by PCR. MASA can be performed with a small quantity of target substrate.

The novel part of the current approach is that functional, structural and regulatory genes associated with minisatellites are accessed without screening the conventional cDNA library proving this be highly useful for such genome analysis where prior information is absent or inadequately available. The expression profile of genes based on MASA under normal and abnormal conditions is envisaged to be of great relevance for identification of event/stage specific mRNA transcripts. In the context of comparative genomics, mRNA transcripts commonly expressing in a large number of species may be segregated. Following this approach, genes with highest levels of expression in a given tissue may be easily identified and the information from different breeds of animals may be established. In addition, differential expression of genes accessed by MASA may be used to establish genotype phenotype correlation in the context of genetic diseases, cancer biology, stem cell research, tissue engineering, organ transplantation, animal cloning, characterization of genetic integrity of different cell lines and conducting translational research. Minisatellite sequences 33.6, 33.15 have been widely used to explore bubaline genome (Srivastava et al 2006; 2008; Pathak et al 2010). In addition microsatellite probes (2-6 base pairs) such as (AT)n, (CA)n, (GAA)n, (TCC)n, (GACA)n, (GATA)n, (GGAT)n, (GGCA)n and (TTAGGG) were used to analyze buffalo genome (Rawal et al 2012; kumar et al 2011). Following this approach, additional oligo primers based on VNTR loci may be used to undertake analysis of any desired species, cell lines, biopsied samples and cell lines.

10. Technical approaches and methodologies

We describe some of our works related to characterization of the buffalo genome. Further, in the context of functional and comparative genomics, DNA from across the species were also used. DNA was largely procured from the blood samples though in some cases, solid tissues were also used.

10.1. Collection of blood samples and isolation of genomic DNA

DNA was extracted from peripheral blood of buffalo *Bubalus bubalis*, goat *Cipra hircus* sheep *Ovis aries* tiger *Panthera tigris*, lion *Panthera leo*, humans *Homo sapiens*, langur *Presbytis entellus*, Indian rhinoceros *Rhinoceros unicornis*, fish *Hetropnustes fossilis*, bird *Columba livia*, baboon *Papio hamadryas*, pig *Sus scrofa*, rat *Rattus norvegicus*, jungle cat *Felis chaus*, bonnet monkey *Macaca radiate* and leopard *Panthera pardus*. Intactness of DNA was checked on 1% agarose gel and DNA was PCR amplified using bubaline derived β actin primers and visualized on UV transilluminator.

10.2. RNA isolation and synthesis of cDNA

Using buffalo as an experimental animal, total RNA was extracted from testis, kidney, liver, spleen, lung, heart, ovary, brain and sperm using TRIzol (Molecular Research Center, Inc., Cincinnati, OH) following manufacturer's instructions. To check the contamination of mRNA from the cells other than spermatozoa, RNA extractions from the sperms were tested

by RT-PCR both for the *CDH1* (E-cadherin) and *CD45* (tyrosine phosphatase). Similarly, presence of DNA was ruled out by PCR using β-actin primers. Following this, approximately 10 μg of RNA from different tissues and spermatozoa was reverse transcribed into cDNA using commercially available high capacity cDNA RT kit (Applied Biosystems, USA). The success of cDNA synthesis was confirmed by PCR employing 35 cycles of amplification using buffalo derived β-actin primers.

10.3. Minisatellite Associated Sequence amplification (MASA)

Using oligo primer and cDNA from different tissues and spermatozoa, PCR amplifications were carried out. The reaction conditions involved 95°C denaturation for 5 min followed by 35 cycles each consisting denaturation at 95°C for 1 min, annealing at the optimal temperature for 1.5 min, extension of the primer at 72°C for 1 min and final extension at 72°C for 10 min. Approximately, 25 μl of amplified product was resolved on a 20-cm-long, 3% (w/v) agarose gel in 1× TBE buffer at a constant voltage. The distinct bands were sliced from the gel, purified and cloned into pGEMT-easy vector (Promega, USA). In water buffalo, *Bubalus bubalis* using cDNA from the spermatozoa and eight different somatic tissues and an oligo primer based on two units of consensus of 33.6 repeat loci (5' CCTCCAGCCCTCCTCCAGCCCT 3'), Minisatellite-associated sequence amplification (MASA) identified 29 mRNA transcripts (Figure 2).
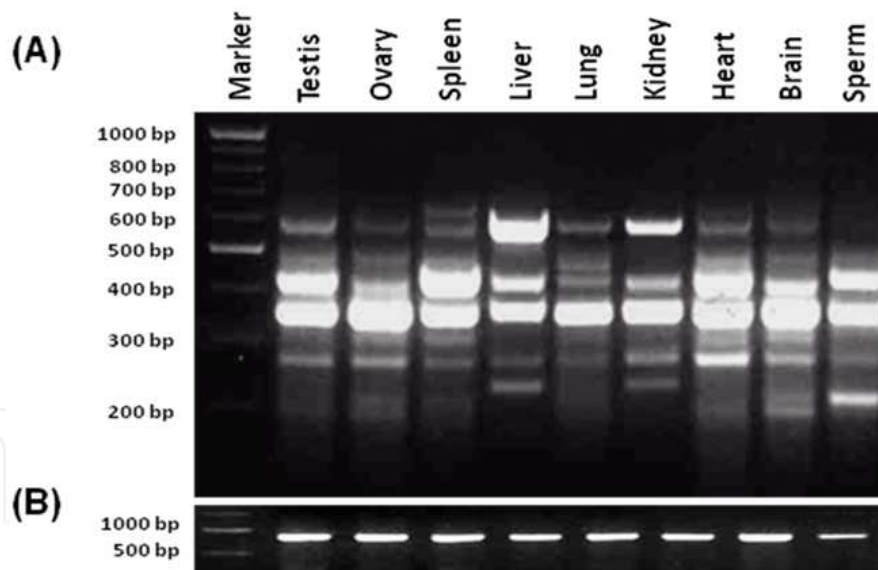

Figure 2. A representative agarose gel showing minisatellite associated sequence amplification (MASA) with cDNA from different somatic tissues and spermatozoa of buffalo as shown on top of the lanes in panel **(A)**. β-actin was used as an internal control **(B)**. M is the molecular marker given in base pairs (bp) (for details, see Pathak et al, 2010).

