Proc. Indian natn Sci Acad.(PINSA) **B66** No.1 pp 49-68 (2000)

Differential Evolution of Coding and Non-Coding Sequences in Related Vertebrates: Implications in Probe Design

SHER ALI and SUPRIYA GANGADHARAN
Molecular Genetics Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi - 110 067

(Received on 20 October 1999; Accepted after revision on 3 February 2000)

Considerable progress has been made in divulging the genome organization of different species by using DNA probe based on coding and non-coding sequences. This encompasses the individualization of the genome to assess the genetic structure of a population and reproductive behavior of animals and birds as well as phylogenetic relationship of different species at the supra-molecular level. Literature survey shows that these sequences originate and undergo evolutionary metamorphosis at the molecular level and are useful as powerful genetic markers in a wide range of species to address a variety of issues related to biology and medicine including forensic sciences. In this paper, we elaborate organizational and evolutionary aspects of some coding and non-coding sequences and their possible use as genetic markers amongst related vertebrate species. The technical procedures, prospects and possible pitfalls are also covered to facilitate the conceptual understanding of the fast emerging technology.

Key Words: DNA fingerprinting, Microsatellite evolution, Genetic distance, Molecular phylogeny

Introduction

The genetic complexity of a species is the molecular percolation of its mating behaviour, male-female ratio, population structure, natural selection, rate of sequence mutation, possible chromosomal rearrangement and several other physical and physiological attributes. Analyses of these aspects and in-depth characterization of a species may be undertaken by employing any of the technical approaches depending upon the aims and objectives. Earlier, genetical studies on animal species were largely based on chromosomal profiles and no attempts were made to uncover the organizational or evolutionary variations of the DNA sequences in the genome nor was there any attempt to develop DNA based marker. Protein(s) or biochemical markers detect reduced level of diversity whereas mtDNA markers have inherent disadvantages like high rate of somatic mutations, maternal inheritance and lack of recombination. Despite the fact that methylation of DNA affects the cleavage sites altering the level of detection of polymorphism, DNA based marker is still considered to be more informative. During earlier studies, attempts have been made to map the

Corresponding author: Tel: 91-11 616 2281 Ext. 453; Fax: 91-11 616 2125; 617 7626; E-mail: sherali@nii.res.in; sheralib5@hotmail.com

Abbreviations: RFLP, Restriction fragment length polymorphism; *DNAF*, DNA finger rinting; *RAPD*, Random amplification of polymorphic DNA; *MASA*, Microsatellite associated sequence amplification; *ALV*, Allele length variation.

bubaline genome using marker based on the homologous loci in cattle (Iannuzzi et al. 1996). However, with respect to organization and evolution of some repeat loci, buffalo may be closer to other species than cattle though morphologically they may appear to be similar. Thus, morphological similarities corroborated by genetic affinities provide a ground for the use of heterologous markers within the related species. In this context, an understanding on the organization and evolution of repeat sequences is envisaged to be of interest.

Repeat Sequence and Their Classification

A significant percentage (about 30-40%) of all mammalian genomes consist of repeated DNA sequences (Britten & Kohne 1968) whose origin, evolution and functions are still being actively pursued to understand their overall biological significance. According to their various characteristic properties like structure, distribution and reiteration frequencies, these sequences are generally classified as tandemly arrayed ones and those that are interspersed in the genome (Jelinek & Schmid 1982). Based on their size, tandemly arrayed repeat units may be further divided into minisatellites and microsatellites. Minisatellites are composed of repeat units that are greater than 7 base pairs while microsatellites have a basic repeat unit of upto 7 bp or less (Bruford & Wayne 1993, Wright 1994). It may be noted that these classification are empirical only to accommodate different types of repeat motifs ranging from 2 to 33 bases or even longer. Repeated DNA sequences may also be divided into short interspersed elements (SINEs) and long interspersed elements (LINEs) (Weiner et al. 1986).

Origin of Interspersed Repeats

It has been hypothesized that interspersed sequences originate from functional genes through a retroposon like mechanism. This requires reverse transcription of RNAs into DNAs, which are then integrated into the genome. The sequences then spread through the genome by several rounds of transcription by RNA polymerase III (to generate SINEs) and RNA polymerase II (to generate LINEs). These sequences then undergo *de novo* amplification to form several sub-families of SINEs and LINEs.

Of the several types of repetitive DNA, *Alu* is the most abundant one in eukaryotes. The first *Alu* like family that was identified as descended from tRNA is the C family of artiodactyles in cow and goat (Lawrence et al. 1985, Rogers 1985). This family showed conserved sequence with 65% homology with several tRNA and could be folded into a secondary structure. Later, it was found that each of these *Alu* like families has evolved independently from tRNA (Rogers 1985). SINEs and LINEs occur very abundantly in the animal genomes but interestingly, a given subfamily is usually present only in a moderate number of related species (Deininger & Daniels 1986). This is due to the independent formation of these repeat families after the divergence of various mammalian orders.

Origin of Minisatellites

Unlike SINEs and LINEs, the tandemly arrayed repeat sequences originate due to "random unequal crossover" between duplicated sequences (Smith 1978), slipped strand mis-pairing (Tautz & Renz 1984, Levinson et al. 1985) and aberrant *in-situ* replication (Schimke 1984). Amongst these mechanisms, it is suggested that minisatellite repeat loci evolve mainly through unequal crossover (Jeffreys et al. 1985a, Levinson & Gutman 1987). The random unequal crossover is envisaged to initiate as a result of rare, illegitimate recombination between homologous chromosomes (sister chromatids) during meiosis. Such events may occur by chance with reasonable frequency in regions of non-repetitive DNA. Misalignment and unequal crossover generate one chromatid with a tandemly duplicated segment and another one with a deletion for the same region, giving rise to a mutant allele. This results in substantial differences in the length as well as nucleotide sequences in the homologous chromosomes (Botstein et al. 1980). Following this, additional unequal crossovers occur more readily involving homologous recombination between related sequences of the tandem arrays (Hardman 1986). These unequal crossovers, then

result in random increase (addition) or decrease (deletion) of particular variant repeats. Some variants are lost while the others increase in frequency, eventually replacing all others. This causes genetic drift of alleles at repeat loci leading to "crossover fixation" in the population (Jarman & Wells 1989).

Origin of Microsatellites

Microsatellites consist of stretches of monotonously repeated short nucleotide motifs distributed in interspersed manner. Almost all permutations of mono-, di-, tri- and tetranucleotide motifs can be found as building blocks of these simple repeat sequences. The molecular events, which play a role in the evolution of simple, repeat sequences are gene conversion (Maeda & Smithies 1986) and replication slippage (Levinson & Gutman 1987). Among this, the most common one in the evolution of microsatellites (short tandem repeats) is the strand slippage at dinucleotide repeat loci. In one of the experiments, Levinson and Gutman (1987) introduced simple sequence stretches of GT/CA motifs into the multiple cloning site of M13 phage and observed that slippage can occur *in vivo* between short direct repeats, leading to deletions/expansions of the motifs. This observation was found to be in accordance with earlier studies (Jones & Kafatos 1982, Brown & Piechaczyk 1983).

Possible Biological Functions of the Repeat sequences

The precise role and biological significance of repeat sequences in the eukaryotic genome is still unclear. However, repeat sequences have been implicated in gene regulation (Wang et al. 1979, Hentschel 1982), replication of telomeres (Blackburn & Szostak 1984), selection and transport of mRNA to the cell cytoplasm (Davidson & Britten 1979, Sutcliffe et al. 1984) and as signals for gene conversion and recombination (Jeffreys et al. 1985a). Further, alternating purine and pyrimidine such as (GT)n or (CA)n have been suggested to form left handed Z-DNA (Zimmerman 1982) and tandem repeats are suspected to adopt unusual structures to play functionally important roles (Vogt 1990). Length expansions of trinucleotide repeats have been found to be associated with a number of human genetic diseases (Caskey et al. 1992, Richards & Sutherland 1992, Bates & Lehrach 1994) whereas extreme shrinkage of the simple repeats to the extent of total absence has been reported in tilapia from a geographically isolated population (John et al. 1996). Employing Northern blot analysis, it has been shown that some repetitive sequences are transcribed both in somatic and germline cells (Epplen et al. 1982, Singh et al. 1984, Schäfer et al. 1986). In recent study, bubaline derived satellite DNA transcribing in somatic and germline tissues with varying signal intensity has been found to undergo programmed sequence modulation in semen samples compared to that from blood or other somatic cells (submitted).

