Y-SPECIFIC RESTRICTION FRAGMENT LENGTH POLYMORPHISMS IN SOUTHERN AFRICAN POPULATIONS

AMANDA BARBARA SPURDLE

A thesis submitted to the Faculty of Medicine, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy

Johannesburg, 1992

ABSTRACT

Seven Y chromosome probes and thirteen restriction enzyme digests were used to examine a conservative estimate of 20000bp, and no new Y-specific polymorphisms were revealed by these systems. The Y chromosome probe 49a, which reveals a Y-specific haplotype with *Taq*I, was shown to reveal five new complex polymorphisms with *Bgl*II, *Hin*dIII, *Pst*I, *Pvu*II and *Sst*I. The new polymorphisms exhibit great genetic diversity, and each enzyme reveals numerous haplotypes, which mostly occur infrequently and are population-specific. The haplotypes for a given enzyme do not correlate strictly with those revealed by the other enzymes, including *Taq*I, suggesting that each polymorphism results from a combination of restriction site mutations and rearrangement events. Association between the different 49a polymorphisms occurs only in individuals of recent common genetic origin.

Y-specific 49a/*Taq*I haplotypes were determined for 933 individuals drawn from 23 different African populations. A total of 31 new haplotypes were observed, some of which contained new alleles or allelic variants. Duplication, in addition to C_pG mutation, is implicated in the generation of certain allelic variants. Cluster analysis of genetic distances

between populations was calculated using the 49a/*Taq*I haplotype frequencies. Y-specific 49a/*Taq*I haplotype analysis of individual populations was not sufficiently sensitive to accurately distinguish between the different Bantu-speaking Negroid tribal groups. Cluster analysis of larger groupings was more stable, and with the exception of the Khoisan, resulted in a basic split between African and non-African populations.

The linkage disequilibrium of the XY275 *Msp*I Y-linked polymorphism was determined. The *high* allele was generally found in association with the Y chromosome, but the Y-associated *low* allele was found to occur in Bantu-speaking Negroids, Khoisan-speaking Negroids, the Khoisan, two groups of mixed ancestry, and the Caucasoid South African Asiatic Indian population. The discovery of Y-associated *low* alleles in non-African as well as African populations suggests that more than one Y chromosome gave rise to the present-day non-African population.

The pDP31/*Eco*RI, p21A1/*Taq*I and Y *Alu* polymorphisms were also studied in several southern African populations. The pDP31 duplication occurred at high frequencies in Caucasoids, and could be used to indicate Caucasoid male gene flow into hybrid populations. The p21A1/*Taq*I point mutation showed no

distinct trends in frequency in the different populations, and several *Taq*I mutations are proposed to have occurred in the repeat unit recognized by this sequence. The Y *Alu* polymorphism occurred infrequently in Caucasoids, at intermediate frequency in the Khoisan, and at high frequency in Negroids. The presence of the Y *Alu* insertion in all three major population groups studied is interpreted to suggest that the insert predates the diversification of *Homo sapiens*.

The relationship between the different Y-linked polymorphisms was determined in the populations studied. The Y *Alu* polymorphism is believed to have originated once from sequencing data, but such information is not available for the other Y polymorphisms studied. No absolute relationship was observed between the Y *Alu* polymorphism and the 49a/*Taq*I, XY275 *Msp*I, pDP31/*Eco*RI and p21A1/*Taq*I polymorphisms. It is suggested that the latter polymorphisms have arisen more than once.

DECLARATION

I declare that this dissertation is my own, unaided work. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university.

I declare that this work has been passed by the Ethics committee of the University of the Witwatersrand, and the certificate number is 24/2/89.

Amanda Barbara Spurdle

Amanda Barbara Spurdle

Seventh day of July, 1992.

This work is dedicated to my parents for their encouragement throughout my academic career, and to my husband, Jonathan Irons, for his support in recent years.

ACKNOWLEDGEMENTS

I am indebted to the many individuals who willingly donated blood samples for DNA extraction, for without them this study could not have been initiated. I am also grateful to Sister Esther Zwane, the Namibia Blood Transfusion Service, Windhoek State Laboratories, the Highveld Blood Transfusion Service, and the officials of the Lemba Cultural Association for their assistance in the collection of these samples. The director of the South African Institute for Medical Research kindly made facilities in the Department of Human Genetics available for the research described in this thesis.

My gratitude is extended to Dr J.Rogers for provision of Pygmy DNA samples, to Drs H.Cooke, M.Hammer, D.Page and J.Weissenbach for Y chromosome DNA probes, and to Dr N.Ellis for sequence data used to construct primers for the detection of the XY275 polymorphism, and for probe Hf0.2. The research detailed in this thesis has also benefited from generous personal communications from a number of the forementioned colleagues, and also from D.Bailey, M.Jobling, G.Lucotte, A.Novelletto, A.S.Santachiara Benerecetti, O.Semino, K.Taylor, J.Wolfe and Z.Ulinowski.

Many thanks to the S.A.I.M.R. Photographics and Illustrations Units for their assistance with figures, to colleagues in the Cytogenetics Unit of our Department for supplying invaluable cell lines, to David Dunn for his initial help with the Antana computer analysis, to Jonathan Irons for providing all-encompassing computer support, and to Alan Spurdle for assistance with printing.

I also wish to express my appreciation to my colleagues in the Department of Human Genetics for their friendship and helpful discussion, especially Dr Michele Ramsay. Special thanks are due to my supervisor Professor Trefor Jenkins for his guidance and support throughout this study.

TABLE OF CONTENTS **PAGE**

CHAPTER THREE

SUBJECTS AND METHODS

CHAPTER SIX

THE *Bgl*II, *Hin*dIII, *Pst*I, *Pvu*II AND *Sst*I 49a RFLPS AND THEIR RELATIONSHIP TO THE *Taq*I POLYMORPHISM

CHAPTER SEVEN

THE Y-LINKED XY275 POLYMORPHISM IN SOUTHERN AFRICAN POPULATIONS

CHAPTER EIGHT

THE pDP31, p21A1 AND *ALU* INSERTION POLYMORPHISMS IN SOUTHERN AFRICAN POPULATIONS

CHAPTER NINE

INTERRELATIONSHIPS BETWEEN THE DIFFERENT Y-LINKED POLYMORPHISMS

LIST OF TABLES

PAGE

CHAPTER 3

CHAPTER 5

CHAPTER 6

CHAPTER 8

CHAPTER 9

LIST OF FIGURES

PAGE

CHAPTER 3

CHAPTER 5

CHAPTER 6

LIST OF ABBREVIATIONS

ACD	acid citrate dextrose
ARMS	amplification refractory mutational system
ASO	allele specific oligonucleotide
bp	base pair
BSA	bovine serum albumin
^{o}C	degrees centigrade
cDNA	complementary DNA
CEPH	Centre d'Etude du Polymorphisme Humaine
cm	centimetre
cM	centiMorgan
dCTP	deoxycytidine-5'-triphosphate
DGGE	denaturing gradient gel electrophoresis
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
EDTA	ethylene-diamine-tetra-acetic acid
g	gram
Hepes	N'-2 ethanesulphonic acid
Ht	haplotype
inv(Y)	inverted Y chromosome
kb	kilobase
l	litre
LB	Luria Bertani Medium
M	molar
μg	microgram
μl	microlitre

mg	milligram
min	minute
ml	millilitre
mM	millimolar
MPSI	male procreative superiority index
mt	mitochondrial
ng	nanogram
nM	nanomolar
OD_{550}	optical density at 550nm
PABY	pseudoautosomal boundary of the Y
pM	picomolar
poly A	polymeric deoxyadenosine monophosphate
PCR	polymerase chain reaction
RFLP	restriction fragment length polymorphism
rpm	revolutions per minute
RNA	ribonucleic acid
RNAse	ribonuclease
SDS	sodium dodecyl sulphate
SSCP	single stranded conformational polymorphism
TBE	tris borate
TE	tris EDTA
Tris	tris(hydroxymethyl)methylamine
UV	ultraviolet
VNTR	variable number of tandem repeats
YAP	Y *Alu* polymorphism

