MICROSATELLITE DNA POLYMORPHISMS IN THE PEOPLE OF NORTH KERALA

THESIS SUBMITTED TO THE UNIVERSITY OF CALICUT
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN ZOOLOGY UNDER THE FACULTY OF SCIENCE

By

NARAYANAN MATHATH M.Sc.
DEPARTMENT OF ZOOLOGY, UNIVERSITY OF CALICUT, KERALA

DECEMBER, 2004

DEPARTMENT OF ZOOLOGY
UNIVERSITY OF CALICUT

Dr. K.V. Lazar, Ph.D.
Lecturer

Phone : 0494 2401144 - *419
CALICUT UNIVERSITY P.O.
KERALA 673 635

Date: 23.12.2004.

CERTIFICATE

This is to certify that this thesis is an authentic record of work carried out by Mr. Narayanan Mathath, under my supervision and guidance in partial fulfilment of the Degree of Doctor of Philosophy under the Faculty of Science of the University of Calicut. No part of this thesis has been presented before for any other degree.

Dr. K.V. Lazar

ACKNOWLEDGEMENTS

I have great pleasure to express my sincere gratitude to Dr. K.V. Lazar, Lecturer, Department of Zoology, University of Calicut, for his guidance.

I express my sincere thanks to Dr. U.V.K. Mohamed, Professor and former Head of the Department of Zoology, for providing me the necessary facilities while initiating the present work and for his constant encouragements.

I am thankful to the Head of the Department of Zoology, for providing me necessary facilities in the Department to complete this work.

I am grateful to Dr. K.P. Genesan, Department of Botany, University of Calicut, for his valuable advices and constant encouragements throughout the tenure of this work.

I am indebted to my family for their help and encouragements. To my wife and children, I owe my gratitude and apologies for their sufferings who patiently put up with my absences and preoccupations and without their help and tolerance this could not have been fulfilled.

I wish to extend my gratitude to my nephew, Dr. Dinesan Mathath, and my friends, who have helped a lot in collecting the samples for this work. Also thanks have been extended to Mr. Kunhikrishnan Thottathan, Koothuparamba for assisting in some of the statistical calculations and Mr. Unnikrishnan Nair, C.V, Kottayam for providing some tutorial assistance in the operation of Microsoft Word and Excel.

I wish to record my gratitude to my colleagues and all friends for their help and co-operation to complete this work.

Narayanan Mathath

CONTENTS

6

FOREWORD

Human haploid genome is made up of about 3 billion base pairs. Several of these base pairs experience mutations in each generations, and, as the way in which these mutations accumulate in populations are influenced by how populations expand, contract, split and merge. The study of genetic variation has the potential to yield a great deal of information regarding our history and evolutionary relationship. These genetic variations are seen more frequently in the polymorphic regions of the DNA; an analysis of these polymorphisms is a quick and cost effective way of determining the identity of each individual in a population, their history and evolutionary relationship. Information obtained on the genotyping can be used in resolving ambiguities in evolutionary relationship based currently on other methods such as paleontological, archeological, anthropological and radiocarbon studies. The present study aims at delineating the genealogy of the people of North Kerala and reconstructing their phylogenetic relationship.

1 INTRODUCTION

Origin of the investigation

Paleontology and archaeology are disciplines that traditionally deal with the construction of human origins and history. They address questions such as when and where our species first emerged, how our ancestors spread over the globe and what the major events occurred in their history. Recently, however, molecular genetics has come to make increasing contributions to their area. In particular several data sets indicate that variation of the human gene pool originated in Africa within the last 200,000 years (Armour et al., 1996; Cann et al., 1987). Furthermore, the study of DNA sequences allows the detection or expansions of population size. Using the same techniques it is possible to reconstruct the history of population, their origin and relationship precisely and unambiguously as in the case of parental diagnosis and forensic analysis.

Objectives

1. *To evaluate the relationship between the people of Kasargod, Kannur, Wayanad and Kozhikode comprising the districts of Kannur, Kozhikode and the social groups of Mukkuva and Paniyan.*

2. *To delineate their phylogenetic relationships.*

Background

Genotyping human population

Molecular biology is now a major source of quantitative and objective information about the evolutionary history of human species. It has provided new insights into our genetic divergence from apes and into the way in which humans are related to one another genetically. For example, studies on mitochondrial DNA (mtDNA) from large numbers of people from different ethnic groups representing five broad geographical regions (sub-Saharan Africans, Asians, Caucasians, aboriginal Australians and aboriginal New Guineans) suggest the African origin for present-day humans (Templeton, 1992). Researchers reconstructed a hypothetical 'ancestral'

human mitochondrial DNA and suggested the existence of a mitochondrial 'Eve,' who, they calculate, could have lived in Africa some 200,000 years ago and from whom we are all descended (Cann et al., 1987; Vigilant et al., 1991). Similarly the questions of large-scale human migration into new world by comparative analysis of modern and ancient native American Indians was resolved by genetic profiling of DNA from both within population as well as continental-wide analyses (Edwards et al., 1992).

Human DNA polymorphisms

More recently a wide variety of polymorphic loci in human chromosomes have been used to investigate the extent of human genetic diversity and delineate the relationships between modern human populations (Tishkoff et al., 1996). In addition to mitochondrial variation, nuclear loci used include: classical blood group and serological markers, RFLPs, microsatellites, retrotransposon insertions and haplotypes of closely linked polymorphisms. When different alleles exist in different populations, additional information on population relationships can be obtained if cladistic information can be used to define the evolutionary relationship among the alleles.

Microsatellite repeat variations

Simple tandemly repeated sequences, or microsatellites, are ubiquitous in the genomes of a wide range of organisms, and the number of repeats within many of them is highly variable in the population of a particular species. The advent of the polymerase chain reaction provided the means for rapid and cost effective analysis of the repeat number, and several groups showed that these sequences were likely to provide a rich source of very informative markers for genotyping (Armour et al., 1996; Weber and May, 1989). By virtue of their extensive length variation between different alleles, they have found many applications in genetic analysis, including the establishment of individual identity and family relationships, parental diagnosis and other forensic analysis. The high level of population variability at these polymorphic loci are due to a high rate of germline mutation to new allelic states, at frequencies (up to 15% per gamete) high enough to measure by direct observation in pedigrees and single molecule analysis of germline DNA (Jeffreys et al., 1985,1986).

Rationale

Several researchers have conducted, approximately similar kinds of studies on various other areas. In the past, most of them are through traditional (Thurston and Rangachari, 1905) and conventional (Benciolini *et al.*, 1985; Gaensslen *et al.*, 1987a,b,c; Murch and Budowle, 1986; Roy, 1990; Sharma *et al.*, 1987) ways. Currently, several polymorphic markers as well as highly advanced and comprehensive techniques are available in genotyping and the same has been used in the present study. The short tandem repeats are the latest and have the utmost discriminating power ever obtained for a single system (Agrawal *et al.*, 2002; Bell *et al.*, 2000; Budowle *et al*; 1999; Edwards *et al.*, 1991a,b,1992; Estoup *et al.*, 1995a,b, 1996; Harbison *et al.*, 2002; Paetkau and Strobeck, 1994; Pfitzinger *et al.*, 1995; Polymeropoulos *et al.*, 1991; Tomas *et al.*, 2001; Weber and May, 1989). As the size range of the microsatellite alleles are amenable to PCR amplification, it can be analyzed in just a few hours. Tandemly repeated DNA sequences are widespread throughout the human genome and show sufficient variability among individuals in a population that they have become important in several fields including genetic mapping, linkage analysis, and human identity testing. Several researchers found a high discrimination power and power of exclusion in their works with STRs (Busque *et al.*, 1997; Corte-Real *et al.*, 1999a,b; Pu *et al.*, 1998).

Approach to the problem

Blood samples were collected from random, unrelated individuals from Kasargod, Kannur, Wayanad and Kozhikode districts along with samples from Mahe, a place under govt. of Pondicherry but surrounded by Kerala, representing North Kerala. Buccal swabs or hair were otherwise collected, where people are reluctant to provide blood.

Most of the Kerala coastal area is inhabited by a community of the fishermen, called *Mukkuva* belonging to the religion Hindu. The high altitude, hilly Wayanad district of North Kerala, has *Paniyan* one of the *Adivasi* tribes (Thurston and Rangachari, 1905) traditionally workers in agricultural farms. Special emphasis was given while collecting the samples on these two for the comparative study. The

samples were genotyped by the following nine documented, well established STR loci – CSF1PO, TPOX, THO1, F13A01, FESFPS, vWA, D16S539, D7S820 and D13S317, by following relevant methodologies from various sources. (PROMEGA GenePrint STR Systems (Silver Stain Detection) – Technical Manual, 2001; Sambrook and Russell, 2001; Walsh et al., 1991).

As an overall turnover, the allele size of each locus in subjects under study were compared and analysed through biostatistics using software and manually. The data were evaluated and a phylogenetic tree was constructed.

Presentation of the material

The dissertation commences with a foreword and introduction. The introduction goes through a brief account of the various aspects of DNA polymorphisms followed by the origin of the work, objectives, background, rationale, approach to the problem and presentation of the material. This is followed by four sections comprising review of the literature, materials and methods, results and discussion and conclusions. The dissertation ends with bibliography.

2 REVIEW OF LITERATURE

Polymerase chain reaction (PCR)

Polymerase Chain Reaction (PCR) unveiled at the American Society of Human Genetics Conference in October 1985 (Innis *et al.*, 1992), is an ingenious new tool for molecular biology that has had an effect on research similar to that of the discovery of restriction enzymes and the Southern blot. PCR is so sensitive that a single DNA molecule has been amplified, and single-copy genes are routinely extracted out of complex mixtures of genomic sequences and visualized as distinct bands on agarose gels (Innis *et al.*, 1992). Kary Mullis, a Cetus corporation scientist discovered the PCR, nick named as 'peoples choice reaction' (Gupta, 1999).

Subsequent to the invention of PCR, **Taq DNA polymerase**, the enzyme used in PCR was chosen as the molecule of the year 1989 and its inventor was awarded the Nobel Prize in 1993 (Inman and Rudin, 1997). Enhancements, such as the use of thermostable DNA polymerases and automation of the method invented by Kary Mullis (Erlich, 1989; Innis *et al.*, 1988; Mullis and Faloona 1987; Mullis *et al.*, 1986; Saiki *et al.*, 1985; Saiki *et al.*, 1988) have fostered the development of numerous and diverse PCR applications throughout the research community.

Long PCR

PCR amplification technologies have been continuously improving since the first methods were reported in 1987 (Mullis and Faloona, 1987). PCR reactions generally apply to produce amplifications of short regions, but later advances allow for the amplification of long regions of genomic DNA up to 20 kb in length (Barnes, 1994; Cheng *et al.*, 1994a; Cheng *et al.*, 1994b). Also amplification of DNA of length 42 kb has been reported elsewhere. Long PCR uses an enzyme mixture of Taq DNA polymerase and a thermostable proofreading polymerase under optimized conditions for the amplification of these long products. Long PCR based analyses are addressed in several literature (Kishida *et al.*, 1995; Kishida *et al.*, 1996; Richie *et al.*, 1999). Special PCR enzymes or commercial kits and methods of amplification for longer length DNA are available.

Applications

The number of applications of PCR seems infinite, and is still growing. They include direct cloning from genomic DNA or cDNA, in vitro mutagenesis and engineering of DNA, genetic fingerprinting of forensic samples, assays for the presence of infectious agents, prenatal diagnosis of genetic diseases, analysis of allelic sequence variations, analysis of RNA transcript structure, genomic footprinting, and direct nucleotide sequencing of genomic DNA and cDNA (Ausubel *et al.*, 1999). Several analogous techniques of PCR are developed to meet various research applications. Some of the wide range of techniques are AFLP (Amplified Fragment Length Polymorphism), Alu-PCR, Asymmetric PCR, Colony PCR, DD-PCR (Differential Display – PCR), Degenerate PCR, Hot-Start PCR, In situ PCR, Inverse PCR, Multiplex PCR, Nested PCR, PCR-ELISA (PCR-Enzyme Linked Immunosorbent Assay), PCR-RFLP(PCR-Restriction Fragment Length Polymorphism), PCR-SSCP(PCR-Single Strand Conformation Polymorphism), QC-PCR(Quantitative Competitive-PCR), RACE (Rapid Amplification of cDNA Ends), RAPD (Random Amplified Polymorphic DNA), RT-PCR (Reverse Transcription-PCR), Touchdown –PCR and Real Time PCR.

Human genome project

As a vision for the future of genomics research, it is appropriate to consider the remarkable path that has brought us here. The landmark discoveries in genetics and genomics, begin with Gregor Mendel's discovery of the laws of heredity (Mendel, 1866) and their rediscovery in the early days of the twentieth century. Recognition of DNA as the hereditary material (Avery, 1944), determination of its structure (Watson and Crick, 1953), elucidation of the genetic code (Nirenberg, 1963), development of recombinant DNA technologies (Cohen, *et al.*, 1973; Jackson *et al.*, 1972), and establishment of increasingly automatable methods for DNA sequencing (Intl. Hum. Genome Seq. Consortium, 2001; Maxam and Gilbert 1977; Sanger and Coulson 1975) set the stage for the Human Genome Project (HGP) to begin in 1990. The HGP is an international 13-year effort coordinated by the U.S. Department of Energy and the National Institute of Health, formally begun in October 1990. It was planned to last 15 years, but rapid technological advances have accelerated the speed and the project

completed in April 2003, two year in advance (Collins *et al.*, 2003, HGP Information, 2003; Jasni and Roberts, 2003). At least 18 countries have established human genome programs. Some of the larger programs are in Australia, Brazil, Canada, China, Denmark, European Union, France, Germany, Israel, Italy, Japan, Korea, Mexico, Netherlands, Russia, Sweden, United Kingdom, and the United States of America. Some developing countries are participating through studies of molecular biology techniques for genome research and studies of organisms that are particularly interesting to their geographical regions. The Human Genome Organisation (HUGO) helps to coordinate international collaboration in the genome project. Project goals are to:

- *sequencing and annotating* the human DNA
- *identify* all the approximate 30,000 genes in human DNA
- *store* this information in databases
- *improve* tools for data analysis
- *transfer* related technologies to the private sector
- *address* the ethical, legal, and social issues (ELSI) that may arise from the project.

As part of the HGP, parallel studies are being carried out on selected model organisms to help develop the technology and interpret human gene function. Sequence and analysis of the human genome working draft was published in the year 2001 jointly by the international consortium led by DOE, USA and Craig Ventor of Celera Genomics. Rapid progress in genome science and a glimpse into its potential applications have spurred observers to predict that biology will be the foremost science of the 21st century. Some current and potential benefits as well as applications of genome research include:

bioarchaeology, anthropology, evolution, and human migration in addition to molecular medicine, microbial genomics, risk assessment, DNA forensics (identification), and agriculture, livestock breeding & bioprocessing.

Bioarchaeology, Anthropology, Evolution, and Human Migration will have the following aspects on the study of the population.

- study evolution through germline mutations in lineages

- study migration of different population groups based on female genetic inheritance
- study mutations on the Y chromosome to trace lineage and migration of males
- compare breakpoints in the evolution of mutations with ages of populations and historical events

Understanding genomics will help us understand human evolution and the common biology we share with all of life. Comparative genomics between humans and other organisms such as mice already has led to similar genes associated with diseases and traits. Further comparative studies will help determine the yet-unknown function of thousands of other genes. Comparing the DNA sequences of entire genomes of different microbes will provide new insights about relationships among the three kingdoms of life: archaebacteria, eukaryotes, and prokaryotes.

Insights from the sequence

The working draft sequence of the human genome provides the following details (HGP, Human Genome Project Sequence Analysis, 2003).

- The human genome contains 3164.7 million chemical nucleotide bases.
- The average gene consists of 3000 bases, but sizes vary greatly, with the largest known human gene being dystrophin at 2.4 million bases.
- The total number of genes is estimated at 30,000 to 35,000, much lower than previous estimates of 80,000 to 140,000 that had been based on extrapolations from gene-rich areas as opposed to a composite of gene-rich and gene-poor areas.
- The order of almost all (99.9%) nucleotide bases are exactly the same in all people.
- The functions are unknown for over 50% of discovered genes.
- Less than 2% of the genome encodes for the production of proteins.
- Repetitive sequences are thought to have no direct functions, but they shed light on chromosome structure and dynamics. Over time, these repeats reshape the genome by rearranging it, thereby creating entirely new genes or modifying and reshuffling existing genes.

- During the past 50 million years, a dramatic decrease seems to have occurred in the rate of accumulation of repeats in the human genome.
- The human genome's gene-dense "urban centers" are predominantly composed of the DNA building blocks G and C.
- In contrast, the gene-poor "deserts" are rich in the DNA building blocks A and T. GC- and AT-rich regions usually can be seen through a microscope as light and dark bands on chromosomes, when stained with Geisma.
- Genes appear to be concentrated in random areas along the genome, with vast expanses of noncoding DNA between.
- Stretches of up to 30,000 C and G bases repeating over and over often occur adjacent to gene-rich areas, forming a barrier between the genes and the "junk DNA." These CpG islands are believed to help regulate gene activity.
- Chromosome 1 has the most genes (2968), and the Y chromosome has the fewest (231).
- Unlike the human's seemingly random distribution of gene-rich areas, many other organisms' genomes are more uniform, with genes evenly spaced throughout.
- Humans have on average three times as many kinds of proteins as the fly or worm because of mRNA transcript "alternative splicing" and chemical modifications to the proteins. This process can yield different protein products from the same gene.
- Humans share most of the same protein families with worms, flies, and plants, but the number of gene family members has expanded in humans, especially in proteins involved in development and immunity.
- The human genome has a much greater portion (50%) of repeat sequences than the mustard weed (11%) , the worm (7%), and the fly (3%).
- Although humans appear to have stopped accumulating repeated DNA over 50 million years ago, there seems to be no such decline in rodents. This may account for some of the fundamental differences between hominids and rodents, although gene estimates are similar in these species. Scientists have proposed many theories to explain evolutionary contrasts between humans and other organisms, including those of life span, litter sizes, inbreeding, and

genetic drift.

- Scientists have identified about 1.4 million locations where single-base DNA difference (SNPs) occurs in humans.
- The ratio of germline (sperm or egg cell) mutations is 2:1 in males vs females.

Future of genomics research

The finished sequence produced by the Human Genome Project covers about 99 percent of the human genome's gene-containing regions, and it has been sequenced to an accuracy of 99.99 percent. In addition, to help researchers better understand the meaning of the human genetic instruction book; the project took on a wide range of other goals, from sequencing the genomes of model organisms to developing new technologies to study whole genomes. Besides delivering on the stated ambitious goals which have been met or surpassed (HGP Goals, 2003), the international network of researchers have produced an amazing array of advances that most scientists had not expected until much later. These "bonus" accomplishments include: an advanced draft of the mouse genome sequence, published in December 2002; an initial draft of the rat genome sequence, produced in November 2002; latest information on single nucleotide polymorphisms (SNPs); and the generation of full-length cDNAs for more than 70 percent of known human and mouse genes(HGP Information–HGP Completion,2003).

The HGP's new research strategies and experimental technologies have generated a steady stream of ever-larger and more complex genomic data sets that have poured into public databases and have transformed the study of virtually all life processes. The genomic approach of technology development and large-scale generation of community resource data sets has introduced an important new dimension into biological and biomedical research. Interwoven advances in genetics, comparative genomics, high-throughput biochemistry and bioinformatics are providing biologists with a markedly improved repertoire of research tools that will allow the functioning of organisms in health and disease to be analysed and comprehended at an unprecedented level of molecular detail. Genome sequences, the bounded sets of information that guide biological development and function, lie at the heart of this revolution. In short, genomics has become a central and cohesive discipline of biomedical research. The practical consequences of the emergence of this new field are

widely apparent. Identification of the genes responsible for human Mendelian diseases, once a Herculean task requiring large research teams, many years of hard work, and an uncertain outcome, can now be routinely accomplished with relative ease.

Our ability to explore genome function is increasing in specificity as each subsequent genome is sequenced. Microarray technologies have catapulted many laboratories from studying the expression of one or two genes in a month to studying the expression of tens of thousands of genes in a single afternoon. Clinical opportunities for gene-based pre-symptomatic prediction of illness and adverse drug response are emerging at a rapid pace, and the therapeutic promise of genomics has ushered in an exciting phase of expansion and exploration in the commercial sector (Guttmacher and Collins, 2002). The investment of the HGP in studying the ethical, legal and social implications of these scientific advances has created a talented cohort of scholars in ethics, law, social science, clinical research, theology and public policy.

Genotyping

Several researchers have attempted to genotype the human beings by adopting various means. These include traditional, classical and currently the outputs of the recent advances of the molecular biology. Traditional typing systems are relatively simple and inexpensive and can provide valuable contributions to the study of populations, identity, characterization of evidence materials etc. Studies employing the traditional methods were conducted by researchers like Thurston and Rangachari (1905), for characterizing various populations. Classical approach in genotyping, used the idea of conventional genetic markers to solve the problems of individual identification, which can be traced back to 1902 when Max Richter and Karl Landsteiner, discoverer of the ABO blood group system, first suggested that ABO typing of forensic blood stains could be used to help identify associations or exclusions between blood at the scene of crime and a suspect (Landsteiner and Richter, 1902). Later, several types of conventional genetic markers were used in the study of individuals and populations.

Traditional approach

Traditional techniques were employed by several researchers, for their related

studies on population including Thurston and Rangachari (1905) who worked on characterizing the Castes and Tribes of Southern India.

They have conducted ethnographic survey of South India, the work of which started in 1901. The survey was extended in the area commonly known as Madras Presidency and officially as the Presidency of Fort St. George and its Dependencies. Included therein were the small feudatory States of Pudukottai, Banganapalle and Sandur as well as the Native States of Travancore and Cochin. Briefly the work laid upon him was to record the 'manners and custom' and physical characters of more than 300 castes and tribes, representing more than 40,000,000 individuals and spread over an area exceeding 150,000 square miles. The Native State of Mysore was excluded ethnographically but included for the purpose of anthropometry and it has been found as, nearly all the castes and tribes, which inhabit the Mysore, common to Madras Presidency.

The primary method adopted was to record the measurements of head, chest and foot and the main focus was on Stature, Nasal Index and Cephalic Index, through the measurements of general stature, nose and head. Based on the measurements of a number of subjects in the southern districts of the Madras Presidency, Thurston and Rangachari (1905) had published certain statistics. These figures showed that "the average cephalic index of 639 members of 19 different castes and tribes was 74.1; and that, in only 19 out of the 639 individuals, did the index exceed so. So far then from the Dravidian being separated from the Todas by reason of their higher cephalic index, this index is, in the Todas, actually higher than in some of the Dravidian peoples." Other prominent researchers, like Haeckel, Risley (Thurston and Rangachari, 1905) also were using similar methods, such as nasal index for the comparitive study of various races.

Classical approach

Significant advances in immunology and human genetics have occurred in the past. The growth and development in forensic serology in the past years has revealed a substantial number of genetic marker systems from which routine parentage testing protocols may be constructed (Gaensslen and Camp, 1986; Silver, 1983; Walker,

1983) and also partial individualization of blood and physiological fluid stains done (Culliford 1971; Gaensslen 1983; Gaensslen and Camp, 1984; Lee, 1982; Sensabaugh, 1982; Sensabaugh, 1983). Interpretation of the significance of typing results in criminalistics applications and calculations of the probability of paternity in nonexclusion parentage cases both require knowledge of genotype and phenotypic frequencies in applicable populations. Thousands of frequency studies on various genetic marker systems have been carried out on many populations throughout the world, the most complete compilation of them being the extraordinary work by Mourant *et al.* (1976). A number of studies have appeared which compile or analyze or both, frequency data for many different genetic marker systems for the population of entire countries, such as Japan (Ferrell *et al.*, 1977; Satoh, *et al.*, 1977; Ueda, *et al.*, 1977), Ireland (Walter and Palsson, 1973) and England (Rothwell, 1985; Stedman, 1972, 1975). Genetic marker frequency data for a number of populations of the United States are available (Gaensslen *et al.*, 1987a,b,c). Several conventional genetic marker systems like, ABO, Rh, Lewis, MNSs, Kell-Cellano, P_1, Duffy, Lutheran, Kidd, Gc, Hp, Gm, Km, Tf, α_1-AT, EAP-AcP, PGM1, AK, ADA, EsD, 6-PGDH, GLO-I etc., are generally conducted and some of these are analysed through the technology of isoelectric focusing (Benciolini *et al.*, 1985; Gaensslen *et al.*, 1987 a,b,c; Murch and Budowle, 1986; Roy, 1990).

Kompf *et al.* (1975) first reported a new genetic marker glyoxalase (GLO 1) in human red cells. Sharma, *et al.* (1988), studied the polymorphism of GLO 1 isoenzymes, in Himachal population (India) and the distribution of various GLO 1 isoenzymes are available. Similar types of studies were seen in North Indian population (Ghosh, 1977), Punjabi population (Sehajpal *et al.*, 1983) and Delhi population (Sharma, *et al.*, 1986). A comprehensive study was made by Ross (1986), for several gene markers, in a large series of individuals residing in South Australia. Information on the polymorphism of PGM1 and Hp types for Australian Caucasians is available in the paper of Lai and Goetz (1988). In addition, population studies have been published on haptoglobin among Maltese (Grech, 1979) and Zimbabweans (Raymond and Fletcher, 1983), on phosphoglucomutase among Zimbabweans (Kobus and Raymond, 1982), Paduans (Cortivo *et al.*, 1984) and Egyptians (Sebetan *et al.*, 1986). World population data on the haptoglobin (HP) and group specific component

(GC) systems were compiled and summarized (Walter and Steegmüller, 1969).

DNA approach

Molecular genetics provides a powerful tool to investigate the origin and history of the human species and populations. Combined with solid data analysis, it can yield a truly unique, and indeed emancipatory contribution to the understanding of ourselves. Both, nuclear and mitochondrial DNA have been employed in the studies, related with various applications.

Repetitive DNA

The genome of higher eukaryotes, including humans, may be divided into different classes based very broadly on known functional properties. The so-called "coding regions" contain DNA sequences, which determine primarily the amino acid sequences of the proteins for which they code, and also the degree of expression of the gene in any tissue at any time. "Noncoding" DNA generally contains DNA sequences and such sequences may either be as a single copy (acting as "spacer DNA" between the coding regions of the genome) or exist in multiple copies, thus being repetitive DNA (Beridze, 1986; Deininger and Daniels, 1986; Hardman, 1986; Jelinek and Schmid, 1982; Singer, 1982; Singer and Skowronski 1985; Stephan, 1986). However these repeats actually represent an extraordinary trove of information about biological processes. The repeats constitute a rich palaeontological record, holding crucial clues about evolutionary events and forces. As passive markers, they provide assays for studying processes of mutation and selection. It is possible to recognize a cohort of repeats 'born' at the same time and to follow their fates in different regions of the genome or in different species. Over time, these repeats reshape the genome by rearranging it, thereby creating entirely new genes or modifying and reshuffling existing genes. They also shed light on chromosome structure and dynamics, and provide tools for medical genetic and population genetic studies.

Gene duplications are usually attributed to rare accidents catalyzed by some of the enzymes that mediate normal recombination processes. Higher eucaryotes, however, contain an efficient enzymatic system that joins the two ends of a broken DNA molecule together, so that duplications (as well as inversions, deletions, and

translocations of DNA segments) can also arise as a consequence of the erratic rejoining of fragments of chromosomes that have somehow become broken in more than one place. When duplicated DNA sequences are joined head to tail, they are said to be tandemly repeated. Once a single tandem repeat appears, it can be extended readily into a long series of tandem repeats by unequal crossover events between two sister chromosomes, in as much as the large amount of matching sequence provides an ideal substrate for general recombination.

Repetitive DNA is generally confined to higher eucaryotes (Alberts *et al.*, 1994; Deininger and Daniels, 1986; Hardman, 1986; Jelinek and Schmid, 1982; Singer, 1982; Singer and Skowronski, 1985). By contrast, lower eukaryotes and prokaryotes have smaller genomes, consisting of mostly low or single copy number coding sequences only (Hardman, 1986). A puzzling observation in the early days of molecular biology was that genome size does not correlate well with organismal complexity. For example, *Homo sapiens* has a genome that is 200 times as large as that of the yeast *S. cerevisiae*, but 200 times as small as that of *Amoeba dubia* (Gregory and Hebert, 1999; Li, 1997). This mystery (the C-value paradox) was largely resolved with the recognition that genomes can contain a large quantity of repetitive sequence, far in excess of that devoted to protein-coding genes (Gregory and Hebert, 1999; Hartl, 2000).

The existence of repeated sequences in higher eukaryotes was generally established by one of three methods (Singer, 1982): a fraction of the genome reannealed very rapidly after denaturation; secondly, centrifugation of genomic DNA in cesium chloride (CsCl) gradients (this method separated fractions of DNA that were of different buoyancy to the main band DNA and these so-called "satellites" were found to be highly repeated sequences) and thirdly, restriction of whole genomic DNA by some enzymes, separation of the fragments by electrophoresis in agarose gels, and staining with ethidium bromide. The term "satellite" has become generally adopted to describe *all* tandemly repeated sequences (whether separable by centrifugation or not) (Fowler, 1988). A system of nomenclature and classification of repetitive DNA has evolved, this being largely based on the structural organization and reiteration frequency of each species (Deininger and Daniels, 1986; Hardman, 1986; Jelinek and

Schmid, 1982; Singer, 1982; Singer and Skowronski, 1985).

The human is the first repeat-rich genome to be sequenced, and so, investigated what information could be gleaned from this majority component of the human genome. It has a much greater portion (50%) of repeat sequences than the mustard weed (11%), the worm (7%), and the fly (3%) (HGP, Human Genome Project, Sequence Analysis, 2003). Although some of the general observations about repeats were suggested by previous studies, the draft genome sequence and completed sequence (Collins *et al.*, 2003; HGP, Human Genome Project Sequence Analysis, 2003; International Human Genome Sequencing Consortium, 2001) provides the first comprehensive view, allowing some questions to be resolved and new mysteries to emerge. The completion of Human Genome Project - Draft Sequence and Completed Sequence (Collins *et al.*, 2003; HGP, Human Genome Project Sequence Analysis, 2003; International Human Genome Sequencing Consortium, 2001), provides us a clear picture of our genome and the latest information on repeat sequences with the nomenclature and classification as of below. Also further depicted some of the noteworthy information seemingly appropriate in the area of this dissertation.

In the human, coding sequences comprise less than 5% of the genome, whereas repeat sequences account for at least 50% and probably much more. Broadly, the repeats fall into five classes and the classification of the repeat sequences of human genome is given in Fig. 2.1.

Fig. 2.1.

Transposon-derived repeats

Most human repeat sequence is derived from transposable elements (Prak and Haigh, 2000; Smith, 1999) and about 45% of the genome belongs to this class. Much of the remaining 'unique' DNA must also be derived from ancient transposable element copies that have diverged too far to be recognized as such. It is necessary briefly to review the relevant features of human transposable elements so as to interpret the analyses of interspersed repeats.

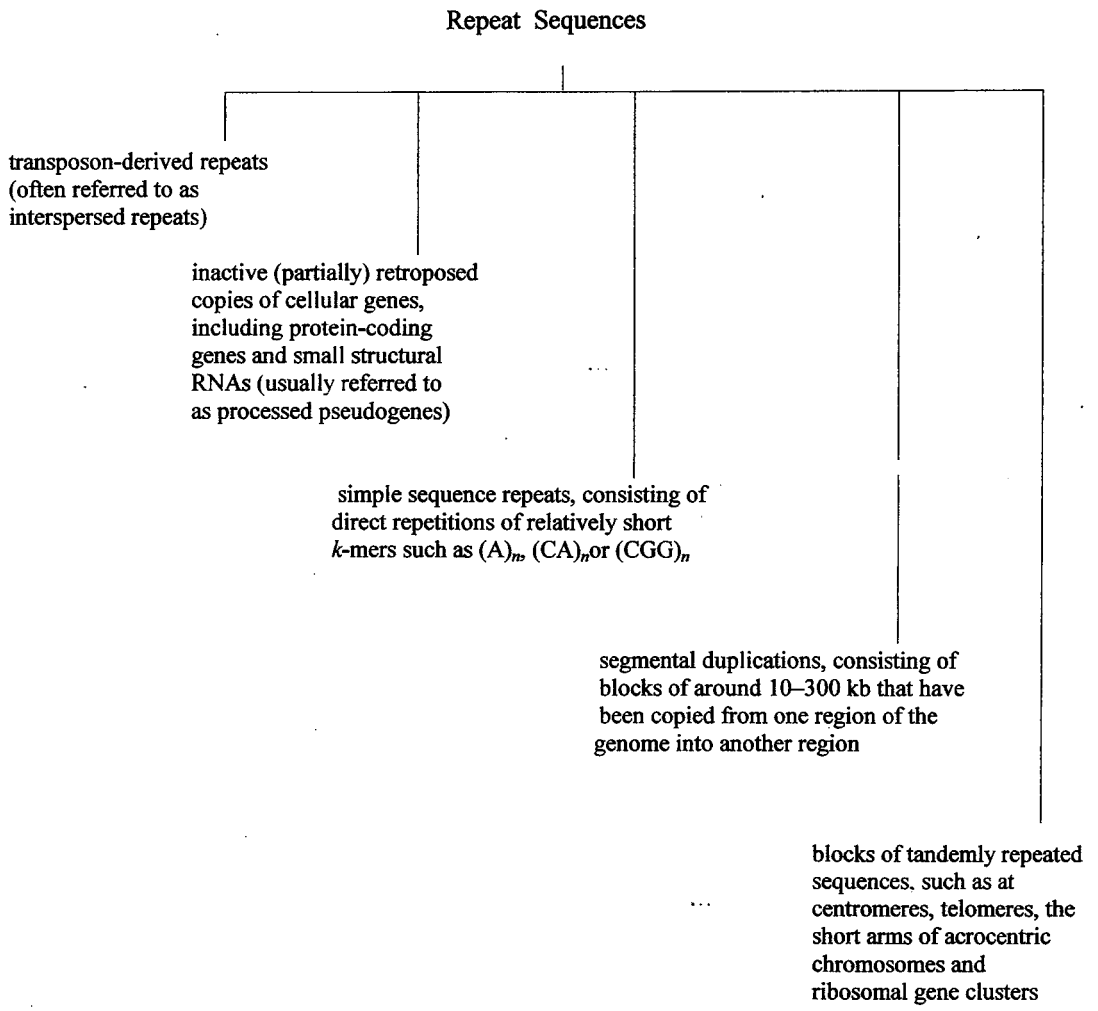

Fig. 2.1. Classification of repeat sequences of human genome

Classes of transposable elements

In mammals, almost all transposable elements fall into one of four types, of which three transpose through RNA intermediates and one transposes directly as DNA. These are long interspersed elements (LINEs), short interspersed elements (SINEs), LTR retrotransposons and DNA transposons.

LINEs are one of the most ancient discoveries in eukaryotic genomes. In humans, these transposons are about 6 kb long, harbour an internal polymerase II promoter and encode two open reading frames (ORFs). A LINE RNA assembles with its own encoded proteins and moves to the nucleus, where an endonuclease activity makes a single-stranded nick and the reverse transcriptase uses the nicked DNA to prime reverse transcription from the 3' end of the LINE RNA. Reverse transcription

frequently fails to proceed to the 5' end, resulting in many truncated, nonfunctional insertions. Indeed, most LINE-derived repeats are short, with an average size of 900 bp for all LINE1 copies, and a median size of 1,070 bp for copies of the currently active LINE1 element (L1Hs). New insertion sites are flanked by a small target site duplication of 7–20 bp. The LINE machinery is believed to be responsible for most reverse transcription in the genome, including the retrotransposition of the non-autonomous SINEs (Okada *et al.*, 1997) and the creation of processed pseudogenes (Esnault *et al.*, 2000). Three distantly related LINE families are found in the human genome: LINE1, LINE2 and LINE3. Only LINE1 is still active.

SINEs are wildly successful freeloaders on the backs of LINE elements. They are short (about 100–400 bp), harbour an internal polymerase III promoter and encode no proteins. These non-autonomous transposons are thought to use the LINE machinery for transposition. Indeed, most SINEs 'live' by sharing the 3' end with a resident LINE element (Okada *et al.*, 1997). The promoter regions of all known SINEs are derived from tRNA sequences, with the exception of a single monophyletic family of SINEs derived from the signal recognition particle component 7SL. This family, which also does not share its 3' end with a LINE, includes the only active SINE in the human genome: the Alu element. By contrast, the mouse has both tRNA-derived and 7SL-derived SINEs. The human genome contains three distinct monophyletic families of SINEs: the active Alu, and the inactive MIR and Ther2/MIR3.

LTR retrotransposons are flanked by long terminal direct repeats that contain all of the necessary transcriptional regulatory elements. The autonomous elements (retrotransposons) contain gag and pol genes, which encode a protease, reverse transcriptase, RNAse H and integrase. Exogenous retroviruses seem to have arisen from endogenous retrotransposons by acquisition of a cellular envelope gene (env) (Malik, *et al.*, 2000). Transposition occurs through the retroviral mechanism with reverse transcription occurring in a cytoplasmic virus-like particle, primed by a tRNA (in contrast to the nuclear location and chromosomal priming of LINEs). Although a variety of LTR retrotransposons exist, only the vertebrate-specific endogenous retroviruses (ERVs) appear to have been active in the mammalian genome. Mammalian retroviruses fall into three classes (I–III), each comprising many families

with independent origins. Most (85%) of the LTR retroposon-derived 'fossils' consist only of an isolated LTR, with the internal sequence having been lost by homologous recombination between the flanking LTRs.

DNA transposons resemble bacterial transposons, having terminal inverted repeats and encoding a transposase that binds near the inverted repeats and mediates mobility through a 'cut-and-paste' mechanism. The human genome contains at least seven major classes of DNA transposon, which can be subdivided into many families with independent origins (Smith, 1996). DNA transposons tend to have short life spans within a species. This can be explained by contrasting the modes of transposition of DNA transposons and LINE elements. LINE transposition tends to involve only functional elements, owing to the cis-preference by which LINE proteins assemble with the RNA from which they were translated. By contrast, DNA transposons cannot exercise a cis-preference: the encoded transposase is produced in the cytoplasm and, when it returns to the nucleus, it cannot distinguish active from inactive elements. As inactive copies accumulate in the genome, transposition becomes less efficient. This checks the expansion of any DNA transposon family and in due course causes it to die out. To survive, DNA transposons must eventually move by horizontal transfer to virgin genomes, and there is considerable evidence for such transfer (Clark and Tidwell, 1997; Haring, *et al.*, 2000; Robertson and Lampe, 1995; Simmons, 1992). Transposable elements employ different strategies to ensure their evolutionary survival. LINEs and SINEs rely almost exclusively on vertical transmission within the host genome (Kordis and Gubensek, 1997; Malik, *et al.*, 1999; Smith, 1996). DNA transposons are more promiscuous, requiring relatively frequent horizontal transfer. LTR retroposons use both strategies, with some being long-term active residents of the human genome and others having only short residence times.

Census of human repeats

A census of the transposable elements in the draft genome sequence has been made available by International Human Genome Sequencing Consortium (2001) using an updated version of the RepeatMasker program (version 09092000) run under sensitive settings, using RepBase Update version 5.08 (Jurka, 2000). Table 2.1 shows the number of copies and fraction of the draft genome sequence occupied by each of

Table 2.1. Number of copies and fraction of genome for classes of interspersed repeat.

	Number of copies x1,000	Total number of bases in the draft genome sequence (Mb)	Fraction of the draft genome sequence (%)	Number of families (subfamilies)
SINEs	1,558	359.6	13.14	
Alu	1,090	290.1	10.60	1(~20)
MIR	393	60.1	2.20	1(1)
MIR3	75	9.3	0.34	1(1)
LINEs	868	558.8	20.42	
LINE1	516	462.1	16.89	1(~55)
LINE2	315	88.2	3.22	1(2)
LINE3	37	8.4	0.31	1(2)
LTR elements	443	227.0	8.29	
ERV-class I	112	79.2	2.89	72(132)
ERV(K)-class II	8	8.5	0.31	10(20)
ERV(L)-class III	83	39.5	1.44	21(42)
MaLR	240	99.8	3.65	1(31)
DNA elements	294	77.6	2.84	
hAT group				
MER1-Charlie	182	38.1	1.39	25(50
Zaphod	13	4.3	0.16	4(10)
Tc-1 group				
MER2-Tigger	57	28.0	1.02	12(28)
Tc2	4	0.9	0.03	1(5)
Mariner	14	2.6	0.10	4(5)
PiggyBac-like	2	0.5	0.02	10(20)
Unclassified	22	3.2	0.12	7(7)
Unclassified	3	3.8	0.14	3(4)
Total interspersed repeats		1,226.8	44.83	

the four major classes and the main subclasses. In the last column, separate consensus sequences in the repeat databases are considered subfamilies, rather than families, when the sequences are closely related or related through intermediate subfamilies. The recognized SINEs, LINEs, LTR retroposons and DNA transposon copies as per draft genome sequence comprise 13%, 20%, 8% and 3% of the sequence, respectively.