Distribution and Evolution of Repeat Requences in the Population

Mechanisms of continuous gain or loss of DNA repeats by unequal crossing over and gene conversion can lead to homogenization (Molecular drive) of any given variant in a sexual population (Cohen et al. 1982, Dover 1986). During the course of evolution of minisatellites by unequal crossing over, some variants may be lost whereas others will persist and continue to increase. Such events will purge all the point mutations and ensure homogeneity of the repeat units. When these cross over fixations accumulate in other tandem repeats and spread through the population, this leads to inter-species variation owing to "concerted evolution" of minisatellite sequences (Dover 1982). These evolutionary changes cause sequence homogeneity in the repeats of an array within a species but heterogeneity in the units of the corresponding array in different species giving rise to inter-specific variations. Careful analysis of such variation enables development of synthetic DNA based genetic markers useful for delineating interspecies genetic variation. Dot blot hybridization of undigested genomic DNA samples with different oligo probes representing consensus sequence of a repeat loci is the fastest and most reliable way to assess the presence or absence

(a) (b)

Figure 1 Dot blot hybridization of heat denatured total genomic DNA (2 μg) from several vertebrate species with oligo probe OAT 24.1 containing six units of GATA repeat motif. The numbers represent spots of DNA samples: 1 and 2- male and 3 and 4- female buffaloes; 5- male and 6- female cattle; 7- male and 8- female goat; 9 and 10 both sheep; 11- male and 12- female horse; 13- male and 14- female pig; 15- male and 16- female rabbit; 17 and 18 both female birds; 19- male and 20- female rat, respectively. Note the difference in the intensity of signals and absence of the same in several samples (a). β-actin cloned probe was used as control to assess the quantity of DNA in the figure 1a (not shown). Dot blot hybridization (same as above) with OAT24.3 oligo probe containing four units of telomeric TTAGGG repeat motif. These sequences are integral part of the end of all the chromosomes in eukaryotes signifying their true evolutionary conservation. Note the distinct signals in all the samples (b).

of a repeat in a given species (figure 1a & b). It is clear from the figure 1a that OAT 24.1 probe $(GATA)_6$ is absent in several vertebrate genomes whereas the evolutionarily conserved telomeric repeat is present in all the species studied *albeit* with copy number variation. It may be noted that no large-scale sequence homogenization occurs by gain or loss due to strand slippage unlike the mechanism involved in microsatellite repeat evolution (Tautz et al. 1986). This is due to the fact that the unit of slippage can occasionally be out of phase with pre-existing motifs. Sequence homogenization is continuously modulated by the *de-novo* generation and recombinatorial reshuffling of long and short-lived motifs differing in length and sequence complexities. The intra-array homogeneities at the minisatellite loci range from highly divergent repetitive sequences to arrays of almost identical repeats as their "core sequence" (Jeffreys et al. 1985a). The rate of mutation in the homogeneous sequences is high, leading to a high level of polymorphism at these loci. Similarly, with the increase in length of the repeat units, the degree of variability increases which paves the way for an increased chance of misalignment and unequal crossovers. Thus, within a population, allele length variation increases with an increase in the rate of unequal crossover (Jarman & Wells 1989).

Hypervariability at Repeat Loci and DNA Probe

There is ample evidence to suggest that hypervariable minisatellite sequences are recombinogenic in mammalian cells (Steinmetz et al. 1986, Meuth et al. 1987). By statistical analysis it was predicted that the rate of unequal exchange within the sequences would have to be about 10 times higher than that of the genomic average to maintain the degree of polymorphism. Such frequent unequal crossovers occurring at the minisatellite regions leads to generation of variability in the number of repeat in the tandem arrays referred to as variable number tandem repeat (VNTR) (Nakamura et al. 1987). These are also called hypervariable regions (HVRs) and were first isolated by chance since they were located near the insulin gene (Bell et al. 1981). Later, minisatellite probe from the myoglobin gene was found to cross-hybridize with many other minisatellites (because of similarities in their core sequences), giving rise to a complex but heritable band pattern dubbed as "DNA fingerprint" (Jeffreys et al. 1985a,b). The cumulative high degree of allelic variability so detected at minisatellite regions made the DNA fingerprinting technology highly useful for identifying individuals or clones of cells having

variety of other applications in forensics, immigration laws, paternity testing, transplant screening and ecological genetics (Ali & Epplen 1991). The hypervariable minisatellites from myoglobin gene are not confined to the human genome and band pattern of similar complexity have been reported in mammals, birds, higher plants, fungi and protozoa using minisatellite probes. Nonetheless, a single minisatellite probe can not be used as universal one owing to undefined sequence complexities with varying rates of mutations amongst highly unrelated genomes. Thus, to cover the larger pool of genetic variability, characterization of individual hypervariable repeat loci and their individual assessment will prove to be informative.

Conventional Restriction Fragment Length Polymorphism (RFLP)

The gain or loss of specific restriction endonuclease cleavage sites in an individual's genome helps in detecting these variants in the form of restriction fragments length polymorphism (Botstein et al. 1980). Human genome contains roughly $3x10^7$ DNA sequence variants (Jeffreys 1979). However, all such variations are not detectable by conventional RFLPs and only about 1% of the vast pool of genetic variability is amicable to this approach (Jeffreys 1979). This is because only a small fraction of the nucleotide sequences are known to have altered restriction sites. Thus, the small differences in fragment length can be overlooked since the probe used may not cover the sequence variants present in the repetitive DNA. Usually most of the RFLPs are diallelic and have limited polymorphic information content (PIC). Owing to limited degree of heterozygosity of diallelic RFLPs, they are not always informative for genome individualization or linkage analysis. Therefore, multilocus RFLPs are considered to be more informative.

Oligonucleotides Probes for DNA Fingerprinting

DNA fingerprinting also referred to as genome individualization was first demonstrated by cloned probes (Jeffreys et al. 1985a). Soon after, synthetic oligo probes were shown to be equally informative in the detection of an individual-specific DNA profile (Ali et al. 1986). Since then, several synthetic oligo probes useful for DNA fingerprinting of a large number of species have been reported (Epplen 1988, Ali & Epplen 1991). Use of different restriction enzymes in combination with the same probe uncovers altered band pattern (figure 2). Thus, even with a single probe, informative band pattern may be obtained by empirically using different restriction enzymes. In the process most informative probe/enzyme combination for a given species may be established.

Cloning and propagation of the repeat fragments often pose technical and logistic problems because they are generally unstable in the prokaryote host and often recombine during

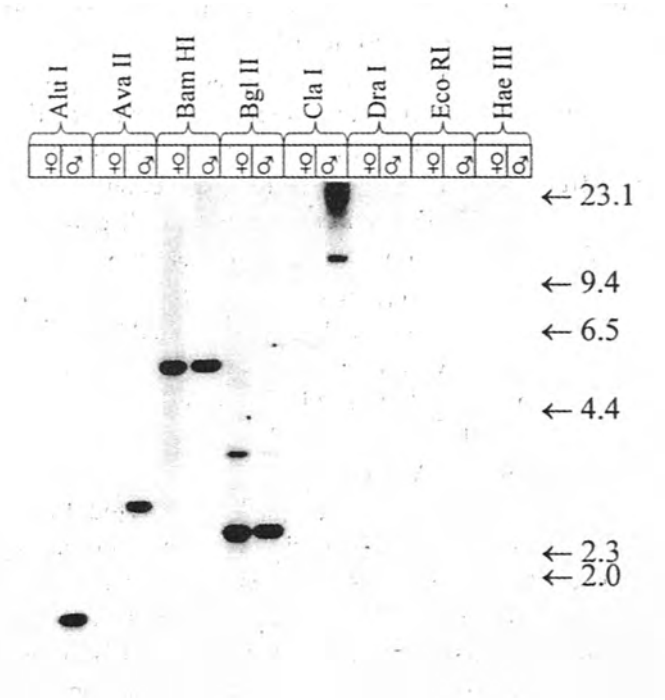


Figure 2 Southern survey of male and female buffalo genomic DNA digested with *Alu* I, *Ava* II, *Bam* HI, *Bgl* I, *Cla* I, *Dra* I, *Eco* RI and *Hae* III enzymes and probed with OAT 24.2 containing six units of GACA repeat motif. Note the difference in the band profile with change of the enzymes used. Male DNA with *Cla* I enzyme shows incomplete digestion whereas *Alu* I shows only one band in the lower molecular weight region. Complete digestion of the DNA is critical for obtaining accurate results. This approach is particularly useful to uncover allele length variation with new probe/enzyme combination. Molecular size marker given in kb is useful for estimating the relative size of the bands

propagation of the clones. Moreover, a majority of such naturally existing tandem repeats do not detect related structures due to lack of complete cross-hybridization which is length dependent (Mariat & Vergnaud 1992). This problem may be circumvented by using "tailor-made" synthetic oligonucleotide probe covering larger pool of genetic variability (Ali & Epplen 1991).

Advantages of Oligonucleotide Probe

Oligonucleotide probes under appropriate hybridization conditions show absolute specificity with the target DNA (Itakura et al. 1984) and mismatched bases do not allow formation of stable duplexes (Wallace et al. 1986). Optimisation of the length of the oligonucleotide for detecting maximum number of variant alleles was first attempted using different repeat unit lengths of (GATA)n in the human genome (Ali et al. 1986). The result showed that an increase in the probe length due to duplex mismatch reduced the overall hybridization. In subsequent studies, additional probes based on 2-6 base repeat motifs were used for hybridization with genomic DNA from different sources (Epplen 1988, Vogel et al. 1988, Weising et al. 1990, Buitkamp et al. 1991). These repeats include (AT)n, (CA)n, (CT)n, (CAC)n, (GAA)n, (TCC)n, (GACA)n, (GATA)n, (GGAT)n, (GGGCA)n, (TTAGGG)n, (TTCCA)n and CACCTCTCCACCTGCC (Epplen et al. 1991, John & Ali 1997, Ali et al. 1999). The number of different oligo probes reported to be useful for all the eukaryotic genomes studied so far, is 20 or less. These probes revealed different band patterns in different species with varying signal intensity (figure 3). The information content of DNA probe is dependent on the average number of discernible polymorphic bands per individual and the respective band frequencies (Jeffreys et al. 1985b). Extensive family studies on human and animal systems using synthetic oligos have shown that bands detected by such probes follow Mendelian inheritance and represent autosomal loci which segregate independently (Epplen et al. 1991). For successful DNA fingerprinting of a species, a panel of different oligo probes may be evaluated by adjusting the length and sequence complexity of the same to establish their reliability and overall informativeness (Epplen et al. 1991).