10.4. Restriction digestion of buffalo genomic DNA

Approximately, 4-5 μg of genomic DNA from buffalo, cattle, goat and sheep were subjected individually to restriction digestion using 4-5 units of *Bam*HI and *Rsa*1enzyme. The digested

DNA fragments were resolved on 0.8% agarose gel in 0.5X TBE for approximately 16-18 hours. In water buffalo, two distinct DNA bands of 1378 and 673 bp with Bam *HI* and four bands of 1331, 651, 603 and 339 base pairs were cut, gel purified (Figure 3). The eluted fragments were cloned and sequenced following standard protocol. For Southern hybridization, DNA was transferred onto Nylon membrane and immobilized by exposure to UV. Membranes were rinsed in 2X SSC, dried and UV cross- linked. Blots were hybridized at 60^{0}C overnight with ^{32}P α-dCTP labeled recombinant plasmid (25 ng) using random priming method (rediprimeTM II kit, Amersham Pharmacia biotech, USA). Washing of the membranes was done using standard protocols and signals were recorded by exposure of the blot to X-ray film (Pathak 2006; 2011).


(a)   (b)
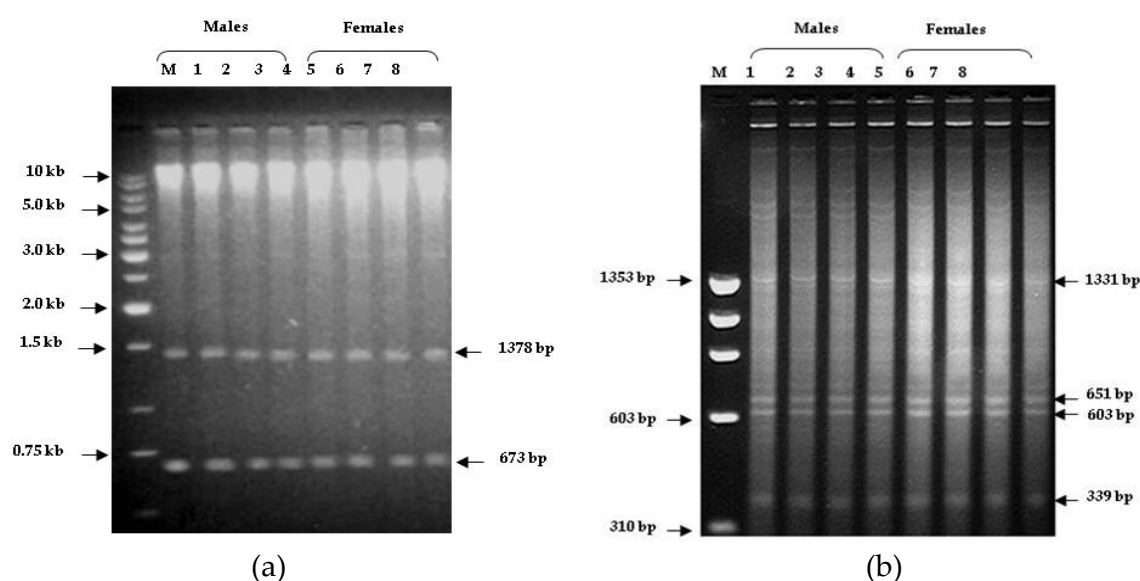
Figure 3. Agarose gel showing restriction digestion of buffalo *Bubalus bubalis,* genomic DNA with *BamH*I **(A)** and *Rsa*I **(B)** enzymes. The two discernible bands 673 bp and 1378 bp with *BamH*I digestion and four bands 1331, 651, 603 and 339 bp in *Rsa*I are highlighted. Molecular weight marker is given on the left in base-pair (bp). Since the patterns are not gender species, this suggests that the bands are originated from the autosomes (for details, Pathak et al 2006; 2011).

10.5. Copy number assessment and relative expression using Real Time PCR

Copy number of desired fragment was calculated based on absolute quantitation assay using SYBR Green dye and Sequence Detection System- 7500 (ABI, USA). The primers specific to fragments, respectively, were designed using Primer Express Software V2.0 (ABI). The standard curve was obtained using 10 folds dilution series of the recombinant plasmids ranging from 30, 00,000 to 30 copies taking 3.36 pg DNA per haploid genome of (assuming haploid genome of farm animals =3.3 pg, wt per base pair = 1.096 × 10^{-21} gm) as standards. The reactions were performed in triplicate using 96 well plates in a 25 µl reaction volume, each having 0.5 ng of buffalo genomic DNA and 50 nM of corresponding primers, employing conditions of 50^{0}C for 2 min, 95^{0}C for 10 min, followed by 40 cycles of 95^{0}C for 10 sec and 60^{0}C for 1 min. Real-time PCR analysis uncovered 1234 and 3420 copies of pDS5 and

pDS4 fragments per the haploid genome and ~2 × 10⁴ copies of pDp1, ~ 3000 copies of pDp2 and pDp3 and ~ 1000 of pDp4 in buffalo, cattle, goat and sheep genomes (Figure 4), respectively. (Pathak et al 2006; 2011) The copy number assessment of these repeats in different known and nondescript breeds of buffalo may enable to establish a correlation, if any, towards the delineation of different breeds.
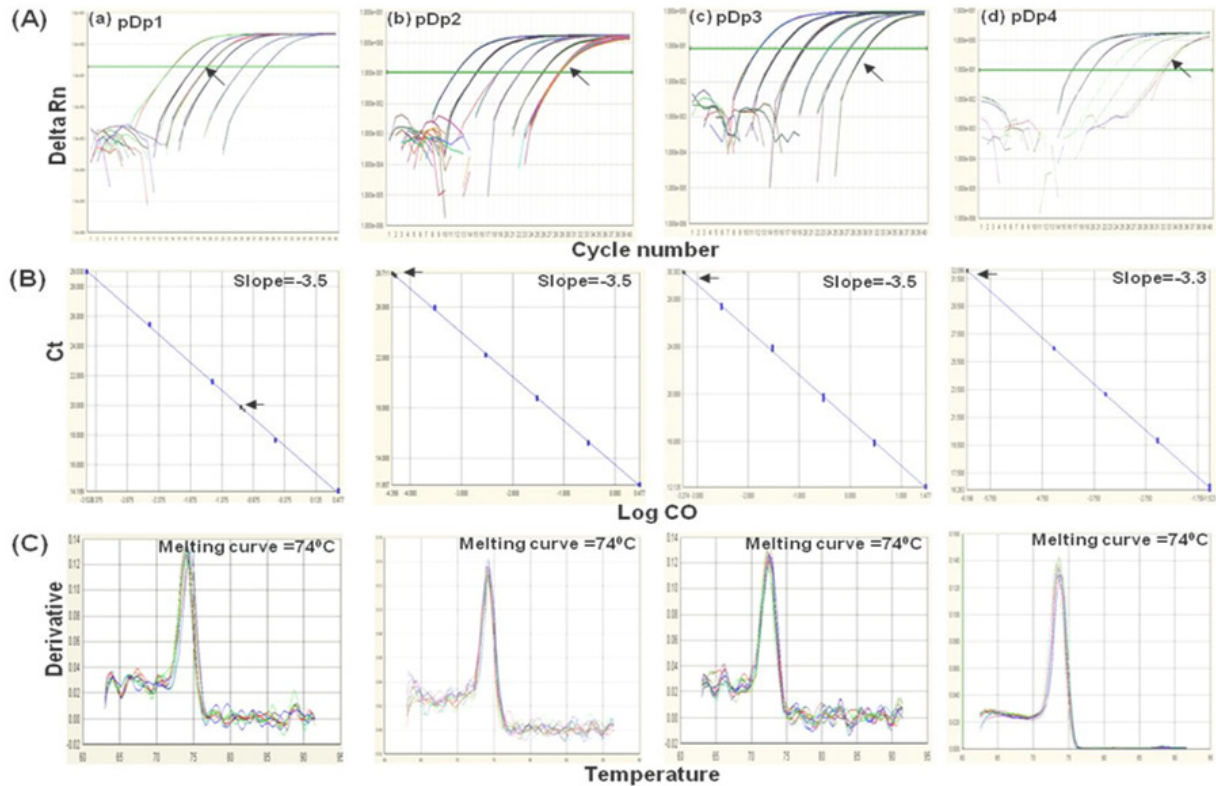
Figure 4. Standard curve based on 10 fold dilution series of pDp1, pDp2, pDp3, pDp4 and genomic DNA from buffalo, cattle, goat and sheep showing the amplification plot (a-d) panel (A), corresponding slopes of -3.3 to -3.5, panel (B) and a single dissociation peak, panel (C), substantiating maximum efficiency of the PCR reaction and high specificity of the primers with target DNA. Arrow indicates genomic DNA from buffalo, cattle, goat and sheep Buffalo, cattle, goat and sheep showed approximately same copy number for pDp1, pDp2, pDp3, pDp4 indicating their conservation across the bovid species (Pathak et al 2011).