PUBLICATIONS

Publications arising from this study:

SPURDLE AB, HAMMER M, JENKINS T. The evolution of Y-specific polymorphisms in African populations (in preparation).

SPURDLE AB, JENKINS T (1990). Candidate TDF probe pDP1007 detects an X polymorphism with *Hind*III at ZFX. *Nucl Acids Res* 18: 3430.

SPURDLE AB, JENKINS T (1992). Y chromosome probe p49a detects complex *Pvu*II haplotypes and many new *Taq*I haplotypes in southern African populations. *Am J Hum Genet* 51: 107-125.

SPURDLE AB, JENKINS T. The Y chromosome as a tool for studying human evolution. *Curr Genetics and Devel*, in press.

SPURDLE AB, JENKINS T. The inverted Y chromosome polymorphism in the Gujerati Muslim Indian population of South Africa has a single origin. *Hum Hered*, in press.

SPURDLE AB, JENKINS T. The Y-specific p21A1/*Taq*I polymorphism occurs in four major population groups. *Hum Hered*, in press.

SPURDLE AB, JENKINS T. The search for Y chromosome polymorphism is extended to Negroids. *Hum Molec Genet*, in press.

SPURDLE AB, JENKINS T. The Y-specific pDP31 duplication polymorphism in southern African populations (in preparation).

SPURDLE AB, JENKINS T. Complex polymorphisms are revealed by Y chromosome probe 49a with *Bgl*II, *Hin*dIII, *Pst*I and *Sst*I (submitted).

SPURDLE AB, MORRIS D, JENKINS T (1989). Y-chromosome probe p49f detects new *Taq*I variants and complex *Pvu*II haplotypes. *Cytogenet Cell Genet* 51: 1084.

SPURDLE A, RAMSAY M, JENKINS T. The Y-associated XY275 *low* allele is not restricted to indigenous African peoples. *Am J Hum Genet*, in press.

CHAPTER ONE

INTRODUCTION

Human population genetics has benefited greatly from the use of restriction fragment length polymorphisms (RFLPs) as an analytical tool (Cavalli-Sforza *et al* 1986, Wainscoat *et al* 1986). Analysis of the frequencies of alleles in different populations permits construction of genetic distance maps between these populations (Cavalli-Sforza 1974, Cavalli-Sforza *et al* 1986). The potential of human Y-linked RFLPs in human population genetic studies has been discussed by various workers, including Casanova *et al* (1985), Ngo *et al* (1986), Breuil *et al* (1987), and Oakey and Tyler-Smith (1990). Since the Y chromosome is a single haploid entity whose markers are not redistributed at meiosis, neutral mutations are transmitted to all male progeny. It represents the specific paternal contribution to the male genome, and thus Y-specific RFLPs are useful for studying male gene flow, especially the direction of such flow. Furthermore, Y-specific gene flow studies complement mitochondrial DNA studies which are specific for maternal gene flow (Cann *et al* 1984, Cann *et al* 1987, De Benedictis *et al* 1989, Semino *et al* 1989).

Southern Africa is a particularly interesting area in which to conduct gene flow studies. Its people represent three of the major races of mankind (Nurse *et al* 1985), namely Negroid, Caucasoid and Khoisan. Present theories on the ancestry of these populations, and on the admixture between them, are based on archeaological, linguistic and some genetic data (Nurse *et al* 1985). The study of male-specific genetic markers in these populations could confirm, refine or refute these theories, since it has been the practice in Africa that unilateral flow of genes occurs from only the females of the subservient group to the dominant group (Jenkins 1982), whereas population groups of the same status interchange genes bilaterally. Unfortunately, relatively few Y-linked RFLPs have been discovered, and there appears to be a general paucity of polymorphism on the Y chromosome relative to other chromosomes (Arnemann *et al* 1985, Jakubiczka *et al* 1989, Malaspina *et al* 1990).

The aims of this study were twofold: to search for new Y-specific polymorphisms in southern African populations, and to use these and previously documented Y chromosome polymorphisms in anthropogenetic studies on southern African populations.

CHAPTER TWO

LITERATURE REVIEW

2.1 GENETIC VARIATION

The human haploid genome is comprised of approximately 3×10^9 nucleotides (Bishop 1974), and contains an estimated 50 000 to 100 000 functional genes (Bishop 1974, Bodmer 1981). A large proportion of the DNA situated between and within genes is thought to be non-functional, and is sometimes referred to as "junk" DNA. Natural variation exists at the genetic level, some of which may be translated into protein variation. The variant forms occurring at a given locus are termed the alleles of that locus.

2.1.1 Polymorphism

Genetic polymorphism was initially defined by Ford (1940, cited in Harris 1980) as the occurrence together in the same locality of two or more discontinuous forms of a species in such proportions that the rarest of them cannot be maintained merely by recurrent mutation. The term was later defined as the occurrence of multiple alleles at a locus, where at least two alleles appear with frequencies greater than one percent (Bodmer and Cavalli-Sforza 1976). Alleles which occur

at frequencies of less than one percent are called rare variants (Meisler 1983). Polymorphism exists throughout the genome. It is generally presumed that the potential for polymorphism is much greater in non-functional DNA, since there are less or no constraints on such DNA. However, this does not preclude the existence of polymorphism in coding and other functional regions of DNA, most of which is neutral, and is not manifested in the form of selective advantage or genetic disease.

2.1.2 Estimates of Polymorphic Variation

The universal occurrence of genetic variation suggests that an individual is likely to be heterozygous at several gene loci. The analysis of protein variants was initially used to determine the magnitude of human variation, and it was estimated from a number of enzyme surveys that an individual would be heterozygous at six percent of enzyme-encoding loci (Harris and Hopkinson 1972). However, this figure did not reflect electrophoretically silent mutations, and Neel (1984) accordingly obtained a higher estimate of heterozygosity, in the region of 12-13%, by using several techniques for improved resolution and subtyping of electromorphs.