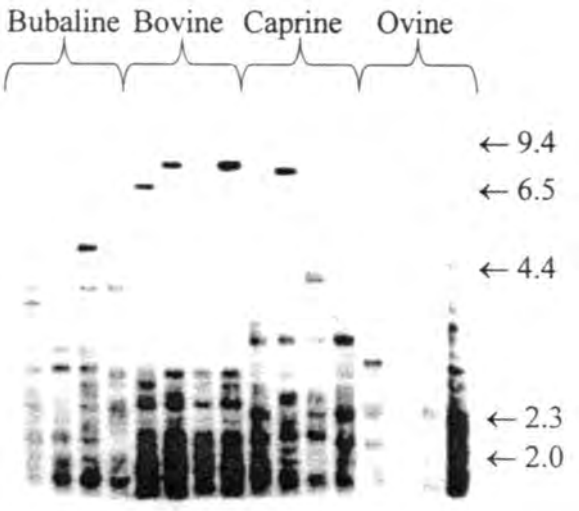


Figure 3 Hybridization profile of DNA samples from four different species of Artiodactyla including buffalo (bubaline), cattle (bovine), goat (caprine) and sheep (ovine), with OMS1/MboI combination. Four random samples (two from each sex of each species) were analysed simultaneously for DNA profiling and subsequently for the construction of phenogram (not shown). Note the multilocus polymorphic band in the lower molecular weight region and a prominent band in some individuals in all the groups except ovine. Weak signals in one lane of ovine are due to less quantity of DNA sample used. Molecular size marker is given in kb

Reliability of Synthetic Oligos as DNA Fingerprint Probes

Soon after the report on DNA fingerprinting by synthetic oligos (Ali et al. 1986), it was shown that a trinucleotide repeat $(CAC)_5$ in combination with *Hin*fI enzyme is the most informative one for human system with a probability of 2.0×10^{-8} for identical band profile between any two random individuals. Initial fingerprint analysis with the synthetic oligos (GATA)n and (GACA)n (Ali et al. 1986) detected less polymorphic variants as compared to $(CAC)_5$ (Schäfer et al. 1988). Hypervariability at repeat loci reflects high rates of *de novo* mutations producing newer length alleles. As expected from the neutral mutation random drift hypothesis, mutation rate increases with variability and becomes significant above approximately 96% heterozygosity level. Germline mutations will produce apparent

exclusions in paternity testing and somatic mutations would produce divergence in the DNA fingerprint pattern of different tissues from the same individual. Thus, before a probe is used in actual case condition, the somatic and germline stability must be known.

Germline Mutation

The germline mutation rate estimated for minisatellite probes in humans was found to be in the range of 0.001-0.004 per gamete per locus (Jeffreys el al. 1985b), whereas for the highly hypervariable marker, lambda MS1, the same was found to be 0.052 per gamete per locus (Jeffreys et al. 1988). The mutation rate detected by oligo probe $(CAC)_5$ in human was well within the range estimated earlier for minisatellite probes. In an attempt to understand the exact mechanisms involved in the germline mutations, Mitani et al. (1990) studied the mutation rate in both mouse and human pedigrees at a minisatellite region. They introduced single base substitutions in the core sequence of this minisatellite and using this as a probe, detected altered hybridization patterns. From the detailed study, they concluded that repeat motif in the tandem array itself contributes to the germline instability.

Mutation that can be identified in pedigrees is limited to a small number of alleles. Therefore, more accurate study on the rate of mutations and the factors causing germline instability of repeat loci was made feasible by sperm typing using PCR technique. Quantitation of mutation load in individual gametes at three hypervariable regions MS32, MS105 and MS31A of human minisatellite loci revealed that there was a preferential gain of a few repeats at one end (polar mutation) on the tandem array (Jeffreys et al. 1994a). In these mutations which appeared to be largely germline specific (Jeffreys et al. 1988), the flanking sequences did not take part. This led to the speculation that mutations could be caused mainly due to complex gene conversion-like events within the same allele or homologous chromosomes (Jeffreys et al. 1994a). The mutational polarity noticed in the study suggested that these events are regulated by elements outside the tandem repeat array such as flanking DNA elements which possibly might be serving as mutation/conversion initiators (Jeffreys et al. 1994a). Further attempts to define these mutation initiator elements are being carried out in transgenic mice harbouring human minisatellite (Jeffreys et al. 1994b) Whatever be the mechanism, the rate of mutation of the hypervariable loci, which are used as markers for genome individualization has great importance in forensic cases and paternity testing. Parentage exclusions with limited number of hypervariable loci would lead to false exclusions of genuine parents if the mutation rate is significantly higher (Gyllensten et al. 1990). Mutation rates of approximately 10^2 per gamete do not significantly interfere with the use of these probes in paternity analysis, provided their rate of occurrence is known and incorporated into the analysis of statistical likelihood ratios (Jeffreys et al. 1991).

Somatic Mutation

Thein et al. (1987) and Thibodeau et al. (1993) reported changes in the fingerprint pattern of the constitutional and tumor DNA from patients. These changes showed alterations in the relative intensities of hybridizing bands and appearance of novel fragments not seen in the corresponding constitutional DNA. Repeat sequence instabilities in the somatic tissues had also been reported in other inherited genetic disorders like myotonic dystrophy, Fragile X syndrome, Kennedy's syndrome and Huntington's disease (Richards & Sutherland 1992). Synthetic oligo probes used for human DNA fingerprinting also detected such somatic instabilities. Both $(CAC)_5$ (Nürnberg et al. 1991) and telomeric repeat $(TTAGGG)_3$ (Hastie et al. 1990) showed expansion and contractions in their repeat length during the progression of glioma tumors (Lagoda et al. 1989). To explain the somatic instability at repeat sequences, several hypotheses have been proposed. Loss of chromosome(s) or chromosomal regions through deletion, mitotic non-disjunction and recombination would lead to the loss of associated minisatellite fragment. Conversely, localized endoreduplication of DNA encompassing a minisatellite would cause specific band amplification (Schimke 1984). Also tissue

or tumor specific changes in DNA methylation affect DNA fingerprints (Goelz et al. 1985). In human, a gene involved in DNA repair located on chromosome 2 has been suggested to be causing instability to mono-, di- and trinucleotide repeats in colon cancer cells (Peltomaki et al. 1993).

Other Methods for Detecting Repeat Sequence Variability

Ligated Oligonucleotide Probe (LOP)

Besides single-stranded synthetic oligo-probes, enzymatically ligated double-stranded probes may be generated for specific purposes (Ali & Wallace 1989). These synthetic repeats consist of head to tail polymerization of an arbitrary oligonucleotide (complementary to VNTR loci) generating tandemly repeated fragments longer than 400 bp that are then used as a probe at higher hybridization stringency. The ligated oligonucleotide probe (LOP) was found to detect polymorphic loci producing highly informative DNA fingerprint pattern in human, compared to its single unit component single-stranded conventional oligo probe (COP). LOPs are more easily generated than cloned VNTR probes (Vergnaud 1989), helps avoiding problems associated with bacterial growth and maintenance as well as *in vitro* labelling.

Polymerase Chain Reaction (PCR)

Amplifications of single locus minisatellite by PCR using flanking sequences as primers and direct sequencing of these amplified alleles have also been used to produce individual specific fingerprints and species specific band pattern. Another approach was to use short tandem repetitive "microsatellite" sequences as primers for PCR (Helminen et al. 1992). This has greatly increased the sensitivity of DNA typing systems and is particularly useful for typing degraded human DNA in forensic cases (Jeffreys et al. 1988, Tautz 1989). Several modifications have been made in conventional PCR techniques to make it more suitable for addressing specific aspects of genome analysis.

Single Stranded Conformational Polymorphism (SSCP)

PCR-SSCP is a widely used technique to detect mutation in genes responsible for various hereditary diseases and somatic mutations of oncogenes and tumor suppressor genes in cancer as well as polymorphisms. The target sequence is amplified and radiolabeled or amplification is conducted with labeled primers. The heat denatured single stranded amplicons are resolved on non-denaturing polyacrylamide gel (PAGE) and the conformational change of the single stranded DNA caused due to point mutation is detected as a mobility shift of the conformational radii (Orita et al. 1989). This method has been used for the linkage mapping of human genome by analysing polymorphisms within Alu repetitive elements. Phylogenetic assessment of length variation at microsatellite locus can also be detected using this technique (Orti et al. 1997). Although PCR-SSCP is highly sensitive, false negative results cannot be excluded easily and some mutations may escape detection under chosen conditions of electrophoresis.