Age distribution

The age distribution of the repeats in the human genome provides a rich 'fossil record' stretching over several hundred million years. The ancestry and approximate age of each fossil can be inferred by exploiting the fact that each copy is derived from, and initially carried the sequence of, a then-active transposon, being generally under no functional constraint, has accumulated mutations randomly and independently of others.

Comparison with other organisms

The International Human Genome Sequencing Consortium, (2001) has analysed the fly, worm and mustard weed genomes for the number and nature of repeats, as well the age distribution and the following conclusions were drawn.

(1) The euchromatic portion of the human genome has a much higher density of transposable element copies than the euchromatic DNA of the other three organisms. The repeats in the other organisms may have been slightly underestimated because the repeat databases for the other organisms are less complete than for the human, especially with regard to older elements; on the other hand, recent additions to these databases appear to increase the repeat content only marginally.

(2) The human genome is filled with copies of ancient transposons, whereas the transposons in the other genomes tend to be of more recent origin. The difference is most marked with the fly, but is clear for the other genomes as well. The accumulation of old repeats is likely to be determined by the rate at which organisms engage in 'housecleaning' through genomic deletion. Studies of pseudogenes have suggested that small deletions occur at a rate that is 75-fold higher in flies than in mammals; the half-life of such nonfunctional DNA is estimated at 12 Myr for flies and 800 Myr for

mammals (Petrov *et al.*, 1996). The rate of large deletions has not been systematically compared, but seems likely also to differ markedly.

(3) Whereas in the human two repeat families (LINE1 and Alu) account for 60% of all interspersed repeat sequence, the other organisms have no dominant families. Instead, the worm, fly and mustard weed genomes all contain many transposon families, each consisting of typically hundreds to thousands of elements. This difference may be explained by the observation that the vertically transmitted, long-term residential LINE and SINE elements represent 75% of interspersed repeats in the human genome, but only 5–25% in the other genomes. In contrast, the horizontally transmitted and shorter-lived DNA transposons represent only a small portion of all interspersed repeats in humans (6%) but a much larger fraction in fly, mustard weed and worm (25%, 49% and 87%, respectively). These features of the human genome are probably general to all mammals. The relative lack of horizontally transmitted elements may have its origin in the well developed immune system of mammals, as horizontal transfer requires infectious vectors, such as viruses, against which the immune system guards.

The differences of transposons in the human and mouse genomes were compared. As with the human genome, care is required in calibrating the substitution clock for the mouse genome. There is considerable evidence that the rate of substitution per Myr is higher in rodent lineages than in the hominid lineages (Li, 1997; Li *et al.*, 1996). The analysis shows that, although the overall density of the four transposon types in human and mouse is similar, the age distribution is strikingly different. Transposon activity in the mouse genome has not undergone the decline seen in humans and proceeds at a much higher rate. In contrast to their possible extinction in humans, LTR retroposons are alive and well in the mouse with such representatives as the active IAP family and putatively active members of the long-lived ERVL and MaLR families. LINE1 and a variety of SINEs are quite active. These evolutionary findings are consistent with the empirical observations that new spontaneous mutations are 30 times more likely to be caused by LINE insertions in mouse than in human (3% versus 0.1%) (Kazazian Jr. and Moran, 1998) and 60 times more likely to be caused by transposable elements in general. It is estimated that around 1 in 600 mutations in human are due to transpositions, whereas 10% of mutations in mouse are due to

transpositions (mostly IAP insertions).

The contrast between human and mouse suggests that the explanation for the decline of transposon activity in humans may lie in some fundamental difference between hominids and rodents. Population structure and dynamics would seem to be likely suspects. Rodents tend to have large populations, whereas hominid populations tend to be small and may undergo frequent bottlenecks. Evolutionary forces affected by such factors include inbreeding and genetic drift, which might affect the persistence of active transposable elements (Malik and Eickbush, 2000).

Variation in the distribution of repeats

Work has been conducted to explore variation in the distribution of repeats across the draft genome sequence, by calculating the repeat density in windows of various sizes across the genome. There is striking variation at smaller scales.

Some regions of the genome are extraordinarily dense in repeats. The densest area appears to be a 525-kb region on chromosome Xp11, with an overall transposable element density of 89%. This region contains a 200-kb segment with 98% density, as well as a segment of 100 kb in which LINE1 sequences alone comprise 89% of the sequence. In addition, there are regions of more than 100 kb with extremely high densities of Alu (>56%at three loci, including one on 7q11 with a 50-kb stretch of >61%Alu) and the ancient transposons MIR (> 15% on chromosome 1p36) and LINE2 (> 18% on chromosome 22q12).

In contrast, some genomic regions are nearly devoid of repeats. The absence of repeats may be a sign of large-scale cis-regulatory elements that cannot tolerate being interrupted by insertions. The four regions with the lowest density of interspersed repeats in the human genome are the four gene clusters, HOXA, HOXB, HOXC and HOXD. Each locus contains regions of around 100 kb containing less than 2% interspersed repeats. The sequence analysis of the four HOX clusters in mouse, rat and baboon shows a similar absence of transposable elements, and reveals a high density of conserved noncoding elements. Other repeat-poor regions, such as a region on chromosome 8q21 (1.5% repeat over 63 kb) containing a gene encoding a

homeodomain zinc-finger protein, a region on chromosome1p36 (5%repeat over 100 kb) with no obvious genes and a region on chromosome18q22 (4%over100kb) containing three genes of unknown function are also worth to be noticed.

Distribution by GC content

The next focus has been on the correlation between the nature of the transposons in a region and its GC content. The team has calculated the density of each repeat type as a function of the GC content in 50-kb windows. As has been reported (Goldman *et al.*, 1984; Manuelidis and Ward, 1984; Meunier-Rotival *et al.*, 1982; Smit, 1999; Soriano *et al.*, 1983), LINE sequences occur at much higher density in AT-rich regions (roughly fourfold enriched), whereas SINEs (MIR, Alu) show the opposite trend (for Alu,up to five fold lower in AT-rich DNA) and generally seen in GC rich regions. LTR retroposons and DNA transposons show a more uniform distribution, dipping only in the most GC-rich regions.

Biases in human mutation

Indirect studies have suggested that nucleotide substitution is not uniform across mammalian genomes (Bains, 1992; Filipski, 1987; Sueoka, 1988; Wolfe *et al.*, 1989). By studying sets of repeat elements belonging to a common cohort, one can directly measure nucleotide substitution rates in different regions of the genome. The team of the International Human Genome Sequencing Consortium (2001), has found strong evidence that the pattern of neutral substitution differs as a function of local GC content. Since the results are observed in repetitive elements throughout the genome, the variation in the pattern of nucleotide substitution seems likely to be due to differences in the underlying mutational process rather than to selection.

Fast living on chromosome Y

The pattern of interspersed repeats can be used to shed light on the unusual evolutionary history of chromosome Y. The analysis by International Human Genome Sequencing Consortium (2001) shows that the genetic material on chromosome Y is unusually young, probably owing to a high tolerance for gain of new material by insertion and loss of old material by deletion. Several lines of evidence support this picture.

Active transposons

Investigation has also been concentrated in identifying the youngest retrotransposons in the draft genome sequence. This set should contain the currently active retrotransposons, as well as the insertion sites that are still polymorphic in the human population. The youngest branch in the phylogenetic tree of human LINE1 elements is called L1Hs (Smit *et al.*, 1995). Active transposons like subsets, refered as Ta and pre-Ta, defined by a diagnostic trinucleotide (Boissinot *et al.*, 2000; Skowronski *et al.*, 1988), of L1Hs family has a great value in population genetics since at least 50% are still segregating as polymorphisms in the human population (Boissinot *et al.*, 2000; Kazazian, Jr. *et al.*, 1988; Moran, 1999) and they provide powerful markers for tracing population history because they represent unique (non-recurrent and non-revertible) genetic events that can be used (along with similarly polymorphic Alus) for reconstructing human migrations.

Transposons as a creative force

The primary force for the origin and expansion of most transposons has been selection for their ability to create progeny, and not a selective advantage for the host. However, these selfish pieces of DNA have been responsible for important innovations in many genomes, for example by contributing regulatory elements and even new genes.

Twenty human genes have been recognized as probably derived from transposons (Jurka and Kapitonov, 1999; Smit, 1999) and the team of International Human Genome Sequencing Consortium (2001) has identified another 27 cases, bringing the total to 47. LINE1 activity clearly has also had fringe benefits. The LINE1 machinery can also cause reverse transcription of genic mRNAs, which typically results in nonfunctional processed pseudogenes but can, occasionally, give rise to functional processed genes. There are at least eight human and eight mouse genes for which evidence strongly supports such an origin (Brosius, 1999). Many other intronless genes may have been created in the same way. Transposons have made other creative contributions to the genome. A few hundred genes, for example, use transcriptional terminators donated by LTR retroposons. Other genes employ regulatory elements derived from repeat elements (Brosius, 1999).

Simple sequence repeats

Simple sequence repeats (SSRs) are a rather different type of repetitive structure that is common in the human genome—perfect or slightly imperfect tandem repeats of a particular k-mer. SSRs contain minisatellites and microsatellites with longer and shorter repeat units, respectively. With the exception of poly (A) tails, SSRs are thought to arise by slippage during DNA replication (Kruglyak *et al.*, 1998; Toth *et al.*, 2000). International Human Genome Sequencing Consortium (2001) has compiled a catalogue of all SSRs over a given length in the human draft genome sequence, and studied their properties. SSRs comprise about 3% of the human genome, with the greatest single contribution coming from dinucleotide repeats (0.5%).

There is approximately one SSR per 2 kb (the number of nonoverlapping tandem repeats is 437 per Mb). The catalogue confirms various properties of SSRs that have been inferred from sampling approaches. The most frequent dinucleotide repeats are AC and AT (50 and 35% of dinucleotide repeats, respectively), whereas AG repeats (15%) are less frequent and GC repeats (0.1%) are greatly under-represented. The most frequent trinucleotides are AAT and AAC (33% and 21%, respectively), whereas ACC (4.0%), AGC (2.2%), ACT (1.4%) and ACG (0.1%) are relatively rare. Overall, trinucleotide SSRs are much less frequent than dinucleotide SSRs (Ellegren, H 2000). SSR alleles/Short tandem repeats (STRs) which are generally less than 350 bp in length, can be successfully amplified by PCR (Saiki *et al.*,1988). Di, tri, tetra and penta nucleotide repeats are routinely analyzed (Bacher and Schumm, 1998; Busque *et al.*,1997; Gaviria *et al.*, 004; Klintschar *et al.*,1999; Pfitzinger *et al.*,1995; Shuster *et al.*,1998)

SSRs have been extremely important in human genetic studies, because they show a high degree of length polymorphism in the human population owing to frequent slippage by DNA polymerase during replication. The SSR catalogue also allowed us to resolve a mystery regarding mammalian genetic maps. Such genetic maps in rat, mouse and human have a deficit of polymorphic (CA)n repeats on chromosome X, documented elsewhere. There are two possible explanations for this deficit. There may simply be fewer (CA)n repeats on chromosome X; or (CA)n repeats may be as dense on chromosome X but less polymorphic in the population. Infact,

analysis of the draft genome sequence shows that chromosome X has the same density of (CA)n repeats per Mb as the autosomes (data not shown). Thus, the deficit of polymorphic markers relative to autosomes results from population genetic forces. Possible explanations include that chromosome X has a smaller effective population size, experiences more frequent selective sweeps reducing diversity (owing to its hemizygosity in males),or has a lower mutation rate (owing to its more frequent passage through the less mutagenic female germline). The availability of the draft genome sequence should provide ways to test these alternative explanations.

Polymorphism in tandemly repeated DNA

All polymorphisms associated with tandemly repeated DNA arise from restriction enzymes creating DNA fragment lengths that contain *variable numbers* of the repeated species (Fowler *et al.*, 1988). There are, in general, two contrasting causes by which this occurs. In one case, the enzyme cuts the DNA *externally* to the block of tandem repeats, and in the other, the DNA is cut *within* the lengths of tandemly repeating units. The consequence of this is that in the first mechanism any enzyme that cuts the DNA in regions flanking (but not within) the tandem repeat will reveal any variation that may exist in the number of repeats, and hence the fragment size. The second mechanism is enzyme specific, polymorphism being generally a result of an enzyme that cuts (or fails to cut) infrequently into the higher order repeat units of each tandem array. Thus restriction enzyme digestion of DNA produces a mixture of restriction fragments, the lengths of which can vary between individuals, and these are therefore referred to as *restriction fragment length polymorphisms* or RFLPs (Ross and Harding, 1989).

Polymorphism arising by the first mechanism have been referred to as "minisatellites" (Jeffreys *et al.*, 1985), hypervariable regions (HVRs) (Jeffreys 1987) and variable number of tandem repeats (VNTRs) (Nakamura *et al.*, 1987). Polymorphisms arising by the second mechanism have been referred to as VNTRs (Caskey, 1987).

Minisatellites

The minisatellite polymorphism is an insertion/deletion mutational event,

thereby lengthening or shortening the overall fragment length. The length of repeat units inserted or deleted is typically between 64 and 10 bp long (Jarman *et al.*, 1986). Tandem arrays of such units may exist at either a unique or a number of dispersed genomic sites. The single most important property of these tandem repeats, first described was their genomic variability (Wyman and White, 1980). The "minisatellites" (Jeffreys, 1985a) have smaller blocks of tandemly repeated DNA and is scattered throughout the genome; however, these loci show a propensity for the telomeric regions of the chromosomes (Royle *et al.*, 1988). The length of the repeat sequence ranged from 16 bp in λ33.15 to 64 bp in λ33.4 (Jeffreys *et al.*, 1985a). International Human Genome Sequencing Consortium (2001) states the minisatellites as having longer repeat units (n = 14-500 bases). The cause of a loss or gain in the number of repeat units at any genomic site may be either (1) integral numbers of unit slippage at replication or (2) unequal recombination between the tandemly repeated sequences or both (International Human Genome Sequencing Consortium, 2001; Jeffreys, 1987). The recombination may indeed be directed by a particular GC rich "core sequence" which (with some sequence diversity) has been found to be embedded in many of the GC rich "minisatellites". Jeffrey's conclusion is that such core sequences, 11 to 16 bp in length in his examples, either promote the initial duplication of DNA immediately adjacent to it or assist in changing the number of repeat units so created by unequal recombination or faulty replication or both (Jeffreys, 1987). Many minisatellites are highly polymorphic due to allelic variation in repeat copy number in the minisatellite and these variants are transmitted in a Mendelian fashion (Jeffreys, 1985a).

Uses of minisatellites

The practical consequence of these details are as follows. If a "polycore" probe (exemplified by Jeffrey's probes 33.15 and 33.6 (Jeffreys *et al.*, 1985a) is hybridized under moderate stringency to suitably restricted DNA, the multiple loci (between 40 to 60) of this "core" sequence are revealed as complex patterns of restriction fragments (genetic "fingerprints"). The DNA 'fingerprints' obtained using these core minisatellite probes are reproducible and are suitable for individual identification, as well as somatically stable, which are completely specific to an individual (or to his or her identical twin), an essential prerequisite for identification purposes (Jeffreys

1985b). This method of individualization has been applied to forensic science samples (Dodd, 1985; Gill *et al.*, 1985, 1987). If the probe, however, contains tandem repeats of both the "core" sequence and flanking DNA specific to particular loci, then under high stringency conditions such a probe will generally detect only two (heterozygous) fragments from a single genomic locus. Probes of this sort have also been used in the analysis of forensic science samples (Ginsti *et al.*, 1986., Kanter *et al.*, 1986). If a "mixture" of single-locus probes is used simultaneously, a combined fragment pattern arises (Wainscoat *et al.*, 1987). In addition to the personal identification, eg. forensic individualization and disputed parentage, the minisatellite polymorphism is valuable in constructing linkage maps, diagnosing genetic disorders etc.(Jeffreys, 1985b; Odelberg and White, 1989).

Jeffreys *et al.* (1985) prepared a pure repeat probe from the human myoglobin minisatellite by purification of a single 33-bp repeat element followed by head-to-tail ligation and cloning of the resulting polymer into pUC 13. Low stringency hybridization of this repetitive insert to human DNA digested with restriction endonuclease EcoRI detected numerous cross-hybridizing DNA fragments, some of which showed signs of polymorphic variation. To improve detection of polymorphism, the hybridization was repeated with human DNA digested with HinfI or HaeIII, both of which cleave at a 4-bp sequence not present in the 33-bp repeat sequence and which should release minisatellites in relatively small DNA fragments whose size will reflect more closely the number of repeats per minisatellite. The repetitive probe detected multiple DNA fragments in human DNA as well as the parent DNA fragment from the human myoglobin gene. Variation was detectable in both HinfI and HaeIII digests of human DNA, consistent with polymorphism resulting from length variation of minisatellite regions, which are apparently transmitted in a Mendelian fashion.

Microsatellites

Microsatellites consist of tandem repeated sequences but with a repeat unit much shorter than seen at minisatellites and typically 2,3 or 4 base pairs long (Jeffreys, 1993) or 2 – 7 bp repeats (Charlesworth *et al.*, 1994; Steffens *et al.*, 1997; Feng-Xia Xiao *et al.*, 1998; Tautz, D 1991) or as 3 – 7 bp repeats (Edwards, A *et al.*, 1991a,b, 1992) or 2 – 6 bp (Thomson *et al.*, 1999).

Microsatellite loci are found in large numbers and relatively evenly spaced throughout the genome. Edwards *et al.* (1991a,b) examined the frequency of 5 microsatellite loci (tri and tetra repeats) on the X chromosome. They found, for either the tri or tetra microsatellite loci, that any given repeat was found every 300 to 500 kb. From this they estimated that for all the 44 possible unique trimeric and tetrameric repeats there are ~ 400,000 loci or about 1 every 10 to 20 kb. Beckmann and Weber (1992) estimated that the STRs occur every 6 to 10 kb in the human geneome. Of the class of loci examined by Edwards *et al.* (1991a,b) ~ 50% are polymorphic. Further, most of these loci are selectively neutral which makes them compatible with the assumptions of most population genetic theory. In humans, 90% of known microsatellites are found in noncoding regions of the genome. When found in human coding regions, microsatellites are known to cause disease (Bio 2002 'Teaching in the Genome Age' www.woodrow.org/teachers/esi/2002/Biology/Projects). They usually don't have any measurable effect on phenotype, and when they do mutate, it is generally harmful, not beneficial. Microsatellites are considered to have regulatory roles in gene expression (Kashi and Soller, 1999). The completion of the HGP, Human Genome Project Sequence Analysis (2003) states, the microsatellites as reshaping the genome by rearranging it and thereby creating entirely new genes or modifying and reshuffling existing genes.

Duplication history

One form of the mutational processes for DNA molecules is *tandem duplication* in which a stretch of DNA is transformed into two or more adjacent copies. The result of a tandem duplication is termed a *tandem repeat.* Over time, individual copies within a tandem repeat may undergo additional, uncoordinated mutations (including new tandem duplications) so that typically, multiple approximate tandem copies are present. A brief view on the formation of these tandem repeats has been already stated under the heading of "minisatellites". Examination of a tandem repeat often suggests that the sequence was produced by a series of tandem duplications interspersed with point mutations (Benson and Dong, 1997). Tandem repeats differ from other types of duplicated sequences on the fact that new copies of duplication seen situated in the adjacent position on the same sequence. The

boundaries of a duplicated pattern are not always possible to distinguish and the size of the duplication unit can be any multiple of the basic pattern size (Benson and Dong, 1997).

Several mechanisms have been proposed for the production of tandem repeats, including replication slippage and unequal crossing over (International Human Genome Sequencing Consortium, 2001; Levinson and Gutman 1987; Okumura, Kiyama and Oishi, 1987; Schlotterer and Tautz, 1992; Smith 1976; Wells, 1996). Biological studies (Strand *et al.*, 1993; Weitzmann *et al.*, 1997) have already provided support for one or the other of the mechanisms. Mathematical modeling has suggested mechanistic characteristics. Accurate modeling of copy number variation at a polymorphic dinucleotide repeat locus has been obtained with a two-phase model which assumes predominantly single copy changes with rare multi-copy changes (Di Rienzo *et al.*, 1994). Comparison of estimated rates of unequal crossing over and observed rates of microsatellite mutation lead to the conclusion that slipped strand mispairing is the major cause of length polymorphism in microsatellites (Bell and Torney, 1993). Studies of Charlesworth *et al.* (1994) study, employing modeling and simulation, suggest that very low recombination rates (unequal crossing over) can result in very large copy number and higher order repeats.

Advantages and uses

Recently microsatellites have been increasingly used as the marker of choice. There are some advantages in utilizing microsatellites over the other markers, which make them desirable. Technically microsatellites are more desirable than the larger VNTR loci because they can be analysed by PCR (although improvements to the thermal stable polymerase are enabling the amplification of larger fragments of DNA) and the alleles can be unambiguously sized on polyacrylamide gels. Since the size range of the alleles are amenable to PCR amplification, analysis can be done in just a few hours and the amount of DNA input in a PCR reaction is generally 50 to 100 times less than with VNTRs. In addition, samples that are too degraded to yield restriction fragments may still contain sufficient intact DNA for PCR analysis. The lack of exact size resolution of larger VNTR loci has lead to procedures such as binning which reduce the statistical power of the analysis (Budowle *et al.*, 1991). Microsatellites have

been found to be variable even in populations, which have low levels of allozyme and mitochondrial variation (Estoup *et al.*, 1995a,b, 1996; Paetkau and Strobeck,1994).

Microsatellites are useful for a number of analyses. They were originally utilized for genetic mapping (Weissenbach *et al.*, 1992) and have been widely used for linkage analyses. Genetic markers based on SSRs—particularly (CA)n repeats—have been the workhorse of most human disease-mapping studies (Broman *et al.*, 1998). In addition they have proven useful in the analysis of paternity and kinship (Queller *et al.*, 1993; Sigurdur Ingvarsson *et al.*, 2000) and in the probability of sample identity at both the individual / forensic (Corte-Real *et al.*, 1999a,b ; Edwards *et al.*, 1992; Klintschar *et al.*, 1998; Steffens *et al.*,1997; Urquhart *et al.*, 1994) and population levels (Paetkau *et al.*, 1995). It is used in the study of evolution and ancestry, including the study of human migration, rapid evolution of bacterial diseases and the cascade of mutations that lead to cancer (Benson and Dong, 1997). Also it is an effective genetic marker in the various other genetic disorders seen in the medical field and as well in the study of anthropological specimens (Takahashi *et al.*, 1997). In the study of entire populations microsatellites are also very useful (Budowle et. al. 1999; Bruford and Wayne 1993; Corte-Real *et al.*, 1999a,b; Lareu *et al.*, 1994; Pfitzinger *et al.*, 1995). Microsatellite variation has been used to study the amount of hybridization between closely related species (Gottelli *et al.*, 1994; Roy *et al.*, 1994). Comparison of levels of variation between species and populations have also proven useful in the assessment of overall genetic variation (Gottelli *et al.*, 1994; Paetkau and Strobeck, 1994; Taylor *et al.*, 1994). They can be used to estimate effective population size (Allen *et al.*, 1995) and to gain insight into the degree of population substructure including both the amount of migration between subpopulations (Allen *et al.*, 1995; Gottelli *et al.*, 1994; Tahir. *et al.*, 2000) and genetic relationships among the various subpopulations (Bowcock *et al.*, 1994; Estoup *et al.*, 1996; Forbes *et al.*, 1995; Foreman and Lambert, 2000; Karen *et al.*, 2002). The availability of a comprehensive catalogue of SSRs is a boon for human genetic studies (International Human Genome Sequencing Consortium, 2001).

Lorente *et al.*(1994), analysed the (HUMTHO1) allele frequencies in the Spanish population. They found the genotype distribution of the Spanish gene pool

satisfying the Hardy-Weinberg laws and as well compared the Spanish data with that of American Caucasian sample population for homogeneity. Chinese population data in Beijing at 3 STR loci- CSF1PO, HUMTHO1 and TPOX by Fung *et al.*(2001) has found the combined power of discrimination as 0.9966. Analysis of Azores Archipelago (Portugal) with four STR loci by Corte-Real *et al.* (1999b) have found a discrimination power and chance of exclusion of 0.9999 and 0.9534 respectively. They have also studied the population of Madeira Archipelago (Portugal) with six STRs, four of which are the selected loci of this dissertation has the findings of combined discrimination power of 0.9999998 and a chance of exclusion of 0.99597 (Corte-Real *et al.*, 1999a). Allele frequency distribution with four loci in the Asturias (Northern Spain) by Bell *et al.* (2000), six loci in the Aragon (North Spain) by Martinez Jarreta *et al.* (1999) and eight STR loci and five polymarker loci in the population of Bengali and west Punjabi residing in India and Pakistan respectively by Tahir *et al.* (2000) have been described. Genotype survey using six STR loci, HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS and HUMvWA (PROMEGA Technical Manual, GenePrint STR Systems [Silver Stain Detection], 2001) in the French Canadian Caucasian population of Quebec by Busque *et al.* (1997) have found a combined probability of discrimination of 0.9999985 for identity purpose as well as 0.9874 for the probability of paternity exclusion. Their study indicate the six STR loci as good informative genetic markers for identity testing purposes in the French Canadian Caucasian population of Quebec. Pu *et al.*(1998) reports a combined discrimination power of >0.99999999949 and a combined exclusion power of >0.999 for Taiwan population. The research was conducted by using a set of nine STR loci, D3S1358, vWA, FGA, THO1, TPOX, CSF1PO, D5S818, D13S317 and D7S820. Table 2.2. shows the matching probability, paternity indices and power of exclusion obtained for three American populations; African American, Caucasian American and Hispanic American, in the genotyping conducted by using nine STR loci; CSF1PO, TPOX, THO1, F13A01, FESFPS, vWA, D16S539, D7S820 and D13S317. Gamero *et al.* (2000) conducted study with ten STR systems on the relatedness of African immigrant to that of Spanish population data. Similar studies especially in U.S and European countries or with their co-operation by more STRs were conducted and found these markers as highly informative and useful.

Table 2.2. **The Matching Probability, Paternity Indices and Power of Exclusion of the selected STR loci of this research (CSF1PO, TPOX, THO1, F13A01, FESFPS, vWA, D16S539, D7S820 and D13S317) in the three American population (PROMEGA Technical Manual, GenePrint STR Systems [Silver Stain Detection], 2001).**

Population	Matching Probability	Paternity Indices	Power of Exclusion
African American	1 in 5.18×10^9	1233	0.9993
Caucasian American	1 in 1.03×10^9	521	0.9985
Hispanic American	1 in 1.84×10^9	563	0.9986

Other researchers (Agrawal *et al.*, 2002; Budowle *et al.*, 1999; Tomas *et al.*, 2001; Egyed *et al.*, 2000; Farfan *et al.*, 2001; Harbison *et al.*, 2002; Lee *et al.*, 1997; Garofano *et al.*, 1999; Meng *et al.*, 1999; Momhinweg *et al.*, 1998; Pfitzinger *et al.*, 1995; Sanchez-Molina and Calvet, 2000; Soares-Vieira *et al.*, 2001; Yamamoto *et al.*, 1999) made use of STRs for the study of various populations and established databases having several unique characteristics of the populations based upon the statistical aspects of the loci studied.

Single nucleotide polymorphisms

In the 1980s, single nucleotide polymorphisms (SNPs) were detected using restriction enzymes to identify the presence or absence of cutting sites and scored by observing the resulting fragment length variation. In the 1990s, the SNP in the guise of the restriction site polymorphism was largely replaced by the short tandem repeat (STR) as the marker of choice for linkage studies. SNPs regained favour amongst molecular geneticists in the late 1990s (Gray *et al.*, 2000). Single nucleotide polymorphisms (SNPs, pronounced 'snips') are DNA sequence variations that occur when a single nucleotide in the genome sequence is changed (Stoneking, 2001). This change in the genome can be expressed either as a deletion, an insertion, or a substitution (Dickinson, 2001; Halushka *et al.*, 1999). The human draft genome sequence has already allowed the identification of more than 1.42 million SNPs, comprising a substantial proportion of all common human variation (The International SNP Map Working Group, 2001). A public collaboration, the International HapMap Project has been set up in 2002 to characterize the patterns of linkage disequilibrium and haplotypes across the human genome and to identify subsets of SNPs that capture most of the information about the patterns of genetic variation to enable large-scale genetic association studies (Collins *et al.*, 2003).

Description of the SNP map

As per The International SNP Map Working Group (2001), 1,419,190 SNPs have been mapped to unique locations in the 2.7 gigabases (Gb) of assembled human genome sequence, providing an average density of one SNP every 1.91 kb. SNP density is relatively constant across the autosomes. To characterize the distribution of SNPs, the above team has examined 366,192 SNPs that fell within finished sequence. Most of the genome contains SNPs at high density: 90% of contiguous 20-kb windows contain one or more SNPs, as do 63% of 5-kb windows and 28% of 1-kb windows. Only 4% of genome sequence falls in gaps between SNPs of >80kb, and some of these gaps are covered by SNPs that were discovered but not mapped. The team also estimated that 60,000 SNPs fall within exon (coding and untranslated regions), and 85% of exons are within 5 kb of the nearest SNP.

Uses

The SNP polymorphisms are of immediate utility for medical genetic studies. The density of SNPs varies considerably across the genome and sheds light on the unique properties and history of each genomic region, which will greatly contribute on the finding of the origin of the human species (The International SNP Map Working Group, 2001). SNPs can also be used to study linkage disequilibrium in the human genome (Hughes, 1999). It can provide a powerful tool for mapping disease genes (Lander and Schork, 1994) and for probing population history (Tischkoff *et al.*, 1996). There has been considerable controversy concerning the typical distance over which linkage disequilibrium extends in the human genome (Collins *et al.*, 1999). With the collection of SNPs now available, it should be possible to resolve this important issue.

Specialized applications of polymorphic DNA

Forensic biology

Over the past 15 – 20 years, the forensic biology has made great strides. Molecular biology procedures have become available for analysis of forensic biological specimens. All types of organisms can be identified by examination of DNA sequences unique to that species. It has originated from the multilocus probe analysis by restriction fragment length polymorphism (RFLP), developed by Jeffreys *et al.*

(1985 a,b). The detection of variable number of tandem repeat (VNTR) sequences by RFLP analysis is a powerful technique for characterizing forensic biology samples (Budowle and Baechtel, 1990; Budowle *et al.*, 1990). Most of the crime laboratories currently employ the latest PCR based analysis of STR loci and others. Generally for identifying individuals, forensic scientists scan 13 DNA regions that vary from person to person and use the data to create a DNA profile of that individual (Applied Biosystems, AmpFlSTR Cofiler PCR Amplification Kit, 1998; Applied Biosystems, AmpFlSTR Profiler Plus PCR Amplification Kit, 2000; Budowle *et al.*, 1998b; Budowle *et al.*, 1999). Further to the choice of individual institution, more DNA loci are analysed and made use of identification purposes (Applied Biosystems, AmpFlSTR Identifiler PCR Amplification Kit, 2001). However, upon the condition and availability of biological samples, scientists follow different kinds of analysis systems so as to meet the challenge. There is an extremely small chance that another person has the same DNA profile for a particular set of regions.

Some examples of the use of DNA for Forensic analysis are as below (HGP, Human Genome Project Information – DNA Forensics, 2003 – Base URL *http://www.ornl.gov/hgmis).*

- Identify potential suspects whose DNA may match evidence left at crime scenes.
- Exonerate persons wrongly accused of crimes.
- Establish paternity and other family relationships.
- Identify endangered and protected species as an aid to wildlife officials (could be used for prosecuting poachers).
- Detect bacteria and other organisms that may pollute air, water, soil, and food.
- Match organ donors with recipients in transplant programs.
- Determine pedigree for seed or livestock breeds.
- Authenticate consumables such as caviar and wine.

DNA typing

Only 0.1% of DNA differs from one person to the next. Scientists can use these variable regions to generate a DNA profile of an individual, using samples from blood,

bone, hair, and other body tissues and products. In criminal cases, this generally involves obtaining samples from crime-scene evidence and a suspect, extracting the DNA, and analyzing it for the presence of a set of specific DNA markers. A series of probes bound to a DNA sample creates a distinctive pattern for an individual and these patterns are potentially unique to an individual, popularly known as 'DNA Fingerprints' (Jeffreys *et al.*, 1985a,b). Forensic scientists compare these DNA profiles to determine whether the suspect's sample matches the evidence sample. A marker by itself usually is not unique to an individual; if, however, two DNA samples are alike at four or five regions, odds are great that the samples are from the same person. The more probes used in DNA analysis, the greater the odds for a unique pattern and against a coincidental match. Usually four to six probes (single locus probes) are recommended for an analysis (HGP, Human Genome Project Information – DNA Forensics, 2003 - Base URL *http://www.ornl.gov/hgmis*). Finally the assessment of the DNA profile match has been evaluated based on a suitable database (Budowle *et al.*, 1991; Budowle *et al.*, 1999; Busque *et al.*, 1997; Jarreta *et al.*, 1999; Paterson *et al.*, 2000)

A cadre of DNA typing methods based on the PCR are available, such as HLA-DQA1 and Polymarker loci, D1S80, STR, mitochondrial DNA sequencing etc. A concise information on the commonly used DNA technologies/systems in forensic biology may be useful at this juncture.

Restriction fragment length polymorphism (RFLP)

RFLP (Budowle *et al.*, 1990; Gill *et al.*, 1987) is a technique for analyzing the variable lengths of DNA fragments that result from digesting a DNA sample with restriction endonucleases. The presence or absence of certain recognition sites in a DNA sample generates variable lengths of DNA fragments, which are separated using gel electrophoresis. They are then hybridized with DNA probes that bind to a complementary DNA sequence in the sample. Basically, the following analytical steps were used for DNA Fingerprinting utilizing the RFLP analysis: (1) DNA extraction, (2) endonuclease digestion (3) electrophoresis, (4) blotting, (5) hybridization, (6) autoradiography, (7) imaging, and (8) interpretation (Laber *et al.*, 1992).

Wyman and White (1980) were the first to isolate a polymorphic DNA locus characterized by a number of variable-length restriction fragments called *restriction fragment length polymorphisms* (RFLPs). In 1984, while searching for disease markers in DNA, Jeffreys *et al.* (1985a,b) discovered a unique application of RFLP technology to the science of personal identification. His method, which he termed a "DNA fingerprint", has been modified and adopted by crime laboratories. The term "DNA fingerprint" has been replaced with a more descriptive and inclusive term "DNA typing or DNA profiling" (Kirby, 1990; Inman and Rudin, 1997).

Jeffreys and his coworkers (Jeffreys *et al.*, 1985a), while analyzing the human myoglobin gene, discovered a region consisting of a 33 bp sequence repeated four times within an intervening sequence. This tandem repeat was referred to as a *minisatellite* and similar regions as being *hypervariable* because the number of tandem repeats is variable both within a locus and between loci. They also discovered that each repeat unit contains a smaller 16-bp core in common with other minisatellites. When DNA is isolated, cleaved with a specific enzyme, and hybridized under low-stringency conditions with a probe consisting of the core repeat, a complex ladder of DNA fragments are detected and as mentioned before, this profile appears to be unique to each individual. Two versions of the polycore probes (multi locus probes) have been developed that detect nonoverlapping sets of human minisatellites and provide independent DNA fingerprints from the same DNA sample (Jeffreys *et al.*,1985b, 1986).

The largest minisatellite DNA fragments detected in a human DNA fingerprint contain large numbers of tandem repeat units and are therefore likely to represent the most variable minisatellites in the human genome. By selectively cloning these large minisatellites, it has proved possible to isolate some of the most variable loci (Wong *et al.*, 1986, 1987). Under high-stringency hybridization conditions, these cloned minisatellites act as locus-specific probes. Nakamura *et al.* (1987) coined the term *variable number of tandem repeats* (VNTR) to describe individual loci where alleles are composed of tandem repeats that vary in the number of core units. These investigators also isolated and characterized a number of probes suitable for single-locus VNTR profiling. The hybridization patterns detected by these hypervariable

minisatellite probes (single locus probes) are very simple and consist of two (occasionally one) hybridizing DNA fragments per individual (Jeffreys *et al.*, 1989). Single locus analysis offered greater sensitivity, increased species specificity, and standard statistical interpretations compared with the multilocus approach.

HLA DQα / HLA DQA1

Human Leukocyte Antigen DQα locus (HLA DQA1 / HLA DQα) becomes the first PCR-amplified DNA analysis system for forensic biology. The 242 bp and 239 bp located in the second exon of the DQα gene in chromosomal location 6p21.3, show variations (Gyllensten *et al.*, 1988; Comey and Budowle, 1991). Primers that flank this region are used in the amplification, and allele-specific oligonucleotide (ASO) probes have been developed to detect each of the DQα alleles (Higuchi, 1988; Saiki *et al.*, 1989). These probes are fixed to a nylon membrane and DQα alleles are identified by hybridization with the amplified DQα DNA. The system has undergone extensive validation process and has been shown to be reliable on laboratory produced samples. Crime laboratories, worldwide, use the commercially available, Applied Biosystems, AmpliType PM (Polymarker) + DQA1 PCR Amplification and Typing Kit (1995), for the purpose of typing HLA DQA1 locus (Erlich and Bugawan, 1989., Gyllensten and Erlich, 1988) as well as PM (Polymarker) loci.

AmpliTypePM – "Polymarker"

The AmpliTypePM system, generally known as PM, polymarker (Budowle *et al.*, 1995, 1998a), is just an expansion of the technique used in HLA DQα analysis. The combination of these genetic loci with DQA1 systems provides the forensic biology, with an increase in the power of discrimination (PD). The characterization of the following five PM loci, low density lipoprotein receptor (LDLR) chromosomal location 19p13.1-13.3 (Yamamoto *et al.*, 1984), glycophorin A (GYPA) 4q28-31 (Kudo *et al.*, 1989), hemoglobin G gammaglobin (HBGG) 11p15.5 (Slightom *et al.*, 1980), D7S8 in location 7q22-31.1 (Horn *et al.*, 1980) and group-specific component (Gc) 4q11-13 (Reynolds and Sensabaugh, 1990;Yang *et al.*, 1985) in addition to HLA-DQA1 (Erlich and Bugawan, 1989; Gyllensten and Erlich, 1988; Saiki *et al.*, 1989) and D1S80 (Budowle *et al.*, 1991; Kasai *et al.*, 1990) for forensic use are well

documented (Budowle *et al.*, 1998a; Hochmeister *et al.*, 1995; Woo and Budowle, 1995).

D1S80

D1S80, localized by multipoint linkage analysis to the distal end of chromosome 1p and shows co-dominant autosomal segregation (Nakamura *et al.*, 1988). Like RFLP, the D1S80 locus contains repetitive DNA sequences that are arranged as tandem repeat units (VNTRs) (Nakamura *et al.*, 1987). However, because of its relatively smaller size, this fragment is amenable to amplification by PCR and this amplified fragment length polymorphisms, generally known in the forensic science literature as AMP-FLPs, AFLPs, or AMFLPs (Inman and Rudin, 1997). The D1S80 tandem repeat unit is 16 bp in length except for the first repeat unit, which is 14 bp (Inman and Rudin, 1997; Kasai, *et al.*, 1990) and the number of repeats varies from about 14 to 41 (Applied Biosystems AmpliFLP D1S80 PCR Amplification Kit, 1993). The D1S80 is amplified by the primers complementary to the flanking conserved sequences (Kasai *et al.*, 1990). Since the D1S80 alleles differ in the number of tandem repeat units in the VNTR region, the different amplified alleles vary in length and can be identified by gel electrophoresis (Allen *et al.*, 1991; Kasai *et al.*, 1990), followed with silver staining (Applied Biosystems AmpliFLP D1S80 PCR Amplification Kit, 1993). The world population data for D1S80 as well as, HLA-DQA1 and PM is documented (Peterson *et al.*, 2000).

STRs

Short Tandem Repeats (STRs) are similar to the D1S80 system, except that the repeat units are shorter. STR loci consist of short repetitive sequence elements of di, tri, tetra and penta base pairs (Bacher and Schumm, 1998; Edwards *et al.*, 1991a,b,1992; Shuster, *et al.*, 1998). These repeats are well distributed throughout the human genome and are a rich source of highly polymorphic markers, which may be detected using the polymerase chain reaction (Busque *et al.*, 1997; Panneerchelvam *et al.*, 2004., Klintschar *et al.*, 1999; Pfitzinger *et al.*, 1995). Alleles of STR loci are differentiated by the number of copies of the repeat sequence contained within the amplified region and are distinguished from one another using radioactive, silver stain

or fluorescence detection following electrophoretic seperation.