Direct analysis of DNA variation has recently become possible, and this has allowed for estimates of variation in both coding and non-coding DNA to be made. Data from the ß-globin gene cluster region indicated a variability of one site in 100 nucleotides (Jeffreys 1979), although this estimate has been revised to one in 200 nucleotides (Ewens *et al* 1981). Cooper and Schmidtke (1984) used randomly cloned DNA segments to derive a less biased estimate of single copy DNA sequence heterozygosity of two to three variable sites per 100 nucleotides. Expansion of the survey resulted in a revision of this figure to three to four variable sites per 100 nucleotides (Cooper *et al* 1985).

There are a number of potential inaccuracies in the estimation of DNA heterozygosity. Most studies have involved a relatively small number of base pairs within each region, and the number of DNA regions examined has not been extensive. Furthermore, there appears to be a non-uniform distribution of variation over the human genome, with increased variation in the non-coding introns or intergenic regions (Jeffreys 1979, Poncz *et al* 1983). The latter phenomenon provides the best explanation for the tenfold higher DNA heterozygosity value as compared to estimates from protein studies (Jeffreys 1979, Neel 1984).

2.1.3 Interpopulation Variation

The human species may be subdivided into many distinct subpopulations. The study of the distribution of genes in these populations, and the maintenance of change in gene frequencies, is termed population genetics. The major population groups or races have different gene pools, and allele frequencies at many loci vary widely among the populations. Some variants are restricted to members of a single group, and extensive variation can occur within each population.

Although the basis of genetic differences among races and their subpopulations is essentially mutation, two factors are implicated in the frequency differences and their maintenance. Natural selection is proposed to increase the frequency of alleles that are beneficial to an individual (Fisher 1930) and, in addition, alleles of little or no evolutionary value may increase in frequency if they are closely linked to evolutionarily important loci (Maynard Smith and Haigh 1974). Alternatively, shifts in gene frequency may be caused by random genetic drift (Kimura 1968), which would include scenarios such as bottlenecks and founder effect. It is likely that both these elements have played a role in the creation of observed population frequency differences (Harris 1980).

Another possible factor influencing interpopulation variation is heterozygote advantage. There are environmental situations in which heterozygotes for some diseases are at a selective advantage in comparison with both homozygous phenotypes. Even a slight heterozygote advantage can lead to an increase in frequency of a gene that is severely detrimental in homozygotes. The classic example of heterozygote advantage is resistance to malaria by heterozygotes for sickle-cell anaemia (Allison 1954). The sickle cell gene has reached its highest frequency in certain regions of west Africa, where heterozygotes are fitter than either homozygote. Similarly, the high frequency of the allele for Tay-Sachs disease in Ashkenazi Jews has been postulated to be due to historic epidemics of tuberculosis in industrial European cities (O'Brien 1991).

2.2 DNA POLYMORPHISM

2.2.1 Aetiology

The mutations giving rise to DNA polymorphism originate by either of two basic mechanisms - errors introduced during DNA replication, or base changes induced by mutagens. The overall mutation rate as a result of replication errors is remarkably low at 1×10^{-10} per bp per cell division. This is due to the proofreading

capability of the DNA replication machinery, and the recognition and removal of errors by a series of DNA repair enzymes. In contrast, DNA damage caused by spontaneous chemical processes such as depurination or deamination is often permanent. Such mutagenesis is induced by reaction with chemicals (including naturally-occurring substances) in the environment, and by exposure to ultraviolet or ionizing radiation.

2.2.1.1 Basepair Substitution

The nucleotide changes occurring in basepair substitution mutations may be classified as transversions or transitions. Tranversions involve the replacement of a purine by a pyrimidine (or vice versa), while transitions describe the substitution of one purine for the other (A <--> G) or one pyrimidine for the other (T <--> C). The nucleotide substitutions do not occur randomly, and transitions appear to be favoured over transversions (Vogel and Kopun 1977, Cooper and Krawczak 1990). This excess of transitions has been correlated with an increased frequency of the CpG doublet in polymorphic restriction enzyme sites (Barker 1984, Cooper and Schmidtke 1984) and human genetic disease mutations (Cooper and Youssoufian 1988, Cooper and Krawczak 1989, Cooper and Krawczak 1990), and is proposed to occur via methylation-mediated deamination. The major form of DNA modification in the

human genome involves methylation of cytosine residues to form 5-methylcytosine, especially when they are located 5′ to guanine (as occurs in the dinucleotide 5′-CG-3′). Spontaneous deamination of the 5-methylcytosine to thymidine in the CG doublet gives rise to a C -> T or G -> A transition.
The spectrum of point mutations occurring outside the CG dinucleotides is also non-random, at both mononucleotide and dinucleotide levels (Cooper and Krawczak 1990), and the higher rate of transitions is consistent with a physical model of mutation through nucleotide misincorporation as a result of transient misalignment of bases at the replication fork (Kunkel and Soni 1988).

Nucleotide substitutions that occur in genes can have a variety of effects. Missense mutations alter the codon of a coding region so that it is translated into a different amino acid, while nonsense mutations generate a stop codon and thus result in a truncated protein. The destruction of termination codons or RNA splicing sites by mutation will also affect the protein product. In addition, mutations occurring outside the gene itself, within control regions, may have an effect on transcription.

2.2.1.2 DNA Rearrangement

Rearrangement polymorphisms result from the insertion or deletion of one or more basepairs. Very large deletions may be detected cytogenetically, and would be associated with a microdeletion or contiguous gene syndrome if they resulted in the removal of several coding regions (Dobyns *et al* 1991). Smaller deletions of transcribed DNA might involve removal of the entire gene, or only a portion thereof. Frameshift mutations occur when the reading frame of translation is altered by the deletion or insertion of a few basepairs. In either case the alteration must not involve a factor of three bases, since the reading frame would then be unchanged but for the addition or removal of several codons.

Recombination between highly similiar or identical DNA sequences is a frequent cause of mutations involving deletion or duplication. Pairing of homologs (during meiosis) or sister chromatids (in mitosis after replication) may be misaligned when the sequences involved belong to a multigene family, or any other array of related or identical sequences, resulting in unequal crossing over. The evolution of several multigene families has been ascribed to such events (Slightom *et al* 1988, Schughart *et al* 1989, Groot *et al* 1990). Certain "hypervariable" polymorphisms,

characterized by a variable number of tandemly repeated DNA sequences and termed VNTRs, are believed to have originated from unequal sister chromatid exchange or possibly polymerase slippage, but not from unequal crossover of homologous chromosomes (Wolff *et al* 1988, Wolff *et al* 1989, Kasperczyk *et al* 1990). Slippage is the favoured mechanism for the generation of repeated sequences involving small numbers of nucleotides, including the dinucleotide CA repeat sequences (Morral *et al* 1991, Schlotterer and Tautz 1992). Some families of repeated sequences show less rigid organization (Ngo *et al* 1986), and models such as the "onion-skin" model for disproportionate replication (Stark and Wahl 1984) may possibly explain their origin.

2.2.2 Detection

At present the detection of polymorphism is possible using a large number of approaches. The choice depends on the nature of the polymorphism, the availability of information such as DNA sequence, and on technical considerations.