Randomly Amplified Polymorphic DNA (RAPD)

RAPD (Welsh & McClelland 1990, Welsh et al. 1991) also referred to as Arbitrarily Primed Polymerase Chain Reaction (AP-PCR) (Williams et al. 1990) is the DNA amplification of arbitrary sequences directed by a very short single oligonucleotide primer to generate complex but characteristic DNA fingerprints. Polyacrylamide gel electrophoresis and silver staining have been used to resolve the amplicon into detailed and reproducible band patterns (Bassam et al. 1992). However RAPD amplicons may also be resolved on the agarose matrix and bands may be recorded from the ethidium bromide stained agarose gel. The concentration of agarose gel may be adjusted depending upon the overall size of the amplicon. The technique has been proven to be useful in genome individualization and analysis of genetic relationships (Wang et al. 1993), detecting inherited diseases (Rothuizen & Van Wolferen 1994) and identifying strains and constructing genetic maps (Johnson et al. 1994).

Disadvantage(s) of RAPD

A drawback of RAPD is that the nature of the randomly primed template DNA remains unknown. Though the methodology is sensitive enough to detect in principle even single nucleotide insertion and deletion or base substitution, a high level of standardization and precise internal control is necessary to obtain consistently reproducible profile (Rothuizen & Van Wolferen 1994). Absence of the expected Mendelian inheritance of DNA amplified with RAPD has been reported in baboon and human CEPH pedigrees (Riedy et al. 1992) and to a lesser extent in beetles (Scott et al. 1992). These non-paternal bands may arise due to heteroduplex formation of alternate alleles in heterozygotes (Hunt & Page 1992) and can be a potential source of artifactual polymorphism during RAPD analysis. The technique has achieved broad applications in the studies of genetic variations. In natural populations, artifactual variations may lead to over-estimation of level of variation if the bands are interpreted as a true variation and therefore the discrimination between true and artificial variants is critical (Ellsworth et al. 1993) and should be looked into carefully while conducting the experiment.

Amplified Fragment Length Polymorphism (AFLP)

Amplified fragment length polymorphism (AFLP) has been proposed to be yet another tool for genetic typing and mapping, useful particularly for very small amounts of DNA. The technique is based on the selective PCR amplification of target DNA, which involves restriction digestion of the same, followed by ligation with the adapter oligonucleotides. The set of oligo primers complimentary to the adapter sequences is then used for amplification (Vos et al. 1995). AFLP can be conducted on DNA of any origin or complexity. However, it suffers from the flaws such as incomplete (partial) amplification of the genomic DNA, reproducibility problem and several cumbersome steps involved in digestion, ligation and amplification. Finally, presence of faint spurious bands from amongst the real one may escape detection.

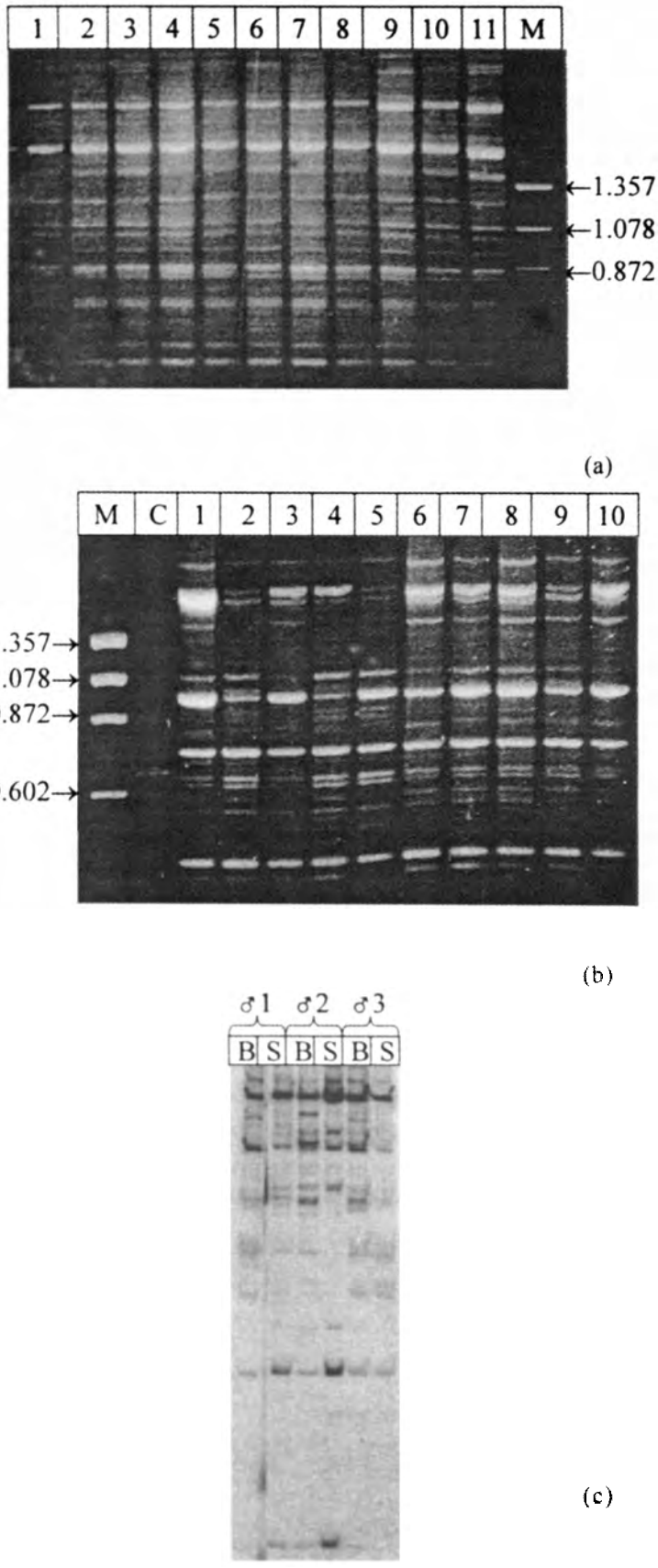
Molecular Basis of Minisatellite associated sequence amplification (MASA)

Minisatellite associated sequence amplification (MASA) is a PCR based reaction analogous to Random Amplification of Polymorphic DNA (RAPD). However, unlike RAPD or AP-PCR, instead of using arbitrary oligo primers, MASA is conducted using strictly well defined oligo primers of 12-16 base residues representing core or consensus sequences of the minisatellite regions, mostly organized in species specific manner. Inter-species utility of MASA is directly related to the conservation status of the sequences used as primer. Thus, primer sequences conserved within the genus will be useful for establishing generic affinities. However, an evolutionarily conserved primer will be useful for a large number of unrelated species encompassing various levels of evolutionary hierarchy (figure 4a, b & c). A few well-documented minisatellite primers used for MASA may not cover the entire pool of clad diversity. Thus, systematic search and identification of additional primer(s), preferably conserved evolutionarily would augment MASA mediated genome analysis of even distantly related clads. Using two different oligo primers of varying length and sequence complexities, species-specific band profiles were recorded in several vertebrate species (Azfer et al. 1999) including highly endangered ones (Ali et al. 1999). Of the RAPD and AFLP approaches, we find that MASA is more promising and highly focussed one for analysis of eukaryotic genome.

Behavioral Ecology and DNA Finger-printing

Evolutionary and population biologists may quantify genetic relationship among individual organisms within and between the species based on coding and non-coding sequences (Mattapallil & Ali 1999). It is also possible to uncover organizational variation of repeat loci between two population of the same species leading to the assessment of environmental effect(s).

Employing DNA fingerprinting, extra-pair paternity and nonspecific nest parasitism in birds have been demonstrated (Burke &Bruford 1987, Birkhead et al. 1990). Detailed studies on the



reproductive behaviour of individual birds and other mammals have also been conducted employing DNA profiling methods (Gullberg et al. 1992). In one such study on "Dunnock" this approach was used to ascertain whether the sub-ordinate male only fed the brood or he also fathered any of the nestling (Burke et al. 1989). DNA fingerprinting approach has also been used in many other species to estimate genetic relatedness (John & Ali 1997). The naked mole rat *Heterocephalus glaber* is unique example among mammals in having a social structure similar to that of social insects. An extreme degree of inbreeding (F= 0.45) the highest recorded among wild animals was noted in this population resulting from consanguineous mating. Similarly, minisatellite band sharing data of African lion revealed kinship within and between lion pride (Packer et al. 1991). Employing DNA fingerprint approach, free ranging chimpanzee population in Africa has been studied and pedigree analysis showed that males were related in the order of half sibling and homozygosity was significantly increased in a kin group proving the kin selection theory. Thus, DNA fingerprinting was found to be useful

← **Figure 4** Minisatellite associated sequence amplification (MASA) mediated DNA band profile of *Capra hircus* with O33.15 primer CACCTCTCCACCTGCC representing consensus sequence of human hypervariable regions. Note the multilocus isomorphic bands in the range of 200 bp to 2 kb and a few variable bands. "M" represents molecular size marker given in kb (a). MASA mediated profile of *Capra hircus* DNA with OAT 15.2 primer representing tetramer repeat GACA. Note more number of polymorphic bands with intense copy number variation in the range of 0.5-1.6 kb. Clearly, 15.2 primer uncovers more variability than O33.15 suggesting that MASA mediated band profile is exclusively based on primer sequences (see text, for details). "M" is the molecular size marker given in kb whereas "C" represents negative control (b). Hot MASA (radioactive reaction) of *Bubalus bubalis* genomic DNA with OAT18.2 primer containing (GACA)4.5 resolved on a 3.5% non-denaturing polyacrylamide gel. "B" and "S" represent DNA from blood and semen samples respectively. Note the difference in the band profile between "B" and "S" of the same individual demonstrating discriminatory power of MASA approach on one hand and indicating possible programmed sequence modulation of the repeat in the semen samples on the other hand. However, the MASA conducted with O33.15 primer failed to uncover the difference between the "B" and "S" samples indicating, once again, the specificity of the primer(s) used (c)

for studying gene flow and evolution of chimpanzee population (Morin et al. 1994). Similar studies on the West Africa *Pantroglodytes verus* showed them to be well-differentiated and independently evolving taxa, with a divergence time of about 1.58 million years from the other two subspecies.

