REVIEW OF LITERATURE
2.0 REVIEW OF LITERATURE

2.1 Repeat sequences and their classification 6

2.2 Origin of interspersed repeats 7

2.3 Origin of minisatellites 7

2.4 Origin of microsatellites 8

2.5 Probable functions of repeat sequences 9

2.6 Evolution and distribution of repeat sequences in the population 10

2.7 Hypervariability at repeat loci 11

A) Minisatellites 11
B) simple sequence repeats 12

2.8 Conventional RFLP vs repeat sequence hypervariability 12

2.9 Simple repeat oligonucleotides as DNA fingerprinting probes 13

2.10 Reliability of synthetic oligos as DNA fingerprinting probes 14

A) Germline mutations 15
B) Somatic mutations 17
2.11 Improved techniques for detecting repeat sequence variability

A) Ligated oligonucleotide probe (LOP)  
B) Polymerase chain reaction (PCR)  
C) Randomly amplified polymorphic DNA (RAPD)

2.12 Population structuring using DNA fingerprinting technique

A) Conservation genetics and behavioural ecology  
B) Phylogenetic analysis  
C) Application of DNA fingerprinting approach in breeding programs

2.13 Other applications of synthetic probes based on repeat sequences

2.14 Ultravariability at repeat loci
REVIEW OF LITERATURE

During the last decade, considerable progress has been made in divulging the genome organisation of different species by using DNA probes based mostly on repetitive DNA. This encompasses the individualisation of genome to assess the genetic structure of populations and reproductive behaviour 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.

2.1 Repeat sequences and their classification:

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

It has been hypothesised that the 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 and reverse transcription by RNA polymerase III (to generate SINEs) and RNA polymerase II (to generate LINEs). These sequences then undergo amplifications to form several sub-families of SINEs and LINEs.

Alu repeats are the most abundant interspersed repetitive DNA sequences in eukaryotes. The first Alu like family, that was identified as descended from tRNA, is the C family of artiodactyles in cow and goat (Rogers, 1985; Lawrence et al., 1985). This family has a conserved sequence that is 65% homologous with several tRNA which can theoretically 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 a given subfamily is usually present only in a moderate number of related species (Deininger and Daniels, 1986). This is due to the independent formation of these repeat families after the divergence of various mammalian orders. These repeats then undergo a very high degree of gene conversion within a species and maintain low level of intraspecies divergence.

2.3 Origin of minisatellites:

Unlike SINEs and LINEs, the tandemly arrayed repeat sequences originate due to "random unequal crossover" events between duplicated sequences (Smith, 1976;
1978), slipped strand mispairing (Kornberg et al., 1964; Streisinger et al., 1966; Tautz and Renz, 1984; Levinson et al., 1985) and aberrant in situ replication (Schimke, 1982). Amongst these mechanisms, it is suggested that minisatellite repeat loci evolve mainly through unequal cross over mechanism (Jeffreys et al., 1985a; Levinson and Gutman, 1987). The random unequal crossover mechanism is envisaged to initiate as a result of rare, illegitimate recombination events that occur 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 at these regions generates one chromatid with a tandemly duplicated segment and another with a deletion for the same region, producing a mutant allele length. 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, where, homologous recombination between related sequences of the tandem arrays takes place (Hardman, 1986). These unequal crossovers, then result in random increase (duplication) 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 and Wells, 1989).

2.4 Origin of microsatellites:
Microsatellites consist of stretches of monotonously repeated short nucleotide motifs which occur as interspersed elements. Almost all permutations of mono-, di-, tri- and tetrancleotide motifs can be found as building blocks of these simple sequences.
Review of literature

The processes which play a role in the evolution of simple repeat sequences are gene conversion (Maeda and Smithies, 1986) and replication slippage (Levinson and Gutman, 1987). Among this, the most common mechanism in the evolution of microsatellites (short tandem repeats) is the strand slippage during the course of DNA replication. The first systematic in vivo studies on slippage at dinucleotide repeat loci were conducted by Levinson and Gutman (1987), who introduced simple sequence stretches with GT/CA motifs into the multiple cloning site of M13 phage. From these experiments, it became clear that slippage can also occur in vivo between short direct repeats, leading to deletions/expansions of the motifs (Efstratiadis et al., 1980; Jones and Kafatos, 1982; Brown and Piechaczyk, 1983).