2.2.2.1 Restriction Fragment Length Polymorphisms

Restriction fragment length polymorphisms (RFLPs) are polymorphisms detectable as a change in DNA fragment length of a particular DNA sequence when it is cleaved

by a specific restriction endonuclease. They may result from DNA substitution mutations within a restriction endonuclease recognition site, leading to formation or loss of that site. Alternatively, DNA rearrangements (such as those described in section 2.2.1.2) which occur between two restriction enzyme sites would alter the length of that fragment. As expected from data on nucleotide substitution rates (Cooper and Krawczak 1990), enzymes which recognize sequences containing the dinucleotide CG (such as *Taq*I and *Msp*I) are effective in revealing RFLPs (Barker 1984, Cooper and Schmidtke 1984, Wijsman 1984, Schumm *et al* 1988).

Mutations resulting in RFLPs were originally detected by Southern blotting of the restriction endonuclease digested DNA, followed by hybridization with DNA sequences complementary to the variant DNA fragment(s). The development of the polymerase chain reaction (PCR) technique for amplification of specific DNA sequences has allowed for alternative approaches to RFLP detection in cases where the appropriate DNA sequence information is available. A small region of DNA can be amplified to reveal the smaller DNA rearrangements, or can be scored for the presence or absence of restriction sites after digestion with the relevant enzyme.

2.2.2.2 Allele-Specific Oligonucleotides

Basepair substitutions which do not result in mutations within restriction endonuclease sites may be detected by hybridization with an oligonucleotide probe specific for the DNA sequence. Sequence information pertaining to the mutated region is required, and oligonucleotides complementary to the sequence of each allele are hybridized in turn to dot or slot blotted DNA. The conditions of hybridization are so stringent that only DNA fragments with exact complementarity will bind to one another. Allele-specific oligonucleotides (ASOs) may also be used to directly detect RFLPs resulting from point mutations within restriction sites.

2.2.2.3 Allele-Specific Amplification

Allele-specific amplification, otherwise known as amplification refractory mutational system or ARMS (Newton *et al* 1989), involves the use of mutation-specific PCR primers to reveal the presence or absence of a particular DNA sequence. PCR primers are constructed in such a way that the position of the basepair substitution mutation is located at the 3′ end of one primer. Any mismatch at this position will prevent DNA extension from the primer, thus amplification of DNA using each of the allele-specific primers will indicate which alleles are present. This

system can be used for any nucleotide substitution, but obviously relies on the knowledge of DNA sequence for the construction of all the primers.

2.2.2.4 DNA Conformational Analyses

Denaturing gradient gel electrophoresis (DGGE) is a technique which may be used to detect genomic DNA polymorphisms that do not reveal RFLPs, without knowledge of sequence information (Noll and Collins 1987). The technique relies on the fact that single base pair differences between otherwise identical DNA molecules can result in altered melting behaviour under suitable conditions. DNA molecules formed by hybridization between genomic DNA and single-stranded probe DNA are held at the temperature of incipient denaturation, and electrophoresed through a polyacrylamide gel containing a gradient of increasing concentrations of urea and formamide. The DNA fragments eventually encounter a concentration of denaturants where an early melting domain, caused by any single mismatch, becomes unstable and melts, producing molecules that are partly disordered and which migrate slowly. The position in the gradient at which migration becomes altered is unique, and characteristic for each individual fragment.

Basepair substitutions can also be detected as single strand conformational polymorphisms (SSCPs). This method is based on the fact that mutations in amplified DNA fragments can be detected as a mobility shift of single stranded DNA electrophoretically separated on non-denaturing polyacrylamide gels (Orita *et al* 1989). However, the effectiveness of this technique in revealing single base substitutions occurring within stable secondary structures has not yet been proven (White *et al* 1992).

An alternative to the DDGE and SSCP analyses for detecting single basepair substitution is the generation of heteroduplex polymorphisms (White *et al* 1992). This technique takes advantage of the formation of heteroduplexes in the PCR products between different alleles from heterozygous individuals. The heteroduplexes can be detected on polyacrylamide gels because they migrate slower than their corresponding homoduplexes. This technique does not rely on secondary structure formation and should thus not be influenced by the position of the point mutations, but it appears that some point mutations are better detected by the SSCP method (White *et al* 1992). It has therefore been suggested that a combination of techniques would facilitate detection of the majority of point mutations (White *et al* 1992).

2.2.2.5 Cleavage

These methods of detection rely on the cleavage of mismatched basepairs ocurring in heteroduplexes. Mismatches at any position in the DNA region studied will be detected due to specific cleavage at that site. RNAse A will cleave RNA:DNA hybrid duplexes (Myers *et al* 1985), while chemical methods are suitable for DNA:DNA duplexes (Cotton *et al* 1988). The latter also have the advantage that specific chemicals cleave certain base mismatches, generating information about the base composition of the mismatch (Cotton *et al* 1988). However, several laborious manipulations are required for each sample, and these methods are thus not practical for screening large numbers of individuals.

2.2.2.6 Sequencing

The determination of the exact sequence of any particular region of DNA should reveal basepair substitutions, as well as small rearrangements. However, the DNA sequencing technique remains a laborious and time-consuming technique, despite the great advances made with the application of PCR (Ruano and Kidd 1991). Its relevance is thus generally in the detection of disease mutation (Cooper and Schmidtke

1989), or in studying DNA regions which mutate rapidly, and thus generate a large amount of information for the area sequenced (Vigilant *et al* 1991).

2.3 APPLICATIONS OF DNA POLYMORPHISM

The existence of DNA polymorphism in all individuals is not merely of academic interest. The mechanism and rate of origin of DNA variation may be of evolutionary significance, but polymorphisms are principally of value for their use as genetic markers throughout the genome. As such, they have been applied in several fields, including medical and population genetics.

2.3.1 Gene Mapping by Linkage Analysis

Genetic linkage analysis may be described as the study of the segregation of inherited traits in pedigrees. It relies on the fact that closely linked loci on a chromosome will commonly be inherited together, since the distance between them will reduce the chance of crossover. Polymorphisms can be used as genetic markers to map disease or other traits caused wholly or partially by a major locus segregating in a pedigree, without knowledge of the biochemical nature of the trait, or of the alterations in the DNA responsible for the trait (Cooper and Schmidtke 1989). No specific gene isolation is required, and the polymorphic markers can

be random sequences functionally unrelated and physically distant from the DNA encoding the locus of interest. It has been proposed that the establishment of a set of well-spaced, highly polymorphic genetic markers covering the entire human genome will allow better definition and substantiation of models of inheritance for the many familial traits which have been refractory to simple genetic analysis in humans (Botstein *et al* 1980). Attempts to reach this goal have been facilitated by the establishment of the CEPH (Centre d'etude de Polymorphisme Humaine) program, which involves the global distribution of DNA from numerous extended pedigrees for linkage mapping of DNA polymorphisms (Dausset *et al* 1990). Linkage maps for several human chromosomes have been made available as a result of such concerted collaborative efforts (Lathrop *et al* 1987, Blanche *et al* 1991, Dracopoli *et al* 1991, Weiffenbach *et al* 1991).