STR typing is more tolerant of the use of degraded DNA templates than other methods of individual identification because the amplification products are less than 400bp long, much smaller than the material detected with AMP-FLP (Budowle *et al.*, 1991) or VNTR (Nakamura *et al.*, 1987) analysis. This format is also amenable to a variety of rapid DNA purification techniques. Like D1S80, STR loci are detected as discrete alleles and thus can be compared directly to an allelic ladder run on the same gel, simplifying comparison and analysis. Though each locus is only moderately polymorphic, many such loci, which exist, can be utilized simultaneously and in this respect, the system behaves like an RFLP. It is possible to choose from the enormous choice of STR loci found in human genome, those in which the distribution of STR alleles in any given population tends to be reasonably well distributed giving advantage over the D1S80 system.

Y – chromosome STR

DNA typing using STR loci has become the method of choice in forensic biology. Autosomal STR loci exhibit a Mendelian type of inheritance pattern with two codominant alleles present at each locus, and are subject to normal recombination and mutation events. In contrast, Y-chromosomal STRs are present only in male individuals with single alleles (Roewer *et al.*, 1992), as they are located on the non-recombining part of the Y chromosome. Thus the Y-STRs segregate as closely linked haplotypes in the male lineage of a family (De-Knijff *et al.*, 1997; Jobling *et al.*, 1997; Kayser *et al.*, 1997).

Y-chromosomal STR loci are of increasing interest in paternity testing, forensic casework, anthropological and evolutionary studies (Hidding and Schmitt, 2000; Junge, and Madea, 1999; Kayser *et al.*, 1997, 2003). The DNA extract is amplified by Y-STR specific primers and analysed manually through electrophoresis in non-denaturing polyacrylamide gel in horizontal or vertical electrophoresis system followed by silver staining (Graw and Seitz, 2000; Schneider *et al.*, 1998) or automatic means (Jiménez *et al.*, 2001; Prinz, *et al.*, 1997). Detailed databases with population

data were reported by several researchers (Da-Costa, 2002; Graw and Seitz, 2000; Jiménez *et al.*, 2001; Junge, and Madea, 1999; Schneider, 1998).

X – chromosome STR

Within the last few years several X chromosome (ChrX) markers have been established for forensic use, representing an efficient supplementation of autosomal and Y-chromosomal STR analyses (Edelmann *et al.*, 2002a; Hering *et al.*, 2001). In particular, ChrX haplotyping can clarify complicated kinship situations (Edelmann *et al.*, 2002b). Multiplex system of X-linked STR that could be useful in paternity studies and other forensic use has been documented (Zarrabeitia *et al.*, 2002). Allele frequency data and parameters of forensic relevance to ChrX markers have been published (Edelmann *et al.*, 2001).

Gender ID

In a forensic sample, it is often necessary to know the sex of the DNA source. The *amelogenin* locus shows a length variation between the sexes (Nakahori *et al.*, 1991) and employed by several crime laboratories (Gomez and Carracedo, 2000). One region of the female form of the gene contains a small deletion (6 bp) in non-essential DNA and produces a shorter product when amplified by PCR. When this region is analyzed, a female with two X chromosomes will show one band. A male with both an X and a Y chromosome will show two bands, one the same size as the female and one slightly larger or in some cases slightly smaller (Inman and Rudin, 1997). Amelogenin combined multiplex amplification STR typing is used in genotyping investigations (PROMEGA Technical Manual, GenePrint STR Systems [Silver Stain Detection], 2001).

Mitochondrial DNA

Mitochondrial DNA (mtDNA), in humans, exists in a high copy number (1000-10,000) copies per cell (Bogenhagen and Clayton, 1974) compared to the nuclear complement of only two copies in each cell. Human mitochondrial DNA, known as "the other human genome" (Johns, 1995) is a closed circular genome, approximately 16.5 Kb in length, including the 1.1 Kb-long noncoding region (control region). It

encodes 13 protein subunits of 4 biochemical complexes and the 24 structural RNAs (2 rRNAs and 22 tRNAs) that are required for the intramitochondrial translation of the protein-coding units. The complete nucleotide sequence of human mtDNA has been reported (Anderson et.al., 1981). It displays considerable sequence variation between individuals (Aquadro and Greenberg, 1983; Cann *et al.*, 1987). Mitochondrial DNA evolves at 5-10 times the speed of nuclear DNA and hence the considerable sequence variation between individuals (Cann *et al.*, 1987). Much of the variation is concentrated within two hypervariable regions (HV1 and HV2) which approximately enclose the non-coding D-loop region (Anderson *et al.*, 1981; Greenberg *et al.*, 1983; Vigilant *et al.*, 1989).

Mitochondrial DNA is useful for forensic examinations (Hagelberg and Sykes, 1989; Piercy *et al.*, 1993; Sullivan *et al.*, 1992) and also has been used extensively in two other major scientific realms. First, there are a number of serious human diseases caused by deleterious mutations in gene-coding regions of the mtDNA molecule, which have been studied by the medical profession to understand their mode of inheritance. Molecular anthropologists have been using mtDNA for almost more than two decades to examine both the extent of genetic variation in humans and the relatedness of populations all over the world. Because of its unique mode of maternal inheritance it can reveal ancient population histories, which might include migration patterns, expansion dates and geographic homelands. Since mtDNA is maternally inherited, these sequence differences enable the maternal lineage of individuals to be determined (Giles et. al. 1980; Gill *et al.*, 1994; Orrego and King, 1990; Weichhold *et al.*, 1998).

While in comparison with nuclear DNA, the number of copies of mitochondrial DNA in cells is much higher. Hence, mtDNA can often be typed more successfully than nuclear DNA, in particular, in stains with a low DNA load (Szibor *et al.*, 2000) or in degraded samples (Carracedo *et al.*, 1998). The investigation of mitochondrial DNA was shown to become more and more important for the identification of skeletal remains like bones (Sullivan *et al.*, 1992; Holland *et al.*, 1993) and teeth (Smith *et al.*, 1993). In forensic casework, mtDNA analysis is particularly important for the individualization of certain types of evidence, notably hair shafts, which contain little

or no genomic DNA (Carracedo *et al.*, 1998). MtDNA variation can be analysed by a variety of strategies (Alonso *et al.*, 1996; Barros *et al.*, 1997; Cann *et al.*, 1987; Stoneking *et al.*, 1991; Thomas *et al.*, 1994) but the combination of PCR amplification with direct DNA sequencing is usually the ultimate choice for identification (Bender *et al.*, 2000; Bertranpetit *et al.*, 1994; Carracedo *et al.*, 1998; Carracedo *et al.*, 2000; Fisher *et al.*, 1993; Sullivan *et al.*, 1994; Szibor *et al.*, 2000). In general the forensic procedure requires the amplification and sequencing of both the HV1 and HV2 regions from the sample and a reference standard before any comparison may be made (Holland *et al.*, 1995; Hopgood *et al.*, 1992; Orrego and King, 1990; Sullivan *et al.*, 1991; Tully *et al.*, 2001; Wilson *et al.*, 1993). Further literature are available on sequencing techniques for analyzing the mitochondrial control region for forensic purposes and population data (Baasner *et al.*, 1998; Ginther *et al.*, 1992; Melton *et al.*, 1997; Old *et al.*, 1993; Parson, *et al.*, 1998; Piercy *et al.*, 1993; Wittig *et al.*, 2000).

Paternity analyses

Conventional markers, ABO, Rh, MNSs, Kell-Cellano, P_1, Duffy, Lutheran, Kidd, Gc, Hp, Gm, Km, Tf, α_1-AT, EAP-AcP, PGM1, AK, ADA, EsD, 6-PGDH and GLO-I systems gave an exclusion probability of 97.32% only (Benciolini *et al.*, 1985). The combined use of RBC and HLA typing results has a mean exclusion capability of approximately 91 to 97% (Hilderson and Henry, 1985). Jeffreys *et al.* (1985a,b), establishes the DNA fingerprints as a single powerful test for positively reporting parentage.

Currently autosomal STRs are the most preferred choice in parentage analysis as well as Y-STRs and Mt-DNA, as situation demands (Baasner *et al.*, 1998; Carracedo *et al.*, 2000; Gill *et al.*, 2001; Ingvarsson *et al.*, 2000; Jobling *et al.*, 1997). Several useful databases are available (Budowle *et al.*, 1999; Busque *et al.*, 1997; Gamero *et al.*, 2000; Graw and Seitz, 2000; Halos *et al.*, 1999; Junge and Madea, 1999; Pu *et al.*, 1999; Tahir *et al.*, 2000; Wittig *et al.*, 2000).

Resolved cases in forensic science

Several cases have been resolved through DNA fingerprinting/typing, since its discovery by Jeffreys *et al.* (1985a,b). The first civil case, involving an immigration

dispute, had been satisfactorily resolved by DNA fingerprinting, in early 1985 (Jeffreys, *et al.*, 1985c). By late 1985, DNA evidence in the first paternity case was accepted by a magistrate's court, and 1986 saw the first application to criminal casework, involving the analysis of semen recovered from a rape victim (Gill and Werrett, 1987; Jeffreys, 1993). The history of the technique in India starts with the identification of body parts of a *Kodava* woman, by DNA fingerprinting, who was murdered on November 28, 1988 (Balakrishna, 1991).

Database

Comparison of a sample without a database, to a suspect utilizes only a small fraction of the potential of DNA typing. If DNA profiles from convicted criminals were stored in a databank, it could be searched for possible perpetrators of suspectless and serial crimes. Legislation mandating the collection and analysis of samples for DNA databanks, required (Herkenham, 2002; Inman and Rudin, 1997). In order for databanks to be most effective, especially on a national level, the DNA system used to create them must be standardized (Budowle *et al.*, 1998b; DNA Forensics Databases, 2003 – Base URL *http://www.ornl.gov/hgmis*). Geneally the data of convicted felons, crime scene specimens, unidentified human remains, relatives of missing persons etc., all find place in the DNA database (Herkenham, 2002).

DNA databanks have also been established in order to determine the frequencies with which forensic markers are found in different population groups. The frequency of occurrence for the DNA profile should be calculated using a scientifically valid method from an established population database (Budowle *et al.*, 1991,1995a,b, 1999; Garofano, *et al.*, 1998; Jankowski *et al.*, 1998; Meng *et al.*, 1999; Monson and Budowle, 1993; Peterson *et al.*, 2000; Pu, *et al.*, 1999; Sovinski, *et al.*, 1996).

Clinical applications

Clinical diagnosis for gene-based pre-symptomatic prediction of illness and adverse drug response are emerging at a rapid pace, and the therapeutic promise of genomics has ushered in an exciting phase of expansion and exploration in the commercial sector (Guttmacher and Collins, 2002). The sequencing of the human genome, along with other recent and expected achievements in genomics, provides an

unparalleled opportunity to advance our understanding of the role of genetic factors in human health and disease, to allow more precise definition of the non-genetic factors involved, and to apply this insight rapidly to the prevention, diagnosis and treatment of disease (Collins *et al.*, 2003).

A key application of human genome research has been the ability to find disease genes of unknown biochemical function by positional cloning (Rieder *et al.*, 1999). Positional cloning relies in the loss of heterozygosity analysis using polymorphic markers. The tedious procedure of finding candidate genes in identified loci in positional cloning has been made easy with the completion of HGP (International Human Genome Sequencing Consortium, 2001). The availability of the genome sequence also allows rapid identification of paralogues of disease genes, which is valuable for two reasons. First, mutations in a paralogous gene may give rise to a related genetic disease. A good example, discovered through use of the genome sequence, is achromatopsia (complete colour blindness). Second, the paralogue may provide an opportunity for therapeutic intervention, as exemplified by attempts to reactivate the fetally expressed haemoglobin genes in individuals with sickle cell disease or –thalassaemia, caused by mutations in the –globin gene.

In the list of advantages been found additionally, many of the diseases which are genetic disorders. Both environmental and genetic factors have roles in the development of any disease. There are four different types of genetic disorders: (1) single-gene, (2) multifactorial, (3) chromosomal, and (4) mitochondrial. Some of the examples are cystic fibrosis, sickle cell anemia, Marfan syndrome, Huntington's disease, hereditary hemochromatosis and Alzheimer's disease.

PCR has helped realize the potential of clinical DNA based diagnosis by producing enough of the target sequence so that simple, rapid, and robust methods for identifying could be employed. Following the first application of PCR in the prenatal diagnosis of sickle-cell anemia (Erlich, 1989), the technique has been widely used in the diagnosis of various diseases and detection of pathogens. A few of these cited are, ß-thalassemia, hemophilia A, Duchenne muscular dystrophy, HLA class II polymorphism, *ras* Oncogenes, the human retroviruses: human immunodeficiency

virus types 1 (HIV-1) and 2 (HIV-2) and human T cell lymphoma/leukemia virus types I (HTLV-1) and II (HTLV-II), hepatitis B virus, cytomegalovirus (CMV), genital human papillomaviruses (HPVs), enteroviruses: polioviruses, coxsackieviruses, echoviruses and hepatitis A (enterovirus 72) etc. (Erlich, 1989; Innis *et al.*, 1992). Automated PCR systems are available for the detection and quantification of infectious agents that cause disease. The PCR detection of disease diagnosis was improved by introducing real-time PCR.

Role of microsatellites

Huntington disease and some others have the involvement of trinucleotide microsatellite repeat expansion (Bio 2002 'Teaching in the Genome Age'2002 – http://www.woodrow.org/teachers/esi/2002/Biology/Projects/p3/disease.htm).
Trinucleotide repeat diseases can be categorized into two subclasses based on the location of the repeats: diseases involving noncoding repeats and coding sequences. In general trinucleotide repeat disorders are either dominantly inherited or X-linked, the one exception being Friedrich's ataxia, which is autosomal recessive (Cummings and Zoghbi, 2000; Goldstein and Schlotterer, 1999).

Huntington's disease is characterized by late onset dementia and loss of motor control, resulting in full-blown chorea after 10-20 years. Motor disorder is often preceded or accompanied by memory deficits, cognitive decline or changes in personality. Juvenile onset is rare and patients show rigidity, bradykinesia, epilepsy, severe dementia and an accelerated disease course. The disease is caused by the expansion of CAG coding repeat in the first exon of the HD gene in chromosome 4. Normal gene contains between 6 and 35 repeats and the affected gene from 36 to 121 repeats. Adult onset typically occurs when the repeat contains 40 – 50 units, whereas alleles containing more than 70 repeats typically result in the more severe juvenile form. The microsatellite adds a string of glutamine amino acids to the huntingtin protein.

Other diseases like Fragile X, Myotonic dystrophy, Spinalbulbar muscular atrophy, Friedrich's ataxia etc., also have the involvement of trinucleotide microsatellite repeat expansion (HGP, Human Genome Project Information, 2004).

Research had been conducted on the information as well as the usage of mononucleotide microsatellites in tumors (Alonso *et al.*, 2001) and di, tri and tetra nucleotide microsatellites in atypical hyperplastic (AH) breast lesions (Rosenberg *et al.*, 1996).

Role of SNPs

SNPs are the bedrock of human genetics; they can be used to track inheritance of any gene, contribute to the traits that make us unique, and underlie our susceptibilities to common diseases such as cancer, diabetes, heart disease etc., and also can be linked to sickle-cell anemia (Czarnik, 1998). They are used for association studies to identify the regions of the genome associated with a disease and as well may help explain why individuals respond differently to drugs (Gray *et al.*, 2000; HGP information, 2001- Base URL: *http://www.ornl.gov/hgmis*).

Role of mitochondria

Our understanding of the role of mitochondrial DNA in certain diseases has evolved rapidly since 1988, when the first mutations in mitochondrial DNA were discovered (Holt *et al.*, 1988; Wallace *et al.*, 1988). Such mutations have subsequently been identified in a variety of diseases (Brown and Wallace, 1994; Howell, 1994; Schon *et al.*, 1994; Wallace, 1992) and the pathogenic role of cumulative mitochondrial-DNA damage is being explored in many common diseases that develop late in life, and even in the aging process itself (Golden and Melov 2001; Shigenaga *et al.*, 1994; Wallace, 1997).

Ancient DNA and human evolution

Molecular biology is now a major source of quantitative and objective information about the evolutionary history of the human species. One of the more curious, but far from least interesting, avenues made accessible by this new molecular biology technology is the molecular analysis of DNA from preserved organisms of the recent or distant past, the ancient DNA studies. It has provided new insights into our genetic divergence from apes (Brown *et al.*, 1982) and into the way in which humans are related to one another genetically (Nei, 1985; Nei and Roychoudhury, 1982). Our picture of genetic evolution within the human species is clouded, because it is based

mainly on comparisons of genes in the nucleus. Mutations accumulate slowly in nuclear genes. In addition, nuclear genes are inherited from both parents and mix in every generation. This mixing obscures the history of individuals and allows recombination to occur. Recombination makes it hard to trace the history of particular segments of DNA unless tightly linked sites within them are considered.

World-wide survey of mtDNA by Cann *et al.*, (1987) adds to knowledge of the history of the human gene pool in three ways. First, mtDNA gives a magnified view of the diversity present in the human gene pool, because mutations accumulate in this DNA much faster than in the nucleus (Wilson *et al.*, 1985). Second, because mtDNA is inherited maternally and does not recombine (Olivo *et al.*, 1983), it is a tool for relating individuals to one another. Third there are about 10^{16} mtDNA molecules within a typical human and they are usually identical to one another (Brown, 1980). Typical mammalian females consequently behave as haploids, owing to a bottleneck in the genetically effective size of the population of mtDNA molecules within each oocyte (Hauswirth and Laipis, 1986). This maternal and haploid inheritance means that mtDNA is more sensitive than nuclear DNA to severe reductions in the number of individuals in a population of organisms (Wilson *et al.*, 1985). A pair of breeding individuals can transmit only one type of mtDNA. The fast evolution and peculiar mode of inheritance of mtDNA provide new perspective on how, where and when the human gene pool arose and grew.

Cann *et al.*, (1987), has studied mtDNAs from 147 people, drawn from five geographic populations through analysis of restriction mapping. The geographic regions consist of Africans (representing the sub-Saharan region), Asians (originating from China, Vietnam, Lacs, Philippines, Indonesia and Tonga), Caucasians (originating from Europe, North Africa, and the Middle East), aboriginal Australians and aboriginal New Guineans. Evolutionary tree has been built by the *parsimony* method based on these mtDNA data and arrived at the finding that, all these mitochondrial DNAs stem from one woman who is postulated to have lived about 200,000 years ago, probably in Africa. All the populations examined except the African population have multiple origins, implying that each area was colonized repeatedly. The findings of the above study fits with one view of the fossil record: that

the transformation of archaic to anatomically modern forms of Homo sapiens occurred first in Africa (Brauer, 1984; Rightmire, 1984), about 100,000 – 140,000 years ago and that all present-day humans are descendants of that African population.

The molecular study of genes from different organisms is helping biologists to build trees relating these organisms to one another. These trees contain information about both the order of branching of lineages linking ancestors to modern descendants and the approximate times at which the branching events occurred. The success of the molecular tree approach to the study of evolution results from two major findings: molecular evolution is dominated by mutations that are inconsequential or nearly so from the standpoint of natural selection (Kimura, 1983) and these mutations accumulate at fairly steady rates on surviving lineages (Zuckerkandl and Pauling, 1965). Greenberg *et al.* (1983), sequenced the large noncoding region, which includes the displacement loop (D loop), from four Caucasians and three Black Americans. A *parsimony tree* for these seven D loop sequences, rooted by the midpoint method, was constructed by Cann *et al.* (1987) and the tree indicated (1) a high evolutionary rate for the D loop (at least five times faster than other mtDNA regions), (2) a greater diversity among Black American D loop sequences, and (3) that the common ancestor was African.

The reproducible retrieval of ancient DNA sequences became possible with the invention of the polymerase chain reaction (Mullis and Faloona 1987; Pääbo *et al.*, 1989). However, theoretical considerations, (Lindahl 1993a; Pääbo and Wilson 1991) as well as empirical studies (Hōss *et al.*, 1996a; Pääbo 1989), show that DNA in fossil remains is highly affected by hydrolytic as well as oxidative damage. Therefore, the retrieval of DNA sequences older than about 100,000 years is expected to be difficult, if not impossible, to achieve (Pääbo and Wilson, 1991). It is noteworthy, though, that even among remains that are younger than 100,000 years most fail to yield amplifiable DNA sequences (Hōss *et al.*, 1996b). In addition, contamination of ancient specimens and extracts with modern DNA poses a serious problem (Handt *et al.*, 1994a) that requires numerous precautions and controls. Therefore, a number of criteria need to be fulfilled before a DNA sequence determined from extracts of an ancient specimen can be taken to be genuine (Handt *et al.*, 1994a; Handt *et al.*, 1996; Lindahl 1993b; Pääbo

et al., 1989). MtDNA sequences have been recovered from an approximately 5000 year-old mummified human body found in the Tyrolean Alps (Handt *et al.*, 1994b), animal bones dating back at least 90,000 years (Loreille *et al.*, 2001) and perhaps 150,000 years or more (Jones, 2001) but these remains came from either cave or permafrost contexts where low temperatures helped preserve the DNA (Handt *et al.*, 1994b). Similarly, the Neanderthal bones that have yielded mtDNA sequences have all come from relatively cool European caves (Smith *et al.*, 2001) and scientists believe, temperture is not the only factor, that influences the survival of DNA (Ovchinnikov, *et al.*, 2001).

Studies of Vigilant *et al.*(1991) depicts (i) the results of a study of sequences of two hypervariable mtDNA segments from 189 individuals, including 121 native Africans; (ii) a tree relating these sequences to a published chimpanzee sequence, thereby permitting the use of a superior method (that is, the outgroup method) for placing the common human mtDNA ancestor on the tree; (iii) more rigorous statistical tests of the geographic origin of the mtDNA ancestor (including a new test); and (iv) a new estimate, based on a comparison of the chimpanzee and human sequences, of when the human mtDNA ancestor lived. Summary of their work and some of the important findings (Ingman, 2001; Ingman *et al.*, 2000) are given below.

The control region is a 1122-base pair segment of noncoding DNA and is the most rapidly evolving and polymorphic region of the human mtDNA genome. Sequence analysis of this region therefore affords the maximum resolution for distinguishing among even very closely related mtDNAs. The 189 individuals studied included 121 native Africans from the following sub-Saharan populations (and places): 25 Kung (Botswana and Namibia), 27 Herero (Botswana), 1 Naron (Botswana), 17 Hadza (Tanzania), 14 Yorubans (Nigeria), 20 Eastern Pygmies (Zaire), and 17 Western Pygmies (Central African Republic). The 68 additional individuals include 20 Papua New Guineans, 1 native Australian, 15 Europeans, 24 Asians, and 8 African Americans. The New Guineans are from various parts of coastal and highland Papua New Guinea; the Australian is from Perth. The Europeans, Asians, and African Americans are heterogeneous with respect to geographic origin.

Comparison of approximately 610 nucleotides from each individual revealed substitution at 179 sites and length changes at 22 sites. Each unique sequence is termed a mtDNA type; these 201 polymorphic sites defined 135 types among the 189 individuals. Sixteen mtDNA types were found to occur more than once in the sample of 189 individuals. People with identical mtDNA types were not found among the geographically heterogeneous Europeans and Asians, but only within individual populations of Africans or within the Papua New Guinea population. There was no sharing of mtDNA types among people from different populations, with one apparent exception. This exception involves a Yoruban and an African American with identical mtDNA types and thus is consistent with the view that African American, stem mainly from West Africa. Even groups as closely related as the Eastern and Western Pygmies did not share any mtDNA types (Vigilant *et al.*, 1989). Based on the above findings, they conclude that mtDNA types defined by control region sequences display strong geographic specificity, in agreement with other studies of human mtDNA variation that used techniques capable of distinguishing among closely related mtDNAs (Cann *et al.*, 1987; Stoneking, 1990; Vigilant *et al.*, 1989). An additional implication is that most women have moved their home bases extremely slowly, especially in nonagricultural populations (Di Rienzo and Wilson, 1991; Vigilant *et al.*, 1989). This encourages the expectation, in principle, of determining where the last common ancestor of extant mtDNAs lived and for which a tree relating mtDNAs was built.

A genealogical tree relating the 135 mtDNA types was built using the *parsimony* method (Swofford and Olsen, 1990), in which a branching network is constructed in an effort to minimize the number of mutations required to relate the types. To convert the resulting network into a tree, the ancestor or root must be placed, which requires additional information or assumptions. Cann *et al.* (1987) used the midpoint method of rooting their tree, assuming that the rate of evolution has been the same in all lineages. If, however, mtDNA evolution were faster in Africans, then the deep African lineages revealed by midpoint rooting of their tree would not indicate an African origin. Instead, these ostensibly deep African lineages would actually be shallow lineages along which more mutations had accumulated. Hence, the tree might not yield any information regarding the geographic origin of the mtDNA ancestor, suggested by some researchers elsewhere. The outgroup method (Swofford and Olsen,

1990) is a preferable method of rooting a tree because it does not rely on assumption that the rate of evolution is the same in all lineages. This method used a sequence from another species (the "outgroup"), such as an African ape, to place the human mtDNA ancestor on the network. The outgroup attaches to the network relating the human mtDNA types at the position that minimizes the total number of mutations in the tree. The point of attachment is the position of the human mtDNA ancestor on the tree. Though Cann *et al.* (1987), could not use the outgroup method due to the unavailability of high-resolution restriction maps of African ape mtDNA, an already published control region sequence from a common chimpanzee had been used to root the tree in the figure constructed by Vigilant *et al.* (1991). A perceived shortcoming of the restriction map study (Cann *et al.* 1987) concerned the use of 18 African Americans as sources of authentic African mtDNAs (Excoffier and Langaney, 1989; Spuhler, 1988). Evidence supporting the African classification of these types has already been discussed (Cann *et al.*, 1987). Other studies also favor the authentication of African American, considered as the sources of African mtDNA (Vigilant *et al.*, 1989).

The above study (Vigilant *et al.*, 1991) concludes on the inferred place of origin of the last common ancestor of human mtDNA in Africa, which was the hypothesis reached by the restriction analysis study by Cann *et al.* (1987). Statistical significance as well as mtDNA sequence differences were taken into consideration to reach to the conclusion. Regarding the date for the common mtDNA ancestor, it is suggested the range of 166,000 to 249,000 years ago which is consistent with the range of 140,000 to 290,000 years ago (Cann *et al.*, 1987), 172,000 years ago (Kocher and Wilson, 1991) and 280,000 years ago (Vigilant *et al.*, 1991). Their findings strongly supports the contention that all the mtDNAs found in contemporary human populations stem from a single ancestral mtDNA that was present in an African population approximately 200,000 years ago. The mtDNA evidence is thus consistent with an origin of anatomically modern humans in Africa within the last 200,000 years, with subsequent migrations out of Africa that established human populations in Eurasia (Cann *et al.*, 1987; Stoneking and Cann, 1989). Evidence from chromosomal genes and their products are also consistent with the African origin hypothesis. The absence of human mtDNAs that diverged from one another more than about 280,000

years ago leads to the inference that the migrating human populations from Africa probably replaced the resident Eurasian populations that descended from earlier migrations of *Homo erectus* from Africa 800,000 to 1,000,000 years ago (Wolpoff, 1980). However, there are other interpretations of the fossil record that do support the African origin hypothesis (Brauer, 1989; Stringer, 1988).

Several other studies are published on the origin of our species. Attempts have been made to date the greatest divergence between mtDNAs of all humans currently existing in order to estimate the date of the maternal MRCA (most recent common ancestor). Using restriction site data (Cann *et al.*, 1987), DNA sequence data (Horai *et al.*, 1995; Vigilant *et al.*, 1991) and different evolutionary rate calibrations (Stoneking *et al.*, 1992), dates on the order of 200,000 years ago have been obtained. Using particular models (Pesole *et al.*, 1992), or slowly occurring transversional substitutions (Wills, 1995), the date may be as far back as 300,000 – 800,000 years, but the majority of analyses still point to dates on the order of 200,000 years ago (Haeseler *et al.*, 1996).

Initial analyses of mitochondrial genomes from around the world indicated that the greatest divergence in a phylogenetic tree was between a branch leading to exclusively African mitochondrial types and branches leading to African as well as other types (Cann *et al.*, 1987; Vigilant *et al.*, 1991). Such a branching structure of the tree would indicate that the ancestor lived in Africa. However, reanalyses of the nucleotide sequence data showed little statistical support for this conclusion (Hedges *et al.*, 1992; Maddison, 1991; Maddison *et al.*, 1992; Templeton, 1992). Various novel methods were used to analyse these data for a third time. Penny *et al.*(1995), located the root among African sequences by using an approach that allows many trees of equal quality to be evaluated, giving more weight to nucleotide positions that are less affected by multiple substitutions. Furthermore, a new tree reconstruction program that allows a *maximum likelihood approach* to be used on large data sets (Strimmer and Haeseler, 1996) shows that the root supports an African origin with 85% reliability as does a principal components analysis of the data (D'Andrade and Morin, 1996). Finally, an *outgroup* was provided that is less affected by multiple substitutions than the chimpanzees by isolating a partial mitochondrial control region sequence that was

inserted in chromosome 11 shortly before the origin of the modern human mitochondrial gene pool. This sequence again indicates that the root lies in Africa. Thus, after the third generation of analyses of the mitochondrial data, the evidence suggests that the best placement of the root is in Africa (Haeseler *et al.*, 1996).

Nei and Takezaki (1996), reanalyzed five different data sets based on frequencies of microsatellites, restriction fragment length polymorphisms, protein polymorphisms and insertions of *Alu* elements in different human populations, using chimpanzees as an outgroup. In all five cases, the roots of the trees relating the human population to each other fell on the branch connecting Africans and non-Africans (Fig. 2.2), indicating that the former have the longest independent history. In four of the data sets, support for this conclusion was statistically significant. Furthermore, in an attempt to date the divergence between African and non African populations based on microsatellites, Goldstein *et al.*(1995) have dated the divergence to 56,000 – 213,000 years by assuming that microsatellites evolve by the gain and loss of single repeat units. This also lends further support to the recent African origin of our gene pool.

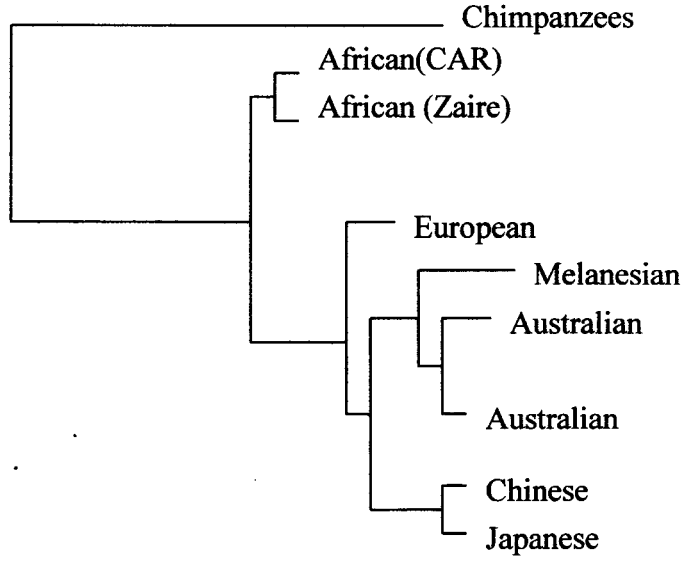

Fig. 2.2. A population tree relating population based on restriction fragment length polymorphisms (Nei and Takezaki, 1996). The root of the tree is determined by the chimpanzee and falls between African and non-African populations, indicating that Africans have the longest independent history among the populations analysed (From Haeseler *et al.*, *Nat. Genet.* 14: 135-140, 1996).

Several models for the origin of *Homo sapiens* have been proposed. The "multiregional origin" model suggests that there was no single origin for all modern

humans (Wolpoff, 1989; Wolpoff, 1992). After the radiation of *Homo erectus* from Africa into Europe and Asia 800,000 to 1.8 million years ago (Stringer, 1992). There was a continuous transition among regional populations from *H. erectus* to *H. sapiens*. Such "parallel evolution" among geographically dispersed populations could have been achieved by considerable amounts of gene flow between populations (Wolpoff, 1989; Wolpoff, 1992). By contrast, the "out of Africa" model suggests that all non-African human populations descend from an anatomically modern *H. sapiens* ancestor that evolved in Africa approximately 100,000 to 200,000 years ago and then spread and diversified throughout the rest of the Earth, supplanting any *Homo* populations still present outside of Africa (Stinger and Andrews, 1988; Swisher lll *et al.*, 1994; Wolpoff, 1989). As already seen through various sources, it was proposed through the study of mtDNA analysis that all modern mtDNA can be traced back through the maternal lineage to a single ancestor that existed in Africa between 100,000 to 300,000 years ago (Cann *et al.*, 1987; Nei and Livshits, 1989; Vigilant *et al.*, 1991). As well, several mtDNA studies (Horai *et al.*, 1995; Penny *et al.*, 1995; Ruvolo *et al.*, 1991; Tamura and Nei, 1993), Y chromosome (Dorit *et al.*, 1995; Hammer, 1995; Underhill *et al.*, 2000; Whitfield *et al.*, 1995) support the original findings of a recent origin of all modern humans. Nuclear autosomal genome data study also strongly supports the out of Africa model of human origins and the recency of the emigration from Africa (Tishkoff *et al.*, 1996).

Scientists have now succeeded in studying the entire sequence of the mtDNA molecule, in a worldwide sample of 53 people (Ingman *et al.*, 2000). As with previous mtDNA studies, this also found, Africans to have the most diverse and therefore the most ancient, mtDNA sequences. It also determined the interesting discovery that African and non-African populations diverged not 100,000 years ago, as some other genetic studies have suggested, but as recently as 52,000 years ago. The results also indicated that after their split from Africans, non-Africans began expanding in population about 38,500 yeas ago.

Neandertals and modern humans

Neandertals are a group of extinct hominids that inhabited Europe and western Asia from about 300,000 to 30,000 years ago. During part of this time they coexisted

with modern humans. Based on morphological comparisons, it has been variously claimed that Neandertals: (1) were the direct ancestors of modern Europeans; (2) contributed some genes to modern humans; or (3) were completely replaced by modern humans without contributing any genes (Brauer and Stringer 1997). Analyses of molecular genetic variation in the mitochondrial and nuclear genomes of contemporary human populations have generally supported the third view, that, Neandertals were a separate species that went extinct without contributing genes to modern humans (Armour *et al.*, 1996; Cann *et al.*, 1987; Hammer, 1995; Tishkoff *et al.*, 1996; Vigilant *et al.*, 1991). However, these analyses rely on assumptions, such as the absence of selection and a clock-like rate of molecular evolution in the DNA sequences under study, whose validity has been questioned (Wolpoff, 1989; Templeton, 1992). An additional and more direct way to address the question of the relationship between modern humans and Neandertals would be to analyze DNA sequences from the remains of Neandertals.

Based on its "classical" morphology, the undated Neandertal fossil is thought to be between 30,000 and 100,000 years old. Several modifications are done to the experiment to conclude the mitochondrial sequence as endogenous to the Neandertal fossil. Since 1991, the Neandertal-type specimen, found in 1856 near Düsseldorf, Germany, has been the subject of an interdisciplinary project of the Rheinisches Landesmuseum Bonn, initiated and led by Schmitz, R.W., (Schmitz, 1996; Schmitz *et al.*, 1995). As a part of this project, a sample was removed from the Neandertal specimen, subjected to mtDNA analysis (Krings *et al.*, 1997) and were able to amplify a sequence of 379 bases. Krings *et al.*, then compared this sequence against a database of 994 different mtDNA sequences from modern humans and the results put the Neandertal genome well outside the limits of modern humans. Another interesting result is that the mtDNA sequence seemed equally distant from all modern groups of humans. In particular, it did not seem to be more closely related to Europeans, something that might have been expected if, as some scientists think, Neandertals were at least partly ancestral to them.

Phylogenetic analyses conducted by the same researchers, agree with the pairwise comparisons of sequence differences in placing the Neandertal mtDNA

sequence outside the variation of modern human mtDNA. The study also yielded a date of 550,000 to 690,000 years before present for the divergence of the Neandertal mtDNA and contemporary human mtDNAs. When the age of the modern human mtDNA ancestor is estimated, a date of 120,000 to 150,000 years is obtained, in agreement with previous estimates (Cann *et al.*, 1987; Vigilant *et al.*, 1991). These studies indicate that the age of the common ancestor of the Neandertal sequence and modern human sequence is about four times greater than that of the common ancestor of modern human mtDNAs. Again through phylogenetic tree, they suggest an African origin of the human mtDNA gene pool. The view that Neandertals would have contributed little or nothing to the modern human gene pool is gaining support from studies of molecular genetic variation at nuclear loci in humans (Armour, *et al.*, 1996; Hammer 1995; Tishkoff *et al.*, 1996). It is also in agreement with assessments of the degree of morphological difference between Neandertal skeletal remains and modern humans (Hublin *et al.*, 1996; Rak, 1993; Schwartz and Tattersall, 1996; Zollikofer *et al.*, 1995) that would classify Neandertals and modern humans as separate species.

In 1999, the same workers successfully extracted a second sequence of 340 base pairs of mtDNA from the same Neandertal fossil (Krings *et al.*, 1999). This study confirmed the results of the first. Again, in the same year scientists have successfully extracted a 345 base pair sequence of mtDNA from a second Neandertal, a 29,000 year-old fossil of a baby recently discovered in Mesmaiskaya cave in south-western Russia (Höss 2000; Ovchinnikov *et al.*, 2000). The results of this study were similar to the previous ones, putting the Mezmaiskaya specimen outside the range of modern human mtDNA. In 2000, scientists announced the sequencing of a third Neandertal mtDNA specimen from a cave at Vindija, Croatia (Krings *et al.*, 2000). When the three Neandertals are compared with modern humans, all three of them cluster together, and apart from all modern humans. This conclusion has been reinforced by a study by Knight (2003). His study strongly confirmed earlier ones showing deeply divergent histories for modern human mtDNA lineages and the known Neandertal ones.

Foley (2003), after reviewing several scientific literatures, some of which are already noted in this context, narrates the findings as follows. The studies of Neandertal mtDNA do not show that Neandertals did not or could not interbreed with

modern humans. However, the lack of diversity in Neandertal mtDNA sequences, combined with the large differences between Neandertal and modern human mtDNA, strongly suggest that Neandertals and modern humans developed separately, and did not form part of a single large interbreeding population. The Neandertal mtDNA studies will strengthen the arguments of those scientists who claim that Neandertals should be considered a separate species, which did not significantly contribute to the modern gene pool.

Population studies

Several valuable population studies have been conducted by various researchers using a wide variety of polymorphic loci (Baasner *et al.*, 1998; Budowle et. al. 1995a, 1999; Bruford and Wayne, 1993; Corte-Real *et al.*, 1999a,b; Da-Costa 2002; Graw and Seitz, 2000; Jiménez *et al.*, 2001; Junge and Madea, 1999; Melton *et al.*, 1997; Pfitzinger *et al.*, 1995; Schneider, 1998). Kivisild *et al.*(2003) studied the Indian tribal and caste populations regarding their genetic heritage using mtDNA, Y chromosome and one autosomal locus. Studies of Bamshad *et al.* (1996), showed that the caste populations of Andhra Pradesh, cluster more often with Africans than with Asians or Europeans which is suggestive of admixture with African populations. Employing the examination of SNP sites in combination with STRs, on three caste groups (Vizag Brahmins, Peruru Brahmins, Kammas), three tribes (Bagata, Poroja, Valmiki) and an additional group (the Siddis) of African ancestry, all inhabitants of Andhra Pradesh, finds gene flow among these populations, (Ramana *et al.*, 2001).

Studies on genetic heritage of Indian populations have found evidence to support either a European or an Asian origin of Indian caste populations, with occasional indications of admixture with African or proto-Australoid populations (Bamshad *et al.*, 1996,1997; Chen *et al.*, 1995; Kivisild *et al.*, 1999; Majumder *et al.*, 1999; Mountain *et al.*, 1995; Passarino *et al.*, 1996; Quintana-Murci *et al.*, 1999). Bamshad *et al.*(2001) employing mtDNA, Y-chromosome and autosomal loci polymorphisms have demonstrated that the upper castes of India have a higher affinity to Europeans than to Asians, and the upper castes are significantly more similar to Europeans than are the lower castes. Also it shows a trend toward upper castes being more similar to Europeans, whereas lower castes are more similar to Asians. As a

conclusion they suggest, Indian castes are most likely to be of proto-Asian origin with West Eurasian admixture resulting in rank-related and sex-specific differences in the genetic affinities of castes to Asians and Europeans.