Relative expression using Real Time PCR was carried out for the desired fragments with Sybr Green assay using cDNA from different tissues and spermatozoa. The primers were designed using Primer Express 2.0 (Applied Biosystems) software. The cyclic conditions were same as that used for copy number calculations. The reaction was performed following standard protocol (Sriavastava et al 2006). The specificity of each primer pair and the efficiency of the amplification were tested by assaying serial dilutions of the cDNA hybridized with oligonucleotides specific for target and normalization control (*GAPDH*). The difference in the Ct value between the target cDNA from different tissues and the control samples (the tissue showing least expression) was used for calculation. The expression level of the desired fragments was calculated using the formula: expression =

(1+E) $^{-\Delta Ct}$, where E is the efficiency of the PCR and ΔCt = difference in threshold cycle value between the test sample and endogenous control. To achieve the maximum (one) efficiency of the Real Time PCR, the amplicon size was kept small (70-150 bp) so that the expression level of the test gene remains 2 $^{-\Delta Ct}$. Each experiment was repeated three times to ensure consistency of the results. Maximum expression of pDS5 and pDS4 was seen in the spleen and liver, respectively. pDp1 showed maximum expression in lung, pDp2 and pDp3 both in Kidney, and pDp4 in ovary. Nine, 33.6 MASA amplified transcripts showed highest expression in spermatozoa and one each in liver and lung (Figure 5).
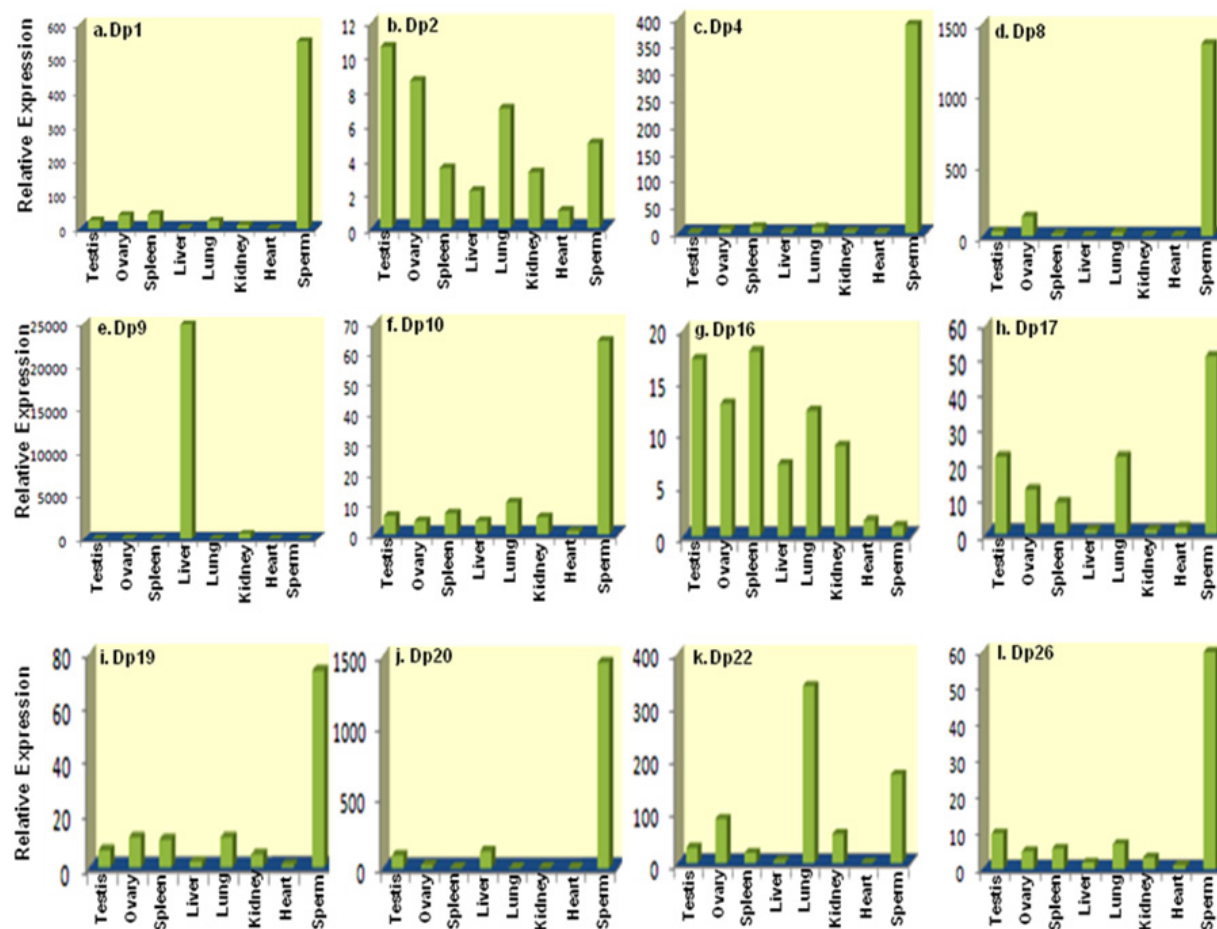
Figure 5. Expressional analyses of the representative 33.6 tagged mRNA transcripts. a-l represents different tissues, gonad, and spermatozoa. Note the maximum expression of some representative mRNA transcripts in the spermatozoa corresponding to Dp1, 4, 8, 10, 17,19, 20, and 26 shown in a, c, d, f, h, i, j, and l), respectively, and exclusive expression of Dp9 in liver (e). Bars represent relative expression of the transcript(s) in folds. Transcript IDs are mentioned on top left corner and tissues, below the panels (Pathak et al 2010)