2.3.2 Prenatal and Preclinical Diagnosis

Polymorphic markers which are closely linked to a genetic disease may be used to trace the inheritance of the defect in a particular family. This is of particular interest when the gene responsible for the disease has not been cloned or characterized (Gusella *et al* 1983, Janssen *et al* 1990, Mulley *et al* 1990), or may be extremely heterogeneous with respect to the

molecular basis of the defect (Tuddenham *et al* 1991). The ability to recognize the abnormal gene in a family suffering from a genetic disorder allows for carrier detection and prenatal diagnosis, and for preclinical diagnosis in the case of late onset disorders (Cooper and Schmidtke 1989). Diagnosis using linked markers is subject to an error rate dependant on the probability of crossover between the linked marker and the genetic defect. However, as the search for more closely linked markers intensifies after the discovery of disease gene loci, this factor becomes negligible (Gilliam *et al* 1987, Smith *et al* 1988, Richards *et al* 1991, Skraastad *et al* 1991). This approach is also limited by the need for informative polymorphic markers in the parents, and information from affected individuals and other family members, in order to establish the linkage phase between the marker allele and the disease gene.

2.3.3 Human Identification

Polymorphic markers, especially the hypervariable regions detected by VNTRs and CA repeats (see section 2.2.1.2), have been applied very successfully to problems of human identification encountered in paternity testing (Jeffreys *et al* 1985b, Jeffreys *et al* 1991), and forensic investigations (Gill *et al* 1985, Wong *et al* 1987). Some markers reveal single loci (Nakamura *et al* 1987, Wong *et al* 1987), whilst others

can be used under conditions of low stringency hybridization as multi-locus probes (Jeffreys *et al* 1985a). The extreme interindividual variation detected by these hypervariable regions suggests that any two individuals are extremely unlikely to be indistinguishable. The cumulative probability of false inclusion of a non-father in a paternity dispute is reported to be 1×10^{-7}, using data from 6 minisatellite probes (Wong *et al* 1987), while DNA fingerprinting studies suggest that the probability of chance identity is very low, at considerably less than 1×10^{-7} per probe (Jeffreys *et al* 1991).

2.3.4 Anthropological Studies

Anthropology is defined as the study of man, his origins, institutions, religious beliefs and social relationships. The application of interpopulation genetic variation to the fields of population and evolutionary genetics has greatly expanded anthropological research. The analysis of allele frequencies in different populations permits the construction of genetic distance maps between the populations (Cavalli-Sforza 1974, Cavalli-Sforza *et al* 1986), and allows genetic relationships between populations to be determined. The occurrence of gene flow, the slow diffusion of genes across a racial barrier (Cavalli-Sforza and Bodmer 1971) can also be

detected by noting changes in the gene frequencies of the original parent population(s). Numerous studies in population genetics have been carried out, using both serogenetic (Piazza *et al* 1988, Barbujani *et al* 1990, Bertranpetit and Cavalli-Sforza 1991) and nuclear and mitochondrial DNA polymorphisms (Bowcock *et al* 1987, Summers *et al* 1987, De Benedictus *et al* 1989, Semino *et al* 1989, Fey *et al* 1990, Morris *et al* 1991).

Evolutionary genetic research is based on an extension of population studies, with determination of the genetic distinction between the major human racial groups. At present there are two competing models for recent human evolution, popularly known as the "Out of Africa" and the multiregional continuum models. The "Out of Africa" hypothesis proposes that modern humans evolved in Africa, and subsequently dispersed throughout Europe and Asia, replacing all other hominids (Cann *et al* 1987, Vigilant *et al* 1991). As an expansion of this theory, distinction at the genetic level has been implicated in the lack of interbreeding between the archaic humans occupying Europe and Asia, and the invading African ancestor (Stringer and Andrews 1988). The multiregional continuum model postulates that modern humans evolved gradually and simultaneously from local forms of *Homo erectus* which had themselves

spread from Africa to Europe and Asia (Wolpoff 1989); the unity of the species is presumed to have been maintained by gene exchange between the isolates.

Research in the field of human genetic variation has ushered in a new era in evolutionary research, and various molecular studies have been undertaken in an attempt to resolve uncertainties resulting from different interpretations of paleoanthropological data. These studies range from the pioneering serogenetic (Cavalli-Sforza and Edwards 1967) and HLA (Piazza *et al* 1975) studies to DNA-based restriction fragment length polymorphism (RFLP) analysis of ß-globin markers (Wainscoat *et al* 1986), mitochondrial (mt) DNA (Cann *et al* 1987) and nuclear gene markers (Bowcock *et al* 1987, Cavalli-Sforza *et al* 1988, Bowcock *et al* 1991). More recently, a comprehensive study involving DNA sequencing of the mtDNA D-loop region has been presented (Vigilant *et al* 1991). The molecular approach to evolutionary studies is based on the assumption that if modern man arose in Africa, then African and non-African humans would be most genetically distinct, and African populations would exhibit more within-group genetic variation than the non-African populations which stemmed from an African migration. With little exception, the molecular data confirm the African/non-African split, in support of the "Out of Africa" model.

2.4 Y CHROMOSOME POLYMORPHISM

2.4.1 Paucity of Y Chromosome Polymorphism

Several studies have been undertaken to search for conventional RFLPs on the Y chromosome (Jakubiczka *et al* 1989, Malaspina *et al* 1990). Despite these systematic searches, a paucity of polymorphism exists, with reported levels of polymorphism as low as <1/18000 nucleotides (Jakubiczka *et al* 1989) and <1/46515 nucleotides (Malaspina *et al* 1990). This scarcity of polymorphism is unexpected in view of the origin of the Y (Charlesworth 1991). The plethora of pseudogenes, rearranged genetic material or retrotransposed sequences originating from the X or autosomes should, in all likelihood, have resulted in an abundance of polymorphism on this haploid chromosome unrestrained by genetic recombination. However, both Jakubiczka *et al* (1989) and Malaspina *et al* (1990) intimate that the role of genetic recombination in maintaining genetic variation should not be underestimated. Furthermore, Goodfellow *et al* (1985) suggest that the human Y chromosome is of recent creation, in which case there may have been insufficient time for the accumulation of significant polymorphism.

2.4.2 Applications of Y Polymorphism

Many researchers have recognized the potential of Y chromosome population and evolutionary studies for providing a record of male-specific gene flow and human evolution. Preceding studies in this field have used autosomally-inherited nuclear DNA markers and maternally-inherited mitochondrial (mt) DNA polymorphisms. The latter have proved to be especially informative because of the approximately five to ten times higher rate of mutation of mtDNA as compared with nuclear DNA. The use of Y-specific markers, which represent the paternal line of inheritance, should complement or expand studies using mtDNA or nuclear polymorphisms. The paucity of Y chromosome polymorphism (see section 2.4.1) has, however, hindered research in this area although several recent studies have intimated the value of such research (Ellis *et al* 1990a, Torroni *et al* 1990, Spurdle and Jenkins 1992).

There is also a restricted purpose for Y-specific polymorphisms in forensic identification of males, and paternity testing of male children, but reports of such studies have been limited (Guerin *et al* 1988a).