Molecular systematics

The genome is a dynamic structure, continually subjected to modification by the forces of evolution. The genomic variation seen in humans represents only a small glimpse through the larger window of evolution, where hundreds of millions of years of trial-and-error efforts have created today's biosphere of animal, plant and microbial species. A complete elucidation of genome function requires a parallel understanding of the sequence difference across species and the fundamental processes that have sculpted their genomes into the modern-day forms. The study of inter-species sequence comparisons is important for identifying functional elements in the genome. Beyond this illuminating role, determining the sequence differences between species will provide insight into the distinct anatomical, physiological and developmental features of different organisms that will help to define the genetic basis for speciation and will facilitate the characterization of mutational processes (Collins *et al.*, 2003).

Various types of molecular data provide a wealth of information with which to address problems at all taxonomic levels. Studies addressing species-level problems have involved the analysis of nuclear, chloroplast and mitochondrial DNA at different levels of sequence resolution using RFLPs, RAPD, DNA sequencing, single copy DNA-DNA hybridization and others. Several of these recent advances in nucleic acid technology has been used in the taxonomic studies of living organisms (Claridge *et al.*, 1997). Susequent to the discovery of a group of living organisms, the "archaea" in 1977, a major modification occurred in the classification of living cells, division into three domains as, *bacteria, archaea* and *eukarya.* The complete genome sequencing of an archaeon, *Methanococcus jannaschii* in1996,strongly prove the above(Gupta 1999).

Assignment of species to individual viruses can be done after the examination of the entire sequence of the virus (Van Regenmortel, 1997). The motive force in contemporary bacterial systematics owes much to developments in molecular biology. These advances underpin an impressive array of methods, which can be used to derive

taxonomically useful information from analyses of nucleic acids. The estimation of the mean overall base composition of DNA, indirect comparisons of nucleotide sequences by DNA: DNA pairing and sequence analyses of conserved homologous genes, notably 16S rRNA, all provide grist for the circumscription of bacterial species (Goodfellow *et al.*, 1997). One of the great strengths of sequence data is that it accumulates, so that old and new sequences can be directly compared, to produce an increasingly comprehensive taxonomic analysis of the microbial world. However, this can only be done if the stretches of sequence from different studies cover the same parts of the molecule. Species concepts of further living organisms like algae, fungi, lichens, plants, marine invertebrates, and several others have been addressed (Claridge *et al.*, Eds.1997). Recent molecular studies have resolved phylogenetic relationships between the four extant Arthropod subphyla, Hexapoda, Myriapoda, Chelicerata and Crustacea (Hwang *et al.*, 2001) and several other works enhance the phylogeny of Arthropoda.

The mitochondrial cytochrome b (COB) and cytochrome c oxidase subunit II (COII) genes have been used in several recent molecular systematic studies of rodents (DeWalt *et al.*, 1993; Ma *et al.*, 1993; Thomas and Martin, 1993), ungulates (Irwin *et al.*, 1991; Irwin and Wilson, 1992; Miyamoto *et al.*, 1994), marine mammals (Irwin and Arnason, 1994), primates (Adkins and Honeycutt, 1994; Disotell *et al.*, 1992; Ruvolo *et al.*, 1991) and eutherian mammal orders (Adkins and Honeycutt, 1991, 1993; Honeycutt and Adkins, 1993). The evolution of COB and COII has been examined in several eutherian mammal orders, with special emphasis on the orders Artiodactyla and Rodentia and discussed the implications of these findings for the molecular systematics of mammals (Honeycutt *et al.*, 1994). Hoss *et al.* (1996b) amplified DNA from an extinct sloth, *Mylodon darwinii*, and sequenced the mitochondrial rDNA. Phylogenetic analyses using homologous sequences from all extant edentate groups suggest that *M. darwinii* was more closely related to the two-toed than the three –toed sloths and thus, that an arboreal life-style has evolved at least twice among sloths. The divergence of *Mylodon* and the two-toed sloth furthermore allows a date for the radiation of armadillos, anteaters and sloths to be estimated. They also concluded that the edentates differ from other mammalian orders in that they contain lineages that diverged before the end of the Cretaceous Period.

Sample collection

Samples were collected from subjects representing the districts of Kasargod, Kannur, Wayanad and Kozhikode, which comprise most of the North Kerala (Fig. 3.1) and from Mahe, an area of Pondicherry State but within the study area. The condition of anonymity was adhered to while procuring the samples.

Fig. 3.1. Map of Kerala State to Show the Districts: Kasargod, Kannur, Wayanad and Kozhikode Comprising Most of the North Kerala. Also Noted, Mahe of Pondicherry State (From Arya, R.P., Kerala Road Atlas & Distance Guide, 2002)

Two categories of collection have been performed.

(I) The samples which have been used for the purpose of general study.

(II) The samples of two entirely different social communities and its study.

3 MATERIALS AND METHODS

Category I. The samples were collected from healthy individuals at random without any consideration on religion, caste, sex or age. The general procedure of sample collection and processing was according to Busque, et al. (1997); Klintschar, et al. (1998) and Vargas, et al. (2001). Most of the samples are buccal epithelium collected through sterile cotton swabs from the donor's mouth and kept back in its individual container. A pair each buccal swab were collected and wherever possible about 2ml of liquid blood through venipuncture by the assistance of a doctor. The blood was mixed thoroughly with the potassium EDTA, through gentle shaking of the bottle by hand for approximately 1 - 2 minutes. Plucked head hairs having good roots, about 4 to 6, were collected from the individuals reluctant to give buccal swabs or liquid blood, and kept in the small polythene bags. Extreme care has been taken to avoid any kind of cross contamination from any source. The details of the first category are given in Table 3.1 (the no. of donors coincide with the type of sample).

Table 3.1. Details of the samples of Category I

District	Liquid Blood	Buccal Swab	Hair	Total Donor
Kasargod	2	10	1	13
Kannur	8	22	3	33
Wayanad	2	11	6	19
Kozhikode	11	44	5	60
Mahe*	—	2	—	2

Grand Total Donors: 127

* *Mahe* – Not a district of Kerala State. It is an area coming under Pondicherry State, but within the study area.

Category II. Collection is at random with same specifications as of category I, except these were from two entirely different social communities (Corte-Real *et al.*, 1999), *Mukkuvas* (also called as *Dheevaras*) residing in Kozhikode and Kannur districts and *Paniyan* of Wayand (Table 3.2).

Mukkuvas: They mainly reside in the coastal area of Kerala and are traditionally engaged in fishing in the sea. They also have an isolated community of their own with arranged marriages among themselves, though they join hands in normal cultural

activities with the general local population. Genotyping of samples collected from the *Mukkuva* individuals residing the coastal belt of approximately 25 kilometers, spread South-North from Vadakara to Thalassery in Kozhikode and Kannur districts were done.

Paniyan: They are one of the *adivasi* tribes residing in the high-range district, Wayanad and can be seen throughout and are traditionally labourers in plantations, estates, paddy fields and other crops. They lead an isolated life from other social groups / tribes of the district.

Table 3.2. Details of the samples of Category II

Social Group	Liquid Blood	Buccal Swab	Hair	Total Donor
Mukkuva	2	23	7	32
Paniyan	—	29	2	31

Sample processing

Storage

Liquid blood samples: All of the samples were stored in appropriate conditions soon after reaching the laboratory. 500 µl of liquid blood spotted on a 4cm x 3cm sterile cotton cloth stapled on a plain paper card 10cm x 6 cm with details noted and kept for drying in room temperature in a fume hood. Once, dried completely, mostly by overnight, transferred in individual small polythene bags, sealed and kept in –20°C freezer. Rest of the liquid blood kept as such in refrigerator.

Buccal swabs: Dried in room temperature, returned in the individual container and stored in –20°C freezer.

Hair: Stored in –20°C freezer.

DNA extraction

The DNA was extracted through Chelex method according to Applied Biosystems, AmpliType User's Guide, Version 2, (1993) and Walsh *et al.*(1991). The extracted DNA was stored in –20°C until used in PCR applications.

DNA quantitation

The extracted DNA was assayed (Fig. 3.12) for its DNA content using QuantiBlot Human DNA quantitation kit of Applied Biosystems (APPLIED BIOSYSTEMS QuantiBlot Human DNA Quantitation Kit Protocol, 2000).

Fig. 3.12.

Concentration of DNA

When the yield of DNA was not enough, it was concentrated in order to achieve the required quantity of DNA for amplification in the recommended volume using Centricon-100 ultrafiltration, according to the manufacturer's protocol (Millipore Centricon Centrifugal Filter Devices, User's Guide 2000). The samples were concentrated to about 15 to 25 µl in a Dynac Centrifuge with fixed angle rotor (APPLIED BIOSYSTEMS AmpliType User's Guide, Version 2, 1993).

Amplification by PCR

Amplification kits, GenePrint STR Multiplex Systems (Silver Stain Detection) – CSF1PO, TPOX, THO1 (CTT Multiplex), F13A01, FESFPS, vWA (FFv Multiplex), D16S539, D7S820, D13S317 (SilverSTR III Multiplex) obtained from PROMEGA Corporation, contain all the nine selected STR loci used in this study (PROMEGA GenePrint STR Systems [Silver Stain Detection] – Technical Manual, 2001). GenePrint Sex Identification Monoplex System –Amelogenin (Silver Stain Detection) from the same source was used to find the sex of few samples. Each of the above multiplex systems contain the appropriate locus-specific primer pairs and allelic ladder in addition to STR 10X Buffer, K562 DNA, STR 2X Loading Solution and pGEM DNA Markers. The required primer pairs are as a mixture for simultaneous amplification of more than one locus and also has a mixture of the allelic ladders for the same set of loci.

The CTT, FFv and SilverSTR III multiplex amplification was done according to the manufacturer's protocol (PROMEGA GenePrint STR Systems [Silver Stain Detection] – Technical Manual, 2001) using a Perkin-Elmer – DNA Thermal Cycler 480 (APPLIED BIOSYSTEMS DNA Thermal Cycler 480, User's Manual, 1995). The

amplified samples were stored in –20°C. The details of the amplification protocols are given below.

System	Initial Incubation	Cycling for First 10 Cycles	Cycling for Last 20 Cycles	Extention Step	Hold Step
CTT Multiplex with or without Amelogenin	96°C for 2 minutes	94°C, 1 minute 64°C, 1 minute 70°C, 1.5 minute	90°C, 1 minute 64°C, 1 minute 70°C, 1.5 minute	None	4°C
FFv or SilverSTR III Multiplex	96°C for 2 minutes	94°C, 1 minute 60°C, 1 minute 70°C, 1.5 minute	90°C, 1 minute 60°C, 1 minute 70°C, 1.5 minute	60°C for 30 minutes	4°C

Selected STRs – locus specific information

The selected STRs locus specific information (CSF1PO, TPOX, THO1, F13A01, FESFPS, vWA, D16S539, D7S820, D13S317 and Amelogenin) are given in Figs. 3.2 – 3.11 (PROMEGA GenePrint STR Systems [Silver Stain Detection] – Technical Manual, 2001 and Chromosome Ideogram information from NCBI Map Viewer, 2004).

Figs. 3.2 – 3.11.

PAGE

Polyacrylamide gel preparation

6% denaturing polyacrylamide gel (17cm wide x 32cm high x 0.4mm thick) suitable for GIBCOBRL/LIFE TECHNOLOGIES, SA32 sequencing gel apparatus (GIBCOBRL/LIFE TECHNOLOGIES Instruction Manual - Model SA Adjustable Sequencing Gel Electrophoresis System, 1998 and Product brochure, Gel-Mix 6, 2000) and used for electrophoresis. Prerun of the gel was done 40 w until the gel surface temperature reached ~50°C.

Sample preparation

Preparation of samples, pGEM DNA Markers and Ladders conducted as per PROMEGA GenePrint STR Systems [Silver Stain Detection]–Technical Manual (2001).

CSF1PO

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
5q33.3-34 (Fig. 3.2)	HUMCSF1PO, Human c-fms proto-oncogene for CSF-1 receptor gene	AGAT
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
295-327	7,8,9,10,11,12,13,14,15	6

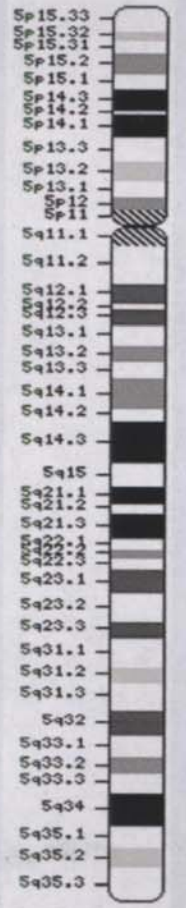

Fig. 3.2. Ideogram of Chromosome #5 Showing the Chromosomal Location of CSF1PO

TPOX

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
2p25.1-pter(Fig. 3.3)	HUMTPOX, Human thyroid peroxidase gene	AATG
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
224-252	6,7,8,9,10,11,12,13	None

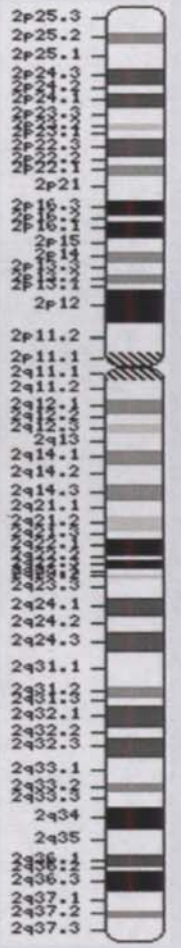

Fig. 3.3. Ideogram of Chromosome # 2 Showing the Chromosomal Location of TPOX

THO1

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
11p15.5(Fig. 3.4)	HUMTHO1, Human tyrosine hydroxylase gene	AATG
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
179-203	5,6,7,8,9,10,11	9.3

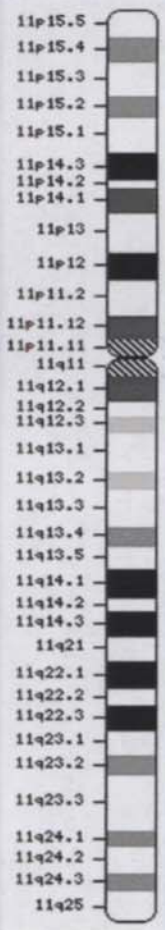

Fig. 3.4. Ideogram of Chromosome # 11 Showing the Chromosomal Location of THO1

F13A01

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
6p24.3-p25.1(Fig. 3.5)	HUMF13A01, Human coagulation factor XIII a subunit gene	AAAG
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
283-331	4,5,6,7,8,9,11,12,13,14,15,16	3.2,10

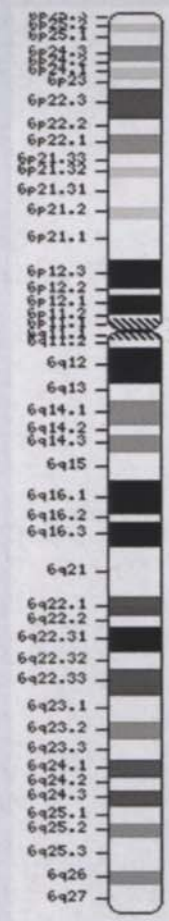

Fig. 3.5. Ideogram of Chromosome # 6 Showing the Chromosomal Location of F13A01

FESFPS

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
15q25-qter(Fig. 3.6)	HUMFESFPS, Human c-fes/fps proto-oncogene	AAAT
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
222-250	7,8,9,10,11,12,13,14	None

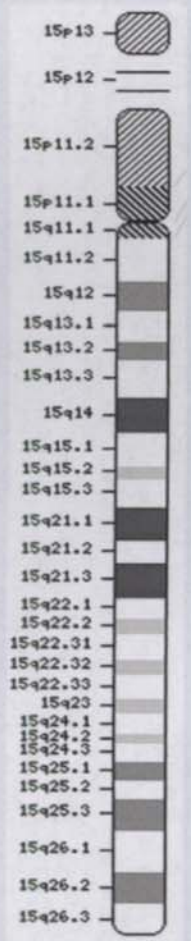

Fig. 3.6. Ideogram of Chromosome # 15 Showing the Chromosomal Location of FESFPS

vWA (formerly vWF)

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
12p12-pter(Fig. 3.7)	HUMVWFA31, Human von Willebrand factor gene	AGAT
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
139-167	13,14,15,16,17,18,19,20	11,21

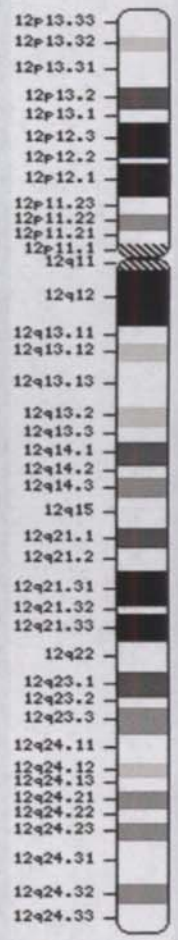

Fig. 3.7. Ideogram of Chromosome # 12 Showing the Chromosomal Location of vWA

D16S539

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
16q24-qter(Fig. 3.8)	NA	AGAT
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
264-304	5,8,9,10,11,12,13,14,15	None

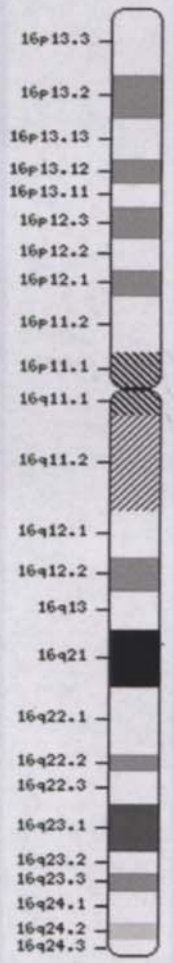

Fig. 3.8. Ideogram of Chromosome # 16 Showing the Chromosomal Location of D16S539

D7S820

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
7q11.21-22(Fig. 3.9)	NA	AGAT
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
215-247	6,7,8,9,10,11,12,13,14	None

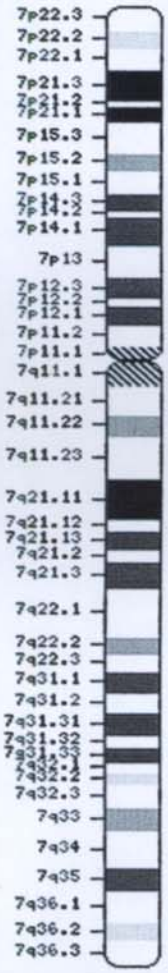

Fig. 3.9. Ideogram of Chromosome # 7 Showing the Chromosomal Location of D7S820

D13S317

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
13q22-q31(Fig. 3.10)	NA	AGAT
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
165-197	7,8,9,10,11,12,13,14,15	None

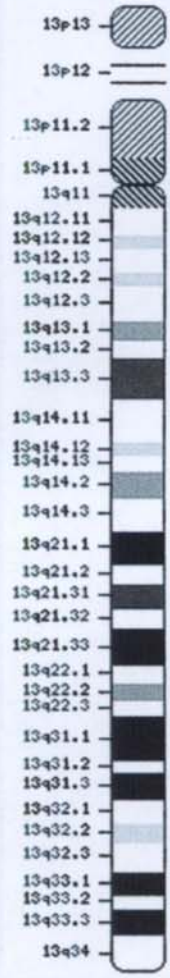

Fig. 3.10. Ideogram of Chromosome # 13 Showing the Chromosomal Location of D13S317

Amelogenin

Chromosomal Location	GenBank Locus and Locus Definition	Repeat Sequence 5'-3'
Xp22.1-22.3 and Yp11.2 (Fig. 3.11)	HUMAMEL, Human Y chromosomal gene for amelogenin-like protein	NA
Allelic Ladder Size Range (bases)	STR Ladder Alleles (# of repeats)	Other Known Alleles (# of repeats)
212-218	NA	None

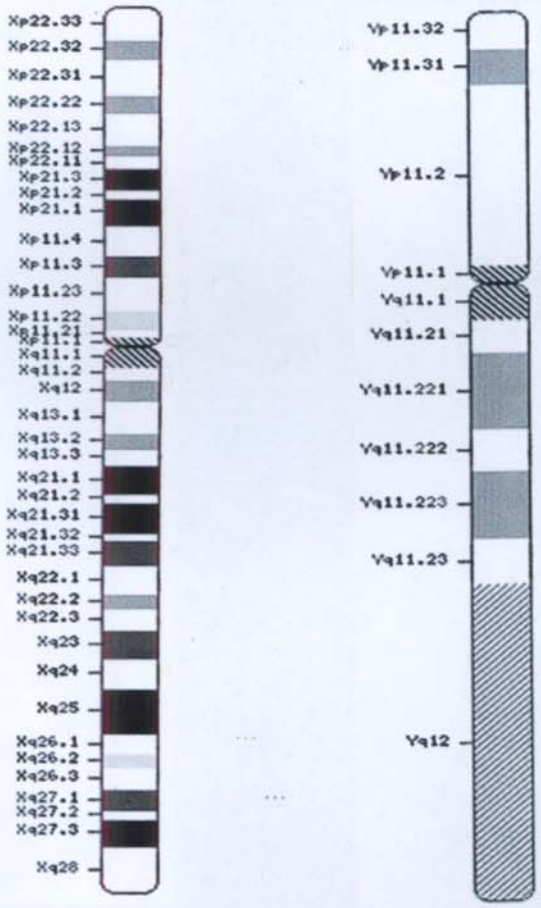

Fig. 3.11. Ideogram of Chromosome X and Y Showing the Location of Amelogenin

Sample loading and electrophoresis

1 – 2 µl samples were loaded onto the respective wells formed by the 49 point, sharkstooth comb after heat denaturation at 95 °C and subsequent chilling in ice. The gel was run using the same conditions as in gel pre-electrophoresis. The electrophoresis was stopped as per the migration characteristics of the dyes (Sambrook and Russell, 2001).

Silver staining

At the end of PAGE, the buffer chambers were emptied and the shorter glass plate with strongly affixed gel separated from the longer glass plate. The gel was placed in a plastic wash tub (54 x 43 x 13cm), and separated DNA fragments were visualized by silver staining (Allen *et al.*, 1989; Bassam *et al.*, 1991).

Exposure of film

A direct image is produced using Automatic Process Compatible (APC) film. Processing was done in a dark room equipped with photograph developing facilities. The dried stained gel was placed on a white light box and the APC film positioned, emulsion side down over the gel to be copied for about 30 seconds to 2 minutes.

Reference photographs of Quantiblot, CTT, FFv and SilverSTR III results are presented below (Figs.3.12 – 3.15).

Figs. 3.12 – 3.15

Statistical analysis

Allele designations were determined by comparison of the sample fragments with those of the allelic ladders. The data were analysed using the population genetic software, Tools for Population Genetic Analyses-TFPGA (Miller, 1997) in addition to the facilitation by Windows Excel and manual calculation. The frequency of each allele for each locus was calculated from the numbers of each genotype in the sample set. The *unbiased estimate of heterozygosity* was calculated according to Nei and Roychoudhury (1974) as per the formula:

Fig. 3.12. QuantiBlot results (using colorimetric [TMB] detection method). The quantity of samples are assayed by the comparison of intensities with the DNA standards.

Slots: (A1 –A6 and B1) – DNA standards representing quantities of 10, 5, 2.5, 1.25, 0.625, 0.3125 and 0.15625 ng, respectively.
(B2-B3) – DNA calibrators 1 and 2.
(B4-E3) – DNA test samples.
E4 – Blank (spotting solution only). Rest empty.

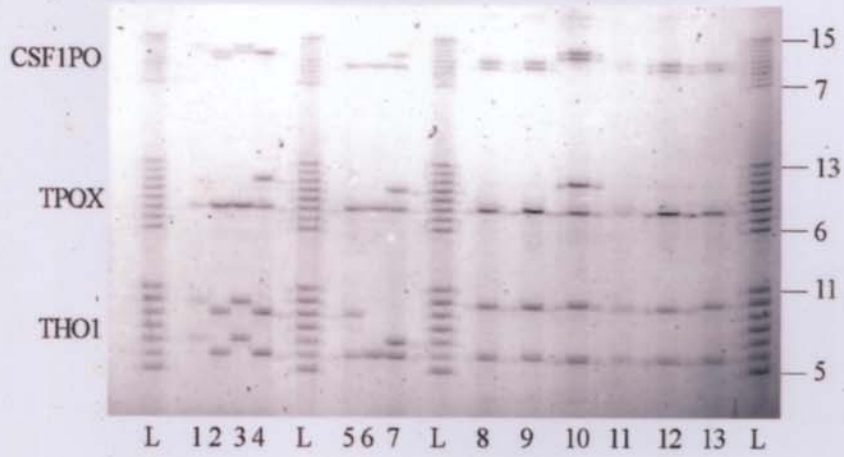

Fig. 3.13. Individual genomic DNA samples (lanes 1-13) amplified using GenePrint STR – CTT (CSF1PO, TPOX, THO1) Multiplex, separated in a 6% polyacrylamide denaturing gel. "L" allelic ladders for the respective loci. The no. of repeats in the ladder are noted and corresponding nos. in the individual DNA samples through comparison with the respective ladders are recorded. Two bands in a STR locus denotes heterozygous, whereas single band was taken as homozygous for that locus.

Examples:

LANE NO.	CSF1PO	TPOX	THO1
1	12,13	8,8	7,10
2	11,12	8,8	6,9
3	12,13	8,8	7,10
4	12,12	8,11	6,9

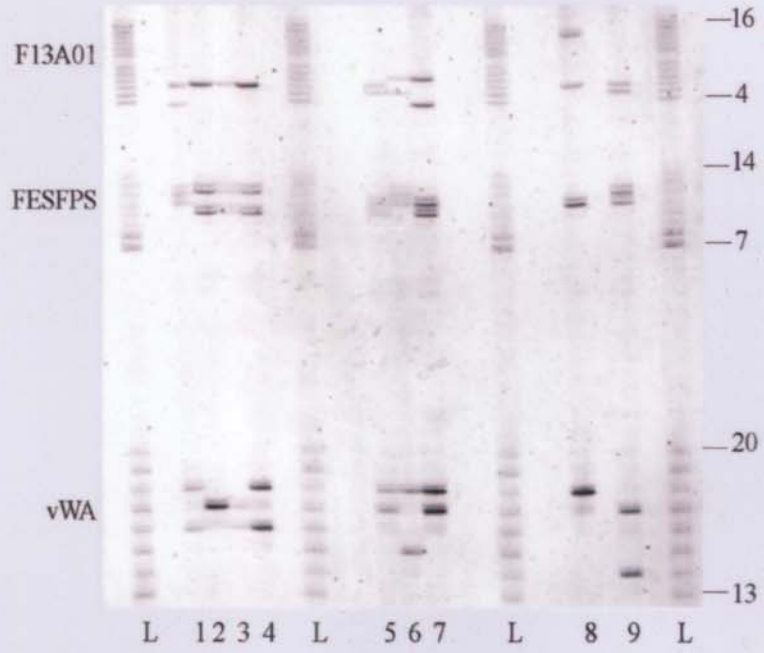

Fig. 3.14. Individual genomic DNA samples (lanes 1-9) amplified using GenePrint STR – FFv (F13A01,FESFPS, vWA) Multiplex, separated in a 6% polyacrylamide denaturing gel. "L" allelic ladders for the respective loci. The no. of repeats in the ladder are noted and corresponding nos. in the individual DNA samples through comparison with the respective ladders are recorded. Two bands in a STR locus denotes heterozygous, whereas single band was taken as homozygous for that locus.

Examples:

LANE NO.	F13A01	FESFPS	vWA
1	3.2,6	11,12	16,18
2	6,6	10,12	16,17
3	6,6	10,12	16,17
4	6,6	10,12	16,18

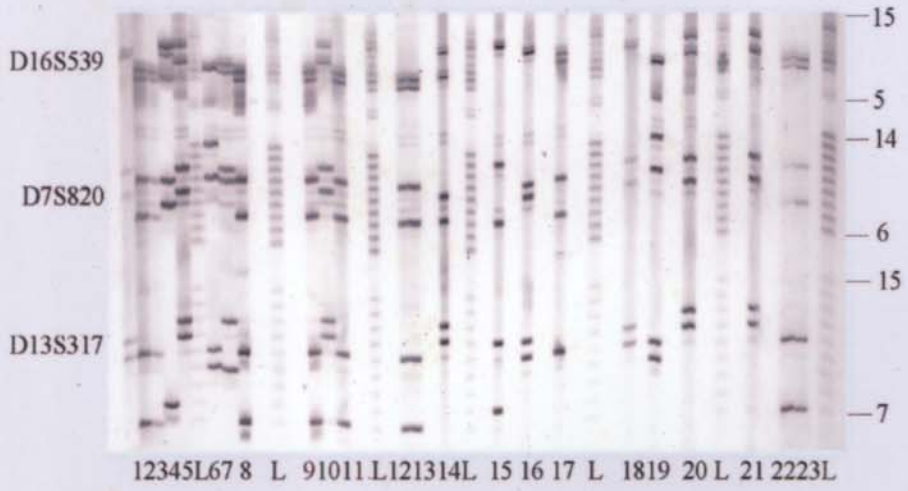

Fig. 3.15. Individual genomic DNA samples (lanes 1-23) amplified using GenePrint STR – SilverSTR III (D16S539, D7S820, D13S317) Multiplex, separated in a 6% polyacrylamide denaturing gel. "L" allelic ladders for the respective loci. The no. of repeats in the ladder are noted and corresponding nos. in the individual DNA samples through comparison with the respective ladders are recorded. Two bands in a STR locus denotes heterozygous, whereas single band was taken as homozygous for that locus.

Examples:

LANE NO.	D16S539	D7S820	D13S317
1	11,11	10,12	11,12
2	8,9	8,11	7,11
3	8,9	8,11	7,11
4	8,11	9,11	7,8

$$h = n(1 - \Sigma x^2) / (n - 1)$$

where x = allele frequencies and n = total number of alleles observed (Lareu *et al.*, 1994; Sueblinvong and Kongsrisook, 1999). The possible divergence from Hardy-Weinberg equilibrium (HWE) was tested through *exact test* by means of conventional Monte Carlo method with 2000 permutations per batch, 10 batches per analysis having a total permutation of 20000 (Guo and Thompson, 1992).

The test of homogeneity between sample populations have been conducted by the contingency *table* [RxC test] (Sokal and Rohlf 1995), *genetic distances* (Nei and Roychoudhury, 1974) and *population tree* constructed by the UPGMA [unweighted pair group-method with arithmetic mean] (Swofford and Olsen, 1990; Mount, 2001).

The potential usefulness of the combination of nine systems was assessed by calculating the discrimination power (DP) for identity testing and power of exclusion (PE) for paternity testing.

The DP gives the percentage probability that two individuals chosen at random from a population will have different phenotypes and is a useful guide to the potential usefulness of a new polymorphism in forensic analyses. The DP was calculated using the formula:

$DP = 1 - \Sigma (Pi)^2$, where Pi represents the frequency of each expected phenotype (Fisher, 1951; Lareu *et al.*, 1994; Sueblinvong and Kongsrisook, 1999).

The PE gives the percentage probability of excluding a wrongly named man as the father in a paternity analysis. The computation of average probability of exclusion was facilitated by Windows Excel from the formula:

$$PE = \Sigma Pi(1 - Pi)^2 + \Sigma (Pi Pj)^2(3Pi + 3Pj - 4)$$

Where Pi = most common allele, Pj = next most common allele (Hummel and Gerchow, 1981; Lareu *et al.*, 1994; Sueblinvong and Kongsrisook, 1999).

The combined DP and PE have been calculated following the formula (APPLIED BIOSYSTEMS, AmpliType PM + DQA1 PCR Amplification and Typing Kits, User's Manual, 1995)

1. Combined power of discrimination, $PD = 1 - [(P_i)_1 (P_i)_2 \ldots..(P_i)_n]$, where *n* is the number of combined markers and P_i is $1 - PD$ for each number.
2. Combined power of exclusion, $PE = 1 - [(P_i)_1 (P_i)_2 \ldots..(P_i)_n]$, where *n* is the number of combined markers and P_i is $1 - PE$ for each number.

The allele sizes obtained on genotyping of the subjects of Kasargod, Kannur, Wayanad and Kozhikode comprising North Kerala with nine STRs, HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31 (vWA), D16S539, D7S820 and D13S317 are given in tables 4.1 to 4.4. Individuals of Mahe also are included to completely satisfy the sampling of North Kerala, since this small place is surrounded by Kannur and Kozhikode districts and results are given in table 4.5. Tables 4.6 and 4.7 represent the genotyping results of Mukkuva and Paniyan respectively, the two social groups selected for the study.

Genotypes of the populations

Table 4.1. Genotypes for the STR loci: HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31(vWA), D16S539, D7S820 and D13S317 in 13 samples collected from unrelated individuals of *Kasargod District*.

Sample No.	STR								
	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
1	10,11	8,10	9,9	4,6	11,12	14,15	8,9	8,11	8,12
2	11,12	8,11	6,8	4,6	10,11	16,17	12,12	7,11	11,12
3	10,11	11,11	8,9	4,7	11,12	14,15	8,10	8,13	12,13
4	11,11	8,8	9,10	4,8	11,13	16,17	12,12	8,11	8,12
5	12,12	8,11	10,10	5,6	10,11	16,17	9,13	8,11	12,14
6	11,11	8,8	9,9	4,4	10,13	14,16	9,12	12,12	8,12
7	12,12	8,10	6,6	6,8	10,12	16,17	10,12	11,12	10,14
8	12,13	8,10	6,9.3	4,5	12,13	17,19	10,11	8,12	12,12
9	11,13	11,11	9,10	6,7	12,13	14,17	10,12	8,13	11,12
10	12,12	9,11	6,7	5,5	11,11	15,17	11,12	8,11	8,8
11	8,11	10,11	6,7	7,7	12,13	17,18	10,12	8,11	8,11
12	10,10	8,11	7,7	3.2,5	11,12	17,18	10,12	11,12	11,12
13	12,12	10,11	9,10	5,5	7,11	15,16	9,12	8,10	11,12

Note: The data are the number of repeats in each locus in comparison with the respective ladder.

Table 4.2. Genotypes for the STR loci: HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31(vWA), D16S539, D7S820 and D13S317 in 33 samples collected from unrelated individuals of *Kannur District*.

Sample No.	STR								
	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
1	11,12	8,8	6,9	4,5	11,11	14,20	8,10	9,11	9,10
2	12,12	8,12	9,10	4,9	12,12	17,18	9,13	12,12	9,10
3	10,11	11,12	6,9	4,5	11,11	18,19	9,12	12,12	8,12
4	12,13	9,11	6,9.3	5,15	10,11	14,17	10,11	7,11	9,10
5	12,12	8,12	6,10	3.2,5	11,12	17,19	9,12	7,11	11,13
6	11,12	11,11	6,6	5,5	10,11	17,17	11,12	8,10	9,10
7	11,12	11,12	7,7	4,7	10,13	14,19	9,12	8,12	8,10
8	11,11	11,11	6,8	5,5	10,11	17,18	10,11	12,13	8,13
9	11,12	10,11	9,9	5,7	11,13	16,16	11,12	11,13	8,12
10	10,13	11,11	6,6	8,8	11,12	14,17	11,13	8,13	12,12
11	11,12	8,11	7,9	5,16	11,12	14,16	9,11	10,12	11,12
12	8,12	8,8	8,9	5,5	11,14	14,16	9,11	7,11	10,12
13	11,11	8,11	7,9	5,6	11,13	17,18	9,11	10,13	8,9
14	10,12	11,11	9,9	5,7	11,13	15,18	10,14	10,13	8,13
15	11,12	8,10	7,9	5,16	7,11	15,16	9,11	8,12	8,10
16	12,12	8,10	7,10	5,6	12,12	16,18	9,11	12,13	11,12
17	11,11	8,11	10,10	6,7	10,13	17,17	12,13	8,12	9,12
18	12,12	8,8	7,10	5,6	11,14	15,17	10,10	9,12	11,12
19	11,11	8,11	8,10	5,7	10,11	17,19	9,12	10,11	10,11
20	11,11	8,8	6,8	3.2,3.2	10,13	15,16	10,12	8,11	8,8
21	10,11	11,12	9,10	5,7	11,12	14,17	9,10	8,10	9,11
22	12,12	10,11	6,8	5,15	10,12	16,17	9,11	7,10	11,12
23	10,12	9,11	6,6	8,8	11,12	16,18	11,13	8,12	10,10
24	10,12	8,12	7,7	4,13	10,13	14,19	8,11	11,12	9,12
25	12,12	8,11	6,10	3.2,6	10,10	16,17	12,13	10,10	8,12
26	12,12	8,8	6,6	5,5	12,12	15,17	9,10	8,11	8,8
27	11,12	10,11	7,10	5,6	10,11	14,17	10,12	9,11	11,12
28	10,10	11,11	6,9	4,4	11,12	18,18	10,11	9,10	11,12
29	12,12	10,11	6,10	5,6	13,14	15,19	13,13	11,12	8,11
30	10,12	8,11	7,8	4,5	10,12	15,17	9,11	10,13	12,12
31	10,12	11,11	6,9	5,7	10,12	17,18	11,11	10,12	12,12
32	12,12	10,10	8,9	4,5	10,12	16,18	13,14	9,11	9,11
33	12,13	9,11	6,9	5,7	10,12	18,18	9,10	8,8	8,11

Note:The data are the number of repeats in each locus in comparison with the respective ladder.

Table 4.3. Genotypes for the STR loci: HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31(vWA), D16S539, D7S820 and D13S317 in 19 samples collected from unrelated individuals of *Wayanad District*.

Sample No.	STR								
	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
1	10,13	11,12	7,9	5,7	10,11	15,17	10,10	9,12	10,12
2	11,11	8,8	6,10	5,5	10,12	16,16	9,11	11,13	8,9
3	10,10	9,12	6,7	6,6	11,13	16,17	11,13	7,8	8,13
4	11,12	8,11	6,10	3.2,6	11,13	14,17	8,11	10,12	9,12
5	12,12	10,11	7,8	4,5	11,12	15,18	10,14	10,13	9,12
6	10,11	10,11	10,10	4,7	11,13	14,17	12,13	7,12	9,13
7	12,12	9,10	9,9	6,6	11,12	16,18	8,9	11,13	10,12
8	10,11	9,11	6,9	5,15	11,12	16,17	10,11	10,12	10,11
9	11,13	8,8	6,8	5,8	11,13	16,16	10,13	11,11	8,8
10	13,13	8,11	9,10	5,5	12,12	14,18	11,12	8,10	12,12
11	10,12	8,11	6,6	5,15	12,13	17,19	13,13	10,11	8,9
12	11,12	11,11	7,8	4,4	12,13	14,16	11,11	12,12	11,12
13	11,11	8,11	9,9	4,6	12,12	16,17	9,13	10,12	12,13
14	12,12	9,11	9,10	5,8	11,12	16,18	12,12	10,11	8,10
15	12,12	8,11	6,9	5,8	11,12	17,18	9,9	8,8	8,13
16	10,11	8,10	6,8	5,5	10,12	15,17	11,13	10,12	12,14
17	10,12	11,11	9,10	5,8	12,12	16,17	8,12	10,11	8,12
18	11,13	8,12	6,7	5,5	10,11	15,16	9,11	12,13	9,14
19	12,12	9,11	6,9.3	3.2,3.2	10,12	15,18	8,12	8,10	8,10

Note: The data are the number of repeats in each locus in comparison with the respective ladder.

Table 4.4. Genotypes for the STR loci: HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31(vWA), D16S539, D7S820 and D13S317 in 60 samples collected from unrelated individuals of *Kozhikode District*.