These 9 transcripts in the spermatozoa, representing vital genes supports their involvement in sperm development and possibly overall testicular functions. In the context of animal biotechnology, such selective tissue specific expression profile is very important to segregate the genetically superior germplasm or any other physical and physiological attributes. This is true particularly in case of buffalo since this species has several breeds. Clearly,

development of MASA mediated tissue specific transcript profiles is envisaged to go a long way in undertaking molecular characterization of not only the genome of buffalo but also those of other economically important animals.

10.6 Chromosome preparation and Fluorescence *In Situ* Hybridization (FISH)

When once mRNA transcripts are made available, it becomes feasible to fish out full length cDNA clones employing 3' and 5'RACE. This approach enables isolation of genes without screening the cDNA library thus circumventing many arduous steps. When clones representing genes are obtained, these can be used for conducting fluorescence hybridization onto the metaphase chromosomes. We have used some of the clones to successfully conduct FISH on buffalo chromosomes to localize the genes and also uncover the distribution of several species of repeat elements.

For chromosome culture, 2 ml of blood was drawn into heparinized vacutainer tubes with sterile syringe from buffalo, cattle, goat, sheep and human. To sterile tissue culture flask containing 5 ml RPMI- 1640, 20% fetal Bovine serum, 2% Phytohemagglutinin, PHA (2 mg/ml), 5 μl concavalin A (3 μg/ml), 2.5 μl mercaptoethanol (50 μM), 50 μl LPS (10 μg/ml), 2.2 ml Antibiotic/antimycotic (0.15mg/ml), 500 μl of blood was added and whole mixture was incubated for 72 hours in 5% CO_2 at 37^0C. After 70 hours, colcemid (10 μg/μl) was added in culture flasks to arrest cells in metaphase stage and cells were further incubated for 2 hours. The cells were then subjected to 0.56%KCL for 30 mins followed by fixative treatment (Methanol: glacial Acetic acid, 3:1). Few drops of cell suspension were dropped onto pre-cleaned chilled slide and blow-dried. The slides were Giemsa stained (Gibco, BRL) for 20 minutes, washed with PBS / distilled water and observed under microscope to record metaphases.

For probe preparation plasmids containing gene of interest were labeled with desired fluorochromes using Nick Translation Kit from Vysis, (Illinois, USA) following supplier's instructions. Hybridization was carried out in 20 μl volume containing 50% formamide, 10% Dextran sulphate, Cot 1 DNA and 2X SSC, pH 7 for 16 hours at 37^0C in a moist chamber. Post hybridization washes were done in 2X SSC at 37^0C (low stringent condition) and then at 60^0C in 0.1X SSC (under high stringent condition). Slides were counterstained with DAPI, screened under Olympus Fluorescence Microscope (BX51) and images were captured with Olympus U-CMAD-2 CCD camera. Chromosome mapping was done following the International System for Chromosome Nomenclature. The pDS5, representing the 1378-bp fragment, showed FISH signals in the centromeric region of acrocentric chromosomes only, whereas pDS4, corresponding to 673 bp, detected signals in the centromeric regions of all the chromosomes. *Rsa*I derived pDp1, pDp2 and pDp3 showed distribution of repeats to all across the buffalo chromosomes (Pathak et al 2006; 2011) (Figure 6).

Chromosomal mapping of *SARS2* gene, using bovine *SARS2* BAC probe localized (Figure 7) the genes to buffalo metaphase chromosome 18 (Pathak et al 2010).
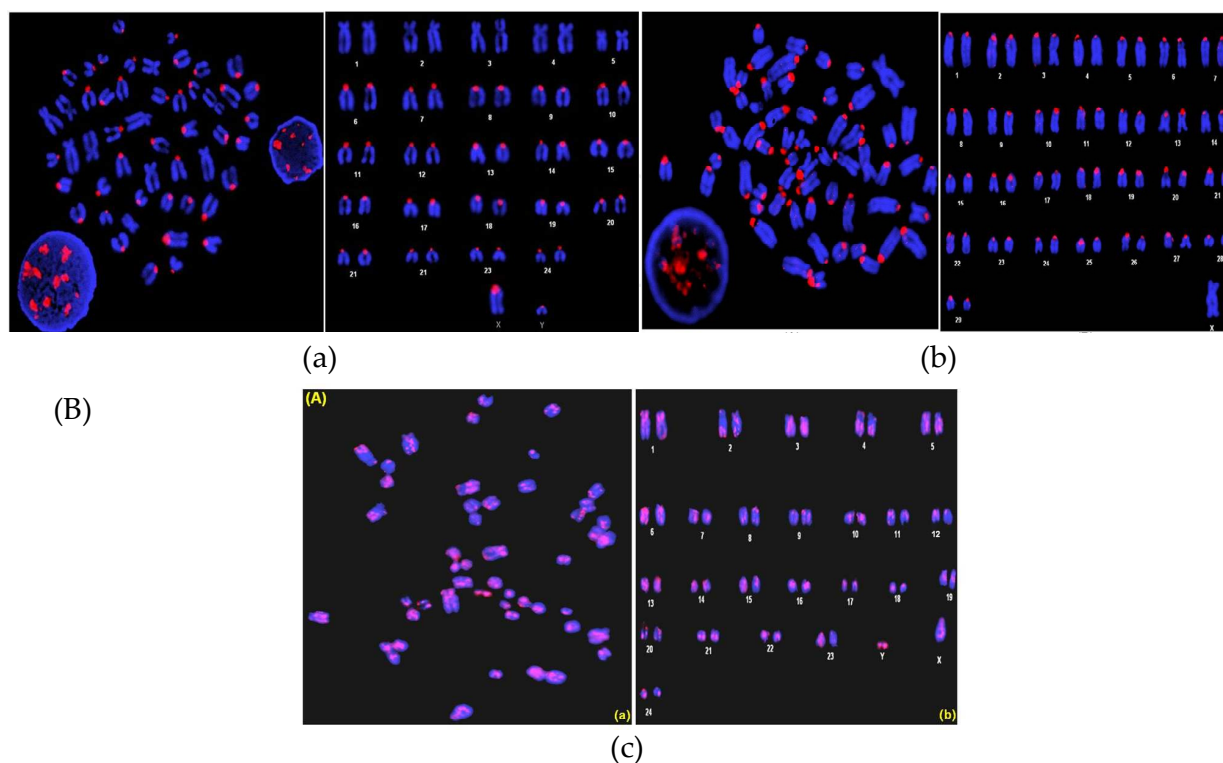
Figure 6. Chromosomal localization of pDS5 clone **(A)** on **(a)** buffalo and **(b)** cattle metaphase chromosomes. Note the absence of signal in all the bi-armed chromosomes. Fluorescence *in situ* hybridization (FISH) of pDp1 **(B)** clone on buffalo metaphase chromosomes **(c)**. Note the dispersed signals over the metaphase chromosomes. (See Pathak et al 2006; 2011).