2.5 Probable functions of repeat sequences:
The precise role of these repeat sequences in the eukaryotic genome is still unclear. In recent studies, repeat sequences have been implicated in gene regulation (Weintraub and Groudine, 1976; Wang et al., 1979; Hentschel, 1982), replication of telomeres (Blackburn and Szostak, 1984), selection and transport of mRNA molecules to the cell cytoplasm (Davidson and Britten, 1979, Sutcliffe et al., 1984) and as signals for gene conversion and recombination (Shen et al., 1981; Goodman 1985; Jeffreys et al., 1985a). Also, it has been suggested that alternating purine and pyrimidine sequences such as (GT)$_n$ or (CA)$_n$ form left-handed Z-DNA (Zimmerman, 1982). Although not proven, it is easy to imagine that these could perform a structural role and that other tandem repeats might also adopt unusual structures which could again be functionally important (Vogt, 1990). Length expansions of simple sequence stretches composed of trinucleotide repeats have been found to be associated with a number of human genetic diseases (Caskey et al., 1992; Bates and Lehrach, 1994; Richards and Sutherland, 1994). It
now appears that repeat sequence mutation is a common cause of human disease particularly of those disorders which follow dominant inheritance. Employing northern blot analysis, it has been shown that some repetitive sequences do transcribe (Epplen et al., 1982; Singh et al., 1984; Schafer et al., 1986).

2.6 Evolution and distribution of repeat sequences in the population:
Mechanisms of continuous gain or loss of DNA repeats by unequal crossing over and gene conversion can lead to homogenisation (molecular drive) of any given variant in a sexual population (Dover, 1982; Coen et al., 1982; Dover and Flavell, 1984; Dover, 1986). During the course of evolution of minisatellites by unequal crossing over, some variants will be lost whereas others will increase in frequency, eventually replacing all others. Such a genetic drift in the alleles of a repeat array 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, generation after generation, leading to an interspecies variation known as "concerted evolution" of minisatellite sequences (Dover, 1982). These evolutionary changes occurring at the minisatellite regions lead to 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 interspecific variations.
At this juncture, it is interesting to note that no large scale homogenisation occurs by gain or loss due to slippage (Tautz et al., 1986), which is the mechanism involved in microsatellite repeat evolution. This is due to the fact that the unit of slippage can occasionally be out of phase with pre-existing motifs. Thus, any homogenisation is continually being obliterated by the generation and combinatorial reshuffling of short and short-lived motifs differing in sequence.
complexities and length. 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 is very high if these sequences are homogeneous and hence the level of polymorphism at these loci is also high. Similarly, the degree of variability increases with the increase in length of the repeat units since it paves the way for an increased chance of misalignment and unequal cross-overs. Thus, within a population, allele size variability increases with an increase in the rate of unequal cross-over (Jarman and Wells, 1989).

2.7 Hypervariability at repeat loci:

A) Minisatellites:

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 of the genomic average, in order to maintain the degree of polymorphism. Such frequent unequal crossovers occurring at the minisatellite regions leads to generation of variability in the number of repeats in the tandem arrays referred to as Variable Number Tandem Repeats (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, it was demonstrated (Jeffreys et al., 1985a) that a minisatellite probe from the myoglobin gene cross-hybridised to many other minisatellites (because of similarities in their core sequences), giving rise to a complex but heritable pattern dubbed as "DNA fingerprint" (Jeffreys et al., 1985b). The cumulative high degree
of allelic variability so detected at minisatellite regions allows DNA fingerprinting technology to be used in identifying individuals or clones of cells, with a variety of applications in forensics, immigration laws, paternity testing, transplant screening and ecological genetics (Hill, 1987). This aspect has also been covered in the introduction.