2.4.3 Documented Y-Linked Polymorphisms

Much of the research on Y chromosome polymorphisms has concentrated on the complex Y-specific *Taq*I RFLP revealed by probes p49a and p49f. The polymorphism found at this locus presents itself as multiple haplotypes (Ngo *et al* 1986), and has thus shown great promise for studies in population genetics. Other RFLPs include a *Taq*I point mutation revealed by p21A1 within a repeated sequence (Jakubiczka *et al* 1989), a deletion/insertion rearrangement polymorphism revealed by p12f2 (Casanova *et al* 1985), and a rearrangement polymorphism, believed to result from a duplication, detected by pDP31 (Page *et al* 1982, Bowcock 1984). None of these polymorphisms has previously been employed in extensive population studies.

The Y *Alu* polymorphism (YAP) is another Y-specific rearrangement polymorphism, and results from the insertion of a 300bp *Alu* element (M.Hammer, personal communication). Limited population screening has shown the insertion to be common in Africans, but also present in non-Africans (M.Hammer, personal communication). Of particular interest is the Y-linked XY275 *Msp*I polymorphism associated with the pseudoautosomal boundary (Ellis *et al* 1990a, Ellis 1991b). Theories concerning the origin and expansion of *Homo sapiens* have been constructed from the frequencies

of the Y-associated alleles in different populations (Ellis *et al* 1990b, Spurdle *et al*, in press). Long-range Y-specific polymorphisms, detected using pulsed-field gel electrophoresis have also been described (Oakey and Tyler-Smith 1990). Results from this study have evolutionary implications, but the system seems inappropriate for smaller-scale population studies.

2.5 ANTHROPOLOGICAL STUDIES IN SOUTHERN AFRICA

Reconstruction of the cultural history of Africa has been possible through the use of several sources of evidence (Murdock 1959). One such source is written records, which are invaluable for north African history, but rarely yield information of any considerable time depth elsewhere on the continent. Other approaches include the direct testimony of materials excavated in archaelogical sites, linguistic relationships, evidence from the distribution of cultivated plants, social organizations (such as the matrilineate and patrilineate forms), and ethnographic distributional analysis. More recently, genetic markers have been studied in African populations in an attempt to determine the affinities between them (Mourant *et al* 1976, Nurse *et al* 1985, Excoffier *et al* 1987).

Africa has been inhabited by representatives of several major racial groups, namely the Khoisan, Negroid (including the Negroid-like Pygmies), Caucasoid and Mongoloid peoples. The latter were latecomers to Africa, and survived only in Madagascar, where they were usually heavily admixed with Negroids (Murdock 1959). At the close of the Paleolithic period, the Khoisan, Negroid and Caucasoid races shared the continent, but subsequent historical movements practically eliminated the Khoisan and Pygmy groups, and even the latecoming Mongoloids. Most territorial gains were accrued to the Negroids, in the Saharan and sub-Saharan regions (Murdock 1959, Excoffier *et al* 1987).

Anthropological studies in southern Africa have concentrated on the Khoisan and the Bantu-speaking Negroid peoples. The Khoisan include the hunter-gatherer San and the pastoralist Khoi, and are believed to have originated in either east Africa (Cooke 1965, Jenkins 1972), or in the Kalahari region (Elphick 1977). Their existence in southern Africa predates the arrival of the Negroids in this area. There are several models, based on linguistic and archaelogical evidence, describing the migration routes of the Bantu-speaking Negroids. However, it is generally accepted that the proto-Bantu-speakers lived in the Middle-Benue valley between Nigeria and Cameroon

(at about 3000 to 5000 years ago), and subsequently spread into sub-Saharan Africa (Oliver and Fagan 1975, Phillipson 1977). The existence of major divergences between the languages of the eastern and western sides of the continent, and their perpetuation in the Bantu languages of southern Africa, has led to the suggestion that an early split into eastern and western Bantu groups must have occurred (Ehret 1973, Phillipson 1977). This theory is supported by archaelogical evidence (Huffman 1982), and a close similarity of the eastern languages to each other implies a relatively recent and rapid expansion in comparison with the western languages (Phillipson 1977, Vansina 1984).

Genetic evidence has confirmed that the Bantu-speaking Negroids are a homogeneous group (Nurse *et al* 1985), who have strong genetic affinities with the Negroids of west and central Africa (Jenkins *et al* 1970, Jenkins 1982). In addition, the Khoisan have been shown to have affinity with southern African Bantu-speaking Negroids, and other sub-Saharan Africans (Zoutendyk *et al* 1955, Nurse *et al* 1985). Despite this, it is accepted that although the Khoisan people belong to the same gene pool as other sub-Saharans, certain morphological and genetic differences point to a lengthy period of differentiation in relative isolation (Tobias 1974, Steinberg *et al* 1975, Nurse *et al* 1985). It has been proposed that, before the onset of the later Stone age,

proto-negriform Africans, with a basically Khoisan-like genome, split into two major branches, which evolved into the Khoisan and Negroids (Tobias 1974, Excoffier *et al* 1987).

Genetic markers have also been used in localized population studies to determine recent gene flow and admixture between population groups. Research based on serogenetic markers indicate that there is considerable admixture of San genes in some of the Bantu-speaking Negroid groups (Jenkins *et al* 1970), and that the enslavement of the Khoisan-speaking Dama group by the Khoi Nama did not involve sexual subjection (Nurse *et al* 1985). Furthermore, the hybrid populations of mixed ancestry are placed genetically between the Negroid and Caucasoid groups (Nurse *et al* 1985).

CHAPTER THREE

SUBJECTS AND METHODS

3.1 SUBJECTS

The populations included in this study may be broadly classified into four groups: Caucasoid, hybrid populations of mixed ancestry, Bantu-speaking Negroid, Khoisan-speaking Negroid, Khoisan and Pygmy.

Pygmy DNA samples were a gift from J.Rogers, and represent unrelated males originating from two areas in Central Africa. The remaining subjects were unrelated males belonging to various southern African populations. Their appropriate geographical areas of origin are shown in Figure 3.1. Peripheral blood samples were collected from consenting individuals contacted at the S.A.I.M.R., at blood donation clinics, and during specific field trips to outlying areas. In the case of Caucasoid and Negroid families used to confirm patterns of Y-specific Mendelian inheritance, blood was collected by prior arrangement.