11. Applications in animal biotechnology

With the availability of human genome sequence much emphases is given to sequence all the potential farm animals. Despite of importance as farm animal research data on the water buffalo is limited. Water buffalo breeders and farmers have been facing many challenges and problems, such as poor reproductive efficiency, sub-optimal production potential, higher than normal incidence of infertility, and lower rates of calf survival. Genome research has created a broad basis for promoting and utilizing gene technologies in many fields of livestock production. Genome biotechnology will provide a major opportunity to advance sustainable animal production systems of higher productivity through manipulating the variation within and between breeds to realize more rapid and better-targeted gains in breeding value. This type of research will also make it possible to distinguish molecular phenotypes and thus improve the use of genetic resources of domestic animals.

To date, researchers have identified several genes or DNA regions that are associated with traits of economic importance including reproduction, growth, lean body, fat quantity, meat

quality, physical traits and disease resistance. For example, the fatty acid composition of dairy and beef food products, increased disease resistance (and thus increased animal welfare). Similarly, decreased methane emissions in cattle help to address the needs of consumers and society for sustainable and cost-effective food production. A number of gene and marker tests are now available commercially from genotyping service companies. Examples are CAST (meat quality), ESR and EPOR (litter size), FUT1 (E. coli disease resistance), HAL (halothane – meat quality, stress), IGF2 (carcase), MC4R (growth and fat), PRKAG3 (meat quality) and RN (meat quality) (Walters, 2011). Many scientists are using genomic information embedded on SNP50 BeadChip, a glass slide containing thousands of DNA markers, to determine disease-resistant genes in cattle, swine, sheep, poultry, fish and then selectively mating the animals in order to create disease resistant animals. Understanding all the expressed genes, their organization and mode of action in bubaline or any other farm animals will positively bridge the gap and facilitate the much needed growth of animal biotechnology.

(a)   (b)   (c)   (d)   (e)
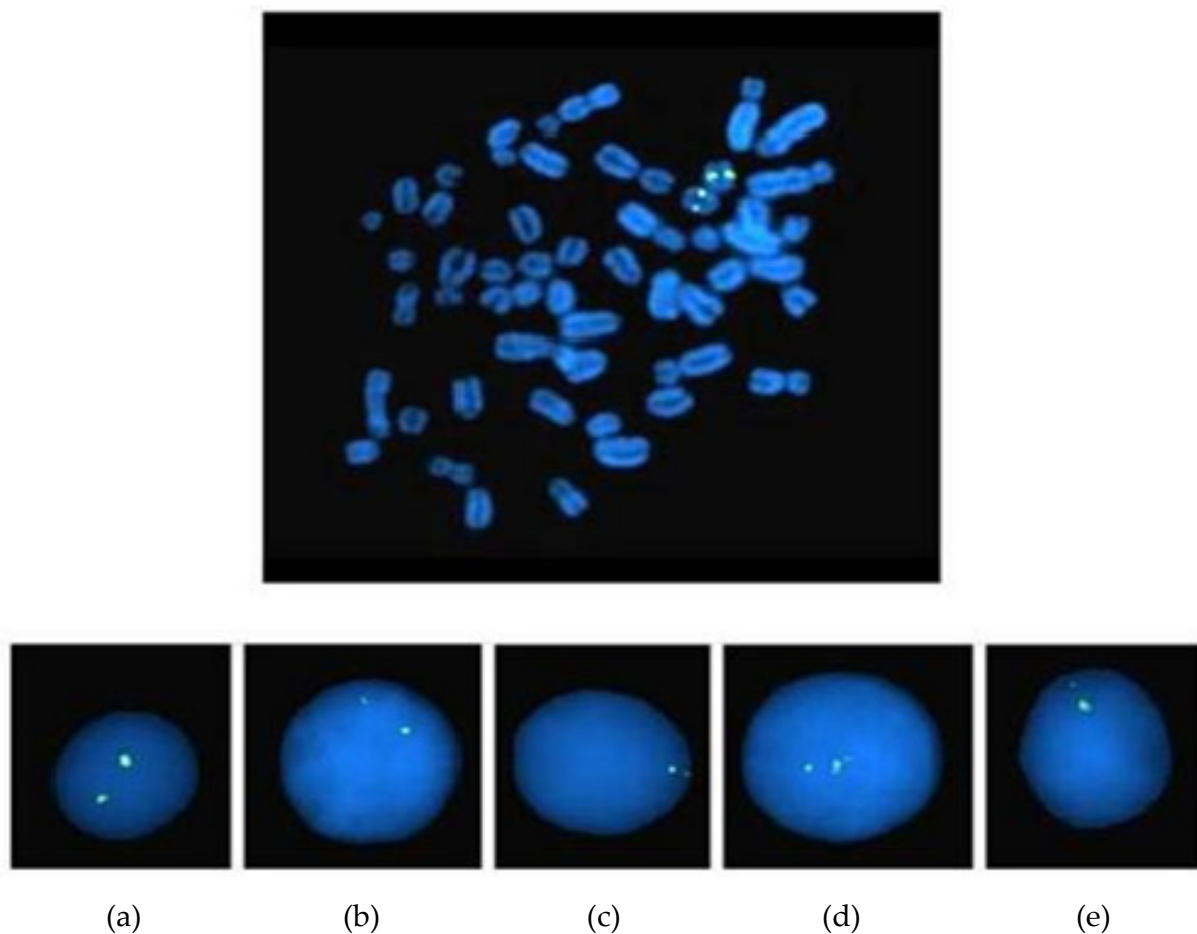
Figure 7. Localization of SARS2 gene on the representative interphase nuclei and metaphase chromosome 18 using FISH.SARS2 BAC probe showing signals on buffalo metaphase chromosome 18 (A) and interphase nuclei (a-e) (B). Note two signals in the interphase nuclei corresponding to those on the homologous chromosomes. (Pathak et al 2010).