The hypervariable minisatellites are not confined to the human genome. Fingerprint patterns of similar complexity have been detected in other mammals, birds, higher plants, fungi and protozoa using the same minisatellite probes derived from human myoglobin gene (Rogstad et al., 1989).

B) Simple sequence repeats:

The internal genomic mechanism of producing and deleting simple sequence repeats (SSRs) (Tautz and Renz, 1984; Tautz et al., 1986), predicts that these stretches should be hypervariable in length and might serve as a polymorphic DNA markers for genome mapping and linkage studies (Tautz, 1989).

2.8 Conventional RFLP vs repeat sequence hypervariability:

Human genome contains roughly 3 X 10^7 DNA sequence variants (Jeffreys, 1979). The gain or loss of specific restriction endonuclease cleavage sites in the genome of individuals helps in detecting these variants in the form of restriction fragments with length polymorphisms (Botstein et al., 1980). But the variability detected by the restriction enzymes is only perhaps 1% of the vast pool of genetic variability (Jeffreys, 1979). The reasons for this limitation are: 1) only a small fraction of the nucleotide sequences are known to have the restriction enzyme cleavage sites (Kessler et al., 1985), 2) the resolution of the RFLP approach is such that small differences in fragment length can be overlooked and, 3) sequence variants present in the repetitive DNA are not easily analysed by conventional
probes. Usually the loci detected by RFLPs are diallelic and thus have limited polymorphic information content (PIC). Because of their limited degree of heterozygosity, diallelic RFLP's are not highly informative with respect to genome individualization or linkage analysis.

2.9 Simple repeat oligonucleotides as DNA fingerprinting probes:
DNA fingerprinting was first demonstrated by genome derived cloned probes (Jeffreys et al., 1985a; 1985b). 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, for DNA fingerprinting, several synthetic oligo probes useful to a large number of species have been reported (Epplen, 1988). Cloning and propagation of the repeat fragments often pose technical problems because they are generally unstable in the prokaryote host and often recombine during propagation of the clones. Moreover, a majority of such naturally existing tandem repeats do not detect related structures due to lack of complete cross-hybridisation which is length dependent (Mariat and Vergnaud, 1992). This problem is easily circumvented by using the synthetic oligonucleotide probes which cover even much larger pool of genetic variability.

The advantage of using synthetic tandem repeat oligonucleotides as probes for detecting allele length variations is that, under appropriate hybridisation conditions, they show absolute specificity. Oligonucleotide probes are absolutely specific since perfectly matched duplexes are formed with the target DNA during the course of hybridisation (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 revealing maximum number of variant alleles was first attempted using different repeat unit lengths of (GATA)$_n$ on the human
Review of literature

genome (Ali et al., 1986). The result showed that an increase in the probe length due to duplex mismatch reduced the overall hybridisation. In subsequent studies, additional probes based on 2-6 base repeat motifs were used for hybridisation with genomic DNA from different sources (Epplen, 1988; Nanda et al., 1988; 1990; Vogel et al., 1988; Arnemann et al., 1989; Weising et al., 1989; 1990; 1991; Buitkamp et al., 1991a; 1991b). These repeats include (AT)$_n$, (CA)$_n$, (CT)$_n$, (CAC)$_n$, (GAA)$_n$, (TCC)$_n$, (GACA)$_n$, (GATA)$_n$, (GGAT)$_n$, (GGGCA)$_n$, and (TTAGGG)$_n$ (Epplen et al., 1991). 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 patterns in different species with respect to bands and signal intensity. The information content of a DNA probe is dependent on the average number of discernible polymorphic bands per individual and the respective band frequencies detected by the probe (Jeffreys et al., 1985b).