Some aspects of behavioral ecology and genetics remain a challenging proposition. A case for example is the infant killing of Hanuman langur by adult. It would be a daunting task to uncover the genetic basis of the aggressive behaviour of an Adult Hanuman Langur leading to infanticide. However, employing DNA analysis, it is possible to establish if the killer has ever been the actual father. In the process, if it is proved that the real father never ever kills his own progeny, troops may be segregated or guarded and invading hostile males may be kept apart. Similarly, phenomenon of allo mothering may be studied on the basis of degree of genetic affinities of the so-called allo-mother with the infant. Generic affinities based on DNA analysis will provide data interpretable in term of behavioral genetics. Probe for DNA profiling of hanuman langur has not been reported but the same may be developed following established approaches (Ali et al. 1986, Ali & Epplen 1991, Epplen et al. 1991). Alternatively, owing to closeness of the two species, probe(s) informative for human DNF may be used for Hanuman langur. This will also enable scoring of all the other beneficial points related to DNF technology including the assessment of population structure and overall sequence variation within and between the troops, species and populations.

Conservation Genetics

Sustainability of fast depleting natural resources has become a cause of global concern and India is no exception. This includes both flora and fauna encompassing a wide spectrum of bio-assets. Owing to the absence of socio-economic parity and fast changing world scenarios, attention may be focused to undertake documentation, management and conservation of endangered species (Ali & Hasnain 1999). In view of the enormous numbers and diversified magnitude of the plant and animal species, molecular characterization of the same is a gigantic proposition. However, it is not impossible to thwart the ever-growing menace of bio-piracy, check the environmental damages and restore the delicate fabrics of ecological balance by undertaking, systematic documentation, identification and genetic characterization of highly endangered animal (Ali et al.1998, 1999) and plant species (Sulaiman et al. 1996). This, in turn, would translate *albeit* partly into conservation of our critical bio-diversity. Similarly, it is possible to identify a successful breeder (male in a species) and the extent to which his own germplasm has been propagated (Mattapallil and Ali 1997). Inbreeding of a confined population will lead to an increased level of genetic homozygosity (loss of hybrid vigour) eventually affecting population dynamics (Ali et al. 1998, Ali et al. 1999). Infusion of newer genetic materials in a population from a small geographical region can not be assessed on the basis of morphological features. However, this may be established on the basis of detailed DNA analysis. Whether, the small population in turn will become more prone to genetic anomaly and consequently less resistant to infectious diseases, affecting eventually its fertility status, is another important aspect manageable to a large extent by DNAF and MASA approaches.

In the wild, several species have become highly endangered due to reasons largely attributable to human greed. It is not surprising therefore to see environmental scientists and field ecologists often at a loss to chalk out appropriate strategy for the conservation of species that are fast becoming extinct. It may be noted that some species have several breeds and are more successful as fast breeder than others. Thus, breed delineation employing molecular marker or marker systems and analysing their fertility and fecundity status will uncover the overall reproductive potential leading to the development of appropriate measure towards their conservation. Breeds of water buffalo, cattle, goat, pig, rabbit and sheep have been characterized on the basis of morphometric attributes and information gleaned from the local resources but not on the basis of in-depth genetical analysis. Breed delineation based on

molecular analysis will provide more authentic documentation of the species and help in developing long term strategy for propagation of the desired germplasm. In earlier studies, we reported breed delineation based on synthetic DNA based genetic markers (John & Ali 1997). This may also be achieved by MASA using appropriate primers.

Phylogenetic Analysis

Using polymorphic markers and statistical methods, taxonomic status, phylogenetic relationship and genetic distances of different species have been studied (Felsenstein 1984). Earlier the marker used in solving the phylogenetic riddles included those that detected isozyme variants, amino-acids substitutions in homologous proteins and karyological relationships (O'brien et al 1985). Homology at β-like globin genes, α-lactalbumin sequences (Easteal 1990) and the fast evolving mt DNA sequences (Moum et al. 1994) were also used for this purpose. In order for the derived topology to be more precise, the loci under consideration should be extremely polymorphic and conserved in a wide range of species (Buchanan et al. 1994). Although the mtDNA sequences are polymorphic and evolve 5-10 times faster as mentioned earlier, they do not provide any information about the extent of nuclear gene flow or variability which is central to the evolution of overall make up of an organism. Owing to ubiquitous distribution across the eukaryotic genomes (Britten 1986), repetitive families offer sufficient fidelity to distinguish populations that are separated by a short evolutionary time of divergence from their common ancestors.

Short interspersed repetitive elements (SINEs) were used as probe to study the phylogenetic relationships and taxonomic classification of closely related species. The phylogenetic assignment of pacific salmon (steelhead trout) was changed from salmon to *Oncorynchus* based on SINEs insertion analysis (Murata et al. 1993). SINEs appear to be inserted irreversibly and therefore provide ideal evolutionary and phylogenetic markers. The stochastic population events like genetic drift and founder effect with respect to minisatellite often lead to concerted evolution (Dover 1982). High level of interpopulation homogeneity was noticed in a pupfish at *Hin*dIII satellite DNA sequence, which is a useful marker to assess the structure of commercially exploited species (Elder and Turner 1994). Similarly, evolutionarily conserved Bkm (Banded Krait minor) sequences have been used to provide phylogenetically useful information about crocodilians (Aggarwal et al. 1994). Based on detailed analysis of several microsatellites, buffalo has been found to be closer to sheep than cattle (Mattapallil & Ali 1999). Extensive genome analysis were carried out to delineate the evolutionary relationships and breed affiliations of economically important farm animals (Ellegren et al. 1993, Buchanan et al. 1994). DNA fingerprinting of the Channel Island fox population of the coast of California, showed very low levels of variations within the population and a high degree of differentiation among populations (Gilbert et al. 1990).

The most widely distributed and highly polymorphic microsatellite repeat sequence (dC-dA)n, (dG-dT)n was first identified independently by two groups of investigators. Since then they have been widely used for a variety of applications. More recently they have been used as linkage markers for gene mapping and interpopulation studies (Chakraborty et al. 1992). Using a set of 30 (CA)n repeat loci, reliable evolutionary relationships were established amongst 14 human populations supporting the hypothesis of an African origin for humans (Bowcock et al. 1994). In another study, characterization of the dinucleotide (CA)n repeat loci in 8 human populations showed extensive diversity (still corroborating with the Hardy-Weinberg equilibrium) and pair-wise genotypic independence across these microsatellite loci (Deka et al. 1994). Despite substantial variation in allele frequency at each locus most frequent alleles were shared by all human populations. Thus, the sequence variations at repeat loci could be used as markers for the evolution of humans as well.

Application of DNA Fingerprinting Approach in Breeding Programs

The information made available by DNA fingerprinting could also be extended to breeding programs of farm animals/birds. The genetic relationships amongst different breeding populations of chicken were established using minisatellite probes (Kuhnlein et al. 1989). Genomic selection of strains/lines of farm animals/birds based on their genetic similarity, for gene introgression in breeding programs (Hillel et al. 1990) and the applicability of this approach to establish evolutionary relationships between species have been demonstrated (Kuhnlein et al. 1989, 1990). DNA fingerprinting approach was used to confirm the origin of somatic cells used for cloning Dolly the sheep (Campbell et al. 1996). On the other hand, a parsimony analysis of the fingerprint patterns of different domesticated equine breeds (Swedish trotters, North Swedish trotters, thoroughbreds and Arabians) using synthetic (TG)n probe clustered them into respective breeds (Ellegren et al. 1993). Similar approach was used for ovine and significant differences were noticed in the allele frequencies at each repeat locus.

Other Applications of Synthetic Probes Based on Repeat Sequences

An oligo probe was used for monitoring success of bone marrow transplantation (BMT) (Yam et al. 1987). Remarkable progress has recently been made in genetic linkage mapping both in human and animal species. There is a growing consensus that systematic approach to dissect out single and multigenic economically important traits of livestocks is possible only with the help of microsatellite markers that are randomly spread throughout the genome of almost all species. Microsatellite mapping of the gene responsible for weaver disease in cattle, which segregates with increased milk production, has been identified. The closely linked marker that allowed selection against weaver disease was used to characterize the role of this locus in milk production (Georges et al. 1993a). Another locus thought to be involved in horn development in *Bos taurus* has been assigned to chromosome 1 by microsatellite mapping (Georges et al. 1993b). Progeny testing in cattle using multilocus fingerprint data has been exploited for mapping loci controlling milk production. Linkage analysis identified five chromosomes (16,9, 10 and 20) giving rise to very strong evidence for the loci controlling milk production (Threadgill & Womack 1990). The Booroola fecundity (FecB) gene responsible for multiple births of sheep has been mapped to chromosome 6 using microsatellite markers (Montgomery et al. 1993). Similarly, "probe walking", by using a battery of polymorphic satellite sequences has been proposed for complete mapping of the animal genome(s) (Washio et al. 1989). These markers may also be used in marker assisted selection (MAS) and breeding program of livestocks.