12. Concluding remarks

Genetic improvement of animals warrants continuous and complex processes of sustained research employing cutting edge tools and techniques of modern biology and recombinant DNA technology. A much deeper and detailed understanding on a given species would eventually prove to be highly useful for possible manipulation of a desired genome. Improvement of domestic animal traits has been the foremost important task for animal breeding. In this pursuit, many techniques have been developed and tested. In recent years, advances in molecular genetics have introduced a new generation of molecular markers for the genetic improvement of the animals. However, utilization of marker-based information for genetic improvement depends on the choice and judicious use of an appropriate marker system for a given application. Selection of markers for different applications is influenced by the degree of polymorphism, reproducibility of the technique, speed of the experiments and cost involved.

As the situation stand now, for a given biological phenomenon where multiple genes are implicated, technical approaches need to be developed to segregate the entire possible genes specific to that phenomenon. A good example is the spermatogenesis that involves putatively close to about 400 plus genes. However, their clear cut involvement and characterization in any species has still not been achieved. When once, such information is made available, this would then provide much needed basis of functional and comparative genomics. Perhaps then, molecular delineation of the "so-called" elite animals or specific breed representing superior germplasm would become feasible.

Author details

Deepali Pathak and Sher Ali
National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi, India

13. References

[1] Ali, S., & Gangadharan, S. (2000). Differential Evolution Of Coding And Non-Coding Sequences In Related Vertebrates: Implications In Probe Design. *Proc. Ind. Nat. Sci. Acad.* 664, 9-67.

[2] Ali, S., & Wallace, R. B. (1988). Intrinsic Polymorphism Of Variable Number Tandem Repeat Loci In The Human Genome. *Nucleic Acids Res.* 16, 8487-8496.

[3] Ali, S., Müller, C.R., & Epplen, J.T. (1986). DNA finger printing by oligonucleotide probes specific for simple repeats. *Hum Genet.* 74(3):239-243.

[4] Amor, D.J., & Choo, K. H. (2002). Neocentromeres: Role In Human Disease, Evolution, And Centromere Study. *Am. J. Hum. Genet.* 71, 695–714.

[5] Amos, W., & Hoelzel, A.R. (1991). Long-Term Preservation Of Whale Skin For DNA Analysis. *Rep. Int. Whal. Commn.* 13, 99–103.

[6] Barros, P., Blanco, M.G., Boán, F., & Gómez-Márquez, J. (2008). Evolution of a complex minisatellite DNA sequence. *Mol Phylogenet Evol.* 49(2):488-494.

[7] Bashamboo, A., & Ali, S. (2001). Minisatellite Associated Sequence Amplification (Masa) Of The Hypervariable Repeat Marker 33.15 Reveals A Male Specific Band In Humans. *Mol. Cell. Probes* 15, 89-92.

[8] Battaglia, E. (1999). The Chromosome Satellite (Navashin's "Sputnik" Or Satelles): A-345-Terminological Comment. *Acta Biologica Cracoviensia, Series Botanica* 41, 15-18.

[9] Bhatnagar, S., Bashamboo, A., Chattopadhyay, M., Gangadharan, S., & Ali, S. (2004). A 1.3 kb satellite DNA from *Bubalus bubalis* not conserved evolutionarily is transcribed. *Z Naturforsch C* 59(11-12):874-879.

[10] Bishop, R., Morzaria, S., & Gobright, E. (1998). Linkage Of Two Distinct At-Rich Minisatellites At Multiple Loci In The Genome Of *Theileria Parva*. *Gene* 216, 245-254

[11] Blott, S.C., Williams, J.L., & Haley, C.S. (1999). Discriminating among cattle breeds using genetic markers. *Heredity (Edinb)* Pt 6:613-619.

[12] Boeva,V., Regnier, M., Papatsenko, D., Makeev, V. (2006). Short fuzzy tandem repeats in genomic sequences, identification, and possible role in regulation of gene expression. *Bioinformatics* 2006 Mar 15; 22(6):676-684.

[13] Bromham, L. (2002).The Human Zoo: Endogenous Retroviruses in the Human Genome. *Trends Ecol. Evol.* 17, 91–97.

[14] Brown, T. A. (2002). The Repetitive DNA Content Of Geno*mes. Genomes* 59-64.

[15] Capy P. (1998). Evolutionary biology. A plastic genome. Nature 396(6711):522-3.

[16] Catasti, P., Chen, X., Mariappan, S.V., Bradbury, E.M. & Gupta, G. (1999) DNA Repeats In The Human Genome. Genetica 106, 15-36.

[17] Cavalier-Smith, T. 1985. Eukaryotic Gene Numbers, Non-Coding DNA, And Genome Size. *In* The Evolution Of Genome Size. *Edited By* T. Cavalier-Smith. *John Wiley And Sons, Chichester, U.K.* Pp. 69–103.

[18] Charlesworth, B., Sniegowski, P., & Stephan, W. (1994). The Evolutionary Dynamics Of Repetitive DNA In Eukaryotes. *Nature* 371, 215-220.

[19] Chattopadhyay, M., Prashant, S. G., Kapur, V., Azfer, Md. A., Prakash, B., & Ali, S. (2001). Satellite Tagged Transcribing Sequences In The Bubaline *Bubalus bubalis* Genome Undergo Programmed Modulation In The Meiocytes: Possible Implication In Transcriptional Inactivation. *DNA Cell Biol.* 20, 587-593.

[20] Debrauwere, H., Gendrel, G.C., Lechat, S., & Dutreix, M. (1997). Differences and similarities between various tandem repeat sequences: Minisatellites and microsatellites. *Biochimie.* 79:577–586.

[21] Dey, I., & Rath, P.C. (2005). A Novel Rat Genomic Simple Repeat DNA With RNA-Homology Shows Triplex (H-DNA)-Like Structure And Tissue-Specific RNA Expression. *Biochem. Biophys. Res. Commun.* 276, 286-228.

[22] Durward, E., Shiu, O.Y., Luczak, B., & Mitchelson, K. R. (1995). Identification Of Clones Carrying Minisatellite-Like Loci In An Arabidopsis Thaliana Yac Library. *Journal Of Experimental Botany* 46, 271-274.

[23] Elgar, G., & Vavouri, T. (2008). Tuning in to the signals: noncoding sequence conservation in vertebrate genomes. *Trends Genet.* **24** (7): 344–52.

[24] Epplen, J.T., Kyas, A. & Maueler, W. (1996) Genomic Simple Repetitive DNAs Are Targets For Differential Binding Of Nuclear Proteins. *Febs Letters* 389, 92-95.