Simple repetitive DNA shows organizational variability with respect to sequence complexity (base composition) and overall genomic distribution within the species. A single oligo or a set of probes informative for one species is not useful for all the species. Thus, for successful DNA fingerprinting of a species, a panel of different oligo probes needs to be evaluated, first for the polymorphic information content (Epplen et al., 1991). Extensive family studies on human and animal systems using synthetic repeat sequence probes have shown that bands detected by synthetic DNA sequence probes follow Mendelian inheritance and represent autosomal loci which segregate independently (Jeffreys et al., 1986).

2.10 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 Hinfl enzyme
Review of literature

is the most informative probe for human with a probability of $2.0 \times 10^{-8}$ for identical band profile between any two random individuals (Schafer et al., 1988b). Initial fingerprint analysis with the synthetic oligos $(GATA)_n$ and $(GACA)_n$ (Ali et al., 1986) detected less polymorphic variants as compared to $(CAC)_5$ (Schafer et al., 1988b). Hypervariability at repeat loci also reflects high rates of de novo mutations producing new length alleles. As expected from the neutral mutation-random drift hypothesis, mutation rate increases with variability and becomes significant above approximately 96% heterozygosity which is again dependent on the effective population size (Jeffreys et al., 1987b; 1988a). Mutations detected by these probes have direct relevance to forensic cases and legal applications. 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.

A) Germline mutations:

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 et al., 1985b), whereas with the highly hypervariable marker, lambda MS1, the same was found to be $0.052$ per gamete per locus (Jeffreys et al., 1988a). The mutation rate detected by oligo probe $(CAC)_5$ was well within the range of that estimated earlier for minisatellite probes (Schafer et al., 1988a; Nurnberg et al., 1989). 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 MO-1 minisatellite regions. They introduced single base substitutions in the core sequence of this minisatellite and using this as a probe, detected altered hybridisation patterns with a low rate of mutation. They concluded that the repeat
Review of literature

motif in the tandem array itself contributes to the germline instability. Mutation analysis in pedigrees is limited by the small number of mutant alleles that can be identified. Therefore, a 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 (Zhang et al., 1994). 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 of the tandem array (Jeffreys et al., 1994a). In these mutation events, which appeared to be largely germline specific (Jeffreys et al., 1988a), the flanking sequences did not take part and hence led to the speculation that the mutations could be 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 i.e. flanking DNA elements which possibly might be serving as mutation/conversion initiators (Jeffreys et al., 1994a; 1994b). Further attempts to define these mutation initiator elements are being carried out in transgenic mice harbouring human minisatellites (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 that their rate of occurrence is known and
Review of literature

can be incorporated into the statistical likelihood ratio analysis (Jeffreys et al., 1991b).

B) Somatic mutations:
Thein et al. (1987) and Thibodeau et al. (1993) reported changes in the DNA fingerprint pattern of the constitutional and tumour DNA from patients. These changes included alterations in the relative intensities of hybridising DNA bands and appearance of novel minisatellite 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 and Sutherland, 1992). Synthetic oligo probes used for fingerprinting in humans also detected such somatic instabilities. Both (CAC)$_5$ (Nurnberg et al., 1991) and telomeric repeat (TTAGGG)$_3$ (Hastie et al., 1990) showed expansions and contractions in their repeat lengths during the progression of glioma tumours (Lagoda et al., 1989).

To explain the somatic instability at repeat sequences several hypotheses have been proposed. Loss of chromosomes or chromosomal regions through deletion, mitotic non-disjunction or mitotic recombination would lead to loss of associated minisatellite fragments. Conversely localised amplification of DNA (Stark and Wahl, 1984, Schimke, 1984) including a minisatellite would cause specific band intensification. Also, tissue- or tumour-specific changes in DNA methylation affect DNA fingerprints (Goelz et al., 1985). Peltomaki et al. (1993) suggested that, in humans, a gene involved in DNA repair mechanism, located on chromosome 2 seems to be causing instability to the mono- di- and trinucleotide repeats in colon cancer cells.
2.11 Improved techniques for detecting repeat sequence variability:

A) Ligated oligonucleotide probe (LOP):

Besides single-stranded synthetic oligo probes, enzymatically ligated double-stranded probes may also be generated for specific purposes (Ali and 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 400bp that are then used at a higher hybridisation stringency.