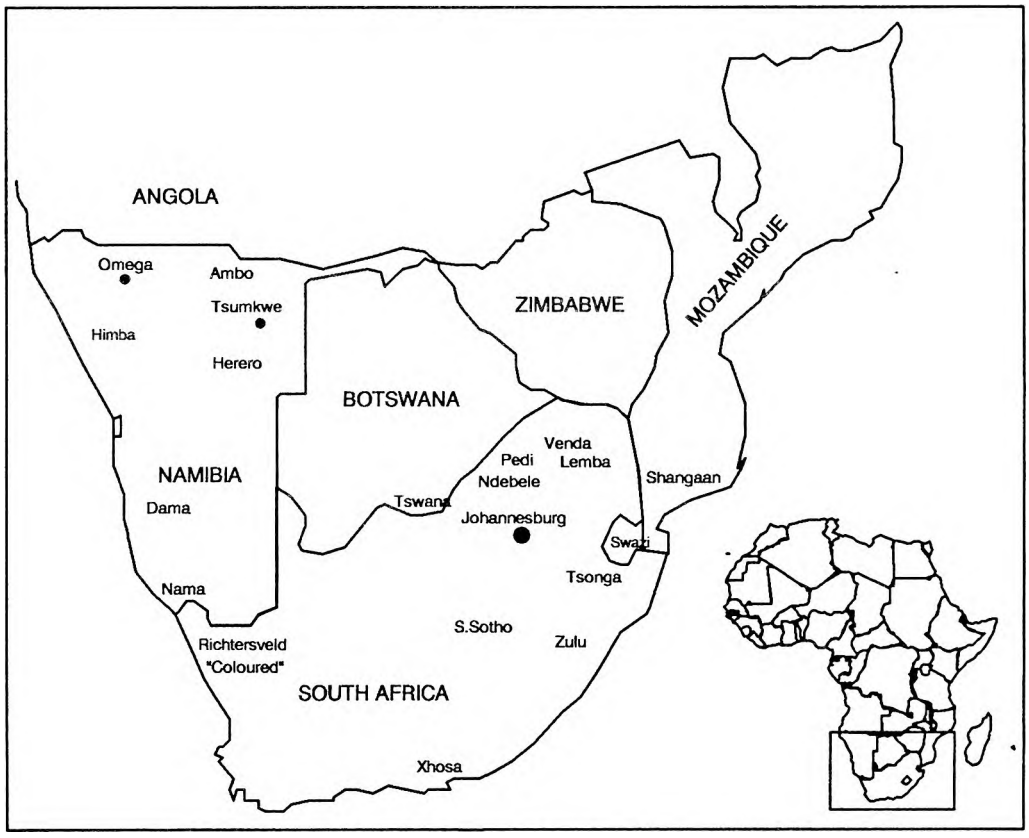

Figure 3.1. Geographical location of the southern African populations studied.

3.1.1 The Caucasoid Population

The South African Caucasoid population was comprised of three groups. The S.A. European group was made up of individuals originating from western Europe, of either historical or recent immigration. The sample included English- and Afrikaans-speaking Caucasoids. The S.A. Asiatic Indian group consists of individuals belonging to the Muslim and Hindu religious groups, originating from at least two migrations from India (Bernstein *et al* 1986). The Jewish sample is comprised mainly of Ashkenazi Jews from eastern Europe.

3.1.2 Hybrid Populations

Two populations of mixed ancestry are represented. The Johannesburg "Coloured" group was collected in the Johannesburg area, and results from the admixture of European Caucasoid, Khoisan, Malay and Bantu-speaking Negroid people (Nurse *et al* 1985). The Richtersveld "Coloured" group has resulted from admixture between European Caucasoid trekboers (farmers) who moved into the northern Cape area in the 18th century, and indigenous Nama women. They have remained relatively isolated over the ensuing years (Nurse *et al* 1985).

3.1.3 Bantu-speaking Negroids

The African Negroids have brown to brownish-black skin colour, and are of medium height. They have lightly curled brownish-black hair, and broad noses. The grouping of the Bantu-speaking Negroids as a cohesive entity is based on a linguistic classification. Under this broad category of Bantu-speaking Negroids there are a number of different chiefdoms, each speaking a different Bantu language. Certain chiefdoms may also be grouped together due to similarities in language and culture. The Zulu, Xhosa, Ndebele and Swazi chiefdoms are classified as the Nguni, whereas the southern Sotho, Pedi (northern Sotho) and Tswana are grouped as the Sotho/Tswana. Likewise the Herero-speaking group includes the Herero and Himba chiefdoms. The Tsonga sample includes individuals from the Tsonga and Shangaan chiefdoms. The migration of the Bantu-speaking Negroids from their postulated area of origin in west-central Africa is believed to have followed at least two routes: a general south-bound course (the Eastern Bantu) and a south-westerly route across central Africa toward the western parts of the continent (the Western Bantu). The migration routes are based on linguistic evidence (Greenberg 1972), but are also supported by archaelogical evidence (Huffman 1982). The Herero and Ambo are the only representatives of Western Bantu populations in the present study.

3.1.4 Enigmatic Bantu-Speakers

The Lemba are Venda-speakers considered by ethnographers to be of alien origin (Van Warmelo 1974). Many factors distinguish them from the other Bantu-speakers (Van Warmelo 1974): many Lemba have a distinctive appearance - angular features with a prominent hooked nose; the men used to wear a long cotton upper garment (*khanzu*) as found along the east coast of Africa; amongst themselves they spoke a language not understood by their hosts in southern Africa; they were metal-workers and proficient traders; marriage was strictly endogamous; certain foods are forbidden e.g. pork, certain other animals and the flesh of cattle not kosher-killed according to their law; circumcision is practised; unintelligible prayers are recited and responded to at certain ritual ceremonies. Although the Lemba themselves claim to belong to one of the lost tribes of Israel, certain facts about them suggest that they are descendants of Semitic traders (presumably Arabs) from the East Coast. One such fact was the observation that their ritual prayers may represent mangled *suras* from the Koran (Van Warmelo 1974).

3.1.5 Khoisan-Speaking Negroids

There are several Negroid populations in southern Africa which speak a Khoisan language (Nurse *et al* 1985), and whose origins are a source of perplexity. The Dama population of Namibia is one such group. Individuals have typical Negroid features (Malan 1980), but culturally and linguistically do not reveal any similarities to the Bantu-speaking Negroids. They speak the language of the Khoi Nama, and are hunter-gatherers. It has been suggested that immigrant Negroids may have lost their cattle to infection, and unable to support themselves by systematic agriculture in the harsh climate, they penetrated into the desert, learning skills and language from San or Khoi hunter-gatherers (Nurse *et al* 1985). This theory is not favoured by linguists, and an alternative hypothesis is that ancestors of the Dama were historically enslaved by the Khoi, and modern Dama people represent the descendants of runaway slaves who escaped into the desert (Nurse *et al* 1985).

3.1.6 The Khoisan Population

The Khoisan group is composed of the Khoi (formerly referred to as "Hottentots") and the San (formerly "Bushmen"). These medium-statured people have a yellow skin colour, flat triangular face, and steatopygia (Malan 1980), and they speak languages composed of clicks and other guttural sounds (Nurse and Jenkins 1977). The San and Khoi have largely been differentiated by cultural differences. The San exist as hunter-gatherers, while the Khoi are pastoralists. The !Kung Sekele San group were collected at the Omega military camp in northern Namibia originate from southern Angola, while the Tsumkwe San are resident in the Tsumkwe region of the Kalahari (see Figure 3.1). The Khoi are represented by the Nama people from Namibia.

3.1.7 The Pygmy Population

The Pygmy are tropical forest dwellers characterized by short stature, brownish-black skin colour, and peppercorn hair. They are traditionally hunter-gatherers, although they are progressively shifting to agricultural practices, and in some cases to partial dependence on settled farmers with whom they exchange game and labour for pottery, tools and certain foods. They are believed to have common ancestry with the Negroids, and indeed show some genetic similarities to them (Nurse *et al* 1985). The Pygmy sample employed in this study was drawn from two areas in Central Africa (Bowcock *et al* 1987). It includes the Biaka (or Aka) Pygmies from Bagandu (also referred to as Babinga) in the south-western area of the Central African Republic, and the Mbuti Pygmies from the Ituri forest in north-east Zaire. The Biaka sample may represent hybridization between African Pygmy and Central African farmers (Bowcock *et al* 1987).