Ultravariability at Repeat Loci

Error-prone allele length estimates and electrophoretic "band shifts" can occasionally lead to apparent exclusions between 'matching' DNA profile databases. More recently new dimensions had been added to the DNA fingerprinting approach employing internal mapping procedure (Jeffreys et al. 1990). Though the bands migrating at the same level are considered to be the same allele, it was clear that repeat units within an allele, although very similar length-wise, show subtle variations in their sequence. They presumably arise by point mutations and hence generate new restriction sites. Jeffreys and co-workers (1990) tried to divulge this variability within an allele by PCR. After an initial amplification of the repeat loci (DIS8) using oligonucleotides with complementary sequences to repeat units (MS32) as primers, aliquots were end labelled and then digested with *Hinf* I or *Hae* III enzymes. This generated a population of DNA molecules in which each possible cutting site is represented by end-labelled fragments of a discrete length from the 5' end. This new technique gives an exact measure of the length of an allele and also the internal mapping of each allele.

Since this method is limited to alleles, small enough to be amplified by PCR and provides mutational bias with deletions more likely to

occur at the 3' end, another approach of assaying these sequence variations in minisatellite alleles was developed (Jeffreys et al. 1991). This technique, minisatellite variant repeat-polymerase chain reaction (MVR-PCR) uses two different tagged primers for the PCR, which can amplify both the variants of the alleles separately, thus obviating the need for a second step of restriction digestion. With an increase in the variability and heterozygosity detected using this approach, the mutation rate detected at the same loci (D1S8) was found to be higher in the CEPH pedigrees. This ultravariability reveals the direct evidence of minisatellite instability due to interallelic recombination/gene conversion.

Future Directions and Envisaged Frontiers

Technology begets technology leading to generation of newer information culminating eventually in its application. Development of minisatellite associated sequence amplification (MASA) took place in our laboratory during the study of one horn rhino *Rhinoceros unicornis* genome (Ali et al. 1999). Soon it was noticed that this approach is equally useful for unequivocal clad identification (Azfer et al. 1999). Although, thus far we have used only two different oligo primers of varying length and sequence complexities for MASA, we believe that universal primer based on telomeric repeat $(TTAGGG)_3$ may be used for generating wealth of information such as clad identification, establishing generic affinities, phylogenetic analysis and even identifying loss or gain of alleles in tumor tissues and finally assessing programmed sequence modulation in semen DNA samples compared to that from somatic ones. The scope of clad identification may be broadened by including more number of species employing MASA approach. Following this, MASA may be put to actual use that would prove to be a decisive factor in the context of illegal trades in wildlife, a major concern for the conservation of endangered species. There are 25 mammalian species that belong to schedule I and II of Wildlife Protection Act (1972). The annual illegal trade is of 5 billion US $ which includes skin, ivory, horn, antler, bone live animals, feather, nail, claws and pods. Thus, development of technology useful for genome individualization and individual clad identification of these species will strengthen the wildlife forensic capabilities. Minisatellite associated sequence amplification (MASA), in addition to clad identification, has been found to uncover sex specific band pattern in human, rhino and Hanuman Langur (under preparation). Thus, MASA proves to be informative not only for clad but also for sex identification of some vertebrates. It would be of interest to analyse more number of species for sex identification using this approach. This would positively broaden its utility for the management of critically endangered species. Finally MASA mediated co-segregating prominent bands may be cloned independently and their inter clonal variation and polar and non-polar mutation may be ascertained. Thus intrinsic polymorphism of MASA mediated allele would be yet another robust approach useful for wildlife and forensic science.

Concluding Remarks

In this article only limited applications have been touched upon and many issues are still to be addressed for which additional markers will be required. We hope that this article will provide the basic background information towards the initiation and development of additional marker systems encompassing larger spectrum of flora and fauna and poorly managed but rich biodiversity of our country.

Acknowledgments

Part of the work reported here was supported by a DST grant no. SP/SO/D-51/93 to SA and an Institutional core grant from the Department of Biotechnology, Government of India. SG acknowledges the award of a Senior Research Fellowship from the Council of Scientific and Industrial Research (CSIR), New Delhi. We thank Shri Khem Singh Negi for technical assistance.

References

Aggarwal R K, Majumdar K C, Lang J K and Singh L 1994 Generic affinities among crocodilians as revealed by DNA fingerprinting with Bkm derived probe; *Proc. Natl. Acad. Sci. USA* **91** 10601-10605

Ali S, Müller C R and Epplen J T 1986 DNA fingerprinting by oligonucleotides specific for simple repeats; *Hum. Genet.* **74** 239-243

_____ and Wallace R B 1989 Enzymatic synthesis of DNA probes complementary to a human variable tandem repeat loci; *Anal. Biochem.* **179** 280-283

_____and Epplen J T 1991 DNA fingerprinting of eukaryotic genomes by synthetic oligonucleotide probes; *Indian J. Biochem. Biophys.* **28** 1-9

_____, Ansari S, Ehtesham N Z, Azfer M A, Homkar U, Gopal R and Hasnain S E 1998 Analysis of the evolutionarily conserved repeat motifs in the genome of the highly endangered central Indian Swamp deer *Cervus duvauceli branderi; Gene* **223** 361-367

_____, Azfer M A, Bashamboo A, Mathur P K, Malik P K, Mathur V B, Raha A K and Ansari S 1999 Characterization of a species specific repetitive DNA from a highly endangered wild animal *Rhinoceros unicornis* and assessment of genetic polymorphism by microsatellite associated sequence amplification (MASA); *Gene* **228** 33-42

_____ and Hasnain E 1999 Conservation, management and documentation of livestock genetic resources and biodiversity: A biotechnology perspective; *Curr. Sci.* **77** 1154-1158

Azfer M A, Bashamboo A, Ahmed N and Ali S 1999 Random amplification of polymorphic DNA with conserved sequences reveals genome-specific monomorphic amplicons: Implications in clad identification; *J. Biosci.* **24** 35-41

Bassam B J, Caetano-Anolles G and Gresshoff P M 1992 DNA amplification fingerprinting and its potential for genome analysis. In *Current topics in plant molecular biology: Plant biotechnology and Development,* ed. P M Gresshoff, CRC press, Boca Raton, 1-9

Bates G and Lehrach H 1994 Trinucleotide repeat expansions and human genetic disease.; *Bioessays* **4** 2777-284

Bell G L, Karam J H and Rutter W J 1981 Polymorphic DNA region adjacent to the 5' end of the human insulin gene; *Proc. Natl. Acad. Sci. USA.* **78** 5759-5753

Birkhead T R, Burke T, Zann R, Hunter F M and Krupa A P 1990 Extra-pair paternity and intra-specific brood parasitism in wild zebra fishes *Taeniopygia guttata* revealed by DNA fingerprinting; *Behav. Ecol. Sociobiol.* **27** 315-324

Blackburn E and Szostak J A 1984 The molecular structure of centromeres and telomeres; *Rev. Biochem.* **53** 163-172

Botstein D, White R L, Skolnick M and Davis R W 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms; *Am. J. Hum. Genet.* **32** 314-331

Bowcock A M, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd J R and Cavalli-Sforza L L 1994 High resolution of human evolutionary trees with polymorphic microsatellites; *Nature* **368** 455-457

Britten R J and Kohne D E 1968 Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms; *Science* **16** 529-540

_____ 1986 Rates of DNA sequence evolution differ between taxonomic groups; *Science* **231** 1393-1398

Brown S D M and Piechaczyk M 1983 Insertion sequences and tandem repetitions as sources of variations in a dispersed repeat family; *J. Mol. Biol.* **165** 249-256

Bruford M W and Wayne R K 1993 Microsatellite and their application to population genetic studies; *Curr. Op. Genet. Dev.* **3** 939-943

Buchanan F C, Adams L J, Littlejohn R P, Maddox J F and Crawford A M 1994 Determination of evolutionary relationships among sheep breeds using microsatellites; *Genomics* **22** 397-403

Buitkamp J, Ammer H and Geldermann H 1991 DNA fingerprinting in domestic animal species; *Electrophoresis* **12** 169-174

Burke T and Bruford M W 1987 DNA fingerprinting in birds; *Nature* **327** 149-152

_____, Davies N B Bruford M W and Hatchwell B J 1989 Parental care and mating behaviour of polyandrous dunnocks *Prunella modularis* related to paternity by DNA fingerprinting; *Nature* **338** 249-251

Campbell K H, McWhir J, Ritchi W A and Wilmut I 1996 Sheep cloned by nuclear transfer from cultured cell line; *Nature* **380** 64-66

Caskey C T, Pizzuti A, Fu Y H, Fenwick R G and Nelson D L 1992 Triplet repeat mutations in human disease; *Science* **256** 784-789

Chakraborty R, Deka R, Jin L and Ferrel R E 1992 Allele sharing at 6 VNTR loci and genetic distances among 3 ethnically defined human population; *Am. J. Hum. Biol.* **4:** 387-397