The ligated oligonucleotide probe (LOP) was found to detect multiple polymorphicoci producing highly informative DNA fingerprint patterns, when compared to its unit component single-stranded conventional oligo probe (COP). LOPs are more easily generated than cloned VNTR probes (Vergnaud, 1989) and avoid problems associated with cloned probes including bacterial growth and maintenance as well as in-vitro labelling.

B) Polymerase chain reaction (PCR):

Amplifications of single locus minisatellites by PCR using specific flanking primers and mapping of repeated sequences within these amplified minisatellite alleles have also been used to produce individual specific fingerprints and species specific patterns (Heath et al., 1993). 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 the ability to type degraded human DNA (Jeffreys et al., 1988b; Boerwinkle et al., 1989; Litt and Luty, 1989; Tautz, 1989; Weber and May, 1989; Horn et al., 1989; Hagelberg
C) Randomly amplified polymorphic DNA (RAPD):

DNA amplification fingerprinting (DAF) is the enzymatic amplification of
arbitrary sequences directed by a very short single oligonucleotide primer to
generate complex but characteristic DNA fingerprints. It uses the most permissive
reaction conditions and a small oligo primer for amplification. Polyacrylamide
gel electrophoresis and silver staining can adequately resolve DAF product into
detailed and reproducible band patterns (Bassam et al., 1992). This amplification
fragment length polymorphism (AFLP) analysis has been proposed to be the
alternative targeting tool for genetic typing and mapping, useful particularly for
very small amounts of DNA (Williams et al., 1990; Welsh and McClelland, 1990;
Welsh et al., 1991; Caetano-Anolles et al., 1991a; 1991b).

The technique had also been proven to be useful in genome individualisation and
analysis of genetic relationships (Hedrick, 1992; Wang et al., 1993; Dinesh et al.,
1993), detecting inherited diseases (Rothuizen and Van Wolferen, 1994) and
identifying strains and constructing genetic maps (Johnson et al., 1994).

A drawback of RAPD is that the nature of the randomly primed template DNA
remains unknown. Though the methodology is sensitive enough to detect even
single nucleotide insertions, deletions or substitutions, a high level of
standardisation and precise internal control is necessary if the DNA profile
obtained has to be consistently reproducible, resolvable and scorabble (Rothuizen
and 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 and Page 1992; Ayliffe et al., 1994) and can be a potential source of artifactual polymorphism during RAPD analysis. Apart from the heteroduplex formation of the alternate alleles, the artifactual bands can also be due to the base composition of the oligonucleotide primers and hence each primer requires independent empirical optimisation of appropriate reaction conditions.

As the technique has achieved broad applications in the studies of genetic variations in natural populations, artifactual variations may lead to over-estimation of levels of variation if the band is interpreted as a true variation and hence the discrimination of true and artifactual variants is critical (Ellsworth et al., 1993) and should be looked into carefully while conducting the experiment.

2.12 Population structuring using DNA fingerprinting technique:
Evolutionary and population biologists need to be able to quantify genetic relationships among individual organisms at many different levels; from close familial relationships to evolutionarily distant phylogenetic ones.

A) Conservation genetics and behavioural ecology:
As mentioned earlier, DNA fingerprinting has proved to be an extremely powerful tool for paternity analysis (extra-pair paternity and maternity), estimation of relatedness and assessing kin selection in many wild animals and birds (Burke et al., 1991).