[25] Feng, Y., Yang, W., Ryan, U., Zhang, L., Kvác, M., Koudela, B., Modry, D., Li, N., Fayer, R., & Xiao, L. (2011). Development of a multilocus sequence tool for typing Cryptosporidium muris and Cryptosporidium andersoni. *J Clin Microbiol.* 49(1):34-41.

[26] Fondon, J.W., & Garner, H.R. (2004) Molecular Origins Of Rapid And Continuous Morphological Evolution. *Proc. Natl. Acad. Sci. USA* 101, 18058–18063.

[27] Furano, A. V. (2000). The Biological Properties And Evolutionary Dynamics Of Mammalian Line-1 Retrotransposons Prog. *Nucleic Acid Res.* 64,255-294.

[28] Georges, M., & Andersson, L. (1996). Livestock Genomics Comes Of Age. Genome Res. 6,907-921.

[29] Georges, M., Gunawardana, A., Threadgill, D.W., & Lathrop, M. (1991). Characterization Of A Set Of Variable Number Of Tandem Repeat Markers Conserved In Bovidea. *Genomics* 11, 24–32.

[30] Gur-Arie, R., Cohen, C.J., Eitan, Y., Shelef, L., Hallerman, E.M. & Kashi, Y. (2000) Simple Sequence Repeats In Escherichia Coli: Abundance, Distribution, Composition, And Polymorphism. *Genome Res.* 10, 62–71.

[31] Haber, J.E., & Louis, E.J. (1998). Minisatellite Origins In Yeast And Humans. *Genomics* 48, 132-135.

[32] Han, J.S., Boeke, J.D. (2005). LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays* **27:**775-784.

[33] Han, J.S., Szak, S.T., & Boeke, J.D. (2004). Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* 429:268-274.

[34] Harris, A.S., & Wright, J.M. (1995). Nucleotide sequence and genomic organization of cichlid fish minisatellites. *Genome* 38(1):177-184.

[35] Henikoff, S., Ahmad, K., & Malik, H. S. (2001). The Centromere Paradox: Stable Inheritance With Rapidly Evolving DNA. *Science* 293, 1098–1102.

[36] Hochgeschwender, U., & Brennan, M.B. (1991) Identifying Genes Within The Genome: New Ways For Finding The Needle In A Haystack. *Bioessays* 13, 139-144.

[37] International Human Genome Sequencing Consortium. (2004) Finishing the euchromatic sequence of the human genome. Nature 431(7011):931-945.

[38] Jarman, A.P., & Wells, R.A. (1989). Hypervariable Minisatellites: Recombinators Or Innocent Bystanders? *Trends Genet.* 5,367–371.

[39] Jeffreys, A. J., Neil, D. L., & Neumann, R. (1998). Repeat Instability At Human Minisatellites Arising From Meiotic Recombination. *Embo J.* 17, 4147–4157.

[40] Jeffreys, A.J., & Wilson, V., & Thein, S.L. (1985). Hyper Variable 'Minisatellite' Regions In Human DNA. *Nature* 314, 67-73.

[41] Jobling, M.A., & Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet.* 4(8):598-612.

[42] Jordan, I. K., Rogozin, I. B., Glazko, G.V. & Koonin, E.V.(2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends in Genetics 19, 68–72.

[43] Kapur, K., Prasanth, S. G., O'ryan, C., Azfer, Md, A., & Ali, S. (2003). Development Of A DNA Marker By Minisatellite Associated Sequence Amplification (Masa) From The Endangered Indian Rhino (*Rhinoceros Unicornis*). *Mol. Cell. Probes* 17, 1-4.

[44] Kelly, M.K., Alver, B., & Kirkpatrick, D.T. (2011). Minisatellite alterations in ZRT1 mutants occur via RAD52-dependent and RAD52-independent mechanisms in quiescent stationary phase yeast cells. DNA Repair (Amst). 10(6):556-566.

[45] Kim, M. & Mullet, J.E. (1995) Identification Of A Sequence-Specific DNA Binding Factor Required For Transcription Of The Barley Chloroplast Blue Light-Responsive Psbd-Psbc Promoter. *Plant Cell* 7, 1445-1457.

[46] Kit, S. (1961). Equilibrium Sedimentation In Density Gradients Of DNA Preparations From Animal Tissues. *J. Mol. Biol.* 3,711-716.

[47] Kumar, S., Gupta, R., Kumar, S., & Ali, S. (2011). Molecular mining of alleles in water buffalo *Bubalus bubalis* and characterization of the TSPY1 and COL6A1 genes. *PLoS One* 6(9):e24958.

[48] Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., et al. (2001). Initial Sequencing and Analysis of the Human Genome. *Nature* 409, 860–921.

[49] Legendre, M., Pochet, N., Pak, T., & , K.J. (2007). Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome Res.* 17(12):1787-1796.

[50] Lopes, J., Ribeyre, C., & Nicolas, A. (2006). Complex minisatellite rearrangements generated in the total or partial absence of Rad27/hFEN1 activity occur in a single generation and are Rad51 and Rad52 dependent. *Mol Cell Biol.* 26(17):6675-6689.

[51] Lue, N. F., Buchman, A. R., & Kornberg, R. D. (1989) Activation Of Yeast RNA Polymerase Ii Transcription By A Thymidine-Rich Upstream Element In Vitro. *Proc. Natl. Acad. Sci. USA* 86, 486-490.

[52] Martienssen, R. A., & Colot, V. (2001). DNA Methylation And Epigenetic Inheritance In Plants And Filamentous Fungi. *Science* 293, 1070-1074.

[53] Miller, W. J., & Capy, P., eds. (2004), Mobile genetic elements: protocols and genomic applications, *Humana Press*, 289.

[54] Mudgal, V.O. (1988). Comparative Efficiency For Milk Production Of Buffaloes And Cattle In The Tropics. *Proceedings Of Ii World Buffalo Congress.* New Delhi, India, Vol Ii, Part Ii, 454–462.

[55] Nakamura, Y., Leppert, M., O'connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. et al. (1987).Variable Number Of Tandem Repeat (Vntr) Markers For Human Gene Mapping. *Science* 235, 1616–1622.

[56] Norman, A., D. (2001). APPENDIX 1B Overview of Human Repetitive DNA Sequences. Current Protocols in Human Genetics.

[57] Ohno, S. (1972). So Much `Junk' In Our Genomes. *Brookhaven Symp. Biol.* 23, 366-370.

[58] 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.

[59] Pathak, D., Srivastava, J., Premi, S., Tiwari, M., Garg, L.C., Kumar, S., & Ali, S. (2006). Chromosomal localization, copy number assessment, and transcriptional status of *BamH*I repeat fractions in water buffalo Bubalus bubalis. *DNA Cell Biol.* 25(4):206-214.