_____1994 Evolutionary considerations of DNA fingerprinting data; in *III rd Intl Conf DNA fingerprinting,* India, p9

Cohen E S, Thoday J M and Dover G 1982 Rate of turnover of structural variants in the rDNA gene family of Drosophila melanogaster; *Nature* **295** 564-568

Davidson E H and Britten R J 1979 Regulation of gene expression: possible role of repetitive sequences; *Science* **204** 1052-1059

Deininger P L and Daniels G R 1986 The recent evolution of mammalian repetitive DNA elements; *Trend. Genet.* **2** 76-80

Deka R, Shriver M D, Yu L M, Jin L, Aston C E, Chakraborty R and Ferrell R E 1994 Conservation of human chromosome 13 polymorphic microsatellite (CA)n repeats in chimpanzees; *Genomics* **22** 226-230

Dover G A 1982 Molecular drive-a cohesive mode of species evolution; *Nature* **299** 111-117

_____ 1986 Molecular drive in multigene families: how biological novelties arise, spread and are assimilated; *Trend Genet.* **2** 159-165

_____ 1987 DNA turnover and the molecular clock; *J. Mol. Evol.* **26** 47-58

Easteal S 1990 The pattern of mammalian evolution and relative rate of molecular evolution.; *Genetics* **124** 165-173

Elder jr J F and Turner B J 1994 Concerted evolution at the population level: Pupfish Hind III satellite DNA sequences; *Proc. Natl. Acad. Sci. USA* **91** 994-998

Ellegren H, Davies C J and Andersson L 1993 Strong association between polymorphism in an intronic microsatellite and in the coding sequence of the BoLA-DRB 3 gene; implications for microsatellite stability and PCR based DRB 3 typing; *Anim. Genet.* **24** 269-275

Ellsworth D L Rittenhouse K D and Honeycutt R L 1993 Artificial variation in randomly amplified polymorphic DNA banding patterns; *Biotechniques* **14** 214-215

Epplen J T, McCarrey J R, Sutou S and Ohno S 1982 Base sequence of a cloned snake W-chromosome DNA fragment and identification of a male specific putative mRNA in the mouse; *Proc. Natl. Acad. Sci. USA.* **79** 3798-3802

_____1988 On simple repeated GATA/GACA sequences: A critical reappraisal; *J. Hered.* **79** 409-417

_____, Ammer H and Epplen C 1991 Oligonucleotide fingerprinting using simple repeat motifs; *A convenient ubiquitously applicable method to detect hypervariability for multiple purposes* pp. 50-69 eds T. BURKE., G. DOLF, A. J. JEFFREYS, and R. WOLFF, DNA fingerprinting approaches and applications. Birkhauser, Switzerland.

Felsenstein J 1984 Distance methods for inferring phylogenies: a justification;. *Evolution* **38** 16-24

Georges M, Dietz A B, Mishra A, Nielsen D, Sargeant L S, Sorensen A, Steele M R, Zhao X Y, Leipold H, Womack J E and Lathrop M 1993a Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus; *Proc. Natl. Acad. Sci. USA* **90** 1058-1062

_____, Drinkwater R, King R, Mishra A, Moore S S Neilsen D Sargeant L S, Sorensen A, Steele M R Zhao X Y et al. 1993b Microsatellite mapping of a gene affecting horn development in Bos taurus; *Nature Genet.* **4** 206-210

Gilbert D A, Lehman N, O'Brien S J and Wayne R K 1990 Genetic fingerprinting reflects population differentiation in the California Channel Island Fox; *Nature* **344** 764-767

Goelz S E, Vogelstein B, Hamilton S R and Feinberg A P 1985 Hypomethylation of DNA from benign and malignant colon neoplasms; *Science* **228** 187-190

Gullberg A, Tegelstrom H and Gelter H P 1992 DNA fingerprinting reveals multiple paternity in families of great and blue tits (*Parus major* and *Parus caeruleus*); *Heriditas* 117 103-108

Gyllensten U B, Jakobsson S and Temrin H 1990 No evidence for illegitimate young in monogamous and polygynous warbles; *Nature* **343** 168-170

Hardman N 1986 Structure and function of repetitive DNA in eukaryotes; *Biochem. J.* **234** 1-11

Hastie N D, Demster M, Dunlop M G, Thompson A M, Green D K and Allshire R C 1990 Telomere reduction in human colorectal carcinoma and with ageing; *Nature* **346** 866-868

Helminen P, Sajantila A, Johnsson V, Lukka M, Ehnholm C and Peltonen L 1992 Amplification of three hypervariable DNA regions by polymerase chain reaction for paternity determinations: Comparison with conventional methods and DNA fingerprinting; *Mol. Cell Probe* **6** 21-26

Hentschel C C 1982 Homocopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S_1 nuclease; *Nature* **295** 714-716

Hillel J, Schaap T, Haberfeld A, Jeffreys A J, Plotzky Y, Cahaner A and Lavi U 1990 DNA fingerprinting applied to gene introgression in breeding programs; *Genetics* **124** 783-789

Hunt G J and Page R E 1992 Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the honey bee; *Theor. Appl. Genet.* **85** 15-20

Iannuzzi L, Dimeo G P, Gallagher D S, Ryan A M, Ferrara L and Womack J M 1996 Gene mapping in river buffalo chromosomes by fluorescence in-situ hybridisation (FISH. International symposium on buffalo products; Paestum, Italy, December 1-4, (1994). Wageningen Press, wageningen 1996. EAAP publication **82** 255-259

Itakura K, Rossi J J and Wallace R B 1984 Synthesis and use of synthetic oligonucleotides; *Ann. Rev. Biochem.* **53** 323-356

Jarman A P and Wells R H 1989 Hypervariable minisatellites: recombinators or innocent bystanders; *Trend Genet* **5** 367-371

Jeffreys A J 1979 DNA sequence variants in the Ggg, Agg, dd and bb globin genes of man.; *Cell* **18** 1-10

_____, Wilson V and Thein S L 1985a Individual specific DNA fingerprints of human DNA; *Nature* **316** 76-79

Jeffreys A J, Brookfield J F Y and Semeonoff R 1985b Positive identification of a immigration test case using human DNA fingerprints; *Nature* **317** 818-819

_____, Royle N J, Wilson V and Wong Z 1988 Spontaneous mutation rates of new length alleles at tandem repetitive hypervariable loci in the human DNA; *Nature* **332** 278-281

_____, Neumann R and Wilson V 1990 Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis; *Cell* **60** 473-487

_____, MacLeod A, Tamaki K, Neil D L and Monckton D G 1991 Minisatellite repeat coding as a digital approach to DNA typing; *Nature* **354** 204-209

_____, Tamaki K, MacLeod A, Monckton D G, Neil D L and Armour J A L 1994a Complex gene conversion events in germline mutation at human minisatellites; *Nature Genet* **6** 136-145

Jeffreys A J, Armour J A L, Collick A, Neil D L, Allen M J, Fretwell N, Guram T, Neumann R, May C A and Jobling M 1994b The basis of DNA fingerprinting: variation and mutation at human minisatellites, in *III$_{rd}$ Intl. Conf. on DNA fingerprinting*, India $_{pp}$ 16

Jelinek W R and Schmid C W 1982 Repetitive sequences in eukaryotic DNA and their expression; *Annu. Rev. Biochem.* **51** 813-844

John M V, Parwez I, Sivaram M V S, Mehta S, Marwah N and Ali S 1996 Analysis of VNTR loci in fish genomes using synthetic oligodeoxyribonucleotide probes; *Gene* **172** 191-197

_____ and Ali S 1997 Synthetic DNA based genetic markers reveal intra and inter-species DNA sequence variability in the *Bubalus bubalis* and related genomes; *DNA and Cell Biology* **16** 369-378

Johnson S L, Midson C N, Ballinger E W and Postlethwait J H 1994 Identification of RAPD primers that reveal extensive polymorphism between laboratory strains of Zebra fish; *Genomics* **19** 152-156

Jones W C and Kafatos F C 1982 Accepted mutations in gene family: evolutionary diversification of duplicated DNA; *J. Mol. Evol.* **19** 87-103

Kuhnlein, U, Dawe Y, Zadworny D and Gavora J S 1989 DNA fingerprinting: a tool for determining genetic distances between strains of poultry; *Theor. Appl. Genet.* **77** 669-672

_____, Zadworny D, Dawe Y, Fairfull R W and Gavora J S 1990 Assessment of inbreeding by DNA fingerprinting: Development of a calibration curve using defined strains of chickens. *Genetics* **125** 161-165

Lagoda P J L, Seitz G, Epplen J T and Issinger O G 1989 Increased detectability of somatic changes in the DNA after probing with "synthetic" and "genome derived" hypervariable multilocus probes; *Hum. Genet.* **84** 35-40

Lawrence, C B, McDonnell D P and Ramsey W J 1985 Analysis of repetitive sequence elements containing tRNA-like sequences; *Nucl. Acid. Res.* **13** 4239-4252

Levinson G, Marsh J L, Epplen J T and Gutman G A 1985 Cross hybridizing snake satellite, Drosophila and mouse DNA sequences may have arisen independently; *Mol. Biol. Evol.* **2** 494-504

_____ and Gutman G A 1987 Slipped strand mis-pairing: a major mechanism for DNA sequence evolution; *Mol. Bio. Evol.* **4** 203-221