Parentage analysis using markers for hypervariable repeat loci (Burke and Bruford, 1987) has allowed the quantification of extra-pair paternity and conspecific nest parasitism in population studies of several species (Wetton et al., 1987; Birkhead et al., 1990; Gibbs et al., 1990; Rabenold et al., 1990; Westneat, 1990; Gullberg et al., 1992). Detailed observations of the reproductive behaviour of individual
Review of literature

birds and other animals have been made on the basis of DNA fingerprinting analysis. In one such study on 'Dunnock', DNA fingerprinting approach was used to ascertain whether the subordinate male simply fed the brood or whether he had actually fathered any of the nestlings (Burke et al., 1989). DNA fingerprinting approach has also been used in many other species to estimate relatedness. The naked mole rat *Heterocephalus glaber* is unique 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 (Reev et al., 1990).

Studies on kinship and social structure in the wild has also been carried out extensively using DNA fingerprinting approach. Minisatellite band sharing data of African lions reveal kinship within and between lion prides (Packer et al., 1991). As the coalition size increases, the proportion of coalitions with non-relatives drops sharply and the members within a coalition group become more homogenous.

Simple sequence repeat (SSR) polymorphisms at eight different loci were studied in the free ranging individuals of chimpanzee population in Africa (Morin et al., 1994). Pedigree relationships showed that the males were related in the order of half-siblings and homozygosity was significantly increased in a kin group proving the kin selection theory. DNA fingerprinting approach was used to study the gene flow and evolution of chimpanzee population (Morin et al., 1994). The West African *Pan troglodytes verus* was found to be a well-differentiated and independently evolving taxa, with a divergence time of about 1.58 million years, from the other two subspecies (*P. t. troglodytes* and *P. t. schweinfurthii*). This data has great implication in evolutionary and conservation biology.
B) Phylogenetic analysis:
Using polymorphic markers and various statistical methods, taxonomic status, phylogenetic relationships and genetic distances of different species have been studied (Sneath and Sokal, 1973; Felsenstein, 1984). Earlier, the markers used in solving the phylogenetic riddles included those that detect isozyme variants, amino acid substitutions in homologous proteins and karyological relationships (O'Brien et al., 1985). Homology at β-like globin genes, α-lactalbumin gene sequences (Easteal, 1990) and the fast evolving mitochondrial DNA sequences (Wenink et al., 1993; Moum et al., 1994) was also used for this purpose.

In order for the derived topology to be more precise, the loci under consideration should be extremly polymorphic and conserved in a wide range of species (Buchanan et al., 1994). Although the mitochondrial DNA sequences are polymorphic and evolve 5-10 times faster than the nuclear genes, they do not provide any information about the extent of nuclear gene flow or variability which is central to the evolution of the overall make-up of an organism. The hypervariable repetititve DNA families have been proven to be useful for studies on molecular evolution (Britten, 1986) since these sequences are distributed across eukaryotic genomes. They offer sufficient fidelity to distinguish populations that are seperated by a short evolutionary time of divergence from their common ancestry.

Interspecifically conserved short interspersed repetitive elements (SINEs) were used as probes 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 Oncorhyncus based on the SINE insertion analysis (Murata et al., 1993). SINEs appear to be inserted
Review of literature

irreversibly and should therefore provide ideal evolutionary and phylogenetic markers (Okada, 1991).

The stochastic population events like genetic drift and founder effect with respect to minisatellites often lead to concerted evolution (Dover, 1982). High levels of intrapopulation homogeneity was noticed in pup fish at Hind III satellite DNA sequences which is a useful marker to assess the stock structure of commercially exploited species (Elder jr 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 the qualitative and quantitative differences in the repeat copy number of Bkm sequences, the different genera of crocodilians were differentiated into two major groups.