[60] Pathak, D., Srivastava, J., Samad, R., Parwez, I., Kumar, S., & Ali, 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. *Chromosome Res.* 18(4):441-458.

[61] Patience, C., Takeuchi, Y., & Weiss, R. A. (1997). Infection Of Human Cells By An Endogenous Retrovirus Of Pigs. *Nat. Med.* 3, 276-282.

[62] Proudfoot, N.J., Gill, A., & Maniatis, T. (1982). The Structure Of The Human Zeta-Globin Gene And A Closely Linked, Nearly Identical Pseudogene. *Cell* 31, 553-563.

[63] Rawal, L., Ali, S., & Ali, S. (2012). Molecular mining of GGAA tagged transcripts and their expression in water buffalo Bubalus bubalis. *Gene* 492(1):290-295.

[64] Reed, J.M., Fleischer, R. C., Eberhard, J., & Oring, L.W. (1996). Minisatellite DNA Variability In Two Populations Of Spotted Sandpipers Actitis Macularia In Minnesota, U.S.A. Wader Study Group Bull. 79, 115-117.

[65] Richard, C. & Mark, B. (2009). The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* 10 (10): 691–703.

[66] Roy-Engel A, M., Carroll, M.L., Vogel, E., et al. (2001). Alu insertion polymorphisms for the study of human genomic diversity. *Genetics* 159 (1): 279–290.

[67] Royle, J.R., Clarkson, R.E., Wong, Z., & Jeffreys, A.J. (1988). Clustering Of Hypervariable Minisatellites In The Proterminal Regions Of Human Autosomes. *Genomics* 3, 352–360.

[68] Sethi, R. K. (2003). Improving Rivering And Swamp Buffaloes Through Breeding. 4th Asian Buffalo Congress Lead Papers, 50.

[69] Shapiro, J.A., & Von Sternberg, R.(2005). Why repetitive DNA is essential to genome function. *Biol Rev Camb Philos Soc.*80(2):227-50

[70] Sinden, R.R. (1999). Biological Implications Of The DNA Structures Associated With Disease-Causing Triplet Repeats. *Am. J. Hum. Genet.* 64, 346–353.

[71] Skuce, R. A., Mccorry, T. P., Mccarroll, J. F., Roring, S.M.M., Scott, A.N., Brittain, D., Hughes, S.L., Hewinson, R.G., Sydney, D., & Neil, L. (2002). Discrimination Of Mycobacterium Tuberculosis Complex Bacteria Using Novel Vntr-Pcr Targets. *Microbiology* 148, 519-528.

[72] Slamovits, C.H., & Rossi, M.S. (2002). Satellite DNA: Agent Of Chromosomal Evolution In Mammals. Mastozoología Neotropical 9, 297-308.

[73] Srivastava, J., Premi, S., Kumar, S., & Ali, S. (2008). Organization and differential expression of the GACA/GATA tagged somatic and spermatozoal transcriptomes in Buffalo Bubalus bubalis. *BMC Genomics* 9:132.

[74] Srivastava, J., Premi, S., Pathak, D., Ahsan, Z., Tiwari, M., Garg, L.C., & Ali, S. (2006). Transcriptional status of known and novel genes tagged with consensus of 33.15 repeat loci employing minisatellite-associated sequence amplification (MASA) and real-time PCR in water buffalo, Bubalus bubalis. *DNA Cell Biol.* 25 (1):31-48.

[75] Stoker, N.G., Cheah, K.S.E., Griffin, J.R., Pope, F.M., & Solomon, E. (1985). A Highly Polymorphic Region 3′ To The Human Type Ii Collagen Gene. *Nucleic Acids Res.* 13, 4613–4622.

[76] Subramanian, V.M., Madgula, R., George, R.K., Mishra, M.W., Pandit, C.S. Kumar, & L. Singh. (2003). *Bioinformatics* 19, 549–552.

[77] Sykorová, E., Fajkus, J., Mezníková, M., Lim, K.Y., Neplechová, K., Blattner, F.R., Chase, M.W., & Leitch, A.R. (2006). Minisatellite telomeres occur in the family Alliaceae but are lost in Allium. *Am J Bot.* 93(6):814-823.

[78] Tang, S.J. (2011). Chromatin Organization by Repetitive Elements (CORE): A Genomic Principle for the Higher-Order Structure of Chromosomes. *Genes* 2011, *2*(3), 502-515

[79] Tautz, D. (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* 17(16):6463-6471.

[80] Tautz, D. (1993). Notes On The Definition And Nomenclature Of Tandemly Repetitive DNA Sequences. In DNA Fingerprinting: State Of The Science. 21–28.

[81] Tautz, D., & Renz, M. (1984). Simple Sequences Are Ubiquitous Repetitive Components Of Eukaryotic Genomes. Nucleic Acids Res. 25, 4127-4138.

[82] Tomilin, N.V. (2008). Regulation Of Mammalian Gene Expression By Retroelements And Non-Coding Tandem Repeats. *Bioessays* 30, 338–348.

[83] Toth, G., Gaspari, Z. & Jurka, J. (2000). Microsatellites In Different Eukaryotic Genomes: Survey And Analysis. *Genome Res.* 10, 967–981.

[84] Tourmente, S., Deragon, J.M., Lafleuriel, J., Tutois, S., Pelissier, T., Cuvillier, C., Espagnol, M.C., & Picard, G. (1994). Characterization Of Minisatellites In Arabidopsis Thaliana With Sequence Similarity To The Human Minisatellite Core Sequence. Nucleic Acids Research 22, 3317–3321.

[85] Ugarkovic, D., & Plohl, M. (2002). Variation In Satellite DNA Profiles-Causes And Effects. *EMBO* 21, 5955-5959.

[86] Vermaak, D., Bayes, J.J., & Malik, H.S. (2009). A surrogate approach to study the evolution of noncoding DNA elements that organize eukaryotic genomes. *J Hered.* 100(5):624-636.

[87] Vinces, M.D., Legendre, M., Caldara, M., Hagihara, M., & , K.J. (2009). Unstable tandem repeats in promoters confer transcriptional evolvability. *Science* 324(5931):1213-1216.

[88] Walters, R. (2011). More commercial benefits on horizon as pig genome project nears completion. Information from pig genome already being used in the industry. *Pig International* , 41, 2:16

[89] Wilson, G.A., & Strobeck, C. (1999). The isolation and characterization of microsatellite loci in bison, and their usefulness in other artiodactyls. *Anim Genet.* 30(3):226-227.

[90] Zhang, L., Yuan, D., Yu, S., Li, Z., Cao, Y., Miao, Z., Qian, H., & Tang, K. (2004). Preference of simple sequence repeats in coding and non-coding regions of Arabidopsis thaliana. *Bioinformatics* 20(7):1081-1086.