Sample No.	STR								
	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
1	11,13	8,8	6,10	4,5	10,11	14,16	10,12	8,12	10,11
2	10,10	8,10	7,10	6,7	10,12	16,17	12,13	8,11	9,10
3	11,12	10,11	6,8	4,6	11,12	16,17	11,12	10,12	8,12
4	12,13	8,11	6,10	4,5	11,11	14,17	10,13	11,12	8,10
5	11,12	10,11	6,8	4,6	13,14	15,16	9,11	7,9	8,8
6	11,12	9,11	7,9	3.2,5	11,11	17,17	11,11	10,11	12,12
7	11,12	8,11	7,8	5,7	10,12	15,18	9,11	10,13	8,13
8	11,11	10,11	10,11	4,5	10,12	16,18	9,10	8,9	12,12
9	10,12	8,9	6,10	5,7	11,12	15,16	12,13	8,8	8,8
10	11,12	10,12	6,6	3.2,5	11,11	15,18	9,12	8,11	8,12
11	11,12	11,11	8,9	4,5	11,12	15,17	9,10	11,12	8,10
12	9,10	8,11	6,9.3	3.2,5	10,12	14,16	11,11	10,10	11,13
13	9,10	8,11	6,10	4,5	10,11	17,17	10,10	11,12	8,14
14	10,11	9,11	6,9	5,14	10,10	14,15	10,13	8,11	11,12
15	11,12	11,11	6,9	4,4	10,12	15,15	9,11	7,10	9,12
16	11,13	10,10	8,9	4,5	11,11	14,17	11,13	10,11	11,12
17	12,13	8,11	6,10	4,5	10,11	17,17	10,15	8,11	11,14
18	10,11	9,11	6,8	4,15	11,12	15,16	11,12	9,11	13,13
19	12,13	7,8	7,8	5,5	10,10	15,17	11,13	7,9	8,8
20	11,11	11,11	6,10	3.2,6	11,12	15,18	9,10	9,11	8,8
21	10,10	9,11	7,9.3	4,15	11,11	16,16	12,13	8,12	8,12
22	10,10	11,11	7,9.3	9,15	13,13	17,17	10,13	11,13	12,12
23	10,12	7,8	5,8	5,5	11,12	17,19	9,12	8,12	12,12
24	12,13	9,11	6,9	5,14	12,12	16,18	11,11	8,8	10,11
25	11,11	9,12	7,9	5,6	11,12	14,16	9,10	11,11	12,13
26	10,12	8,10	6,7	4,5	10,12	17,18	11,13	8,12	11,12
27	10,11	11,11	6,9	5,8	11,12	18,18	11,11	8,9	8,11
28	10,12	8,11	6,6	4,8	11,13	16,18	9,11	8,10	9,14
29	11,12	11,12	6,8	4,15	12,13	16,16	9,11	10,10	8,9
30	12,12	11,11	9,10	3.2,5	11,12	17,19	11,12	8,11	11,12
31	11,12	11,11	7,9	4,6	11,12	15,16	12,12	11,11	11,11
32	11,12	9,10	7,8	4,5	12,12	16,17	10,11	12,12	12,13
33	11,13	8,8	7,9	3.2,3.2	12,13	16,16	12,13	8,11	12,12

34	12,13	9,10	6,10	3.2,6	12,12	17,18	11,13	12,12	10,12
35	12,12	10,11	9,9	5,6	11,12	14,16	10,12	8,12	10,11
36	11,13	11,11	9,10	5,5	11,12	16,17	9,10	10,12 *	11,13
37	10,11	8,12	8,9	4,5	11,12	16,17	9,12	10,10	8,13
38	11,13	8,11	6,7	7,7	12,12	17,17	8,9	11,11	8,8
39	11,11	8,11	6,9	5,6	11,13	16,18	13,13	8,10	12,12
40	10,12	9,12	6,6	4,4	12,13	16,16	10,13	10,12	8,12
41	11,12	8,8	9,10	5,5	9,11	15,16	11,11	10,12	11,14
42	13,13	10,11	10,10	3.2,3.2	10,12	17,17	9,13	8,12	12,14
43	11,14	8,9	9,9	4,7	12,12	16,18	13,13	12,13	11,12
44	11,12	11,11	6,7	5,7	12,12	17,18	11,12	11,12	8,11
45	12,12	8,8	6,8	6,8	12,12	17,19	9,9	10,10	12,12
46	12,13	11,11	8,10	5,7	10,10	16,18	11,12	8,11	10,10
47	10,12	8,10	9,9	5,5	11,11	15,17	10,11	10,11	8,10
48	10,12	8,8	10,10	5,5	10,11	14,18	8,9	10,10	9,12
49	12,12	8,8	9,10	5,6	12,13	16,17	11,13	11,12	10,12
50	12,12	9,10	6,9	5,6	12,12	16,17	11,12	8,10	8,11
51	11,12	8,11	6,6	4,7	10,11	16,17	8,13	10,11	8,10
52	11,11	8,10	9,10	5,14	12,12	14,16	12,12	11,12	8,11
53	11,11	9,11	9,9	5,11	12,13	14,17	11,12	11,12	8,8
54	10,12	11,12	7,9	4,5	11,13	14,17	9,11	8,8	8,13
55	11,12	8,11	7,8	3.2,5	10,11	14,17	9,13	10,12	11,12
56	10,12	9,11	6,9	4,5	12,13	16,19	12,12	7,10	12,12
57	10,12	10,11	9,9	4,15	11,12	14,16	13,13	8,10	8,14
58	12,12	8,11	9,9	5,6	12,12	14,18	8,10	8,11	9,10
59	10,11	8,11	6,9	5,6	12,12	15,17	9,9	10,13	8,12
60	11,12	8,9	6,9	5,7	11,12	17,17	8,9	10,12	9,12

Note: The data are the number of repeats in each locus in comparison with the respective ladder.

Table 4.5. Genotypes for the STR loci: HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31(vWA), D16S539, D7S820 and D13S317 in 2 samples collected from unrelated individuals of *Mahe*.

Sample No.	STR								
	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
1	12,13	8,11	6,8	4,15	11,12	16,17	8,11	8,10	8,14
2	11,12	9,11	7,7	4,5	10,12	18,20	8,9	8,12	8,12

Note: The data are the number of repeats in each locus in comparison with the respective ladder.

Table 4.6. Genotypes for the STR loci: HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31(vWA), D16S539, D7S820 and D13S317 in 32 samples collected from unrelated individuals of *Mukkuva Community*.

Sample No.	STR								
	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
1	15,15	8,8	6,6	4,5	10,11	16,18	9,12	11,11	8,12
2	12,12	8,11	8,9.3	3.2,5	10,10	17,18	9,12	8,8	9,12
3	11,12	8,9	5,9	4,6	10,10	14,15	9,12	8,8	8,12
4	11,12	8,8	6,9	5,7	11,12	14,16	11,12	8,12	9,12
5	11,12	11,11	7,8	4,5	11,12	17,18	10,11	8,11	9,14
6	12,12	8,11	7,9	5,7	11,11	14,19	11,11	10,13	9,13
7	12,12	8,8	6,9.3	3.2,3.2	12,13	16,16	11,13	10,11	9,12
8	10,12	8,11	9,9.3	4,5	12,13	17,19	9,9	8,10	8,12
9	10,11	8,9	6,6	5,6	10,11	16,18	11,12	8,10	9,12
10	9,10	8,11	5,9	4,6	10,11	16,17	9,11	8,12	12,13
11	12,12	9,11	9,9	5,6	10,10	18,18	9,13	8,11	11,12
12	11,12	8,10	9.3,9.3	5,14	11,11	15,19	9,13	10,12	12,13
13	12,12	9,11	7,9	4,4	13,14	17,18	11,11	7,8	11,13
14	10,12	8,8	6,9.3	5,5	11,11	16,17	12,13	10,11	8,9
15	12,12	8,8	9.3,9.3	4,5	12,12	17,17	9,11	10,12	8,8
16	11,11	9,11	6,7	5,14	10,11	15,18	13,13	8,12	11,13
17	11,12	8,11	6,6	4,15	10,11	17,18	12,13	12,12	8,9
18	12,12	9,11	6,7	4,5	10,12	16,18	9,12	8,8	9,12
19	12,12	8,9	8,9	3.2,6	10,12	16,19	11,13	10,11	11,12
20	12,12	11,11	6,7	4,5	13,13	17,17	9,11	8,8	8,9
21	10,12	8,11	7,8	3.2,5	11,12	16,17	10,13	10,12	12,12
22	9,12	8,8	6,8	5,6	11,12	15,18	9,11	8,8	8,13
23	10,12	8,11	6,7	5,5	10,12	14,18	12,13	8,8	8,13
24	12,12	8,11	6,7	4,6	12,12	15,17	11,13	8,10	8,12
25	11,13	8,9	6,8	4,4	10,13	15,16	8,12	10,11	12,12
26	9,12	8,9	8,9.3	5,7	11,12	16,18	12,12	8,11	9,11
27	11,12	8,8	6,7	5,14	10,12	17,18	11,11	12,12	8,8
28	11,12	8,11	6,6	4,15	10,11	17,18	9,12	8,11	9,11
29	12,12	9,11	9,9.3	5,5	10,12	17,18	11,13	8,10	11,13
30	11,12	8,8	6,9	3.2,5	11,12	14,17	9,12	10,12	11,13
31	12,12	11,11	8,9	4,6	10,12	16,17	12,13	7,8	8,8
32	12,12	9,11	9,9	9,15	12,12	17,18	8,12	7,8	8,8

Note: The data are the number of repeats in each locus in comparison with the respective ladder.

Table 4.7. Genotypes for the STR loci: HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31(vWA), D16S539, D7S820 and D13S317 in 31 samples collected from unrelated individuals of *Paniyan Community*.

Sample No.	STR								
	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
1	11,12	8,8	6,9	5,5	11,11	14,16	12,12	10,10	9,9
2	12,12	8,11	6,9	5,5	10,10	13,15	9,11	7,8	8,8
3	10,10	8,9	7,9	7,14	12,13	14,15	11,11	8,11	11,11
4	12,12	8,11	6,6	5,6	10,12	17,18	9,12	8,10	8,8
5	11,11	8,11	6,9	5,14	11,12	15,17	11,12	8,11	8,12
6	11,12	8,8	6,6	5,6	11,11	14,16	11,12	8,8	8,11
7	10,12	9,11	6,9	5,5	11,11	14,17	9,12	8,10	8,8
8	12,12	8,11	7,9	5,6	10,11	16,17	9,11	8,11	8,12
9	11,12	8,8	6,7	5,7	11,11	16,17	9,13	7,11	9,12
10	11,12	8,8	9,9	5,5	11,12	13,19	9,11	8,8	8,8
11	10,12	8,9	6,9	6,6	10,11	15,16	10,10	9,11	8,11
12	10,10	8,8	9,9	5,5	10,10	16,17	11,12	10,11	8,8
13	10,11	8,11	6,9	5,6	10,12	16,16	10,11	10,10	9,9
14	12,12	9,11	9,9	5,5	11,11	18,19	11,12	10,12	9,11
15	11,12	8,8	6,10	5,7	11,13	15,17	9,11	10,11	8,11
16	12,15	8,11	6,6	5,14	11,13	14,17	11,12	8,10	8,11
17	10,12	8,9	6,7	5,6	10,12	14,15	11,11	10,11	8,8
18	10,12	9,11	8,10	5,7	11,11	16,16	9,12	7,8	11,11
19	11,11	9,11	6,6	5,6	10,11	17,18	11,11	10,10	9,11
20	12,12	8,11	6,6	5,5	10,11	18,19	10,11	8,11	8,11
21	12,12	11,11	8,10	5,5	10,11	14,18	11,12	8,10	8,12
22	11,12	8,11	6,10	4,5	7,8	16,17	9,12	8,10	8,12
23	11,12	8,11	9,9	5,5	11,12	17,18	12,12	10,12	8,8
24	10,12	8,9	6,6	4,5	10,11	17,17	10,11	10,10	8,9
25	10,11	8,11	6,7	7,14	10,11	16,17	11,12	8,10	8,12
26	12,15	8,8	6,10	5,6	10,10	14,16	9,9	8,11	9,11
27	12,12	8,11	6,9	5,5	12,13	15,16	11,11	9,11	11,12
28	11,12	11,11	7,9	5,6	12,13	13,15	9,11	7,11	9,12
29	12,15	9,11	6,7	5,5	10,11	16,16	9,9	10,12	8,11
30	12,12	8,11	6,10	5,6	10,11	17,17	11,11	8,10	8,9
31	10,12	8,11	6,6	6,6	11,12	15,16	12,12	10,11	11,12

Note: The data are the number of repeats in each locus in comparison with the respective ladder.

850

4 RESULTS AND DISCUSSION

C1

Genealogy analysis

Genotyping with the STR loci – CSF1PO, TPOX, THO1, F13A01, FESFPS, vWA D16S539, D7S820 and D13S317 has been documented in various populations and genealogical studies are well established in resolving forensic and paternity cases (Aaspollu *et al.*, 2000; Amar *et al.*, 1999; Andersen *et al.*, 1996; Balding *et al.*, 1996; Budowle *et al.*, 1997; Caskey and Hammond, 1992; Corte-Real *et al.*, 1999 a,b; Edwards *et al.*, 1991a,b; Evett *et al.*, 1996; Ghosh *et al.*, 2004; Gill and Evett, 1995; Panneerchelvam, 2004; Regueiro *et al.*, 2004; Zupanic *et al.*, 1998). Here the genealogical analysis using the above STRs were done to evaluate the phylogenetic relationships of the subjects under study.

Arrangement of data

Data were arranged in four populations to reach the first objective of *evaluating the relationship between different social groups in North Kerala* comprising people of Kannur, Kozhikode and social groups comprising, *Mukkuva* and *Paniyan*. A total of 53 individuals from Kannur comprising samples of Kasargod district, once a part of Kannur district (table 4.1), 33 samples of Kannur district (table 4.2), 5 samples from Wayanad district once a part of Kannur and Kozhikode districts (table 4.3) and 2 samples of Mahe, a small place surrounded by Kannur and Kozhikode districts (table 4.5)., a total of 74 individuals from Kozhikode comprising all the samples of Kozhikode district (table 4.4) and 14 samples from the Wayanad district collected from the area related to Kozhikode district (table 4.3)., 32 individuals from *Mukkuva* (table 4.6) living in the coastal belt and 31 individuals from *Paniyan* (table 4.7) were subjected to statistical analysis. Samples of Kasargod, Kannur, Wayanad and Mahe were pooled to a sample size of 30 or more which is the minimum number required for statistical interpretation (Mahajan, 1999). To evaluate the *genotypes of the people of North Kerala,* a total of 127 individuals from Kannur and Kozhikode have been pooled to create a North Kerala population that covers almost all area from Kozhikode district towards east and north of Kerala state (Fig. 3.1).

Population genetic considerations

A genetically stable population should be homogeneously mixed. In 1908, G.H.

Hardy, an English mathematician and W. Weinberg, a German physician, independently discovered that an equilibrium is established between frequencies of alleles in a random mating population. It was also shown that the relative gene frequencies remained unaltered from one generation to the next, regardless of their dominance and recessive relationship. A mathematical relationship called Hardy-Weinberg Theorem was developed to explain this equilibrium, which is now known as Hardy-Weinberg Law (Gupta, 1994). The Hardy-Weinberg law is a cornerstone in the development of population genetics (Guo and Thompson, 1992). The law states that in a large random mating population with no selection, mutation, or migration, the allele frequencies and the genotype frequencies are constant from generation to generation and that, furthermore, there is a simple relationship between the allele frequencies and the genotype frequencies. Because of its importance, testing of the hypothesis that a population exhibits Hardy-Weinberg proportions (HWP) has drawn a lot of attention during past decades, though the problem itself seems to be very simple from a statistical viewpoint.

The technical measure of mixing is that each locus should be in Hardy-Weinberg equilibrium, the main parameter to be satisfied and the other related parameters will be seen in right direction, once the population is in Hardy-Weinberg equilibrium (Lander, 1989). This is nothing more than a restatement of the fact that each person's alleles must represent a random selection from the overall pool. In classical statistics, data are used to test hypotheses about specified values of some parameters. Originally this meant that a test statistical calculation was made from the data and a hypothesis was either rejected or not rejected. The decision was based on comparing the statistical analysis with a tabulated set of critical values. For categorical data, the simplest test is the *chi-square goodness-of-fit* test.

The methods proposed so far for testing HWP can be categorized into two groups. One consists of large-sample goodness-of-fit tests such as Pearson's χ^2, likelihood ratio statistic G^2, and conditional χ^2 test (Li, 1955). These tests have one characteristic in common: they lean heavily on asymptotic results. The other approach involves exact tests (Chapco, 1976; Fisher, 1935; Haldane, 1954; Levene, 1949). These tests usually involve more computing and were, in the past, thus restricted to

diallelic locus with small sample sizes due to lack of computing power. It has long been recognized that the standard goodness-of-fit tests can sometimes lead to false rejection or acceptance of HWP when the sample sizes are small and / or some cell frequencies are small or zero (Emigh, 1980). Although various corrections for small sample sizes are made (Elston and Forthofer, 1977; Emigh and Kempthorne, 1975; Smith, 1986), it is found that they usually do not greatly improve the results obtained from the traditional goodness-of-fit tests (Emigh, 1980; Hernández and Weir, 1989). Evett and Weir (1998) found that, though the test is easy to apply, it can give misleading results when expected counts are small. There have been several ad-hoc rules put forward to reduce the chance of spurious significant results, but a better procedure is to avoid the chi-square goodness-of-fit test whenever small expected counts occur. Thus an exact test is preferred when the sample size is small and / or some cell frequencies are small or zero. Genotyping with VNTRs/STRs etc., though the sample size is moderately large, the number of genotypes is so large that some sample genotype frequencies will be zero, especially when the corresponding population allele frequencies are low. As a result, the adequacy of applying classical goodness-of-fit tests of HWP is questionable and use of the exact test is desirable (Guo and Thompson, 1992). Briefly, these tests assume that the hypothesis is true and calculate the probability of the observed outcome or a more extreme (less probable) outcome and report as *P*-values instead of reject/not reject statements. Low values of this probability suggest that the hypothesis is not true (Evett and Weir, 1998).

Exploration of social groups

Tables 4.8 to 4.11 represents the allele frequencies and Hardy-Weinberg statistical parameters of 9 STR: HUMCSF1PO, HUMTPOX, HUMTHO1, HUMF13A01, HUMFESFPS, HUMVWFA31(vWA), D16S539, D7S820 and D13S317 loci in Kannur, Kozhikode, *Mukkuva* and *Paniyan* populations., the 2 assigned populations and 2 social groups, respectively of North Kerala under study. The people comprising the Kannur and Kozhikode have been selected as two social groups for the purpose of study, though these are not documented, so far. The histograms depicting comparison of allele frequencies of all 9 STR loci of Kannur,

Kozhikode, *Mukkuva* and *Paniyan* population are presented in Figs. 4.1 – 4.3. Table 4.12 shows the most and least frequent alleles in the above four populations.

Table 4.8. Allele frequencies and Hardy-Weinberg statistical parameters of 9 STR loci of Kannur population (N = 106).

Allele	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
3.2	--	--	--	0.0566	--	--	--	--	--
4	--	--	--	0.1792	--	--	--	--	--
5	--	--	--	0.3774	--	--	--	--	--
6	--	--	0.2830	0.1415	--	--	--	--	--
7	--	--	0.1887	0.1226	0.0189	--	--	0.0566	--
8	0.0189	0.3302	0.1038	0.0566	--	--	0.0660	0.2170	0.2264
9	--	0.0566	0.2358	0.0094	--	--	0.1981	0.0566	0.1132
9.3	--	--	0.0189	--	--	--	--	--	--
10	0.1604	0.1321	0.1698	--	0.2170	--	0.1981	0.1509	0.1132
11	0.3019	0.4057	--	--	0.3302	--	0.2170	0.1981	0.1604
12	0.4528	0.0755	--	--	0.2642	--	0.1981	0.2170	0.3113
13	0.0660	--	--	0.0094	0.1415	--	0.0943	0.1038	0.0472
14	--	--	--	--	0.0283	0.1321	0.0283	--	0.0283
15	--	--	--	0.0283	--	0.1226	--	--	--
16	--	--	--	0.0189	--	0.1981	--	--	--
17	--	--	--	--	--	0.3019	--	--	--
18	--	--	--	--	--	0.1604	--	--	--
19	--	--	--	--	--	0.0660	--	--	--
20	--	--	--	--	--	0.0189	--	--	--
HE	0.6734	0.7001	0.7887	0.7827	0.7529	0.8067	0.8211	0.8266	0.7974
HU	0.6798	0.7067	0.7962	0.7901	0.7601	0.8144	0.8289	0.8345	0.8050
H	0.5283	0.6792	0.7170	0.7358	0.8679	0.8868	0.8868	0.9057	0.8491
P	0.2004	0.4074	0.7126	0.3310	0.5888	0.6755	0.2887	0.3748	0.5392
S.E.	0.0078	0.0129	0.0128	0.0117	0.0137	0.0135	0.0077	0.0137	0.0142

N: number of chromosomes analyzed., HE: expected heterozygosity under H-W equilibrium., HU: heterozygosity (unbiased)., H: heterozygosity (direct count)., *P*: Hardy-Weinberg exact test., S.E.: standard error of *P*

Results over all loci:
Average expected heterozygosity under H-W equilibrium: 0.7722
Average heterozygosity (unbiased) : 0.7795
Average heterozygosity (direct count) : 0.7841
Percentage of polymorphic loci over all loci (no criterion) : 100.0000

Table 4.9. Allele frequencies and Hardy-Weinberg statistical parameters of 9 STR loci of Kozhikode population (N = 148).

Allele	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
3.2	--	--	--	0.0878	--	--	--	--	--
4	--	--	--	0.2027	--	--	--	--	--
5	--	--	0.0068	0.3919	--	--	--	--	--
6	--	--	0.2770	0.1149	--	--	--	--	--
7	--	0.0135	0.1149	0.0743	--	--	--	0.0338	--
8	--	0.2838	0.1149	0.0473	--	--	0.0541	0.2027	0.2568
9	0.0135	0.1284	0.2905	0.0068	0.0068	--	0.1892	0.0405	0.0676
9.3	--	--	0.0270	--	--	--	--	--	--
10	0.1892	0.1284	0.1622	--	0.1486	--	0.1284	0.2432	0.1149
11	0.3108	0.3986	0.0068	0.0068	0.2973	--	0.2432	0.2297	0.1419
12	0.3649	0.0473	--	--	0.4324	--	0.1892	0.2095	0.2838
13	0.1149	--	--	--	0.1081	--	0.1892	0.0405	0.0811
14	0.0068	--	--	0.0203	0.0068	0.1149	--	--	0.0541
15	--	--	--	0.0473	--	0.1216	0.0068	--	--
16	--	--	--	--	--	0.2905	--	--	--
17	--	--	--	--	--	0.2973	--	--	--
18	--	--	--	--	--	0.1419	--	--	--
19	--	--	--	--	--	0.0338	--	--	--
20	--	--	--	--	--	--	--	--	--
HE	0.7211	0.7252	0.7853	0.7739	0.6907	0.7779	0.8140	0.7987	0.8062
HU	0.7260	0.7301	0.7907	0.7792	0.6954	0.7832	0.8195	0.8041	0.8116
H	0.7027	0.7297	0.7838	0.7703	0.6757	0.8108	0.7568	0.7838	0.7432
P	0.1926	0.7182	0.1828	0.4160	0..3812	0.8590	0.4861	0.5501	0.5040
S.E.	0.0051	0.0085	0.0090	0.0085	0.0139	0.0081	0.0088	0.0093	0.0130

N: number of chromosomes analyzed., HE: expected heterozygosity under H-W equilibrium., HU: heterozygosity (unbiased)., H: heterozygosity (direct count)., *P*: Hardy-Weinberg exact test., S.E.: standard error of *P*

Results over all loci:
Average expected heterozygosity under H-W equilibrium: 0.7659
Average heterozygosity (unbiased) : 0.7711
Average heterozygosity (direct count) : 0.7508
Percentage of polymorphic loci over all loci (no criterion) : 100.0000

Table 4.10. Allele frequencies and Hardy-Weinberg statistical parameters of 9 STR loci of *Mukkuva* population (N = 64).

Allele	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
3.2	--	--	--	0.0938	--	--	--	--	--
4	--	--	--	0.2500	--	--	--	--	--
5	--	--	0.0313	0.3750	--	--	--	--	--
6	--	--	0.3125	0.1250	--	--	--	--	--
7	--	--	0.1563	0.0469	--	--	--	0.0469	--
8	--	0.4844	0.1250	--	--	--	0.0313	0.4063	0.2656
9	0.0469	0.1719	0.2188	0.0156	--	--	0.2188	--	0.1875
9.3	--	--	0.1563	--	--	--	--	--	--
10	0.0938	0.0156	--	--	0.2969	--	0.0313	0.2031	--
11	0.1875	0.3281	--	--	0.2813	--	0.2656	0.1563	0.1250
12	0.6250	--	--	--	0.3125	--	0.2500	0.1719	0.2656
13	0.0156	--	--	--	0.0938	--	0.2031	0.0156	0.1406
14	--	--	--	0.0469	0.0156	0.0781	--	--	0.0156
15	0.0313	--	--	0.0469	--	0.0938	--	--	--
16	--	--	--	--	--	0.2031	--	--	--
17	--	--	--	--	--	0.2969	--	--	--
18	--	--	--	--	--	0.2656	--	--	--
19	--	--	--	--	--	0.0625	--	--	--
20	--	--	--	--	--	--	--	--	--
HE	0.5620	0.6279	0.7891	0.7656	0.7261	0.7813	0.7759	0.7373	0.7881
HU	0.5709	0.6379	0.8016	0.7778	0.7376	0.7937	0.7882	0.7490	0.8006
H	0.5313	0.6563	0.7500	0.8125	0.6875	0.8750	0.8125	0.7188	0.8125
P	0.1164	0.7472	0.3076	0.2673	0.3420	0.8730	0.6282	0.2034	0.2628
S.E.	0.0053	0.0131	0.0132	0.0069	0.0094	0.0102	0.0062	0.0052	0.0109

N: number of chromosomes analyzed., HE: expected heterozygosity under H-W equilibrium., HU: heterozygosity (unbiased)., H: heterozygosity (direct count)., *P*: Hardy-Weinberg exact test., S.E.: standard error of *P*

Results over all loci:
Average expected heterozygosity under H-W equilibrium: 0.7281
Average heterozygosity (unbiased) : 0.7397
Average heterozygosity (direct count) : 0.7396
Percentage of polymorphic loci over all loci (no criterion) : 100.0000

Table 4.11. Allele frequencies and Hardy-Weinberg statistical parameters of 9 STR loci of *Paniyan* population (N = 62).

Allele	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
3.2	--	--	--		--	--	--	--	--
4	--	--	--	0.0323	--	--	--	--	--
5	--	--	--	0.6129	--	--	--	--	--
6	--	--	0.4677	0.2097	--	--	--	--	--
7	--	--	0.1129	0.0806	0.0161	--	--	0.0645	--
8	--	0.5000	0.0323	--	0.0161	--	--	0.2903	0.4355
9	--	0.1452	0.2903	--	--	--	0.2258	0.0323	0.1774
9.3	--	--	--	--	--	--	--	--	--
10	0.1935	--	0.0968	--	0.2903	--	0.0806	0.3548	--
11	0.2258	0.3548	--	--	0.4355	--	0.4032	0.2097	0.2419
12	0.5323	--	--	--	0.1613	--	0.2742	0.0484	0.1452
13	--	--	--	--	0.0806	0.0484	0.0161	--	--
14	--	--	--	0.0645	--	0.1290	--	--	--
15	0.0484	--	--	--	--	0.1452	--	--	--
16	--	--	--	--	--	0.2742	--	--	--
17	--	--	--	--	--	0.2581	--	--	--
18	--	--	--	--	--	0.0968	--	--	--
19	--	--	--	--	--	0.0484	--	--	--
20	--	--	--	--	--	--	--	--	--
HE	0.6259	0.6030	0.6738	0.5687	0.6930	0.8065	0.7045	0.7383	0.6993
HU	0.6362	0.6129	0.6848	0.5780	0.7044	0.8197	0.7160	0.7504	0.7107
H	0.6129	0.7097	0.6452	0.5806	0.7097	0.8387	0.6452	0.8065	0.6452
P	0.7735	0.4484	0.1053	0.3717	0.1032	0.1908	0.3397	0.2469	0.5928
S.E	0.0139	0.0101	0.0078	0.0077	0.0076	0.0086	0.0143	0.0107	0.0102

N: number of chromosomes analyzed., HE: expected heterozygosity under H-W equilibrium., HU: heterozygosity (unbiased)., H: heterozygosity (direct count)., *P*: Hardy-Weinberg exact test., S.E.: standard error of *P*

Results over all loci:
Average expected heterozygosity under H-W equilibrium: 0.6792
Average heterozygosity (unbiased) : 0.6903
Average heterozygosity (direct count) : 0.6882
Percentage of polymorphic loci over all loci (no criterion) : 100.0000

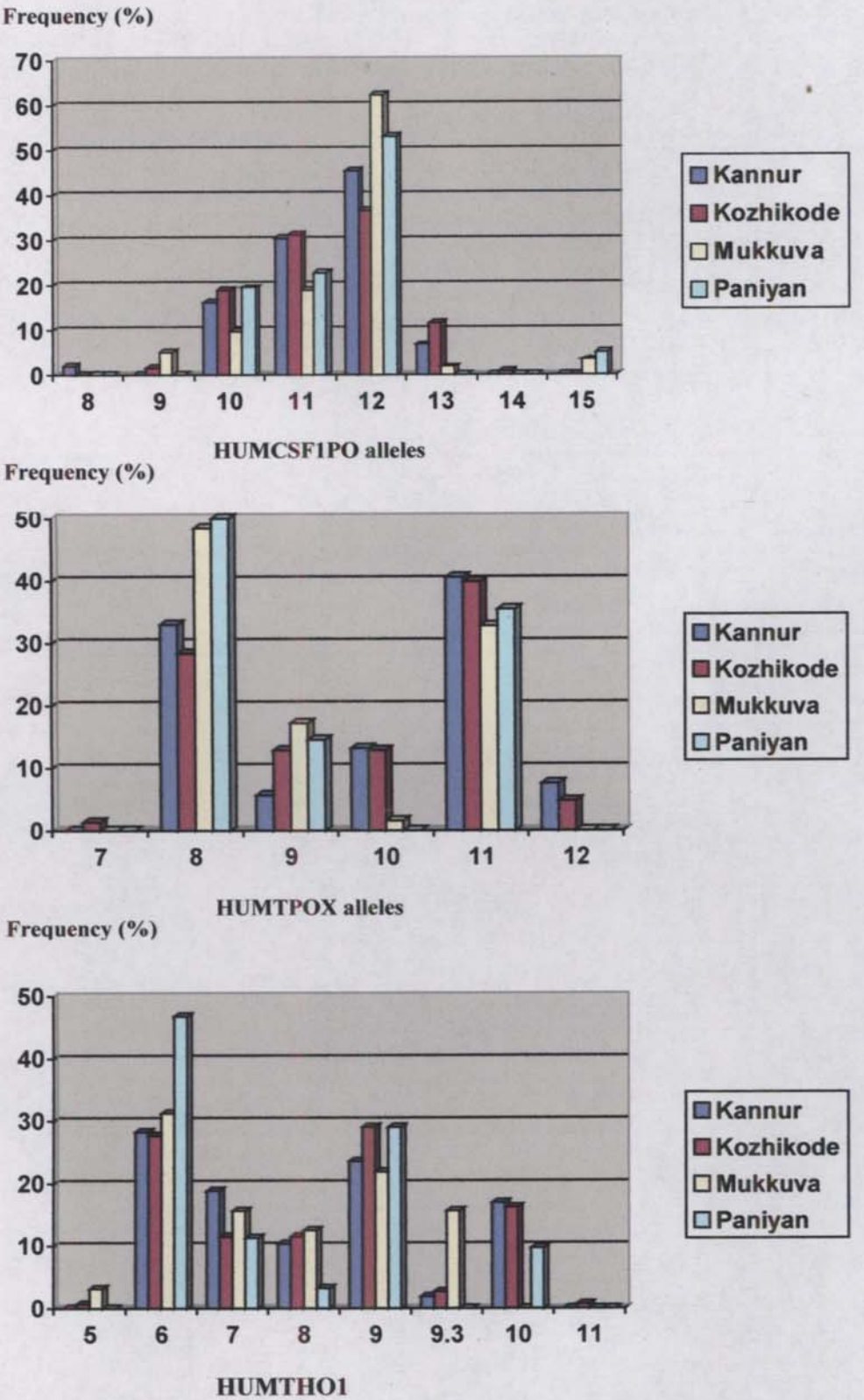

Fig. 4.1. Histograms depicting comparison of HUMCSF1PO, HUMTPOX, HUMTHO1 allele frequencies in Kannur, Kozhikode, *Mukkuva* and *Paniyan* population.

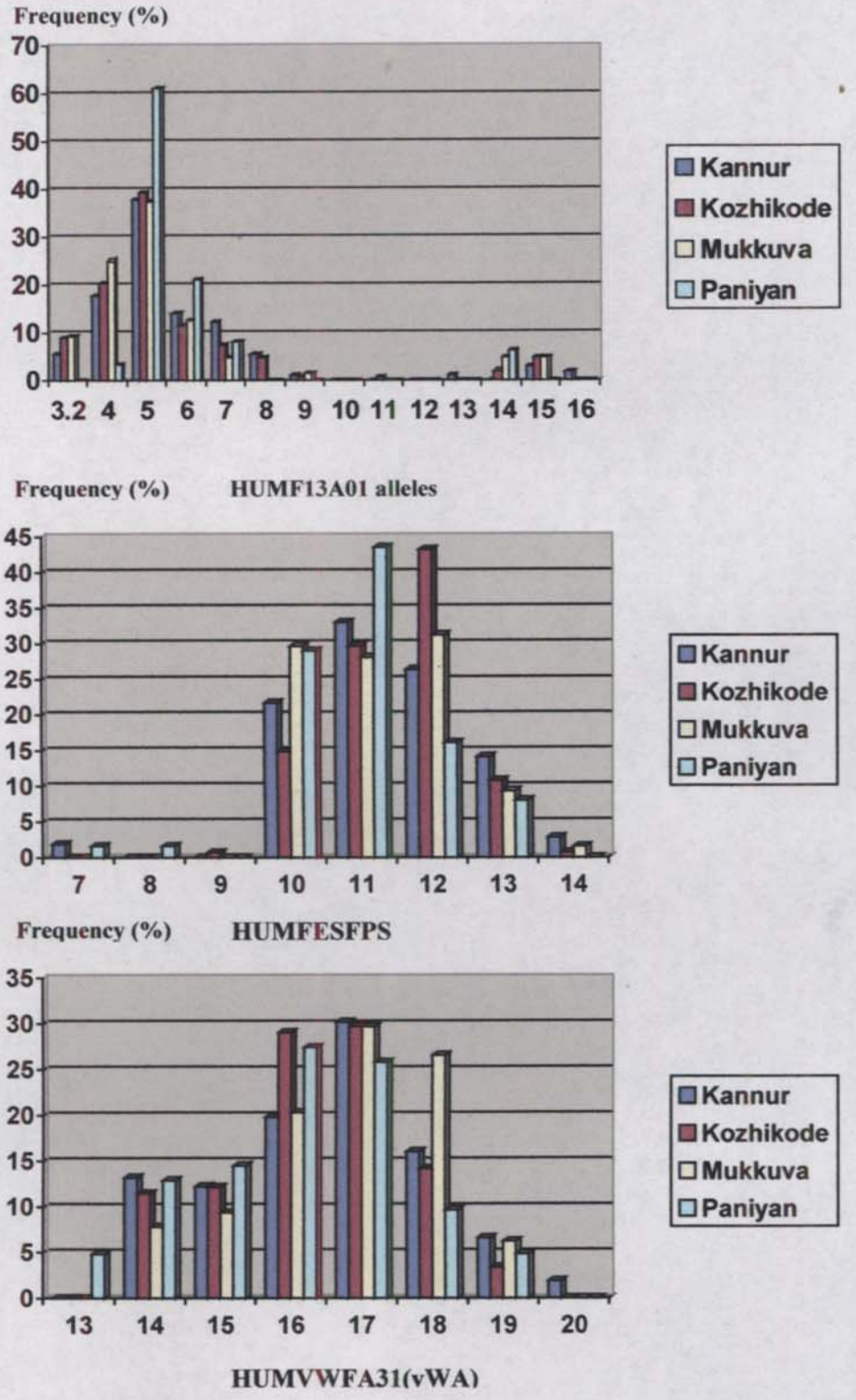

Fig. 4.2. Histograms depicting comparison of HUMF13A01, HUMFESFPS, HUMVWFA31(vWA) allele frequencies in Kannur, Kozhikode, *Mukkuva* and *Paniyan* population.

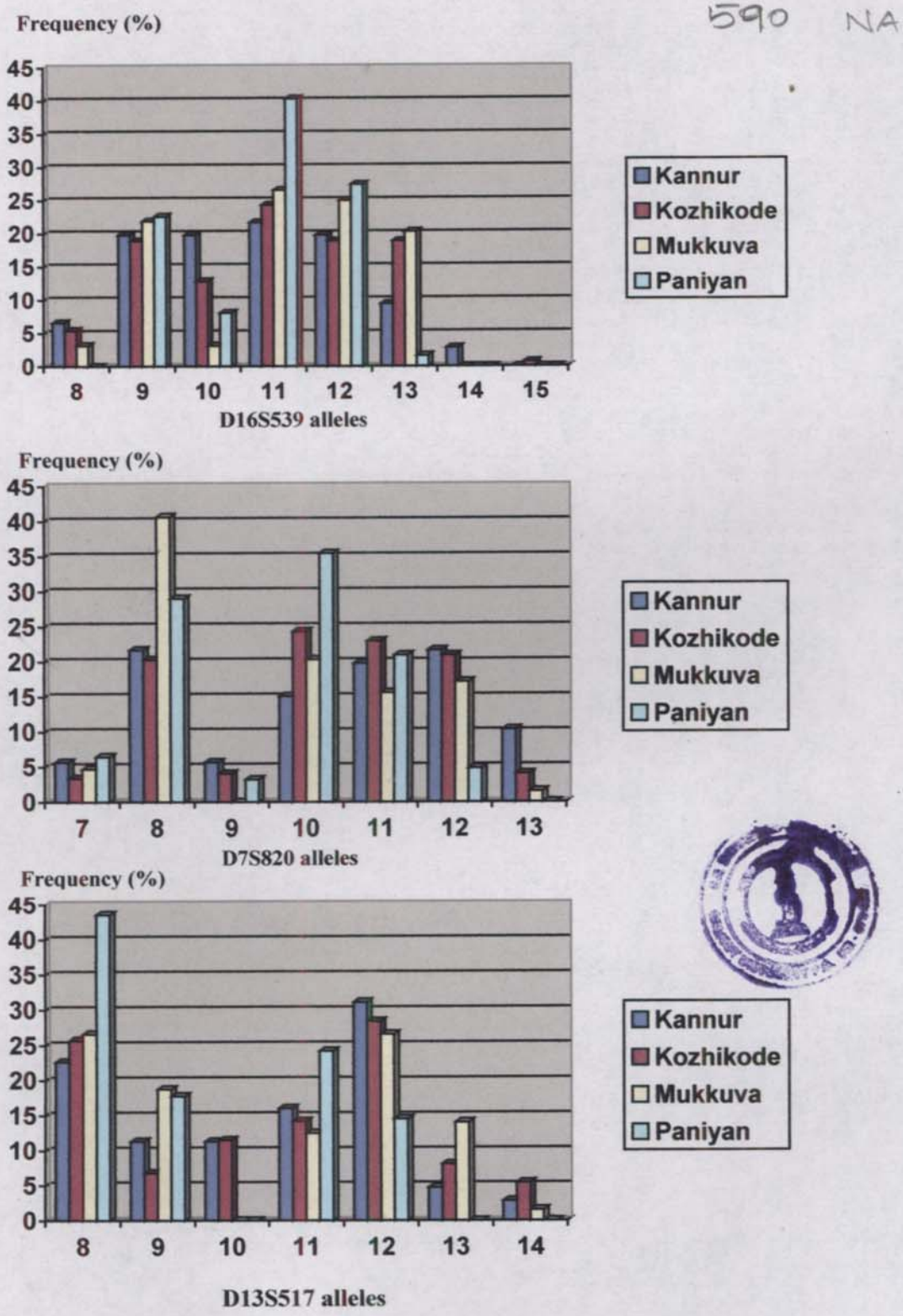

Fig. 4.3. Histograms depicting comparison of D16S539, D7S820, D13S317 allele frequencies in Kannur, Kozhikode, *Mukkuva* and *Paniyan* population.

Table 4.12. Most frequent and least frequent alleles seen in Kannur, Kozhikode, *Mukkuva* and *Paniyan* populations.

POP	ALLELE	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
KA	MF	12	11	6	5	11	17	11	8,12	12
	LF	8	9	9.3	9,13	7	20	14	7,9	14
KO	MF	12	11	9	5	12	17	11	10	12
	LF	14	7	5,11	9,11	9,14	19	15	7	14
MU	MF	12	8	6	5	12	17	11	8	8,12
	LF	13	10	5	9	14	19	8,10	13	14
PA	MF	12	8	6	5	11	16	11	10	8
	LF	15	9	8	4	7,8	13,19	13	9	12

KA: Kannur., KO: Kozhikode., MU: *Mukkuva*., PA: *Paniyan*., MF: Most frequent allele., LF: Least frequent allele.