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 (Litt and Luty, 1989; Weber and May, 1989). Since then they have been widely used for a variety of applications. More recently they have been used as linkage markers for gene mapping (Dietrich et al., 1994; Gyapay et al., 1994) and interpopulation studies (Edwards et al., 1992; Chakraborty et al., 1992). Using a set of 30 (CA)$_n$ repeat loci, reliable evolutionary relationships were established amongst 14 human populations (Bowcock et al., 1994). A significantly higher heterozygosity and number of alleles was noticed in the African population supporting the hypothesis of an African origin for humans. In another study, characterisation 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
Review of literature

microsatellite loci (Deka et al., 1995). Though there is a substantial variation in allele frequency at each locus, at a given locus, most frequent alleles were shared by all human populations. When these loci were used simultaneously, they detected population affinities.

Extensive genome analysis were also carried out to delineate the evolutionary relationships (Gilbert et al., 1990) and breed affiliations of economically important farm animals (Ellegren et al., 1992; Buchanan et al., 1994). Similarly, the DNA fingerprinting of the Channel Island fox populations 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). Geographically and hence reproductively isolated groups developed increased level of homogeneity, due to fixation of alleles in the population. Thus, the sequence variations at repeat loci could be used as markers for evolved breeds of a species.

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., 1992). Similar approach was also used in ovine species using microsatellite repeats. Here, highly significant differences were noticed in the allele frequencies between samples from different breeds (Buchanan et al., 1994). The genetic distances were estimated in the breeds based on the mean allelic frequencies at each repeat locus and from the knowledge on the generation interval, the approximate time of divergence of each breed was also calibrated with precision (Chakraborty, 1994).

C) Application of DNA fingerprinting approach in breeding programs:

The vast amount of information made available by DNA fingerprinting could also
be exploited for the breeding programmes of farm animals/birds. Kuhnlein et al., (1990) measured fingerprint similarities among experimental chicken lines of known pedigree and demonstrated a close and linear relationship between the mean band frequency and the mean inbreeding coefficient. The genetic relationships between different breeding populations of chicken were also analysed using minisatellite probes to delineate their genetic distances (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 (Kuhnlein et al., 1989) were proposed. Kuhnlein et al. (1989) also proposed a linear model for analysing the quantitative traits as dependent variables and fingerprint patterns as independent variables.

2.13 Other applications of synthetic probes based on repeat sequences:
Repeat sequences have also found application in monitoring success of bone marrow transplantations (BMT) (Yam et al., 1987). Using synthetic oligos complementary to tandemly repeated sequences, Yam and coworkers have documented the donor marrow engraftment, mixed hematopoietic chimerism, patient's pre-BMT phenotype and the origin of malignant hematopoietic cells in the patients who developed recurrent hematologic malignancy following BMT. Remarkable progress has recently been made in genetic linkage mapping both in human and animal species. A 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 genomes of almost all species. Microsatellite mapping of the gene responsible for weaver
disease in cattle which segregates with increased milk production, identified a closely linked marker that allowed selection against weaver disease and was used to characterise 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 (1, 6, 9, 10 and 20) giving very strong evidence for the loci controlling milk production (Georges, 1994).

The gene responsible for multiple births, *Booroola* fecundity gene (*FecB*) of sheep has been mapped to chromosome 6 using microsatellite markers (Montgomery et al., 1993).

Similarly, using a battery of these polymorphic satellite sequences in "Probe Walking", a complete mapping of genomes can be achieved at the earliest (Washio et al., 1989). The markers so developed, can thus be used in Marker Assisted Selection (MAS) and breeding programs of livestock.

**2.14 Ultraviability at repeat loci:**

Error-prone allele length estimates and electrophoretic "band shifts" can occasionally lead to apparent exclusions between 'matching' DNA profiles and greatly weaken the potential statistical power for investigating DNA profile databases (Lander, 1989; Budowle et al., 1991). 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 (D1S8) using oligonucleotides with complementary sequences to repeat units (MS32) as primers, aliquots were end labelled and then digested with HinfI or HaeIII 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 of the allele, another approach of assaying these sequence variations in minisatellite alleles was developed (Jeffreys et al., 1991a). This technique, MVR (minisatellite variant repeat) 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 detected reveals the direct evidence of minisatellite instability due to interallelic recombination/gene conversion.