Maeda N and Smithies O 1986 The evolution of multigene families: Human haptoglobin genes; *Ann. Rev. Genet.* **20** 81-108

Mariat D and Vergnaud G 1992 Detection of polymorphic loci in complex genomes with synthetic tandem repeats; *Genomics* **12** 454-458

Mattapallil M J and Ali S 1997 Parentage assessment of an IVF calf from *Bubalus bubalis* by DNA fingerprinting; *Gene* **206** 209-214

_____ and _____ 1999 Analyses of conserved microsatellite sequences suggest closer relationship between water buffalo *Bubalus bubalis* and sheep *Ovis aries; DNA and Cell Biology* 18 513-519

Meuth M, Nalban-talfu J, Pheaar D and Miles C 1987 Molecular basis of genome rearrangements at the hamster *aprt* locus, In: *Banbury report*, 28, Mammalian cell mutagenesis, Cold Spring Harbor Laboratory, p 183

Mitani K, Takahashi Y and Kominami R 1990 A GGCAGG motif in minisatellites affecting their germline instability; *J. Biol. Chem.* **265** 15203-5210

Montgomery G W, Crawford A M, Penty J M 1993 The ovine Booroola fecundity gene (FeeB) is linked to markers from a regions of human chromosome 4q; *Nature Genet.* **4** 410-414

Morin, P A, Moore J J, Chakraborty R, Jin L and Woodruff D S 1994 Kin selection, social structure, gene flow and the evolution of chimpanzees; *Science* **265** 1193-1201

Moum T, Johansen S, Erikstad K E and Piatt J F 1994 Phylogeny and evolution of the auks (sub family Alcinae) based on the mitochondrial DNA sequences; *Proc. Natl. Acad. Sci. USA.* **91** 7912-7916

Murata S, Takasaki N, Saitoh M and Okada N 1993 Determination of the phylogenetic relationships among pacific salmonids by using short interspersed elements (SINEs) as temporal landmark of evolution; *Proc. Natl. Acad. Sci. USA.* **90** 6995-6999

Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E and White R 1987 A variable number of tandem repeat (VN TR) markers for human gene mapping; *Science* **235** 1616-1622

Nurnberg P, Zischler H, Fuhrmann E, Thiel G, Losanova T, Kinzel D, Nisch G, Witkowski R and Epplen J T 1991 Co-amplification of simple repetitive DNA fingerprint fragments and the EGF receptor gene in human gliomas; *Genes Chromosome Cancer* 3 79-88

O'Brien S J, Nash W G, Wildt D E, Bush M E and Benvieste R E 1985 A molecular solution to the riddle of the Giant Panda's phylogeny; *Nature* **317** 140-144

Orita M, Suzuki Y, Sekiya T ad Hayashi K 1989 A rapid and sensitive detection of point mutations and genetic polymorphisms using PCR; *Genomics* **5** 874-879

Orti G, Pearse D E and Avise J C 1997 Phylogenetic assessment of length variation at a microsatellite locus; *Proc. Natl. Acad. Sci. USA.* **94** 10745-10749

Packer C, Gilbert D A, Pusey A E and O'Brien S J 1991 A molecular genetic analysis of kinship and co-operation in African lions; *Nature* **351** 562-565

Peltomaki P, Aaltonen L A, Sistonen P, Pylkkanen L, Mecklin J, Jarvinen H, Green J S, Jass JR, Weber JL, Leach FS, Petersen GM, Hamilton SR, Chapelle A and Vogelstein B 1993 Genetic mapping of a locus predisposing to human colorectal cancer; *Science* **260** 810-812

Richards R I and Sutherland G R 1992 Heritable unstable DNA sequences; *Nature Genet.* **1** 7-9

Riedy M F, Hamilton W J and Adquadro C F 1992 Excess of non-paternal bands in offspring from known primate pedigree assayed using RAPD PCR; *Nucl. Acid. Res.* **20** 918

Rogers J 1985 Origins of repeated DNA; *Nature* **317** 765-766

Rothuizen J and Van Wolferen M 1994 Randomly amplified DNA polymorphism in dogs are reproducible and display mendelian transmission; *Anim. Genet.* **25** 13-18

Schafer R, Botz E, Becker A, Bartels F and Epplen J T 1986 The expression of the evolutionarily conserved GATA/GACA repeats in mouse tissues; *Chromosoma* **93** 496-501

Schimke R T 1984 Gene amplification in cultured animal cells; *Cell* 37 705-713

Scott M P, Haymes K M and Williams S M 1992 Parentage analysis using RAPD PCR; *Nucl. Acid. Res.* **20** 5493

Singh L, Philips C and Jones K W 1984 The conserved nucleotide sequences of Bkm which define sxr in the mouse are transcribed; *Cell* **36** 111-120

Smith G P 1978 What is the origin and evolution of repetitive DNA? *Trends Biochem. Sci.* 3 N34-N36

Steinmetz M, Stephen D and Lindahl K F 1986 Gene organization and recombinatorial hotspots in the murine major histocompatibility complex; *Cell* **44** 895-904

Sulaiman I M and Hasnain S E 1996 Random amplified polymorphic DNA markers reveal genetic homogeneity in the endangered Himalayan species *Meconopsis paniculata* and *M. simplicifolia.; Theor. Appl. Genet.* **93** 91-96

Sutcliffe J G, Milner R J, Gottesfeld J M and Lerner R A 1984 Identifier sequences are transcribed specifically in the brain; *Nature* **308** 237-241

Tautz D and Renz M 1984 Simple sequences are ubiquitous repetitive components of eukaryotic genomes; *Nucl. Acid. Res.* **12** 4127-4138

_____, Trick M and Dover G A 1986 Cryptic simplicity in DNA is a major source of genetic variation; *Nature* **322** 652-656

_____1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers; *Nucl. Acid. Res.* **17** 6463-6471

Thein S L, Jeffreys A J, Gooi H C, Cooter F, Flint J, O'connor N T J, Weatherall D J and Wainscoat J S 1987 Detection of somatic changes in human cancer DNA by DNA fingerprinting analysis; *Br. J. Cancer.* **55** 353-356

Thibodeau S N, Bren G and Schaid D 1993 Microsatellite instability in cancer of the proximal colon; *Science* **260** 816-819

Threadgill D W and Womack J E 1990 Genome analysis of the major bovine milk protein genes; *Nucl. Acid. Res.* **18** 6935-6942

Vergnaud G 1989 Polymers of random short oligonucleotide detect polymorphic loci in the human genome; *Nucl. Acid Res.* **17** 7623-7630

Vogel W, Steinbach P, Djalali M, Mehnert K, Ali S and Epplen J T 1988 Chromosome 9 of Ellobius lutescens is the X chromosome; *Chromosoma* **96** 112-118

Vogt P 1990 Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved "chromatin folding code"; *Hum. Genet.* **84** 301-336

Vos P, Hogers R, Bleeker M, Reijans M, van de lee T, Hornes M, Frijters A, Pot J, Peleman J and Kuiper M 1995. AFLP: a new technique for DNA fingerprinting; *Nucl. Acid. Res.* **23** 4407-4414

Wallace R B, Petz L D and Yam P Y 1986 Amplification of synthetic DNA probes to the analysis of DNA sequence variants in man, *Cold Spring Harbor Symposia Quantitative Biology,* **LI** 257-261

Wang A H, Quigley G J, Kolpak F J, Crawford J L, van-Boom J H van-der-Marel G and Rich A 1979 Molecular structure of a left handed double helical DNA fragment at atomic resolution; *Nature* **282** 680-686

Wang G, Whittam T S, Berg C M and Berg D E 1993 RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains; *Nucl. Acid. Res.* **21** 5930-5933

Washio K, Misawa S and Ueda S 1989 Probe walking: Development of novel probes for DNA fingerprinting; *Hum. Genet.* **83** 223-226

Weiner A M, Deininger P L and Efstratiadis A 1986 Non-viral retroposons: Genes, Pseudogenes and transposable elements generated by the reverse flow of genetic information; *Ann. Rev. Biochem.* **55** 631-661

Weising K, Fiala B, Ramloch K Kahl G and Epplen J T 1990 Oligonucleotide fingerprinting in angiosperms, *Fingerprint News* **2** 5-8

Welsh J and McClelland M 1990 Fingerprinting genomes using PCR with arbitrary primers; *Nucl. Acid. Res.* **18** 7213-7218

_____, Petersen C and McClelland M 1991 Polymorphisms generated by arbitrarily primed PCR in the mouse application to strain identification and genetic mapping; *Nucl. Acid. Res.* **19** 303-306

Williams J G K, Kubelik A R, Livak K J, Rafalski J A, Tingey S V 1990 DNA polymorphism amplified by arbitrary primers are useful as genetic markers; *Nucl. Acid. Res.* **18** 6531-6535

Wright J M 1994 Mutations at VNTRs: are minisatellites the evolutionary progeny of microsatellites? *Genome* **37** 345-347

Yam P, Petz L D, Ali S, Stock A D and Wallace R B 1987 Development a single probe for documentation of chimerism following bone-marrow transplantation; *Am. J. Hum. Genet.* **41** 867-881

Zimmerman S B 1982 The three dimensional structure of DNA; *Ann. Rev. Biochem.* **51** 395-427