Sampled populations were observed to be in Hardy-Weinberg (HWE) for all the analysed markers, $P > 0.05$ in the 9 systems. $P>0.05$ indicates that the population does not deviate from Hardy-Weinberg equilibrium and hence not significant (Sueblinvong and Kongsrisook, 1999). The genetic variability of a population is usually measured by the average heterozygosity per locus (Nei and Roychoudhury, 1974). The average heterozygosity (direct count) for 9 systems of Kannur population was more than 78% and for Kozhikode it has been seen as more than 75%; the same for *Mukkuva* was 73% and for *Paniyan* 68% (tables 4.8 – 4.11). The average heterozygosity (direct count), average expected heterozygosity under H-W equilibrium and average heterozygosity showed no significant differences. The data confirms the heterogeneity of the sample populations, one of the related factors of Hardy-Weinberg equilibrium.

Phylogenetic analyses

Phylogenetic study was performed through three types of methods:

1. *Contingency Table:* A contingency table (RxC test) was done to determine if significant differences in allele frequencies exist among the four populations (Sokal and Rohlf 1995). This was accomplished by a pairwise analysis of all populations (tables 4.13 – 4.18) through the exact test for population differentiation (Raymond and Rousset 1995) Fisher's combined probability test was also done to see the over all significance of the complete analysis of 9 STR loci (Fisher, 1954; Manly, 1991; Sokal and Rohlf, 1995). The overall *P* for Kannur vs. Kozhikode was found as 0.1716,

whereas Kannur vs. *Mukkuva*, Kannur vs. *Paniyan*, Kozhikode vs. *Mukkuva*, Kozhikode vs. *Paniyan* and *Mukkuva* vs. *Paniyan*, the P = 0.0000 and hence significant. The result clearly indicates that the populations of Kannur and Kozhikode are same or homogeneously mixed whereas other populations of *Mukkuva* and *Paniyan* are different from the Kannur and Kozhikode. It is also clear from the contingency table that the genotypes of *Mukkuva* and *Paniyan* populations are quite different.

Exact test for population differentiation with specifications:

# of dememorization steps: 1000
# of batches : 30
# of permutations per batch : 2000

Table 4.13. *Pairwise analysis of Kannur vs. Kozhikode population*

Locus	*P*	S.E.
CSF1PO	0.2445	0.0158
TPOX	0.3196	0.0133
THO1	0.7543	0.0152
F13A01	0.5197	0.0178
FESFPS	0.0275	0.0053
vWA	0.4046	0.0145
D16S539	0.1292	0.0124
D7S820	0.2508	0.0134
D13S317	0.7108	0.0140

Overall: P = 0.1716

Table 4.14. *Pairwise analysis of Kannur vs. Mukkuva*

Locus	*P*	S.E.
CSF1PO	0.0069	0.0015
TPOX	0.0001	0.0001
THO1	0.0002	0.0002
F13A01	0.0958	0.0082
FESFPS	0.6582	0.0082
VWA	0.6567	0.0108
D16S539	0.0086	0.0019
D7S820	0.0229	0.0026
D13S317	0.0165	0.0023

Overall: P = 0.0000

Table 4.15. *Pairwise analysis of Kannur vs. Paniyan*

Locus	*P*	S.E.
CSF1PO	0.0301	0.0027
TPOX	0.0002	0.0001
THO1	0.0598	0.0057
F13A01	0.0000	0.0000
FESFPS	0.1798	0.0088
vWA	0.2957	0.0102
D16S539	0.0033	0.0010
D7S820	0.0008	0.0007
D13S317	0.0001	0.0001

Overall: P = 0.0000

Table 4.16. *Pairwise analysis of Kozhikode vs. Mukkuva*

Locus	*P*	S.E.
CSF1PO	0.0002	0.0001
TPOX	0.0035	0.0010
THO1	0.0010	0.0010
F13A01	0.7038	0.0117
FESFPS	0.1230	0.0063
vWA	0.2495	0.0125
D16S539	0.3442	0.0152
D7S820	0.0550	0.0043
D13S317	0.0057	0.0018

Overall: P = 0.0000

Table 4.17. *Pairwise analysis of Kozhikode vs. Paniyan*

Locus	*P*	S.E.
CSF1PO	0.0006	0.0002
TPOX	0.0015	0.0007
THO1	0.1014	0.0076
F13A01	0.0000	0.0000
FESFPS	0.0013	0.0013
vWA	0.2508	0.0122
D16S539	0.0003	0.0002
D7S820	0.0169	0.0036
D13S317	0.0000	0.0000

Overall: P = 0.0000

Table 4.18. *Pairwise analysis of Mukkuva vs. Paniyan*

Locus	*P*	S.E.
CSF1PO	0.2091	0.0078
TPOX	0.9413	0.0024
THO1	0.0001	0.0001
F13A01	0.0000	0.0000
FESFPS	0.1592	0.0105
vWA	0.1047	0.0083
D16S539	0.0049	0.0013
D7S820	0.0524	0.0050
D13S317	0.0031	0.0009

Overall: P = 0.0000

2. *Genetic Distances*: The genetic difference between two populations may be measured by the genetic distance proposed by Nei and Roychoudhury (1974). Nei's (1972/1978) *Identities / Distances* and Nei's (1972/1978) *Minimum Distances* were conducted to check the genetic difference between the four selected populations (tables 4.19 and 4.20). All types of distance measures by Nei and Roychoudhury are non-negative, and thus the sampling variation of gene frequencies may produce non-zero estimates of distance even if the two populations under comparison are identical (Nei and Roychoudhury, 1974). The results of the present study was in tune with the R x C test as there is a homogeneity of Kannur and Kozhikode populations, non-relation of Kannur & Kozhikode with that of Mukkuva and Paniyan as well as non-relation of Mukkuva with Paniyan. Further Kannur population keeps little less distance to Mukkuva and Paniyan than that of Kozhikode population.

Table 4.19. Nei's (1972/1978) Identities / Distances

Populations Compared	*Distance*	*Identity*	*Unbiased Distance*	*Unbiased Identity*
Kannur *vs.* Kozhikode	0.0286	0.9718	0.0010	0.9990
Kannur *vs. Mukkuva*	0.0765	0.9264	0.0384	0.9624
Kannur *vs. Paniyan*	0.0990	0.9058	0.0649	0.9372
Kozhikode *vs. Mukkuva*	0.0863	0.9173	0.0533	0.9480
Kozhikode *vs. Paniyan*	0.1165	0.8900	0.0876	0.9161
Mukkuva vs. Paniyan	0.0960	0.9085	0.0566	0.9450

Table 4.20. *Nei's (1972/1978) Minimum Distances*

Populations Compared	*Original Minium*	*Unbiased Minimum*
Kannur *vs.* Kozhikode	0.0065	0.0003
Kannur *vs. Mukkuva*	0.0193	0.0098
Kannur *vs. Paniyan*	0.0294	0.0202
Kozhikode *vs. Mukkuva*	0.0216	0.0132
Kozhikode *vs. Paniyan*	0.0336	0.0254
Mukkuva vs. Paniyan	0.0280	0.0167

3. *Population Tree:* A population tree based on the genetic distances was constructed (Fig. 4.4) by the UPGMA (unweighted pair group-method with arithmetic mean) (Swofford and Olsen, 1990; Mount, 2001). The phenogram indicates that the Kannur and Kozhikode populations are very closely related and the *Paniyan* is most distantly related to the Kannur and Kozhikode population. Kannur and Kozhikode keeps next relation to *Mukkuva.*

Difference between populations will tend to diminish over time as a consequence of *gene flow.* This may come about because of migration of people between populations, or because of marriage between people from different populations. The homogeneous mix up of Kannur and Kozhikode populations can be attributed to the above fact (Evett and Weir, 1998). The individual identity and distance kept by the other two populations, *Mukkuva* and *Paniyan* can be due to their social identity and matrimonial alliances exclusively within their community. The distance seen by Kannur and Kozhikode populations with *Paniyan* and also the considerable distance by *Mukkuva* to *Paniyan,* points on the probability of *Paniyan* tribe as one of the earliest type of the population of North Kerala.

Fig. 4.4.

Probing North Kerala population

An unbiased North Kerala population was created by pooled genotypes of Kannur and Kozhikode populations, i.e., all the individuals of Kasargod, Kannur, Wayanad, Kozhikode and Mahe except the two selected populations, *Mukkuva*

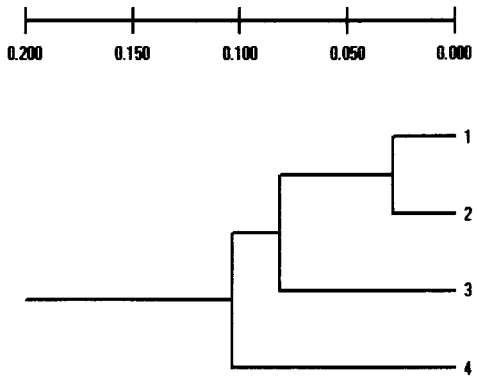

Fig. 4.4. UPGMA (unweighted pair group-method with arithmetic mean) to show the relationship of the populations. 1. Kannur, 2. Kozhikode, 3. *Mukkuva*, 4. *Paniyan*

community and *Paniyan* tribe. Allele frequencies and Hardy-Weinberg statistical parameters of North Kerala for the complete nine systems are given in table 4.21. The most and least frequent alleles for the nine STR systems are noted in table 4.22. A total of 69 alleles for all studied STRs could be observed in the population. The population was observed to be in Hardy-Weinberg equilibrium (HWE) for all the analysed markers ($P > 0.05$). The set of STR loci examined shows a high degree of genetic polymorphism in all cases having heterozygosity (direct count) values ranging from 62.99% to 84.25% (table 4.21). Average heterozygosity (direct count) seen as more than 76%. The average heterozygosity (direct count), average expected heterozygosity under H-W equilibrium and average heterozygosity showed no significant differences.

The nine loci showed a combined discrimination power (DP) of 0.9999999999 and a combined probability of exclusion (PE) 0.9997 and the marker D16S539 being the most informative. The results can be used as a population database for the North Kerala population, which is highly useful for forensic individualization and paternity tests.

Table 4.21. Allele frequencies and Hardy-Weinberg statistical parameters of 9 STR loci of North Kerala population (N = 254).

Allele	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
3.2	--	--	--	0.0748	--	--	--	--	--
4	--	--	--	0.1929	--	--	--	--	--
5	--	--	0.0039	0.3858	--	--	--	--	--
6	--	--	0.2795	0.1260	--	--	--	--	--
7	--	0.0079	0.1457	0.0945	0.0079	--	--	0.0433	--
8	0.0079	0.3031	0.1102	0.0512	--	--	0.0591	0.2087	0.2441
9	0.0079	0.0984	0.2677	0.0079	0.0039	--	0.1929	0.0472	0.0866
9.3	--	--	0.0236	--	--	--	--	--	--
10	0.1772	0.1299	0.1654	--	0.1772	--	0.1575	0.2047	0.1142
11	0.3071	0.4016	0.0039	0.0039	0.3110	--	0.2323	0.2165	0.1496
12	0.4016	0.0591	--	--	0.3622	--	0.1929	0.2126	0.2953
13	0.0945	--	--	0.0039	0.1220	--	0.1496	0.0669	0.0669
14	0.0039	--	--	0.0118	0.0157	0.1220	0.0118	--	0.0433
15	--	--	--	0.0394	--	0.1220	0.0039	--	--
16	--	--	--	0.0079	--	0.2520	--	--	--
17	--	--	--	--	--	0.2992	--	--	--
18	--	--	--	--	--	0.1496	--	--	--
19	--	--	--	--	--	0.0472	--	--	--
20	--	--	--	--	--	0.0079	--	--	--
HE	0.7040	0.7167	0.7889	0.7791	0.7255	0.7925	0.8208	0.8139	0.8040
HU	0.7068	0.7196	0.7920	0.7821	0.7283	0.7956	0.8240	0.8171	0.8071
H	0.6299	0.7087	0.7559	0.7559	0.7559	0.8425	0.8110	0.8346	0.7874
P	0.3311	0.3898	0.1103	0.1035	0.6385	0.7189	0.7605	0.8181	0.3363
S.E.	0.0143	0.0113	0.0078	0.0078	0.0146	0.0067	0.0116	0.0059	0.0126
DP	0.8608	0.8745	0.9236	0.9266	0.9337	0.9271	0.9425	0.9387	0.9350
PE	0.4712	0.4966	0.6066	0.6112	0.5016	0.6157	0.6618	0.6488	0.6366

N: number of chromosomes analyzed; HE: expected heterozygosity under H-W equilibrium; HU: heterozygosity (unbiased); H: heterozygosity (direct count); *P*: Hardy-Weinberg exact test; S.E.: standard error of *P*; DP: discrimination power; PE: Probability of exclusion

Results over all loci:
Average expected heterozygosity under H-W equilibrium: 0.7717
Average heterozygosity (unbiased) : 0.7747
Average heterozygosity (direct count) : 0.7647
Percentage of polymorphic loci over all loci (no criterion) : 100.0000

Table 4.22. Most frequent and least frequent alleles seen in North Kerala population in the 9 STR loci.

Allele	CSF1PO	TPOX	THO1	F13A01	FESFPS	vWA	D16S539	D7S820	D13S317
MF	12	11	6	5	12	17	11	11	12
LF	14	7	5,11	11,13	9	20	15	7	14

MF: most frequent allele., LF: least frequent allele

5 CONCLUSIONS

CONCLUSIONS

A very large number of studies have been carried out among Indian ethnic groups concerning classical polymorphisms such as blood groups, serum proteins, and red cell enzymes. The present study with the STR polymorphisms of the social groups of North Kerala comprising people from Kannur and Kozhikode and social groups of *Mukkuva* and *Paniyan* indicated the following phylogenetic relationships:

1. Kannur and Kozhikode population are homogeneously mixed.
2. *Paniyan* are far distantly related to Kannur and Kozhikode population.
3. *Mukkuva* show distance in relation to *Paniyan.*
4. *Mukkuva* population are more related to Kannur and Kozhikode than *Paniyan.*
5. Kannur population are more closely related to *Mukkuva* and *Paniyan* than Kozhikode population on the basis of genetic distance.

The existence of an appropriate population database is essential for a biometric interpretation of results obtained in forensic genetics; this study fulfils the establishment of a population database concerning the nine STR polymorphisms observed in North Kerala. The database showed no deviation from Hardy-Weinberg equilibrium and holds a high DP and PE, and therefore highly useful for scientific law enforcement authorities for forensic individualization and paternity analysis. The data generated in this study to a certain degree can also serve the study of ancient peoples' migration pattern within North Kerala and can be used for various phylogenetic studies.

6 BIBLIOGRAPHY

BIBLIOGRAPHY

Aaspollu, A., Sumeri, I. and Kelve, M. (2000). Allele distribution at nine STR loci – D3S1358, vWA, FGA, THO1, TPOX, CSF1PO, D5S818, D13S317 and D7S820 – in the Estonian population. *Progress in Forensic Genet.*, **8**: 129-131.

Adkins, R.M. and Honeycutt, R.L. (1993). A molecular examination of archontan and chiropteran monophyly. In: MacPhee R.D.E., (ed.), *Primates and their relatives in phylogenetic perspective*. Plenum Press, New York, pp. 227-249.

Adkins, R.M. and Honeycutt, R.L. (1994). Evolution of the primate cytochrome c oxidase II gene. *J. Mol. Evol.*, **38**: 215-231.

Adkins, R.M. and Honeycutt, R.L. (1991). Molecular phylogeny of the super-order Archonta. *Proc. Natl. Acad. Sci.*, USA, **88**: 10317-10321.

Agrawal, S., Muller, B., Bharadwaj, U., Bhatnagar, S. and Khan, F. (2002). Distribution of allele frequencies of six STR markers in North Indians. *J.Forensic Sci.*, **47**: 682-685.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1994). *Molecular Biology of The Cell, 3rd ed.*, Garland Publishing Co., New York.

Alford, R.L., Hammond, H.A., Coto, I., and Caskey, C.T. (1994). Rapid and efficient resolution of parentage by amplification of short tandem repeats. *Am.J.Hum.Genet.*, **55**: 190-195.

Allen, P.J., Amos, W., Pomeroy, P.P. and Twiss, S.D. (1995). Microsatellite variation in grey seals (Halichoerus grypus) shows evidence of genetic differentiation between two British breeding colonies. *Mol. Ecol.*, **4**: 653-662.

Allen, R., Budowle, B., Chakraborty, R., Giusti, A. and Eisenberg, A. (1991). Analysis of the VNTR locus *AmpliFLP* D1S80 by the PCR followed by high resolution PAGE. *Am. J Hum. Genet.*, **48**: 137-144.

Allen, R., Graves, G. and Budowle, B. (1989). Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver. *BioTechniques* **7**: 736-744.

Alonso, A., Andelinov, S., Martin, P., Sutlovic, D., Erceg, I., Huffine, E., De Simón, L.F., Albarrán, C., Definis-Gojanovic, M., Fernandez-Rodriguez, A., García, P., Drmic, I., Rezic B., Kuret, S., Sancho, M. and Primorac, D. (2001). DNA typing

from skeletal remains: Evaluation of multiplex STR systems on DNA isolated from bone and teeth samples. *Profiles in DNA*, **4**: 3-8.

Alonso, A., Castro, A., Fernandez, I., Gomez de Cedron, M., Garcia-Orad, A., Meyer, E. and Martinez de Pancorbo, M. (1995). Population study of 3 STR loci in the Basque Country (Northern Spain). *Int. J.Leg. Med.*, **107**:239-245.

Alonso, A., Martin, P., Albarran, C., Garcia, O. and Sancho, M. (1996). Rapid detection of sequence polymorphisms in the human mitochondrial DNA control region by polymerase chain reaction and single-strand conformation analysis in mutation detection enhancement gels. *Electrophoresis,* **17**: 1299-1301.

Alonso, M., Hamelin, R., Kim, M., Porwancher, K., Sung, T., Parhar, P., Miller, D.C. and Newcomb, E.W. (2001). Microsatellite instability occurs in distinct subtypes of pediatric but not adult central nervous system. *Cancer Res.*, **61**: 2124.

Amar A., Brautbar, C., Motro, U., Fisher, T., Bonne-Tamir, B. and Israel, S. (1999). Genetic variation of three tetrameric tandem repeats in four distinct Israeli ethnic groups. *J.Forensic Sci.*, **44**: 983-986.

Anderson, J.F., Greenhalgh, M.J., Butler, H.R., Kilpatrick, S.R., Piercy, R.C., Way, K.A., Myhill, H.S., Wright, J.C., Hallett, R. and Parkin, B.H. (1996). Further validation of a multiplex STR system for use in routine forensic identity testing. *Forensic Sci. Int.*, **78**: 47-64.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. *Nature,* **290**: 457-465.

APPLIED BIOSYSTEMS, *AmpFlSTR Cofiler PCR Amplification Kit, User's Manual.* (1998). Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS, *AmpFlSTR Identifiler PCR Amplification Kit, User's Manual* .(2001). Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS, *AmpFlSTR Profiler Plus PCR Amplification Kit, User's Manual.* (2000). Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS, *AmpliFLP D1S80 PCR Amplification Kit, Part No. N808-0054/Manual.*(1993). The Perkin-Elmer Corporation, currently Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS, *AmpliTypePM + DQA1 PCR Amplification and Typing Kit, Part No. N808-0094 /Manual.*(1995). The Perkin-Elmer Corporation,currently Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS, *AmpliType User's Guide, Version 2* (1993). The Perkin-Elmer Corporation,currently Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS (2001). *Brochure on ABI PRISM 7000 Sequence Detection System.* Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS(2004). Product News, Announcing the family of Real-Time PCR platforms from Applied Biosystems. *Biosystems Solutions, European Edn., Issue 10 Summer* - Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS (2001). *Product Catalogue.* Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

APPLIED BIOSYSTEMS, *QuantiBlot Human DNA Quantitation Kit Protocol.* (2000). Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

Aquadro, C.F. and Greenberg, B.D. (1983). Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics,* **103**: 287-312.

Armour, J.A.L., Anttinen, T., May, C.A., Vega, E.E., Sajantila, A., Kidd, J.R., Kidd, K.K., Bertranpetit, J., Paabo, S. and Jeffreys, A.J. (1996). Minisatellite diversity supports a recent African origin for modern humans. *Nature Genet.,* **13**: 154-160.

Arya, R.P, editor.(2002). Kerala Road Atlas & Distance Guide. *Indian Map Service, Jodhpur (Raj.). India.*

Asmundo, A. and Crino, C.(1998). Population study of the short tandem repeat polymorphisms HumTHO1, HumvWA31, HumFESFPS and HumF13A01 in Sicily (Southern Italy). *Int. J.Leg. Med.,* **111**: 281-283.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1999). *Short Protocols in Molecular Biology.* A compendium of methods from Current Protocols in Molecular Biology, John Wiley & Sons, Inc. NY.

Avery, O.T., MacLeod, C.M. and McCarty, M. (1944). Studies of the chemical nature of the substance inducing transformation of pneumococcal types. Induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* Type III. *J. Exp. Med.*, **79**: 137-158.

Ayres, K.L., Chaseling, J., Balding, D.J. (2002). Implications for DNA identification arising from an analysis of Australian forensic databases. *Forensic Sci. Int.*, **129**: 90-98.

Baasner, A., Schafer, C., Junge, A. and Madea, B. (1998). Polymorphic sites in human mitochondrial DNA control region sequences: population data and maternal inheritance. *Forensic Sci. Int.*, **98**:169-178.

Bacher, J. and Schumm, J.W. (1998). Development of highly polymorphic pentanucleotide tandem repeat loci with low stutter. *Profiles in DNA* **2**: 3. (www.promega.com/profiles/202/202-03).

Bains, W. (1992). Local sequence dependence of rate of base replacement in mammals. *Mutat. Res.*, **267**: 43-54.

Balakrishna, J. (1991). Accuracy of DNA fingerprinting. Science Express-Window on Science, Technology & Medicine. *Indian Express*, Tuesday, May 7, 1991.

Balding, D.J., Greenhalgh, M. and Nichols, R.A. (1996). Population genetics of STR loci in Caucasians. *Int. J.Leg. Med.*, 108: 300-305.

Bamshad, M., Bhaskara, R.B., Naidu, J.M., Prasad, B.V.R., Watkins, S. and Jorde, L. (1997). Letters to the editor. *Hum. Bio.*, 69: 432-435.

Bamshad, M., Fraley, A.E., Crawford, M.H., Cann, R.L., Busi, B.R., Naidu, J.M.and Jorde, L.B (1996). mtDNA variation in caste populations of Andhra Pradesh, .India. *Hum. Bio.*, **68**: 1-28.

Bamshad, M., Kivisild, T., Watkins, W.S., Dixon, M.E., Ricker, C.E., Rao, B.B., Naidu, J.M., Ravi Prasad, B.V., Reddy, P.G., Rasanayagam, A., Papiha, S.S., Villems, R., Redd, A.J., Hammer, M.F., Nguyen, S.V., Carrol, M.L., Batzer, M.A. and Jorde, L.B. (2001). Genetic evidence on the origins of Indian caste populations. *Genome Res.*, **11**: 994-1004.

Barnes, W.M. (1994). PCR amplification of up to 35 kb DNA with high fidelity and high yield from λ templates. *Proc. Natl. Acad. Sci., USA,* **91**: 2216-20.

Barros, F., Lareu, M.V., Salas, A. and Carracedo, A.(1997). Rapid and enhanced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products. *Electrophoresis,* **18**: 52-54.

Bassam, B.J., Caetano-Anolles, G. and Gresshoff, P.M. (1991). Fast and sensitive silver staining of DNA in polyacrylamide gels. *Anal. Biochem.,* **196**: 80.

Battacharyya, N.P., Basu, P., Das, M., Pramanik, S., Banerjee, R., Roy, B., Roychoudhury, S. and Majumder, P. (1999). Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res.*, **9**: 711-719.

Beckmann J.S. and Weber J.L. (1992). Survey of human and rat microsatellites. *Genomics,* **12**: 627-631.

Bell, B., Nievas, P., Abecia, E., Jarreta, M-B., Hinojal, R. and Cordero, M-A. (2000). Population genetics of the STR loci HUMCSF1PO, HUMF13A01, HUMFES/FPS and D12S391 in Asturias (northern Spain). *Forensic Sci. Int.,* **113**: 21-23.

Bell, G. and Torney, D. (1993). Repetitive DNA sequences: some considerations for simple sequence repeats. *Comp. Chem.,* **17**: 185-190.

Benciolini, P., Biasiolo, M., Cortivo, P. and Caenazzo, L.. (1985). Problems in probability of paternity interpretation. *Forensic Sci. Int.,* **29**: 199-206.

Bender, K., Schneider, P.M. and Rittner, C. (2000). Application of mtDNA sequence analysis in forensic casework for the identification of human remains. *Forensic Sci. Int.,* **113**: 103-107.

Benson, G. and Dong, L. (1997). Reconstructing the Duplication History of a Tandem Repeat. Dept. of Biomathematical Sciences, The Mount Sinai School of Medicine, New York, NY 10029-6574.

Beridze, T .(1986). *Satellite DNA.* Springer-Verlag, Berlin

Bertranpetit, P., Sala, J., Calafell, F., Underhill, P.A., Moral, P. and Comas, D. (1994). Human mitochondrial variation and the origin of Basques. *Ann. Hum. Genet.,* **59**:63-81.

Bio .(2002). 'Teaching in the Genome Age'. *In: Function of Microsatellites,* http://woodrow.org/teachers/esi/2002/Biology/Projects.

Bjerre, A., Court, D.S., Lincoln, P. and Morling, N. (1997). A report of the 1995 and 1996 paternity testing workshops of the English speaking working group of the International Society of Forensic Haemogenetics. *Forensic Sci. Int.,* **90**: 41-55.

Blab, G., Jaubert, B., Oesterreich, W., Ackermann, E., Pieper, A. and Herrmann, S. (1996). Use of PCR in forensic casework in the area Belin-Brandenburg: allele frequency distribution of six microsatellites. *Adv. in Forensic Haemogenetics,* **6**: 287-289.

Blake, E., Mihalovich, J., Higuchi, R., Walsh, P.S. and Erlich, H. (1992). Polymerase Chain Reaction (PCR) amplification and human leukocyte antigen (HLA)-DQα oligonucleotide typing on biological evidence samples: casework experience. *J.Forensic Sci.,* **37**: 700-726.

Bogenhagen, D. and Clayton D.A. (1974). The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. *J. Biol. Chem.,* **249**: 7991-7995.

Bohossian, H. B., Skaletsky, H. and Page, D. C. (2000). Unexpectedly similar rates of nucleotide substitution found in male and female hominids. *Nature,* **406**: 622-625.

Boissinot, S., Chevret, P. and Furano, A. V. (2000). L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.,* **17**: 915-928.

Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J.R. and Cavalli-Sforza, L.L.(1994). High resolution of human evolution with polymorphic microsatellites. *Nature,* **368**: 455-457.

Brauer, G. (1984). In: *Origins of Modern Humans: A World Survey of the Fossil Evidence* (Eds. Smith F.H and Spencer, F), Liss, New York,pp. 327-410

Brauer, G. (1989). In: *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans.* Mellars, P and Stringer, C., Eds., Edinburgh Univ. Press, Edinburgh, pp. 123-154.

Brauer, G. and Stringer, C.B. (1997). Models, polarization and perspectives on modern human origins. In: *Conceptual issues in Modern Human Origins Research.* Clark, G.A and Willermet, C.M., Eds., New York: de Gruyer.

Brinkmann, B., Sajantila, A., Goedde, H.W., Matsumoto, H., Nishi, K. and Wiegand, P. (1996). Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci. *Eur. J.Hum. Genet.*, **4**: 175-182.

Broman, K. W., Murray, J. C., Sheffield, V. C., White, R. L. and Weber, J. L.(1998). Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.*, **63**: 861-869.

Brosius, J. (1999). Genomes were forged by massive bombardments with retroelements and retrosequences. *Genetica*, **107**: 209-238.

Brown, M.D. and Wallace, D.C .(1994). Molecular basis of mitochondrial DNA disease. *J Bioenerg. Biomembr.*, **26**: 273-289.

Brown, W.M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natn. Acad. Sci.*, U.S.A, **77**: 3605-3609.

Brown, W.M., George, M. and Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci., USA* **76**: 1967-1971.

Brown, W.M., Prager, E.M., Wang, A. and Wilson, A.C.(1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution. *J. Molec. Evol.*, **18**: 225-239.

Bruford, M.W. and Wayne, R.K. (1993). Microsatellites and their application to population genetic studies. *Curr. Opin. Genet, Dev.*, **3**:939-943.

Budowle, B. and Baechtel, F.S. (1990). Modifications to improve the effectiveness of Restriction Fragment Length Polymorphism Typing. *Appl. and Theoret. Elec.*, **1**: 181-187.

Budowle, B., Baechtel, F.S. and Fejeran, R. (1998a). Polymarker, HLA-DQA1, and D1S80 allele frequency data in Chamorro and Filipino populations from Guam. *J.Forensic Sci.*, **43**: 1195-1198.

Budowle, B., Baechtel, F.S., Giusti, A.M., and Monson, K.L. (1990). Applying highly polymorphic variable number of tandem repeats loci genetic markers to identity testing. *Clinical Biochem.*, **23**: 1-7.

Budowle, B., Baechtel, F.S., Smerick, J.B., Presley, K.W., Giusti, A.M., Parsons, G., Alevy, M.C. and Chakraborty, R. (1995a). D1S80 population data in African Americans, Caucasians, Southeastern Hispanics, Southwestern Hispanics, and Orientals. *J.Forensic Sci.*, **40**: 38-44.

Budowle, B., Chakraborty, R., Giusti, A.M., Eisenberg, A.J. and Allen, R.C. (1991). Analysis of the VNTR locus D1S80 by the PCR followed by High-Resolution PAGE. *Am. J Hum. Genet.,* **48**: 137-144.

Budowle, B., Giusti, A.M., Waye, J.S., Baechtel, F.S., Fourney, R.M., Adams, D.E., Presley, L.A., Deadman, H.A. and Monson, K.L. (1991). Fixed-Bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR loci, for use in forensic comparisons. *Am. J. Hum. Genet.,* 48: 841-855.

Budowle, B., Lindsey, J.A., DeCou, J.A., Koons, B.W., Giusti, A.M. and Comey, C.T. (1995b). Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci). and HLA-DQα using multiplex amplification and typing procedure. *J.Forensic Sci.,* **40**: 45-54.

Budowle, B., Masibay, A., Anderson, S.J., Barna, C., Biega, L., Brenneke, S., Brown, B.L., Cramer, J., DeGroot, G.A., Douglas, D., Duceman, B., Eastman, A., Giles, R., Hamill, J., Haase, D.J., Janssen, D.W., Kupferschmid, T.M., Lawton, T., Lemire, C., Llewellyn, B., Moretti, T., Neves, J., Palaski, C., Schueler, S., Sgueglia, J., Sprecher, C., Tomsey, C. and Yet, D. (2001). STR primer concordance study. *Forensic Sci. Int.,* **124**: 47-54.

Budowle, B., Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A. and Keys, K.M.(1999). Population Data on the thirteen CODIS core short tandem repeat loci in African American, U.S. Caucasians, Hispanics, Bahamians, Jamaicans and Trinidadians. *J. Forensic Sci.*, **44**: 1277-1286.

Budowle, B., Moretti, T.R., Keys, K.M., Koons, B.W. and Smerick, J.B.(1997). Validation studies of the CTT STR multiplex system.*J. Forensic Sci.*, **42**: 701-707.

Budowle, B., Moretti, T.R., Niezgoda, S.J. and Brown, B.L. (1998b). CODIS and PCR-Based Short Tandem Repeat Loci: Law Enforcement Tools. In: *Second European Symposium on Human Identification,* PROMEGA Corporation, Madison, Wisconsin, pp.73-88.

Budowle, B., Waye, J.S., Shutler, G.G., and Baechtel, F.S. (1990). "Hae III – A suitable restriction endonuclease for restriction fragment length polymorphism analysis of biological evidence samples." *J.Forensic Sci.,* 35(2): 530-536.

Busque, L., Desmarais, D., Provost, S., Schumm, J.W.,Zhong, Y. and Chakraborty, R. (1997). Analysis of allele distributions for six short tandem repeat loci in the French Canadian population of Quebec. *J. Forensic Sci.*, **42**: 1147-1153.

Cann, R.L., Stoneking, M. and Wilson, A.C. (1987). *Mitochondrial DNA and human evolution. Nature*, **325**: 31-36.

Carracedo, A., Aloja, E.D., Dupuy, B., Jangblad, A., Karjalainen, M., Lambert, C., Parson, W., Pfeiffer, H., Pfitzinger, H., Sabatier, M., Court, D.S., Vide, C. (1998). Reproducibility of mtDNA analysis between laboratories: a report of the European DNA profiling group (EDNAP). *Forensic Sci. Int.*, **97**: 165-170.

Carracedo, A., Bar, W., Lincoln, P., Mayr, W., Morling, N., Olaisan, B., Schneider, P., Budowle, B., Brinkman, B., Gill, P., Holland, M., Tully, G. Wilson, M. (2000). DNA commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Sci. Int.*, **110**: 79-85.

Caskey, C.T. (1987). Disease diagnosis by recombinant DNA methods. *Science*, **236**: 1223-1229.

Caskey, C.T and Hammond, H.A. (1992). Forensic use of short tandem repeats via PCR. *Adv. in Forensic Haemogenetics*, 4: 18-25.

Chapco, W. (1976). An exact test of the Hardy-Weinberg law. *Biometrics*, **32**: 183-189.

Charlesworth, B., Sniegowski, P., and Stephan, W.(1994). The evolutionary dynamics of repetitive DNA in Eukaryotes. *Nature*, **371**: 215-220.

Chen, Y.S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S. and Wallace, D.C. (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.*, **57**: 133-149.

Cheng, S., Fockler, C., Barnes, W.M. and Higuchi, R. (1994a). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proc. Natl. Acad. Sci.*, USA, **91**: 5695-5699.

Cheng, S., Sheng-Yung,C.,Gravit, P. and Respess, R. (1994b). Long PCR .*Nature* ,**369**: 684-685.

Claridge, M.F., Dawah, H.A. and Wilson, M.R., Eds. (1997). *Species – The units of biodiversity*. Chapman and Hall.

Clark, J. B. and Tidwell, M. G. (1997). A phylogenetic perspective on P transposable element evolution in Drosophila. *Proc. Natl Acad. Sci.* USA, **94**: 11428-11433.

Clayton, T.M., Whitaker, J.P. and Maguire, C.N. (1995). Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. *Forensic Sci. Int.,* **76**: 7-15

Cohen, S.N., Chang, A.C., Boyer, H.W and Helling, R.B. (1973). Construction of biologically functional bacterial plasmids *in vitro. Proc. Natl Acad. Sci. USA,* **70**: 3240-3244.

Collins, A., Lonjou, C. and Morton, N. E.(1999). Genetic epidemiology of single-nucleotide polymorphisms. *Proc. Natl. Acad. Sci.*USA, **96**:15173-15177.

Collins, F. and Galas, D. (1993). A new five-year plan for the US Human Genome Project. *Science,* **262**: 43-46.

Collins, F.S., Green, E.D., Guttmacher, A.E. and Guyer, M.S. (2003). A vision for the future of genomics research. *Nature,* **422**: 835-847.

Collins F.S., Morgan, M. and Patrinos, A. (2003). The Human Genome Project: Lesson from Large Scale Biology. *Science,* **300**: 286-290.

Comey, C.T and Budowle, B. (1991). Validation studies on the analysis of the HLA DQα locus using the polymerase chain reaction. *J.Forensic Sci.,* **36**: 1633-1648.

Corte-Real, F., Souto, L., Anjos, M.J., Carvalho, M., Vieira, D.N. Carracedo, A. and Vide, M.C. (1999a). Population distribution of six PCR-amplified loci in Madeira Archipelago (Portugal). *Forensic Sci. Int.,* **100**: 93-99.

Corte-Real, F., Souto, L., Anjos, M.J., Carvalho, M., Vieira, D.N., Carracedo, A. and Vide, M.C. (1999b). Population study of HUMTHO1, HUMVWA31/A, HUMF13A1, AND HUMFES/FPS systems in Azores. *J. Forensic Sci.*, **44**: 1261-1264.

Cortivo, P., Biasiolo, M., Scorretti, C., Caenazzo, L. and Benciolini, P. (1984). The PGM1 polymorphism as revealed by ultrathin-layer isoelectric focusing in the population of Padua. *Forensic Sci. Int.,* **24**: 169-172.

Cossutta, F., Perossa, R., Frassanito, F., Altamura, B.M. and Fattorini, P. (2000). Population genetics of fourteen STR loci in North-East Italy. *Progress in Forensic Genet.,***8**: 178-180.

Culliford, B.J. (1971). *The Examination and Typing of Bloodstains in the Crime Laboratory.* U.S. Government Printing Office, Washington, DC, 1971.

Cummings, C.J. and Zoghbi, H.Y. (2000). Trinucleotide repeats: mechanism and pathophysiology. *Ann. Rev. Genomics Hum. Genet.*, **1**: 281-328.

Czarnik, A.W. (1998). Illuminating the SNP genomic code. *Modern Drug Disc.* Nov/Dec: 49-54.

Da-Costa, A.N., Silva, R. and Moura-Neto, R.S. (2002). Y-chromosome variation in a Rio de Janeiro, Brazil, population sample. *Forensic Sci. Int.*, **126**: 254-257.

D'Andrade, R. and Morin, P.A. (1996). A principal components and individual-by-site analysis of chimpanzee and human mitochondrial DNA. *Am. Anthropol.*, **98**:2-20.

Da Silva, L.A.F., Pimental, B.J., De Azevedo, D.A., Da Silva, E.N.P. and Dos Santos, S.S. (2002). Allele frequencies of nine STR loci – D16S539, D7S820, D13S317, CSF1PO, TPOX, THO1, F13A01, FESFPS and vWA – in the population from Alagoas, northeastern Brazil. *Forensic Sci. Int.*, **130**: 187-188.

Deininger, P.L. and Daniels, G.R. (1986). The recent evolution of Mammalian Repetitive DNA Elements. *Trends in Genet.*, **2**: 76-80.

De-Knijff, P., Kayser, M., Caglia, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Penacino, G., Perez-Lezaun, A., Piccinini, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichhold, G. and Roewer, L. (1997). Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int. J. Leg. Med.*, **110**: 134-149.

DeWalt, T.S., Sudman, P.D., Hafner, M.S. and Davis, S.K. (1993). Phylogenetic relationships of pocket gophers (Cratogeomys and Pappogeomys) based on mitochondrial DNA cytochrome b sequences. *Mol. Phyl. Evol.*, **2**: 193-204.

Dickinson, K. (2001). Single Nucleotide Polymorphisms: Their affect on Human Genomic Research and Biotechnological Advances. http://www.*Single Nucleotide Polymorphisms.htm.*

Di Rienzo, A., Peterson, A., Garza, J., Valdes, A., Slatkin, M. and Freimer, N. (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci., USA*, **91**: 3166-3170.

Discovery health.com., (2001). Use of DNA forensic science in military matters. In: *Forensic Scicene and Genetic Variation.* Copyright 2001, Discovery

Communications Inc.

Disotell, T.R., Honeycutt, R.L. and Ruvolo, M. (1992). Mitochondrial DNA phylogeny of the Old-World monkey tribe Papionini. *Mol. Biol. Evol.*, **9**: 1-13.

DNA Amplification System. (1989). Order No. L-1072B, *Perkin Elmer Cetus (Perkin-Elmer Corp.)* – currently, Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

DNA Forensics Databases. (2003) In: *DNA Forensics*. Human Genome Project Information, base URL: *http://www.ornl.gov/hgmis*.

DNA Thermal Cycler 480, *User's Manual.* (1995). *Perkin-Elmer Corporation* – currently , Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

Dodd, B.E. (1985). DNA Fingerprinting in the matters of family and crime. *Nature,* 318: 506-507.

Domenici, R., Bibbiani, R., Fornaciari, S., Nardone, M., Rocchi, A., Spinetti, I., Venturi, M. and Bargagna, M.(1998). Application of conventional and DNA polymorphisms to paternity cases. *Progress in Forensic Genet.,* **7**: 89-91.

Domenici, R., Fornaciari, S., Nardone, M., Rocchi, A., Spinetti, I., Venturi, M. and Presciuttini, S. (2000). Allele frequencies of the HUMTHO1, HUMVWA31/A, HUMFES/FPS, HUMF13A1 loci in Tuscany (Italy). *Progress in Forensic Genet.,* **8**: 164-166.

Donald R. Johns. (1995). Mitochondrial DNA and Disease. *N. Eng. J. Med.,* **333**: 638-644.

Dorit, R.L., Akashi, H., Gilbert, W. (1995). Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science*,**268**: 1183.

Edelmann, J., Deichsel, D., Hering, S., Plate, I. and Szibor, R (2002a). Sequence variation and allele nomenclature for the X-linked STRs. *Forensic Sci. Int.,* 129: 99-103.

Edelmann, J., Hering, S., Kuhlisch, E. and Szibor, R. (2002b). Validation of the STR DXS7424 and the linkage situation on the X-chromosome. *Forensic Sci. Int.,* 125: 217-222.

Edelmann, J., Hering, S., Michael, M., Lessing, R., Deichsel, D., Meier-Sundhausen, G.,Roewer, L., Plate, I. and Szibor, R. (2001). 16 X-chromosome STR loci frequency data from a German population. *Forensic Sci. Int.,* 124: 215-218.

Edwards, A., Civitello, A., Hammond, H.A. and Caskey, C.T. (1991a). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.*, **49**, 746-756.

Edwards, A., Hammond, H.A., Chakraborty, R. and Caskey, C.T. (1991b). DNA typing with trimeric and tetrameric tandem repeats: polymorphic loci, detection systems, and population genetics. In: *The Second International Symposium on Human Identification, Scottsdale, AZ.* **31(52).** 1-9.

Edwards A., Hammond, H.A., Jin, L.,Caskey, C.T. and Chakraborty, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics,* **12**: 241-253.

Egyed, B., Furedi, S., Angyal, M., Boutrand, L., Vandenberghe, A., Woller, J. and Padar, Z.S.(2000). Analysis of eight STR loci in two Hungarian populations. *Forensic Sci. Int.,* **113**: 25-27.

Ellegren, H.(2000). Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genet.,* **24**: 400-402.

Elston, R.C. and Forthofer, R. (1977). Testing for Hardy-Weinberg equilibrium in small samples. *Biometrics,* **33**: 536-542.

Emigh, T.H. (1980). A comparison of tests for Hardy-Weinberg equilibrium. *Biometric,s* **36**: 627-642.

Emigh, T.H. and Kempthorne, O. (1975). A note on goodness-of-fit of a population to Hardy-Weinberg structure. *Am. J. Hum. Genet.,* **27**: 778-783.

Entrala, C., Lorente, J.A., Lorente, M., Alvarez, J.C., Budowle, B. and Villanueva, E. (1999). Spanish population data on the loci D13S317, D7S820 and D16S539 generated using silver staining (SilverSTR III Multiplex). *J.Forensic Sci.,* **44**: 1032-1034.

Erlich, H.A. (1989). ed., *PCR Technology: Principles and Applications for DNA Amplification.* Stockton Press Inc., NY.

Erlich, H.A. and Bugawan, T.L. (1989). HLA class II gene polymorphism: DNA typing, evolution, and relationship to disease susceptibility. In: *PCR Technology, Principles and Applications for DNA Amplification,* ed. Erlich, H.A., Stockton Press Inc., NY, pp.193-208.

Esnault, C., Maestre, J. and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.,* **24**: 363-367 .

Estoup, A., Garnery, L., Solignac, M., and Cornuet, J.M. (1995b). Microsatellite variation in honey bee populations: Hierarchical genetic structure and test of the IAM and SMM. *Genetics,* **140**: 679-695.

Estoup, A., Solignac, M., Cornuet, J.M, Goudet, J. and Scholl, A. (1996). Genetic differentiation of continental and island populations of Bombus terrestris in Europe. *Mol. Ecol.*, **5**: 19-31.

Estoup, A., Tailliez, C., Cornuet, J.M., and Solignac, M. (1995a). Size homoplasy and the mutational processes of interrupted microsatellites in two bee species, Apis mellifera and Bombus terrestrus (Apidae). *Mol. Biol. Ecol.,* **12**:1074-1084.

Evett, I.W., Lambert, J.A., Knight S.D., Fairley, M. and Lee, L.D.(1996). A study of independence between STR and conventional blood type loci. *Forensic Sci. Int.,* **79**: 163-166.

Evett, I.W. and Weir, B.S. (1998). Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. Sinauer Associates, Inc., U.S.A.

Excoffier, L. and Langaney, A.(1989). Origin and differentiation of human mitochondrial DNA. *Am. J. hum. Genet.,* **44**: 73-85.

Farfan ,M.J., Prieto, V., Torres, Y., Lopez-Soto, M. and Sanz, P. (2001). STR data for the APPLIED BIOSYSTEMS, AmpFlSTR Profiler Plus and Cofiler loci from the Maghreb (North Africa). *Forensic Sci. Int.,* **121**: 199-200.

Feng-Xia Xiao., Anja Gilissen., Jean-Jacques Cassiman. and Ronny Decorte. (1998). Quadruplex fluorescent STR typing system (HUMVWA, HUMTHO1, D21S11 and HPRT) with sequence-defined allelic ladders. Identification of a new allele at D21S11. *Forensic Sci. Int.,* **94**: 39-46.

Ferrel, R.E., Ueda, N., Satoh, C., Tanis, R.J., Neel, J.V., Hamilton, H.B., Inamizu, T. and Baba, K. (1977). The frequency in Japanese of Genetic Variants of 22 proteins. I. Albumin, Ceruloplasmin, Haptoglobin and Transferrin. *Annals of Hum. Genet.,* **40**: 407-418.

Filipski, J. (1987). Correlation between molecular clock ticking, codon usage fidelity of DNA repair, chromosome banding and chromatin compactness in germline cells. *FEBS Lett.*, **217**: 184-186.

Fisher, D.L., Holland, M.M., Mitchell, L., Sledzik, P.S., Wilcox, A.W., Wadhams, M. and Weedn, V.W. (1993). Extraction, Evaluation, and Amplification of DNA

from decalcified and un-decalcified United States Civil War Bones. *J.Forensic Sci.*, **38**: 60-68.

Fisher, R. (1951). Standard calculations for evaluating a blood group system. *Heredity* **5**: 95-102.

Fisher, R.A. (1954). *Statistical methods for research workers.* 12th ed., Oliver and Boyd, Edinburg.

Fisher, R.A. (1935). *The logic of scientific inference. J. Roy. Stat. Soc.* **98**: 39-54.

Foley, J. (2003). Fossil Hominids: mitochondrial DNA. http://www.talkorigins.org.

Forbes, S.H., Hogg, J.T., Buchanan, F.C., Crawford, A.M. and Allendorf, F.W. (1995). Microsatellite evolution in congeneric mammals: Domestic and Bighorn sheep. *Mol. Biol. Evol.*, **12**: 1106-1113.

Foreman L.A. and J.A. Lambert. (2000). Genetic differentiation within and between four UK ethnic groups. *Forensic Sci. Int.*, 114: 7-20.

Fowler, J.C.S., Burgoyne, L.A., Scott, A.C. and Harding, H.W.J. (1988). Repetitive Deoxyribonucleic Acid (DNA) and Human Genome Variation – A Concise Review Relevant to Forensic Biology. *J.Forensic Sci.*, **33** : 1111-1126.

Francke, U. (1999). Williams-Beuren syndrome: genes and mechanisms. *Hum. Mol. Genet.*, **8**: 1947-1954.

Fregeau, C.J, Bowen, K.L. and Fourney, R.M. (1999). Validation of highly polymorphic fluorescent multiplex short tandem repeat systems using two generations of DNA sequencers. *J.Forensic Sci.*, **44**: 133-166.

Fung, W.K., Liu, Y., Hou, Z., Tang, H. and Jia, S. (2001). Chinese population data in Beijing at 3 STR loci - CSF1PO, HUMTHO1, and TPOX. *J.Forensic Sci.*, **46**: 1000.

Fung, W.K., Ye. J., Hu, L., Zhao, X., Liu, B., Wong, D.M. and Law, M.Y. (2001). Allele frequencies for nine STR loci in Beijing Chinese. *Forensic Sci. Int.*, **121**: 207-209.

Gaensslen, R.E. and Camp, F.R. (1984). Forensic Serology: Analysis of Bloodstains and Body Fluid Stains. In: *Forensic Sciences*, Vol. 2, C.H. Wecht, Ed., Matthew Bender, New York.

Gaensslen, R.E. and Camp, F.R. (1986). Forensic Serology: Parentage Testing. In: *Forensic Sciences*. C.H. Wecht, Ed., Matthew Bender, New York, Vol. 2.

(originally as Camp, F.R. and Gaensslen, R.E., Forensic Serology: Paternity Testing and Transfusion Reactions).

Gaensslen, R.E. (1983). *Sourcebook in Forensic Serology, Immunology and Biochemistry.* U.S. Government Printing Office, Washington, DC.

Gaensslen, R.E., Bell, S.C. and Lee, H.C. (1987a). Distributions of Genetic Markers in United States Populations: I. Blood Group and Secretor Systems. *J. Forensic Sci.*, **32**: 1016-1058.

Gaensslen, R.E., Bell, S.C. and Lee, H.C. (1987b). Distributions of Genetic Markers in United States Populations: II. Isoenzyme Systems. *J.Forensic Sci.*, 321348-1381.

Gaensslen, R.E., Bell, S.C. and Lee, H.C. (1987c). Distributions of Genetic Markers in United States Populations: III. Serum Group Systems and Hemoglobin Variants. *J. Forensic Sci.*, 32 : 1754-1774.

Gamero J.J., Romero J.L., Gonzalez J.L., Arufe M.I. Cuesta M.I., Corte-Real F., Carvalho M., Anjos M.J., Vieira D.N. and Vide M.C. (2000). A study on ten short tandem repeat systems: African immigrant and Spanish population data. *Forensic Sci. Int.*, **110**: 167-177.

Gamero, J.J., Souto, L., Vieira, D.N., Vizcaya, M.A., Arufe, M.I., Feliu, M.M., Vide, M.C. and Romero, J.L (1998). Population study of the HUMVWA, HUMTHO1, HUMFES, HUMF13A1 STR polymorphisms in the south-west of Spain. *Progress in Forensic Genet.*, **7**: 366-368.

Garofano, L., Lago, G., Vecchio, C., Pizzamiglio, M., Zanon, C., Virgili, A., Albonici, L., Manzari, V. and Budowle, B. (1998). Italian population data on the polymarker system and on the five short tandem repeat loci CSF1PO, TPOX, THO1, F13B, and vWA. *J.Forensic Sci.*, **43**: 837-840.

Garofano, L., Pizzamiglio, M., Bizzaro, G.P., Donao, F., Rossetti, M. and Budowle, B. (1999). Italian population data on two new short tandem repeat loci: D6S477 and D19S433. *Forensic Sci. Int.*, **101**:203-208.

Garofano, L., Pizzamiglio, M., Vecchio, C., Lago, G., Floris, T., D'Errico, G., Brembilla, G., Romano, A. and Budowle, B. (1998). Italian population data on thirteen short tandem repeat loci: HUMTHO1, D21S11, D18S51, HUMVWFA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358. *Forensic Sci. Int.*, **97**: 53-60.

Gaviria, A., Ibarra, A.L., Jaramillo, N., Palacio, O.D., Acosta, M.A., Brion, M. and

Carracedo, A.(2004). Nineteen autosomal microsatellite data from Antioquia (Colombia). *Forensic Sci. Int.*, **143**, 69-71.

Ghosh, A., Ramesh, A., Raj Kumar, K.A. and Seshadri, M. (2004). STR polymorphism among two tribal populations of India. *J.Forensic Sci.*, **49**: 856-857.

Ghosh, A.K. (1977). Polymorphism of red cell glyoxalase 1. *Hum. Genet.*, **39**: 91.

GIBCOBRL/LIFE TECHNOLOGIES, Instruction Manual. (1998). *Model SA Adjustable Sequencing Gel Electrophoresis System.*GIBCOBRL/LIFE TECHNOLOGIES, USA.

GIBCOBRL/LIFE TECHNOLOGIES, *Product brochure, Gel Mix 6*, Doc.Rev. 09-11-00, Cat. No. 15543010. (2000). GIBCOBRL/LIFE TECHNOLOGIES, USA.

Giles, R.E., Blanc, H., Cann, H.M. and Wallace, D.C. (1980). Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci.*, USA, **77**: 6715-6719.

Gill, P. and Evett, I. (1995). Population genetics of short tandem repeat (STR) loci. *Genetica*, **96**: 69-87.

Gill, P. and Werret, D.J. (1987). Exclusion of a man charged with murder by DNA fingerprinting. *Forensic Sci. Int.*, **35**: 145-148.

Gill, P., Brenner, C., Brinkmann, B., Budowle, B., Carracedo, A., Jobling, M.A., De Knijff, P., Kayser, M., Krawczak, M., Mayr, W.R., Morling, N., Olaisen, B., Pascali, V., Prinz, M., Roewer, L., Schneider, P.M., Sajantila, A. and Tyler-Smith, C. (2001). DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Forensic Sci. Int.*, **124**: 5-10.

Gill, P., Ivanov, P.L., Kimpton, C., Piercy, R., Benson, N. and Tully, G. *et al.* (1994). Identification of the remains of the Romanov family by DNA analysis. *Nature Genet.*, **6**: 130-135.

Gill, P., Jeffreys, A.J., and Werret, D.J. (1985). Forensic Application of DNA Fingerprints. *Nature*, **318**: 577-579.

Gill, P., Kimpton, C. and Aliston-Greiner, R. *et al.* (1995). Establishing the identity of Anna Anderson Manahan. *Nature Genet.*, **9**: 9-10.

Gill, P., Kimpton, C.P., Urquhart, A., Oldroyd, N.J., Millican, E.S., Watson, S.K. and Downes, T.J. (1995). Automated short tandem repeat (STR) analysis in forensic casework – a strategy for the future. *Electrophoresis*, **16**: 1543-1552.

Gill, P., Lygo, J.E., Fowler, S.J., and Werret, D.J. (1987). An evaluation of DNA fingerprinting for forensic purposes. *Electrophoresis*, **8**: 38-44.

Gill, P., Sparkes, R. and Kimpton, C. (1997). Development of guidelines to designate alleles using an STR multiplex system. *Forensic Sci. Int.*, **89**: 185-197.

Ginsti, A., Baird, M., Pasquale, S., Balazs, I. and Glassbury, J. (1986). Application of Deoxyribonucleic Acid (DNA) polymorphisms to the analysis of DNA recovered from sperm. *J Forensic Sci.*, **31**: 409-417.

Ginther, C., Issel-Tarver, L. and King, M.C. (1992). Identifying individuals by sequencing mitochondrial DNA from teeth. *Nat Genet.*, **2**: 135-138.

Golden, T.R. and Melov, S. (2001). Mitochondrial DNA mutations, oxidative stress, and aging. *Mech. Ageing Dev.*, **122**: 1577-1589.

Goldman, M. A., Holmquist, G. P., Gray, M. C., Caston, L. A. and Nag, A. (1984). Replication timing of genes and middle repetitive sequences. *Science,* **224**: 686-692.

Goldstein, David, B. and Schlotterer, C. (1999). *Microsatellites: Evolution and Applications.* Oxford University Press, NY.

Goldstein, D.B., Linares, A.R., Cavalli-Sforza, L.L. and Feldman, M.W. (1995). Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci.*, USA, **92**: 6723-6727.

Gomez, J. and Carracedo, A. (2000). The 1998-1999 collaborative exercise and proficiency testing program on DNA typing of the Spanish and Portuguese Working Group of the International Society for Forensic Genetics (GEP-ISFG). *Forensic Sci. Int.,* **114**: 21-30.

Goodfellow, M., Manfio, G.P. and Chun, J. (1997). Towards a practical species concept for cultivable bacteria. In: *Species-The units of biodiversity.* Eds., Claridge, M.F., Dawah, H.A., Wilson, M.R. Chapman and Hall, pp. 25-50.

Goodman, M. *et al.*, (1998). Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol. Phylogenet., Evol.*, **9**: 585-598.

Gottelli, D., Sillero-Zubiri, C., Applebaum, G.D., Roy, M.S., Girman, D.J. Garcia-Moreno, J., Ostranders, E.A., and Wayne, R.K. (1994). Molecular genetics of the most endangered canid: the Ethiopian wolf Canis simensis. *Mol. Ecol.*, **3**:301-312.

Graw, M. and Seitz, T. (2000). Y chromosomal short tandem repeat (STR) loci in a representative group of males living in South Württemberg: a database for application in forensic medicine. *Forensic Sci. Int.*, **113**: 43-46.

Gray, I.C., Campbell, D.A. and Spurr, N.K. (2000). Single nucleotide polymorphisms as tools in human genetics. *Hum. Mol. Genet.*, **9**: 2403-2408.

Grech, J.L. (1979). Haptoglobins in the Maltese population. *Med. Sci. and the Law*, **19**: 159-164.

Greenberg, B.D., Newbold, J.E. and Sugino, A.(1983). Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene*, **21**: 33-49.

Gregory, T. R. and Hebert, P. D. (1999). The modulation of DNA content: proximate causes and ultimate consequences. *Genome Res.*, **9**: 317-324.

Guo, S.W. and Thompson, E.A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics,* **48** : 361-372.

Gupta, P.K. (1999). *Cell and Molecular Biology,* Rastogi Publications, Meerut , India.

Guttmacher, A.E. and Collins, F.S. (2002). Genomic medicine – A primer. *N. Engl. J. Med.*, 347: 1512-1520.

Gyllensten, U.B. and Erlich, H.A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQα locus. *Proc. Natl. Acad. Sci.*, USA, **85**: 7652-7656.

Haeseler, A.V., Sajantila, A., Paabo, S. (1996). The genetical archaeology of the human genome. *Nat.Genet.*, **14**: 135-140.

Hagelberg, E. and Sykes, B. (1989). Ancient bone DNA amplified. *Nature*, **342**: 485-488.

Haldane, J.B.S. (1954). An exact test for randomness of mating. *Journal of Genetics,* **52**: 631-635.

Hallenberg, C and Morling, N. (2001). A report of the 1997, 1998 and 1999 paternity testing workshops of the English speaking working group of the International Society for Forensic Genetics. *Forensic Sci. Int.*, **116**: 23-33.

Halos, S.C., Chu, J.Y., Ferreon, A.C.M. and Magno, M.M.F.(1999). Philippine population database at nine microsatellite loci for forensic and paternity applications. *Forensic Sci. Int.*, **101**: 27-32.

Halushka, M.K., Fan, J-B., Bentley, K., Hsie, L., Shen, N., Weder, A., Cooper, R., Lipshutz, R. and Chakravarti, A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.*, **22**: 239-247.

Hammer, M.F. (1995). A recent common ancestry for human Y chromosomes. *Nature*, **378**: 376-378.

Handt, O., Hōss, M., Krings, M., and Pāābo, S. (1994a). Ancient DNA: methodological challenges. *Experientia*, **50**: 524-529.

Handt, O., Krings, M., Ward, R.H. and Pāābo, S. (1996). The retrieval of ancient human DNA sequences. *Am. J. Hum. Genet.*, **59**: 368-376.

Handt, O., Richards, M., Trommsdorff, M., Kilger, C., Simanainen, J., Georgiev, O., Bauer, K., Stone, A., Hedges, R., Schaffner, W., Utermann, G., Sykes, B. and Pääbo, S. (1994). Molecular genetic analyses of the Tyrolean Ice Man. *Science*, **264**: 1775-1778.

Hardman, N. (1986). Structure and function of Repetitive DNA in Eukaryotes. *Biochem. Journal*, **234**: 1-11.

Haribson S.A, Stanfield A.M, Buckleton J.S. and Walsh, S.J. (2002). Allele frequencies for the four major sub-populations of New Zealand at three STR loci-HUMTHO1, HUMTPOX and CSF1PO. *Forensic Sci. Int.*, **126**: 258-260.

Haring, E., Hagemann, S. and Pinsker, W. (2000). Ancient and recent horizontal invasions of Drosophilids by P elements. *J. Mol. Evol.*, **51**: 577-586.

Hartl, D. L. (2000). Molecular melodies in high and low C. *Nature Rev. Genet.*, **1**: 145-149.

Hauswirth, W.W. and Laipis, P.J. (1986). In: Achievements and Perspectives in Mitochondrial Research, Vol. 2 *Biogenesis* (eds. E. Quagliariello, E., Slater, E.C., F. Palmieri, F., Saccone, C. and Kroon, A.M.), Elsevier, New York, pp.49-60.

Hedges, S.B., Kumar, S., Tamura, K. and Stoneking, M.(1992). Human origins and analysis of mitochondrial DNA sequences. *Science*, **255**: 737-739.

Hering, S., Kuhlisch, E. and Szibor, R. (2001). Development of the X-linked tetrameric microsatellite marker DXS6789 for forensic purposes. *Forensic Sci. Int.*, **119**: 42-46.

Herkenham, D. (2002). DNA Database legislation and legal issues. *Profiles in DNA*, **5** : 6-7.

Hernándes, J.L. and Weir, B.S. (1989). A disequilibrium coefficient approach to Hardy-Weinberg testing. *Biometrics* **45**: 53-70.

HGP, Human Genome Project Information. (2001). Announcements on the First Analysis of Genome Sequence, February 12 – SNP count up to 1.4 million: Map accelerates discovery of disease and human population history. Base URL: *http://www.ornl.gov/hgmis.*

HGP, Human Genome Project Information. (2003). Base URL: *http://www.ornl.gov/hgmis*.

HGP, Human Genome Project Information .(2003). – *DNA Forensics,* Base URL *http://www.ornl.gov/hgmis.*

HGP, Human Genome Project Information. (2003). – *Gene Therapy.* Base URL *http://www.ornl.gov/hgmis.*

HGP, Human Genome Project Information. (2003).– *Goals*. Base URL: *http://www.ornl.gov/hgmis*.

HGP, Human Genome Project Information. (2003) – HGPcompletionApr2003.htm. Base URL: *http://www.ornl.gov/hgmis*.

HGP, Human Genome Project Sequence Analysis. (2003) – *Insights Learned from the Sequence.* Base URL: *http://www.ornl.gov/hgmis*.

HGP, International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature,* **409**: 860-921.

HGP, Special issue for Human Genome Project. (2001). *Science,* **291**: 5507.

HGP, Special issue for Human Genome Project.(2001). *Nature,* **409**.

Hidding, M. and Schmitt, C. (2000). Haplotype frequencies and population data of nine Y-chromosomal STR polymorphism in a German and a Chinese population. *Forensic Sci. Int.,* **113**: 47-53.

Higgs, D.R., Goodbourn, S.E.Y., Wainscoat, J.S., Clegg, J.B. and Weatherall, D.J. (1981). *Nucleic Acids Res.*, **9**: 4213-4224.

Higuchi, R., von Beroldingen, C.H., Sensabaugh, G.F. and Erlich, H.A. (1988). DNA typing from single hairs. *Nature,* **332**: 543-546.

Hilderson, Y. and Henry, M.R. (1985). *Scientific Testing for Paternity Establishment,* 2nd ed. U.S. Dept. of Health and Human Services, Washington, DC., 15.

Hochmeister, M.N., Budowle, B., Borer, U.V. and Dirnhofer, R. (1995). A method for the purification and recovery of genomic DNA from and HLA DQA1 amplification product and its subsequent amplification and typing with the APPLIED BIOSYSTEMS, AmpliTypePM PCR amplification and typing kit. *J.Forensic Sci.*, **40**: 649-653.

Holland, M.M., Fisher, D.L., Mitchell, L.G., Rodriquez, W.C., Canik, J.J. and Merril, C.R. *et al.*(1993). Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam war. *J. Forensic Sci.*, **38**: 542-553.

Holland, M.M., Fisher, D.L., Roby, R.K., Ruderman, J., Bryson, C. and Weedn, V.W. (1995). Mitochondrial DNA sequence analysis of human remains. *Crime Lab. Digest* **22**: 109-115.

Holt, I.J., Harding, A.E. and Morgan-Hughes, J.A. (1988). Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature*, **331**:717-719.

Honeycutt, R.L. and Adkins, R.M. (1993). Higher level systematics of eutherian mammals: an assessment of molecular characters and phylogenetic hypotheses. *Annu. Rev. Ecol. Syst.*, **24**: 279-305.

Honeycutt, R.L., Nedbal, M.A., Adkins, R.M. and Janecek, L.L. (1994). Mammalian mitochondrial DNA evolution: A comparison of the cytochrome b and cytochrome c oxidase II genes. *J. Mol. Evol.*, **40**: 260-272.

Hopgood, R., Sullivan, K.M. and Gill, P. (1992). Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. *BioTechniques*, **13**: 60-68.

Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. and Takahata, N. (1995). Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci.*, USA, **92**: 532-536.

Horn, G.T., Richards, B., Merrill, J.J. and Klinger, K.W. (1980). Characterization and rapid diagnostic analysis of DNA polymorphisms closely linked to the cystic fibrosis locus. *Clin. Chem.*, **36**: 1614-1619.

Hōss, M., Dilling, A., Currant, A and Pāābo, S. (1996b). Molecular phylogeny of the extinct ground sloth Mylodon darwinii .*Proc. Natl. Acad. Sci.*,USA, **93**:181-185.

Hōss, M., Jaruga, P., Zastawny, T.H., Dizdaroglu, M. and Pāābo, S. (1996a). DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res.*, **24**: 1304-1307.

Höss, M. (2000). Neandertal population genetics. *Nature*, **404**: 453-454.

Howell, N. (1994). Primary LHON mutations: trying to separate "fruyt" from "chaf." *Clin. Neurosci.*, **2**: 130-137.

Hublin, J.J., Spoor, F., Braun, M., Zonneveld, F. and Condemi, S. (1996). A late Neandertal associated with Upper Paleolithic artifacts. *Nature*, **381**: 224-226.

Hughes, A. L. (1999). Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. *J. Mol. Evol.*, **48**: 565-576.

Hummel, K. and Gerchow, J. (1981). *Biomathematical Evidence of Paternity. Springer Verlag,* Berlin.

Hwang, U.W., Friedrich, M., Tautz, D., Park, C.J. and Kim, W.(2001). Mitochondrial protein phylogeny joins myriapods with chelicerates. *Nature*, **413**: 154-157.

Ingman, M., Kaessmann, H., Paabo, S. and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**: 708-713.

Ingman, M. (2001). Mitochondrial DNA Clarifies Human Evolution. *www.actionbioscience.org/evolution/ingman.html.*

Ingvarsson, S., Finnsdottir, V., Sigurdsson, A. and Geirsson, G. (2000). Population studies and validation of paternity determinations by six microsatellite loci. *J.Forensic Sci.*, **45**: 692-695.

Inman, K. and Rudin, N. (1997). *An introduction to Forensic DNA Analysis.* CRC Press, New York.

Innis, M.A., Gelfand, D.H. and Sninsky, J.J., eds. (1992). *PCR protocols: a guide to methods and applications.* Academic Press, San Diego, CA.

Innis, M.A., Myambo,K.B., Gelfand, D.H. and Brow, M.A.D. (1988). DNA sequencing with Thermus aquaticus DNA polymerase and direct sequencing of polymerase chain reaction – amplified DNA. *Proc. Natl. Acad. Sci.,* USA, **85**: 9436 – 9440.

Irvin, D.M. and Arnason, U. (1994). Cytochrome b gene of marine mammals: phylogeny and evolution. *J. Mamm. Evol.*, **2**: 37-55.

Irvin, D.M., Kocher, T.D. and Wilson, A.C .(1991). Evolution of the cytochrome b gene of mammals. *J. Mol. Evol.*, **32**: 128-144.

Ivanov, P.L., Wadhams, M.J., Roby, R.K., Holland, M.M., Weedn, V.W. and Parsons, T.J. (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nature Genet.*, **12**: 417-420.

Jackson, D.A., Symons, R.H. and Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli. Proc. Natl Acad. Sci.*, USA, **69**: 2904-2909.

Jankowski, L.B., Budowle, B., Swec, N.T., Pino, J.A., Freck-Tootell, S., Corey, H.W., Schwartz, R., LaRue, E.J., Rochin, W.L., Kearney, C.J. and Tarver, M. (1998). New Jersey Caucasians, African American, and Hispanic population data on the PCR-based loci HLA-DQA1, LDLR, GYPA, HBGG, D7S8, and Gc. *J.Forensic Sci.*, **43**: 1037-1040.

Jarman, A.P., Nicholls, R.D., Weatherall, D.J., Clegg, J.B., and Higgs, D.R. (1986). Molecular characterization of a Hypervariable region downstream of the Human Alpha-Globin gene complex. *The EMBO Journal* **5**: 1857-1863.

Jarreta, B.M., Roche, P.D. and Abecia, E. (1999). Genetic variation at six STR loci (HUMTHO1, HUMTPOX, HUMCSF1PO, HUMF13A01, HUMFES/FPS, HUMVWFA31) in Aragon (North Spain). *Forensic Sci. Int.*, **100**: 87-92.

Jasni B.R. and Roberts L. (2003). Building on the DNA Revolution. *Science*, **300**:277.

Jeffreys, A.J., Brookfield, J.F.Y. and Semeonoff, R. (1985c). Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, **317**: 818-819.

Jeffreys, A.J. (1993). DNA Symposium, DNA typing: approaches and applications. *J.Forensic Sci., Soc.*, **33**: 204-211.

Jeffreys, A.J. (1987). Highly Variable Minisatellites and DNA Fingerprints. *Biochem. Soc. Trans.*, **15**: 309-317.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985a). Hypervariable 'minisatellite' regions in human DNA. *Nature*, **314**:67-72.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1985b). Individual-specific 'fingerprints' of human DNA. *Nature*, **316**: 76-79.

Jeffreys, A.J., Wilson, V., Thein, S.L., Weatherall, D.J. and Ponder, B.A.J. (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet.*, **39**: 11.

Jeffreys, A.J., Wong, Z., Wilson, V., Patel, I., Neumann, R. Royle, N., and Armour, J.A.L. (1989). Applications of Multilocus and Single-locus Minisatellite DNA Probes in Forensic Medicine. In: *DNA Technology and Forensic Science*, Edited by Ballantyne, J., Sensabaugh, G. and Witkowski, J. Cold Spring Harbor Laboratory Press, New York.

Jelinek, W.R. and Schmid, C.W. (1982). Repetitive sequences in Eukaryotic DNA and their expression. *Ann. Rev.of Biochem.*, **51**: 813-844.

Jiménez, G., Picornell, A., Tomàs, C., Castro, J.A. and Ramon, M.M. (2001). Y-chromosome polymorphism data in Majorcan, Minorcan and Valencian populations (eastern Spain). *Forensic Sci. Int.*, **124**: 231-234.

Jobling, M.A., Pandya, A. and Tyler-Smith, C. (1997). The Y chromosome in forensic analysis and paternity testing. *Int. J. Leg. Med.*, **110**: 118-124.

Johns, D.R. (1995). Mitochondrial DNA and Disease. In: Seminars in Medicine of the Beth Israel Hospital, Boston. *The New Eng. J. Med.*, **333**: 638-644.

Jones, M .(2001). *The molecule hunt: Archaeology and the search for ancient DNA.* Allen Lane / Penguin Press, London.

Jung, J.M., Comey, C.T., Baer, D.B. and Budowle, B. (1991). Extraction strategy for obtaining DNA from bloodstains for PCR amplification and typing of the HLA-DQα gene. *Int. J Leg. Med.*, **104**: 145-148.

Junge, A. and Madea, B. (1999). Population studies of the Y-chromosome specific polymorphisms DYS389 I+II, DYS390 and DYS393 in a Western German population (Bonn area). *Forensic Sci. Int.*, **101**: 195-201.

Jurka, J. and Kapitonov, V. V. (1999). Sectorial mutagenesis by transposable elements. *Genetica*, **107**: 239-248.

Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.*, **16**: 418-420.

Kanter, E., Baird, M., Shaler, R. and Balazs, I. (1986). Analysis of Restriction Fragment Length Polymorphisms in Deoxyribonucleic Acid (DNA) recovered from bloodstains. *J. Forensic Sci.*, **31**: 403-408.

Kasai, K., Nakamura, Y. and White, R. (1990). Amplification of a variable number of tandem repeat (VNTR) locus (pMCT 118) by the polymerase chain reaction (PCR) and its application to forensic science. *J.Forensic Sci.*, **35**: 1196-1200.

Kashi, Y. and Soller, M. (1999). "Functional roles of Microsatellites and Minisatellites." *In: Microsatellites: Evolution and Applications.* Edited by Goldstein and Schlotterer. *Oxford University Press.*

Kayser, M., Brauer, S., Schadlich, H., Prinz, M., Batzer, M.A., Zimmerman, P.A., Boatin, B.A. and Stoneking. M. (2003). Y chromosome STR haplotypes and the genetic structure of U.S populations of African, European, and Hispanic ancestry. *Genome Res.*, (*http://www.genome.org/cgi/doi/10.1101/gr.463003*): 624-634.

Kayser, M., Caglia, A., Corach, D., Fretwell, N. and Gehrig, C. *et al.* (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Leg. Med.*, **110**:125-133.

Kazazian, H. H. Jr. and Moran, J. V. (1998). The impact of L1 retrotransposons on the human genome. *Nature Genet.*, **19**:19-24.

Kimpton, C., Gill, P., D'Aloja, E., Andersen, J.F., Bar, W., Holgersson, S., Jacobsen, S., Johnsson, V., Kloosterman, A.D., Lareu, M.V., Nellemann, L., Pfitzinger, H., Phillips, C.P., Rand, S., Schmitter, H., Schneider, P.M., Sternersen, M. and Vide, M.C. (1995). Report on the second EDNAP collaborative STR exercise. *Forensic Sci. Int.*, **71**: 137-152.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge Univ. Press, New York.

Kirby, L.T. (1990). *DNA Fingerprinting An Introduction.* Stockton Press.

Kishida, T., Tamaki, Y. and Kuroki, K. (1996). Long-distance PCR of VNTR at the D17S74 (CMM86) locus. *Jpn. J. Leg. Med.*, **50**: 174-177.

Kishida, T., Tamaki, Y., Kuroki, K., Fukuda, M. and Wang, W. (1995). PCR amplification of D2S44 (YNH24) alleles. *Jpn. J. Leg. Med.*, **49**: 299-303.

Kivisild, T., Bamshad, M.J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W.S., Dixon, M.E., Papiha, S.S., Mastana, S.S., Mir, M.R., Ferak, V. and Villems, R. (1999). Deep common ancestry of Indian and Western-Eurasian mitochondrial DNA lineages. *Curr. Bio.*, **9**: 1331-1334.

Kivisild, T., Rootsi, S., Metspalu, M., Mastana, S., Kaldma, K., Parik, J., Metspalu, E., Adojaan, M., Tolk, H.V., Stepanov, V., Golge, M., Usanga, e., Papiha,S.S.,

Cinnioglu, C., King, R., Cavalli-Sforza, L., Underhill, P.A. and Villems, R.(2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.*, **72**: 313-332.

Klintschar, M., Al-Hammadi, N.and Reichenpfader, B. (1999). Population genetic studies on the tetrameric short tandem repeat loci D3S1358, VWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317 and D7S820 in Egypt. *Forensic Sci. Int.*, **104**: 23-31.

Klintschar, M., Ricci, U., Hammadi, N.A., Reichenpfader, B., Ebner, A. and Uzielli, M.L.G. (1998). Genetic variation at the STR loci D12S391 and CSF1PO in four populations from Austria, Italy, Egypt and Yemen. *Forensic Sci. Int.*, **97**: 37-45.

Knight, A. (2003). The phylogenetic relationship of Neandertal and modern human mitochondrial DNAs based on informative nucleotide sites. *J. Hum. Evol.*, **44**:627-632.

Kobus, H.J. and Raymond, M.A. (1982). The PGM isoenzyme distribution among negroid people of Zimbabwe. *Forensic Sci. Int.*, **19**: 67-74.

Kocher, T.D. and Wilson, A.C. (1991). In: Evolution of Life: Fossils, Molecules and Culture. Osawa, S and Honjo, T., Eds., Springer-Verlag, Tokyo, pp. 391-413.

Kompf, J., Bissbort, S., Gussman, S., and Ritter, H. (1975). Polymorphism of red cell glyoxalase 1 (EC: 4.4.1.5). A new genetic marker in man. *Humangenetik*, **23**: 139.

Kordis, D. and Gubensek, F. (1997). Bov-B long interspersed repeated DNA (LINE) sequences are present in Vipera ammodytes phospholipase A2 genes and in genomes of Viperidae snakes. *Eur. J. Biochem.*, 246: 772-779.

Krings, M., Capelli, C., Tschentscher, F., Geisert, H., Meyer, S. and Haeseler, A.V. *et al.* (2000). A view of Neandertal genetic diversity. *Nat. Genet.*, **26**: 144-146.

Krings, M., Geisert, H., Schmitz, R.W., Krainitzki, H., and Pääbo, S. (1999). DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc. Natl. Acad. Sci.*, USA, **96**: 5581-5585.

Krings, M., Stone, A., Schmitz, R.W., Krainitzki, H., Stoneking, M., and Pääbo, S. (1997). Neandertal DNA Sequences and the Origin of Modern Humans. *Cell* **90**: 19-30.

Kruglyak, S., Durrett, R. T., Schug, M. D. and Aquadro, C. F.(1998). Equilibrium distribution of microsatellite repeat length resulting from a balance between

slippage events and point mutations. *Proc. Natl Acad. Sci.*, USA **95**: 10774-10778.

Kudo, S. and Fukuda, M. (1989). Structural organization of glycophorin A and B genes. Glycophorin B gene evolved by homologous recombination at *Alu* repeat sequences. *Proc. Natl. Acad. Sci.*, USA, **86**: 4619-4623.

Laber, T.L., O'Connor, J.M., Iverson, J.T., Liberty, J.A and Bergman, D.L. (1992). Evaluation of four deoxyribonucleic acid (DNA) extraction protocols for DNA yield and variation in restriction fragment length polymorphism (RFLP) sizes under varying gel conditions. *J. Forensic Sci.*, **37**: 404-424.

Lai, L and Goetz, R.J (1988). Phosphoglucomutase and haptoglobin phenotypes among Australian Caucasians. *J.Forensic Sci., Soc.*, **28**: 191-197.

Lander, E. S. and Schork, N. J. (1994). Genetic dissection of complex traits. *Science,* **265** :2037-2048.

Lander, E.S. (1989). Population genetic considerations in the forensic use of DNA typing. *Banbury Report 32: DNA technology and forensic science*, Cold Spring Harbor Laboratory Press, 143-155.

Landsteiner, K and Richter, M. (1902). Über die verwertbarkeit individueller blutdifferenzen für die forensische praxis. *Medizinisch Beamte* **16**: 85-89.

Lareu M.V., Phillips, C.P., Carracedo, A., Lincoln, P.J., Syndercombe, D. and Thomson, C.J.A. (1994). Investigation of the STR locus HUMTHO1 using PCR and two electrophoresis formats: UK and Galician Caucasian population surveys and usefulness in paternity investigations. *Forensic Sci. Int.*, **66**: 41-52.

Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R., and Gelfand, D.H. (1989). Isolation, characterization, and expression in Escherichia coli of the DNA polymerase gene from Thermus aquaticus. *J. Biol. Chem.*, **264**:6427-6437.

Lee, H.C. (1982). Identification and Grouping of Bloodstains. In: *Forensic Science Handbook,* R. Saferstein, Ed., Prentice Hall, Englewood Cliffs, NJ, pp.267-337.

Lee, J.C., Chia-Hwei, C., Li-Chin, T., Adrian, L. and Chang, J.G. (1997). The screening of 13 short tandem repeat loci in the Chinese population. *Forensic Sci. Int.*, **87**: 137-144.

Levene, H. (1949). On a matching problem arising in genetics. *Annals of Mathematical Statistics,* **20**: 91-94.

Levinson, G. and Gutman, G.A. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.*, **4**, 203 - 221.

Li, C.C. (1955). *Population Genetics*. University of Chicago Press, Chicago.

Li, W.H., Ellsworth, D.L., Krushkal,J., Chang,B.H. and Hewett-Emmett,D.(1996). Rates of nucleotide substitution in primates and rodents and the generation-timeeffect hypothesis. *Mol. PhyloGenet., Evol.*, **5**:182-187.

Li, W. H. (1997). *Molecular Evolution* . Sinauer, Sunderland, Massachusetts.

Litt, M. and Luty, J.A. (1989). A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. of Hum. Genet.*, **44**: 397-401.

Lorente, J.A., Lorente, M., Budowle, B., Wilson, M.R., and Villanueva, E. (1994). Analysis of the HUMTHO1 allele frequencies in the Spanish Population. *J.Forensic Sci.*, **39**: 1270-1274.

Loreille,O., Orlando, L., Patou-Mathis, M., Philippe, M., Taberlet, P. and Hänni, C. (2001). Ancient DNA analysis reveals divergence of the cave bear, *Ursus spelaeus*, and brown bear, *Ursus arctos*, lineages. *Curr. Biol.*, **11**: 200-203.

Lutz, S., Weisser, H.J., Heizmann, J. and Pollak, S. (1996). mtDNA as a tool for identification of human remains identification using mtDNA. *Int. J. Leg. Med.*, **109**: 205-209.

Maddison, D.R.(1991). African origin of human mitochondrial DNA reexamined. *Syst. Zool.*, **40**: 355-363.

Maddison, D.R., Ruvolo, M. and Swofford, D.L.(1992). Geographic origins of human mitochondrial DNA phylogenetic inference from control region sequences. *Syst. Biol.*, **41**: 111-124.

Ma- D-P., Zharkikh A., Graur, D., VandeBerg, J.L. and Li, W-H. (1993). Structure and evolution of opossum, guinea pig, and porcupine cytochrome b genes. *J. Mol. Evol.*, **36**: 327-334.

Mahajan, B.K. (1999). Sampling. *In: Methods in Biostatistics*, Jaypee Brothers Medical Publishers (P) Ltd., New Delhi.

Majumder, P.P., Roy, B., Banerjee, S., Chakraborty, M., Dey, B., Mukherjee, N., Roy, M., Thakurta, P.G. and Sil, S.K. (1999). Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. *Eur. J. Human Genet.*, 7: 435-446.

Malik, H. S. and Eickbush, T. H.(2000). NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from Caenorhabditis elegans.*Genetics,* **154**: 193-203.

Malik, H. S., Burke, W. D. and Eickbush, T. H. (1999). The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.,* **16**: 793-805.

Malik, H. S., Henikoff, S. and Eickbush, T. H. (2000). Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.*,**10**:1307-1318.

Manly, B.F.J. (1991). *Randomization and Monte Carlo methods in biology*. Chapman and Hall, NY.

Manuelidis, L. and Ward, D. C.(1984). Chromosomal and nuclear distribution of the Hind III 1.9-kb human DNA repeat segment. *Chromosoma,* **91**: 28-38.

Markowicz, K.R., Tonelli, L.A., Anderson, M.B., Green, D.J., Herrin, G.L., Cotton, R.W., Gottschall, J.L. and Garner, D.D. (1990). Use of Deoxyribonucleic Acid (DNA) fingerprints for identity determination: Comparison with traditional paternity testing methods-part II. *J.Forensic Sci.,* **35**: 1270-1276.

Mauricio-da-Silva, L., Silva, R.S., Dellalibera, E. and Donadi, E.A. (2000). Population genetics of HPRTB, F13B, and LPL in Pernambuco, Northeast Brazil. *J.Forensic Sci.,* **45**: 684-686.

Maxam, A.M. and Gilbert, W. (1977). A new method for sequencing DNA. *Proc. Natl Acad. Sci.*, USA, **74**:560-564.

Melton, T., Wilson, M., Batzer, M. and Stoneking, M. (1997). Extent of heterogeneity in mitochondrial DNA of European populations. *J. Forensic Sci.,* **42**: 437-446.

Mendel, G. (1866). Versuche über Pflanzen-Hybriden. *Verhandlungen des naturforschenden Vereincs, Abhandlungen, Brünn* **4**: 3-47.

Meng, H.Y., Hou, Y.P., Chen, G.D., Li, Y.B., Wu, J., Walter, H. and Prinz, M. (1999). Frequencies of D8S384 alleles and genotypes in European, African-American, Chinese and Japanese populations. *J.Forensic Sci.,* **44**: 1273-1276.

Meunier-Rotival, M., Soriano, P., Cuny, G., Strauss, F. and Bernardi, G. (1982). Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA. *Proc. Natl Acad. Sci.*, USA, **79**, 355-359.

Miller, M.P. (1997). *Tools for Population Genetic Analyses (TFPGA):* A Windows program for the analysis of allozyme and molecular population genetic data.

Dept. of Biological Sciences- Box 5640, Northern Arizona University, Flagstaff, AZ 86011-5640, U.S.A.

MILLIPORE - *Centricon, Centrifugal Filter Devices, User's Guide.*(2000). *Millipore Corporation*, USA.

Miscicka-Sliwka, D., Czarny, J., Grzybowski, T. and Wozniak, M. (1998). Population genetics of 14 STRs: vWA, THO1, TPOX, CSF1PO, D5S818, D13S317, D7S820, D16S539, F13A1, FES/FPS, F13B, LPL, D3S1358 and FGA in the Pomerania-Kujawy region of Poland. *Adv. in Forensic Genet.*, **7**: 261-263.

Miyamoto, M.M., Allard, M.W., Adkins, R.M., Janecek, L.L. and Honeycutt, R.L. (1994). A congruence test of reliability using linked mitochondrial DNA sequences. *Syst. Biol.*, **43**: 236-249.

Momhinweg, E., Luckenbach, C., Fimmers, R. and Ritter, H. (1998). D3S1358 : Sequence analysis and gene frequency in a German population. *Forensic Sci. Int.*,**95**: 173-178.

Monson, K.L. and Budowle, B. (1993). A comparison of the fixed bin method with the floating bin and direct count methods: Effect of VNTR profile frequency estimation and reference population. *J.Forensic Sci.*, **38**: 1037-1050.

Moran, J. V. (1999). Human L1 retrotransposition: insights and peculiarities learned from cultured cell retrotransposition assay. *Genetica,* **107**: 39-51.

Mountain, J.L., Hebert, J.M., Battacharyya, S., Underhill, P.A., Ottolenghi, C., Gadgil, M. and Cavalli-Sforza, L.L. (1995). Demographic history of India and mtDNA-sequence diversity. *Am. J. Hum. Genet.*, **56**: 979-992.

Mount, D.W. (2001). Phylogenetic Prediction. *In: Bioinformatics – Sequence and Genome Analysis.*Cold Spring Harbor Laboratory Press, New York.

Mourant, A.E., Kopec, A.C. and Domaniewska-Sobczak, K.(1976). *The Distribution of the Human Blood Groups and Other Polymorphisms.* Oxford University Press, London.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symp. Quant. Biol.*, **51**: 263-273.

Mullis, K.B. and Faloona, F.A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.*, **155**: 335-350.

Murch, R.S. and Budowle, B. (1986). Applications of Isoelectric focusing in Forensic Serology. *J.Forensic Sci.*, **31**: 869-880.

Nakahori, Y., Takenaka, O. and Nakagome, Y. (1991). A human X-Y homologous region encodes "amelogenin." *Genomics,* **9**, 264-269.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R. and Holm, T. *et al.*(1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science,* **235**: 1616-1622.

NCBI Map Viewer (2004). *http://www.ncbi.nih.gov/mapview/maps.cgi?*

Nei, M. and Livshits, G. (1989). Classical analyses of blood group and protein polymorphisms also support the out of Africa hypothesis. *Hum Hered.*, **39**: 276.

Nei, M. and Roychoudhury, A.K. (1982). Genetic relationship and evolution of human races. *Evol. Biol.*, **14**: 1-59.

Nei, M. and Roychoudhury, A.K. (1974). Sampling variances of heterozygosity and genetic distance. *Genetics* ,**76**: 379-390.

Nei, M and Takezaki, N. (1996). The root of the phylogenetic tree of human populations. *Mol. Biol. Evol.*, **13**: 170-177.

Nei, M.(1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics,* **89**: 583-590.

Nei, M. (1985). In: *Population Genetics and Molecular Evolution,* eds., Ohta, T. and Aoki, K., Japan Sci. Soc. Press, Tokyo, pp.41-64.

Nirenberg, M.W .(1963). The genetic code:I I. *Sci. Am.*, **208**:80-94.

Odelberg, S.J and White, R. (1989). Tandemly Repeated DNA and Its Application in Forensic Biology. *In: DNA Technology and Forensic Science,* Edited by: Ballantyne, J., Sensabaugh, G. and Witkowsi, J. – Cold Spring Harbor Laboratory Press, New York.

Okada, N., Hamada, M., Ogiwara, I. and Ohshima, K. (1997). SINEs and LINEs share common 3' sequences: a review. *Gene,* **205**: 229-243.

Okumura, K., Kiyama, R. and Oishi, M. (1987). Sequence analyses of extrachromosomal *sau3a* and related family DNA: analysis of recombination in the excision event. *Nuc. Acids Res.*, **15**: 7477-7489.

Olivo, P.D., Van de Walle, M.J., Laipis, P.J. and Hauswirth, W.W. (1983). *Nature*, **306**: 400-402.

Orrego, C. and King, M.C. (1990). Determination of familial relationships. In: Innis, M.A., Gelfand, D.H., Sninsky , J.J., editors, *PCR protocols: a guide to methods and applications.* Academic Press Inc., San Diego, CA., pp. 416-426.

Ovchinnikov, I.V., Götherström, A., Romanova, G.P., Kharitonov, V.M., Lidén, K. and Goodwin, W. (2000). Molecular analysis of Neandertal DNA from the northern Caucasus. *Nature* **404**: 490-493.

Ovchinnikov, I.V., Götherström, A., Romanova, G.P., Kharitonov, V.M., Lidén, K., Goodwin, W.(2001). Reply to Smith, *et al.*, *Nature* **410**: 772.

Pääbo, S. (1989). Ancient DNA: extraction, characterization, molecular cloning and enzymatic amplification. *Proc. Natl. Acad. Sci.*, USA, **86**: 1939-1943.

Pääbo, S. and Wilson, A.C. (1991). Miocene DNA sequences – a dream come true? *Current Biol.*, **1**: 45-46.

Pääbo, S., Higuchi, R.G., and Wilson, A.C. (1989). Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *J. Biol. Chem.*, **264**: 9709-9712.

Paetkau, D. and Strobeck, C. (1994). The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol. Ecol.*, **4**: 519-520.

Pai, C-Y., Hsieh, L-L., Tsai, C-W., Chiou, F-S., Yang, C-H. and Hsu, B-D. (2002). Allelic alterations at the STR markers in the buccal tissue cells of oral cancer patients and the oral epithelial cells of healthy betel quid-chewers: an evaluation of forensic applicability. *Forensic Sci. Int.*, **129**: 158-167.

Panneerchelvam, S., Thevan, K.K., KokFai, L., Saravanakumar, M., Sumathy, V., Yuvaneswari, K.C. and Norazmi, M.N. (2004). Allele frequency distribution for 9 STR loci in the Tamil population of Malaysia. *J.Forensic Sci.*, **49**: 863-864.

Parson, W., Parsons, T.J., Scheithauer, R. and Holland, M.M. (1998). Population data fro 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int. J. Leg. Med.*, **111**: 124-132.

Passarino, G., Semino, O., Bernini, L.F. and Santachiara-Benerecetti, A.S. (1996). Pre-Caucasoid and Caucasoid genetic features of the Indian population revealed by mtDNA polymorphisms. *Am. J. Hum. Genet.*, **59**: 927-934.

Paterson, B.L., Su, B., Chakraborty, R., Budowle, B. and Gaensslen, R.E. (2000). World population data for the HLA-DQA1, PM and D1S80 loci with least and

most common profile frequencies for combinations of loci estimated following NRC II guidelines. *J Forensic Sci.*, **45**: 118-146.

PCR Systems, Reagents & Consumables, *1996-1997 Perkin Elmer Catalogue.* – The Perkin Elmer Corporation - currently Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA.

Penny, D., Steel, M., Waddell, P.J. and Hendy, M.D. (1995). Improved analyses of human mtDNA sequences support a recent African origin of Homo sapiens. *Mol. Biol. Evol.*, **12**: 863-882.

Pesole, G., Sbisa, E., Preparata, G. and Saccone, C. (1992). The evolution of the mitochondrial D-loop region and the origin of modern man. *Mol. Biol. Evol.*, **9**: 587-598.

Pestoni, C., Lareu, M.V., Rodriguez, M.S., Munoz, I., Barros, F. and Carracedo, A. (1995). The use of the STRs HUMTHO1, HUMVWA31/A, HUMF13A1, HUMFES/FPS, HUMLPL in forensic application: validation studies and population data for Galicia (NW Spain). *Int. J.Leg.Med.*, **107**: 283-290.

Peterson, B.L., Su, B., Chakraborty, R., Budowle, B. and Gaensslen, R.E. (2000). World population data for the HLA-DQA1, PM and D1S80 loci with least and most common profile frequencies for combinations of loci estimated following NRC II guidelines. *J.Forensic Sci.*, **45**: 118-146.

Petrov, D. A., Lozovskaya, E. R. and Hartl, D. L. (1996). High intrinsic rate of DNA loss in Drosophila. *Nature*, **384**: 346-349.

Pfitzinger, H., Ludes, B., Kintz, P., Tracqui, A. and Mangin, P. (1995). French Caucasian population data for HUMTHO1 and HUMFES/FPS short tandem repeat (STR) systems. *J.Forensic Sci.*, **40**: 270-274.

Piercy, R., Sullivan, K.M., Benson, N. and Gill, P. (1993). The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int. J. Leg. Med.*, **106**: 85-90.

Polymeropoulos, M.H., Xiao, H., Rath, D.S. and Merril, C.R. (1991). Tetranucleotide repeat polymorphism at the human tyrosine hydroxylase gene (TH). *Nucleic Acids Res.*, **19**, 3753.

Prak, E. L. and Haig, H. K. Jr. (2000). Mobile elements and the human genome. *Nature Rev. Genet.*, **1**: 134-144.

Prinz, M., Boll, K., Howard, B. and Shaler, B. (1997). Multiplexing of Y chromosome specific STRs and performance for mixed samples. *Forensic Sci. Int.*, **85**: 209-218.

PROMEGA *GenePrint STR Systems (Silver Stain Detection) – Technical Manual*, Part # TMD004, revised 4/01 (2001). Promega Corporation, 2800 Woods Hollow Road, Madison, W1 53711-5399, USA.

Pu, C-E., Hsieh, C-M., Chen, M-Y., Wu, F-C. and Sun, C-F. (1999). Genetic variation at nine STR loci in populations from the Philippines and Thailand living in Taiwan. *Forensic Sci. Int.*, **106**: 1-6.

Pu, C-E., Wu, F-C., Cheng, C-L., Wu, K-C., Chao, C-H. and Li, J-M. (1998). DNA short tandem repeat profiling of Chinese population in Taiwan determined by using an automated sequencer. *Forensic Sci. Int.*, **97**: 47-51.

Queller, D.C., Strassmann, J.E. and Hughes, C.R.(1993). Microsatellites and Kinship. *TREE*, **8**:285-288.

Quintana,-Murci, L., Semino, O., Poloni, E.S., Liu, A., Van Gijn, M., Passarino, G., Brega, A., Nasidze, I.S., Maccioni, L. and Cossu, G. (1999). Y-chromosome specific YCAII, DYS19 and YAP polymorphisms in human populations: A comparative study. *Ann. Hum. Genet.*, **63**: 153-166.

Rak, Y. (1993). Morphological variation in Homo neandertalensis and Homo sapiens in the Levant. A biogeographic modern. In : *Species, Species Concepts and Primate Evolution*. Eds., Kimbel, W.H. and Martin, L.B. Plenum Press, New York, pp.523-536.

Ramana, G.V., Su, B., Jin, L., Singh, L., Wang, N., Underhill, P. and Chakraborty, R. (2001). Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.*, **9**: 695-700.

Raymond, M.A. and Fletcher, S. (1983). Haptoglobin typing in Zimbabwe. *Forensic Sci. Int.*, **22**:237-242.

Raymond, M.L. and Rousset, F. (1995). An exact test for population differentiation. *Evolution*, **49**: 1280-1283.

Regueiro, M., Carril, J.C., Pontes, M.L., Pinheiro, M.F., Luis, J.R. and Caeiro, B. (2004). Allele distribution of 15 PCR-based loci in the Rwanda Tutsi population

by multiplex amplification and capillary electrophoresis. *Forensic Sci. Int.,* **143**: 61-63.

Reynolds, R.L. and Sensabaugh, G.F. (1990). Use of the polymerase chain reaction for typing Gc variants. In: *Advances in Forensic Haemogenetics.* Eds., Polesky, H.F. and Mayr, W.R. Springer-Verlag, Berlin Heidelberg, pp.158-161.

Richie, K.L., Goldsborough, M.D., Darfler, M.M., Benzinger, E.A., Lovekamp, M.L., Reeder, D.J. and O'Connell, C.D. (1999). Long PCR for VNTR analysis. *J.Forensic Sci.,* **44**: 1176-1185.

Rieder, M.J., Taylor, S.L., Clark,A.G. and Nickerson, D.A.(1999).Sequence variation in the human angiotensin converting enzyme.*Nature Genet.,* **22**: 59-62.

Rightmire, G.P. (1984). In: *Origins of Modern Humans: A World survey of the Fossil Evidence.* Eds. Smith, F.H and Spencer, F., Liss, New York,pp. 295-326.

Robertson, H. M. and Lampe, D. J. (1995). Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. *Mol. Biol. Evol.*,**12**: 850-862.

Roewer, L., Arnemann, J., Spurr, N.K., Grzeschik, K.H. and Epplen, J.T. (1992). Simple repeat sequences on the Y chromosome are equally polymorphic as their autosomal counterparts. *Hum. Genet.,* **89**: 389-394.

Rosenberg, C.L., Des las Morenas, A., Huang, K., Cupples, L.A., Faller, D.V. and Larson, P.S. (1996). Detection of monoclonal microsatellite alterations in Atypical Breast Hyperplasia. *J. Clin. Invest.*, **98**: 1095-1100.

Ross, A. (1986). Results of a gene frequency survey for a number of blood group systems within the South Australian population. *Proceedings of the 9th Australian Int. Forensic Sci. Symp.*: 1-5.

Ross, A.M. and Harding, H.W.J. (1989). DNA Typing and Forensic Science. *Forensic Sci. Int.,* **41**: 197-203.

Rothwell, T.J. (1985). The frequency of occurrence of various Human Blood Groups in the United Kingdom, with observation on their Regional Variation. *J.Forensic Sci., Soc.,* **25**: 135-144.

Royle, N.J., Clarkson, R.E., Wong, Z. and Jeffreys, A.J. (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics*, **3**: 352.

Roy, R. (1990). Distributions of Genetic Markers in a Nebraska Population. *J.Forensic Sci.*, **35**: 1207-1210.

Roy, R. and Reynolds, R. (1995). APPLIED BIOSYSTEMS, AmpliTypePM and HLA DQα typing from pap smear, semen smear and postcoital slides. *J.Forensic Sci.*, **40**: 266-269.

Ruvolo, M., Disotell, T.R., Allard, M.W., Brown, W.M. and Honeycutt, R.L. (1991). Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. *Proc. Natl. Acad. Sci.*, USA, **88**: 1570.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a Thermostable DNA Polymerase. *Science*, **239**: 487-491.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985). Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, **230**: 1350-1354.

Saiki, R.K., Walsh, P.S., Levenson, C.H. and Erlich, H.A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. *Proceedings of the Natl. Acad. Sci., USA*, **86**: 6230-6234.

Sambrook, J. and Russell, D.W.(2001). *Molecular Cloning: A Laboratory Manual, Third Edition,* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Sanchez-Molina, I. and Calvet, R.(2000). Allelic frequencies for the HLA-DQA1, D1S80, HUMTHO1, HUMTPOX, HUMCSF1PO and HUMVWA loci in Cantabria (Middle North Spain). *J. Fornsic Sci.*, **45**: 167-169.

Sanger, F. and Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.*, **94**: 441-448.

Satoh, C., Ferrell, R.E., Tanis, R.J., Ueda, N., Kishimoto, S., Neel, J.V., Hamilton, H.B., and Baba, K. (1977). The frequency in Japanese of Genetic Variants of 22 Proteins. III. Phosphoglucomutase-1, Phosphoglucomutase-2, 6-Phosphogluconate Dehydrogenase, Adenylate Kinase and Adenosine Deaminase. *Annals of Hum. Genet.*, **41**: 169-183.

Sawaguchi, A., Wang, X., Sawaguchi, T. and Tezuka, Y. (2000). Analysis of STR and application to paternity testing. *Progress in Forensic Genet.*, **8**: 377-380.

Schlotterer, C. and Tautz, D. (1992). Slippage synthesis of simple sequence DNA. *Nucl. Acids Res.*, **20**: 211 - 215.

Schmitz, R.W. (1996). *Das Alt-und Mittelpalãolithikum des Neandertals und benachbarter Gebiete. Doctoral dissertation, Prehistory.* University of Cologne, (Cologne, University Microfilms).

Schmitz, R.W., Pieper, P., Bonte, W. and Krainitzki, H. (1995). New investigations of the Homo sapiens neanderthalensis, found in 1856. In: *Advances in Forensic Sciences, Vol. 7: Forensic Odontology and Anthropology. Proceedings of the 13th Meeting of the International Association of Forensic Sciences,* Düsseldorf, August 22nd to 28th, 1993. Eds., Jacob, B. and Bonte, W., Berlin, Kõster, pp.42-44.

Schneider, P.M., Meuser, S., Waiyawuth, W., Seo, Y. and Rittner, C. (1998). Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci. Int.*, **97**: 61-70.

Schon, E.A., Hirano, M. and DiMauro, S. (1994). Mitochondrial encephalomyopathies: clinical and molecular analysis. *J Bionerg. Biomembr.*,**26**: 291-299.

Schwartz, J. and Tattersall, I.(1996). Significance of some previously unrecognized apomorphies in the nasal region of Homo neandertalensis. *Proc. Natl. Acad. Sci.*, USA, **93**: 10852-10854.

Sebatan, I.M., Funayama, M. and Sagisaka, K. (1986). Red cell phosphoglucomutase (PGM_1) subtypes in Egyptians. *Forensic Sci. Int.*, **31**: 55-59.

Sehajpal, P.K., Mehta, K., Sharma, V.K. and Sandhu, H.S. (1983). *Detection of red cell glyoxalase 1 types in menstrual blood.* 10. Internationale Tangung der Geselleschaft fur forensische Blutagruppenkunde, e.v. Munchen 11-15 September.

Sensąbaugh, G.F. (1982). Isozymes in Forensic Science. In: *Isozymes, Current Topics in Biological and Medical Research,* Vol. 6. Eds., Rattazzi, M.C., Scandalios, J.G.and Whitt, G.S. Alan, R. Liss, New York, pp. 247-282.

Sensabaugh, G.F. (1983). The Utilization of Polymorphic Enzymes in Forensic Science. In: *Isozymes, Current Topics in Biological and Medical Research,* Vol.

11. Eds., Rattazzi,M.C.,Scandalios, J.G.and Whitt, G.S. Alan R. Liss,New York,pp.137-154.

Sharma, A., Arora,V.K. and Bhalla, V. (1986).Determination of glyoxalase 1(GLO 1). polymorphs in Delhites. *Indian J. Criminology and Criminalistics*.

Sharma, A., Arora, V.K. and Bhalla, V (1987). Studies on Glyoxalase 1 Isoenzymes in semen stains: Polymorphism in Himachal (India) population. *Forensic Sci. Int.*, **37**: 201-204.

Shigenaga, M.K., Hagen, T.M., Ames, B.N. (1994). Oxidative damage and mitochondrial decay in aging. *Proc. Natl. Acad. Sci.*, USA, **91**: 10771-10778.

Shuster, C.A., Speed, W.C., Kidd, J.R. and Kidd, K.K. (1998). Short tandem repeat polymorphism evolution in humans. *Eur.J.Hum.Genet.*, **6**, 38-49.

Silver, H. (1983). Ed., *Probability of Inclusion in Paternity Testing*. American Association of Blood Banks, Arlington, VA.

Simmons, G. M. (1992). Horizontal transfer of hobo transposable elements within the Drosophila melanogaster species complex: evidence from DNA sequencing. *Mol. Biol. Evol.*, **9**: 1059-1060.

Singer, M.R. and Skowronski, J.(1985). Making Sense Out of LINES: Long Interspersed Repeat Sequences in Mammalian Genomes. *Trends in Bioch. Sci.*, **10**: 119-112.

Skowronski, J., Fanning, T. G. and Singer, M. F. (1988). Unit-length LINE-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.*, **8**: 1385-1397.

Slightom, J.L., Blechl, A.E. and Smithies, O. (1980). Human fetal $^{G}\gamma$– and $^{A}\gamma$–globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell*, **21**: 627-638.

Smit,A. F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet., Dev.*, **9**: 657-663.

Smit, A. F. (1996). The origin of interspersed repeats in the human genome. *Curr. Opin. Genet., Dev.*, **6**: 743-748.

Smit, A. F., Toth, G., Riggs, A. D. and Jurka, J.(1995). Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.*, **246**: 401-417.

Smith, B.C., Fisher, D.L., Weedn, V.W., Warnock, G.R. and Holland, M.M. (1993). A systematic approach to the sampling of dental DNA. *J.Forensic Sci.*, **38**: 1194-1209.

Smith, C.A.B. (1986). Chi-squared tests with small numbers. *Annals of Hum.Genet.*, **50**: 163-167.

Smith, C.I., Chamberlain, A.T., Riley, M.S., Cooper, A., Stringer, C.B. and Collins, M.J. (2001). Neanderthal DNA: Not just old but old and cold? *Nature*, **410**: 771-772.

Smith, G. (1976). Evolution of repeated DNA sequences by unequal crossover. *Science,* **191**: 528-535.

Soares-Vieira, J.A., Munoz, D.R., Iwamura, E.S.M., Billerbeck, A.E.C.and Otto, P.A. (2001). Allele frequency distribution of three STR loci (CSF1PO, TPOX and THO1) in a Brazilian population sample. *J.Forensic Sci.,* **46**: 996-997.

Sokal and Rohlf. (1995). *Biometry, 3rd edition.* W.H. Freeman and Co., New York.

Soriano, P., Meunier-Rotival,M. and Bernardi, G.(1983). The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes. *Proc. Natl. Acad. Sci.*, USA, **80**: 1816-1820.

Sovinsky, S.M., Baird, L.S., Budowle, B., Caruso, J.F., Cheema, D.P.S., Duncan, G.T., Hamby, P.P., Masibay, A.S., Sharma, V.K. and Tahir, M.A. (1996). Development of a deoxyribonucleic acid (DNA) restriction fragment length polymorphism (RFLP) database for Punjabis in East Punjab, India. *Forensic Sci. Int.*,**79**: 187-198.

Stedman, R. (1985). Blood group frequencies of immigrant and indigenous populations from south east England. *J.Forensic Sci. Soc.*, **25**: 95-134.

Stedman, R.(1972). Human population frequencies in twelve blood grouping systems. *J.Forensic Sci., Soc.*, **12**: 379-413.

Steffens, D.L., Roy, R. and Brumbaugh.(1997). Multiplex amplification of STR loci with gender alleles using infrared fluorescence detection. *Forensic Sci. Int.,* **85**: 225-232.

Stephan,W.(1986). Recombination and the Evolution of Satellite DNA. *Genet. Research (Cambridge)*: **47**: 167-174.

Stoneking, M. and Cann, R.L. (1989). In: *The Human Revolution : Behavioural and Biological Perspectives on the Origins of Modern Humans.* Eds., Mellars, P and Stringer, C. Edinberg Univ. Press, Edinburgh, pp. 17-30.

Stoneking, M., Hedgecock, D., Higuchi, R.G., Vigilant, L. and Erlich, H.A. (1991). Population variation of human mtDNA control region sequences detected by

enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.*, **48**: 370-382.

Stoneking, M., Sherry, S.T., Redd, A.J. and Vigilant, L. (1992). New approaches to dating suggest a recent age for the human mtDNA ancestor. *Phil. Trans. R. Soc. Lond.*, **B 337**: 167-175.

Stoneking, M. (2001). Single nucleotide polymorphisms: From the evolutionary past... *Nature*, **409**: 821-822.

Strand, M., Prolla, T., Liskay, R. and Petes, T. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, **365**: 274-276.

Strimmer, K. and Haeseler, A. V. (1996). Quartet puzzling: A maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.*, **13**: 964-969.

Stringer, C.B, (1992). In: *Continuity or Replacement: Controversies in Homo sapiens Evolution.* Brauer,G and Smith, F.H Eds.(Balkerna, Rotterdam)pp.9-23.

Sueblinvong, T. and Kongsrisook, U. (1999). Population data of 8 short tandem repeat loci in the Thai population. *Forensic Sci. Int.*, **103**: 199-205.

Sueoka, N. (1988). Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad.Sci.*, USA, **85**: 2653-2657.

Sullivan, K., Hopgood, R. and Gill, P. (1992). Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int. J. Leg. Med.*, **105**: 83-86.

Sullivan, K.M., Hopgood, R., Lang, B. and Gill, P. (1991). Automated amplification and sequencing of human mitochondrial DNA. *Electrophoresis,* **12**: 17-21.

Sullivan, K.M., Piercy, B., Benson, N., Ivanov, P., Mannucci, A. and Gill, P. (1994). Solid-Phase sequencing of mtDNA: Toward a fully automated forensic test. In: *Advances in Biomagnetic Separation.* Eds., Uhlen, M., Hornes, E. and Olsvic, O. Eaton Publishing, Natick, MA, USA, pp.135-148.

Swofford, D.L. and Olsen, G.J. (1990). In: *Molecular Systematics.* Eds., Hillis D.M and Moritz. Sinauer, Sunderland, MA, pp. 411-501.

Swofford, D.L. (1990). *PAUP Version 3.0.* Illinois Natural History Survey, Champaign.

Szibor, R., Michael, M., Plate, I. and Krause, D. (2000). Efficiency of forensic mtDNA analysis case examples demonstrating the identification of traces. *Forensic Sci. Int.*, **113**: 71-78.

Tahir, M.A., Herrera, R.J., Khan, A.A., Kashyap, V.K., Duncan, G., Barna, C., Budowle, B., Rowold, D.J., Sinha, S. and Amjad, M. (2000). Distribution of HLA-DQA1, polymarker, CSF1PO, vWA, THO1, TPOX, D16S539, D7S820, D13S317 and D5S818 alleles in East Bengali and West Punjabi populations from Indo-Pak Subcontinent. *J.Forensic Sci.*, **45**: 1320-1323.

Takahashi, M., Kato, Y., Mukoyama, H., Kanaya, H. and Kamiyama, S.(1997). Evaluation of five polymorphic microsatellite markers for typing DNA from decomposed human tissues – Correlation between the size of the alleles and that of the template DNA. *Forensic Sci. Int.*, **90**: 1-9.

Tamura, K. and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *J. Mol. Evol.*, **10**: 512-526.

Tautz, D. (1991). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.*, **17**: 6463-6471.

Taylor, A.C., Sherwin, W.B., and Wayne, R.K.(1994). Genetic variation of microsatellite loci in a bottlenecked species: the northern hairy-nosed wombat Lasiorhinus krefftii. *Mol. Ecol.*, **3**: 277-290.

Templeton, A. R. (1992). Human origins and analysis of mitochondrial DNA sequences. *Science,* **255**: 737-741.

The International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphims. *Nature*, **409**: 928-933.

Thomas, A.W., Morgan, R., Sweeney, M., Rees, A. and Alcolado, J. (1994). The detection of mitochondrial DNA mutations using single stranded conformation polymorphism (SSCP) anlysis and heteroduplex analysis. *Hum. Genet.*, **94**: 621-623.

Thomas, W.K. Martin, S.L. (1993). A recent origin of marmots. *Mol. Phyl. Evol.*, **2**: 330-336.

Thomson J.A., Pilotti, V., Stevens, P., Ayres, K.L. and Debenham, P.G. (1999). Validation of short tandem repeat analysis for the investigation of cases of disputed paternity, *Forensic Sci. Int.*, **100**: 1-16.

Thurston, E. and Rangachari, K. (1905). *Castes and Tribes of Southern India, Vol I – A and B.* Cosmo Publications, Delhi, India.

Tishkoff, S.A., Dietzsch, E., Speed, W., Pakstis, A.J., Kidd, J.R., Cheung, K., Bonne-Tamir, B., Santachiara-Benerecetti, A.S., Moral, P., Krings, M., Paabo, S., Watson, E., Risch, N., Jenkins, T. and Kidd, K.K. (1996). Global Patterns of Linkage Disequilibrium at the CD4 Locus and Modern Human Origins. *Science*, **271**: 1380-1387.

Tomàs, C., Picornell, A., Castro, J.A. and Misericòrdia, M. (2001). STR data for the APPLIED BIOSYSTEMS, AmpFlSTR profiler plus loci from Majorcan, Minorcan and Valencian populations (Eastern Spain). *Forensic Sci. Int.*, **121**: 201-204.

Toth, G., Gaspari, Z. and Jurka, J.(2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.*, **10**: 967-981.

Tully, G., Bär, W., Brinkmann, B., Carracedo, A., Gill, P., Morling, N., Parson, W. and Schneider, P. (2001). Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Sci. Int.*, **124**: 83-91.

Ueda, N., Satoh, C., Tanis, R.J., Ferrell, R.E., Kishimoto, S.,-Neel, J.V., Hamilton, H.B. and Baba, K. (1977). The frequency in Japanese of Genetic Variants of 22 Proteins. II. Carbonic Anhydrase I and II. Lactate Dehydrogenase, Malate Dehydrogenase, Nucleoside Phosphorylase, Triose Phosphate Isomerase, Haemoglobin A and Haemoglobin A_2. *Annals of Hum. Genet.*, **41**:43-52.

Underhill, P.A ., Shen, P., Lin, A.A., Jin, L., Passarino, G., Yang, W.H., Kauffman, E., Bonne'-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M.T., Wells, R.S., Piazza, A., Davis, R.W., Feldman, M.W., Cavalli-Sforza, L.L. and Oefner, P.J. (2000). Y chromosome sequence variation and the history of human populations. *Nature Genet.*, **26**: 358-361.

Urquhart, A., Kimpton, C.P., Downes, T.J. and Gill, P. (1994). Variation in short tandem repeat sequences – a survey of twelve microsatellite loci for use as forensic identification markers. *Int. J. of Leg. Med.*, **107**: 13-20.

Van Regenmortel, M.H.V. (1997). Viral species. In: *Species-The units of biodiversity.* Eds. Claridge, M.F., Dawah, H.A. and Wilson, M.R. Chapman and Hall, pp. 17-23.

Vargas, C.I., Castillo, A., Gil, A.M., Pico, A.L. and García, O. (2001). Profiler data for a Colombian population (Department of North Santander). *Forensic Sci. Int.* 124: 219-220.

Vigilant, L., Pennington, R., Harpending, H., Kocher, T.D. and Wilson, A.C.(1989). Mitochondrial DNA sequences in single hairs from a southern African population. *Proc. Natl. Acad. Sci.*, USA, **86**: 9350-9354.

Vigilant, L.(1990). *Ph.D Thesis.* University of California, Berkeley.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. and Wilson, A.C.(1991). African Populations and the Evolution of Human Mitochondrial DNA. *Science*, **253**: 1503-1507.

Wainscoat, J.S., Pilkington, S., Peto, T.E.A., Bell, J.I. and Higgs, D.R. (1987). Allele-Specific DNA Identity Patterns. *Hum. Genet.*, **75**: 384-387.

Walker, R.H. (1983). Ed., *Inclusion Probabilities in Parentage Testing.* American Association of Blood Banks, Arlington, VA.

Wallace, D.C. (1992). Diseases of the mitochondrial DNA. *Annu Rev Biochem.*, **61**: 1175-1212.

Wallace, D.C. (1997). Mitochondrial DNA in Aging and Diesease. *Sci. Am.* Aug: 22-29.

Wallace, D.C., Singh, G., Lott, M.T. *et al.* (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science*, **242**: 1427-1430.

Walsh, P.S., Metzger, D.A. and Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, **10** : 506-513.

Walter, H. and Steegmüller, H. (1969). Studies on the Geographical and Racial Distribution of the Hp and Gc Polymorphisms. *Hum. Heredity,* **19**: 209-221.

Walton, H. and Palsson, J. (1973). The incidence of some Genetic Markers in Ireland. In: *Genetic Variation in Britain.* Eds., Roberts, D.F. and Sunderland, E. Taylor and Francis, London, pp. 161-170.

Watson, J.D. and Crick, F.H.C. (1953). Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, **171**: 737.

Weber, J.L. and May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.*, **44**: 388-396.

Weichhold, G.M., Berk, J.E., Korte, W., Eisenmenger, W. and Sullivan, K.M.(1998). DNA analysis in the case of Kasper Hauser. *Int. J. Leg. Med.*, **111**: 287-291.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morisette, J. Millaseau, P., Vaysseix, G. and Lathrop, M. (1992). A second generation linkage map of the human genome. *Nature*, **359**: 794-801.

Weitzmann, M., Woodford, K., and Usdin, K. (1997). DNA secondary structures and the evolution of hypervariable tandem arrays. *J. Biol. Chem.*, **272**: 9517-9523.

Wells, R. (1996). Molecular basis of genetic instability of triplet repeats. *J. Biol. Chem.*, **271**: 2875-2878.

Whitfield, L.S., Suston, J.E. and Goodfellow, P.N. (1995). Sequence variation of the human Y chromosome. *Nature*, **378**: 379-380.

Wills, C. (1995). When did Eve live? An evolutionary detective story. *Evolution* **49**: 593-607.

Wills, C. (1998). *Children of Prometheus*: *The Accelerating Pace of Human Evolution.* Perseus Books, Reading, MA.

Wilson, K. and Goulding, K.H. (1989). Editors, Isolation of Nucleic Acids – DNA. In: *A Biologist's Guide to Principles and Techniques of Practical Biochemistry, 3rd edn.* Edward Arnold – A division of Hodder & Stoughton.

Wilson, M.R., DiZinno, J.A., Polanskey, D., Replogle, J. and Budowle, B. (1995). Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int. J. Leg. Med.*, **108**: 68-74.

Wilson, M.R., Stoneking, M., Holland, M.M., DiZanno, J.A. and Budowle, B. (1993). Guidelines for the use of mitochondrial DNA sequencing in forensic science. *Crime Lab. Digest* **20**: 68-77.

Wittig, H., Augustin, C., Baasner, A., Bulnheim, U., Dimo-Simonin, N., Edelmann, J., Hering, S., Jung, S., Lutz, S., Michael, M., Parson, W., Poetsch, M., Schneider, P.M., Weichhold, G. and Krause, D. (2000). Mitochondrial DNA in the central European population Human identification with the help of the forensic mt-DNA D-Loop base database. *Forensic Sci. Int.*, **113**: 113-118.

Wolfe, K. H., Sharp, P. M. and Li, W. H. (1989). Mutation rates differ among regions of the mammalian genome. *Nature*, **337**: 283-285.

Wolpoff, M. (1980). *Paleoanthropology.* Knopf, New York.

Wolpoff, M.H. (1992). In: *Continuity or Replacement: Controversies in Homo sapiens Evolution.* Eds., Brauer,G and Smith, F.H. Balkema, Rotterdam, pp. 25-63.

Wolpoff, M.H. (1989). In: *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans.* Eds.,Mellars, P. and Stringer, C. Edinburgh Univ. Press, Edinburg, pp. 62-108.

Wolpoff, M.H. (1989). In: *The Human Revolution: Multiregional evolution., the fossil alternative to Eden.* Eds.,Mellars, P. and Stringer, C. Edinburgh Univ. Press, Edinburg

Wong, Z., Wilson, V., Jeffreys, A.J. and Thein, S.L. (1986). Cloning a selected fragment from a human DNA "fingerprint": Isolation of an extremely polymorphic minisatellite. *Nucl. Acids Res.*, **14**: 4605.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann. Hum. Genet.*, **51**: 269.

Woo, K.M. and Budowle, B.(1995). Korean population data on the PCR-based loci LDLR, GYPA, HBGG, D7S8, Gc, HLA-DQA1, and D1S80. *J.Forensic Sci.*, **40**: 645-648.

Wyman, A.R. and White, R. (1980). A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci.*, USA, **77**:6754-6758.

Yamamoto, T., Davis, C.G., Brown, M.S., Schneider, W.J., Casey, M.L., Goldstein, J.L. and Russell, D.W. (1984). The human LDL receptor: A cysteine-rich protein with multiple Alu sequences in its mRNA. *Cell*, **39**: 27-38.

Yamamoto, T., Uchihi, R., Nozawa, H., Huang, X-L., Leong, Y-K., Tanaka, M., Mizutani, M., Tamaki, K. and Katsumata, Y. (1999). Allele distribution at nine STR loci – D3S1358, vWA, FGA, THO1, TPOX, CSF1PO, D5S818, D13S317

and D7S820 – in the Japanese population by multiplex PCR and capillary electrophoresis. *J.Forensic Sci.*, **44**: 167-170.

Yang, F., Brune, J.L., Naylor, S.L., Apples, R.L. and Naberhaus, K.H. (1985). Human group specific component (Gc) is a member of the albumin family. *Proc. Natl. Acad. Sci.*, USA, **82**: 7994-7998.

Zarrabeitia, M.T., Amigo, T., Sañudo, C., Zarrabeitia, A., González-Lamuño, D. and Riancho, J.A. (2002). A new pentaplex system to study short tandem repeat markers of forensic interest on X chromosome. *Forensic Sci. Int.*, **129**: 85-89.

Zollikofer, C., Ponce de Leon, M., Martin, R. and Stucki, P. (1995). Neandertal computer skulls. *Nature*, **375**: 283-285.

Zuckerkandl, E. and Pauling, L. (1965). In: *Evolving Genes and Proteins*. Eds., V. Bryson and Vogel, H.J. Academic Press, New York, pp. 97-166.

Zupanic, I., Balazic, J. and Komel, R. (1998). Analysis of nine short tandem repeat (STR) loci in the Slovenian population. *Int. J.Leg. Med.*, **111**: 248-250.