3.2 GENERAL APPROACHES

3.2.1 Screening for Y-Specific Polymorphism

DNA from nine Caucasoid and nine Negroid father and son pairs was digested with each of the following range of enzymes: *Bam*HI, *Bgl*I, *Bgl*II, *Eco*RI, *Hae*III, *Hin*dIII, *Msp*I, *Pst*I, *Pvu*II, *Rsa*I, *Taq*I, *Sst*I and *Xba*I. Fragments were separated by electrophoresis, and transferred to nylon membranes by Southern blotting. Digested DNA was screened for RFLPs by sequential hybridization with the Y-specific DNA probes pDP31, pDP61, pDP105, pDP132, pDP97, pDP1007, pY431-HinfA, and p49a/f (see Table 3.1). Potential Y-specific RFLPs were verified by examination of results in both father and son, to ensure the existence of Mendelian Y-specific inheritance patterns. Potential X-specific or autosomal RFLPs also observed during the screening were validated by extension of the analysis to other family members.

3.2.2 Screening for known Y-specific RFLPs

DNA from individuals belonging to various population groups was digested with the appropriate restriction enzyme, and transferred to nylon membranes after electrophoretic separation. In the case of *Taq*I digests, conditions of electrophoresis were perfected for detection of the 49a/f polymorphism. Blots were screened for documented polymorphism with appropriate Y-specific DNA probes.

3.2.3 Screening for the Pseudoautosomal XY275 Polymorphism

*Msp*I blots with DNA from males and females belonging to several population groups had been prepared for a preceding study in the laboratory. These blots were screened for the pseudoautosomal XY275 polymorphism. PCR analysis of the XY275 locus was used to distinguish between the X- and Y-associated alleles of heterozygote males, and to verify the presence of Y-associated *low* alleles in homozygote males, as identified by hybridization methods. Additional male samples were screened *de novo* by PCR to increase the sample size of certain populations.

3.3 METHODS

Details of media and solutions are listed in Appendix One. All centrifugations took place in a Beckman benchtop centrifuge model TJ-6, unless otherwise stated.

3.3.1 Preparation of DNA Probes

All DNA probes employed in this study were inserted in plasmid vectors. Details of the DNA probes are displayed in Table 3.1.

TABLE 3.1

Details of DNA Probes Employed in the Present Study.

Probe	Locus	Localization	Insert(Kb)	Reference
pDP31	DXYS1Y	Yp11	4.5 *EcoRI*	Page *et al* 1982, 1984
pDP61	DXYS8Y	Yp	1.0 *EcoRI/TaqI*	Bernstein *et al* 1987
pDP105	DYZ4	Yp	5.5 *HindIII*	Bernstein *et al* 1987
pDP132	DXYS23Y	Yp	3.5 *HindIII*	Bernstein *et al* 1987
pDP97	DYZ3	Ycen	5.3 *EcoRI*	Bernstein *et al* 1987, Wolfe *et al* 1985
pDP1007	ZFY	Yp11.3	1.3 *HindIII*	Page *et al* 1987
pY431-HinfA	DYZ2	Yq12	0.8 *PstI*	Bernstein *et al* 1987
pYAP-2.8	YAP	Yq11	2.8 *EcoRV*	M.Hammer, personal communication
Hf0.2	PABY	Yp11.3	0.19 *SmaI*	Ellis *et al* 1989
49a	DYS1	Yq11.1	0.9 *XbaI/BamHI*	Ngo *et al* 1986
49f	DYS1	Yq11.1	2.8 *EcoRI*	Ngo *et al* 1986

3.3.1.1 Plasmid Transformation

Certain DNA probes were obtained as DNA preparations, and not as transformed bacterial stocks. The DNA was transformed into an *E. coli* HB101 bacterial host as follows: HB101 was streaked onto a Luria agar plate, and incubated at 37°C overnight. Two ml of LB was inoculated with a single colony and grown overnight at 37°C. Then 100ml LB was inoculated with 1ml of this overnight culture, and incubated with vigorous shaking at 37°C until the culture had an OD_{550} of 0.5. The bacterial culture was chilled on ice for 10 min, and centrifuged at 3000rpm for 15 min to harvest the cells. The pellet was resuspended in 10ml icecold 50mM $CaCl_2$, 10mM Tris pH 8.0, and centrifuged once again at 3000 rpm for 15 min. The final pellet was resuspended in 1ml of the $CaCl_2$/Tris solution, aliquoted into 0.1ml amounts, and kept on ice until required. An aliquot of competent cells was mixed with 1-2μl of plasmid DNA requiring transformation, and left on ice for 30 min. The mixture was heat-shocked in a 45°C waterbath for 2 min, and returned to ice for 10 min. The transformation mix was then added to 3ml LB, and incubated with shaking at 37°C for 1-3 hrs. Transformant cells were selected by plating out 0.1ml of the broth culture onto the appropriate antibiotic-containing LB plates, and incubated overnight at 37°C.

3.3.1.2 Plasmid DNA Extraction

Methods which isolate plasmid DNA from bacterial cells capitalize on differences in configuration and size between the DNA of the plasmid and the *E. coli* chromosome. Plasmids are small and remain in a covalently closed circular form, while the chromosomal DNA is broken down into linear molecules during the extraction procedure, and will be pelleted down with the cell remnants.

The method for extraction was as follows: Bacterial cells were grown overnight in 100ml LB medium with the appropriate antibiotic (25μg/ml for ampicillin, and 12.5μg/ml for tetracycline). The bacterial culture was harvested by centrifugation at 5000rpm for 20 min, and the pellet resuspended in 6ml lysozyme solution. The mixture was placed on ice for 20 min to allow lysis of the cell walls to occur. Then 12ml 0.2M NaOH, 1% SDS was added, and mixed thoroughly by inversion. Lysis of the cell was allowed to proceed for 10 min, before the addition of 7.5ml 3M NaAcetate pH 4.6. This step aids the collection of debris, including protein and chromosomal DNA. After 20 min, the lysed bacterial culture was centrifuged in a Sorvall RC5C at 15000 rpm for 15 min to pellet the debris. The supernatant was transferred to a fresh tube, and 5μl 10mg/ml RNase A added. The tube was incubated at 37°C for 20 min,

ensuring breakdown of contaminating RNA. Contaminating proteins were removed by extraction twice with phenol:chloroform (1:1) and once with chloroform. (Chloroform in the case of protein extraction by organic solvents means 24:1 chloroform:isoamylalcohol). DNA was precipitated by the addition of 2½ volumes of ice-cold absolute ethanol, and storage at -20°C overnight or at -70°C for 20 min. The precipitated DNA was collected by centrifugation at 3000rpm for 20 min, washed in 70% ethanol to remove precipitated salts, and resuspended in 1ml TE pH 8.

The final step of purification was CsCl gradient centrifugation to separate plasmid DNA from any remaining bacterial DNA contaminants. The dissolved plasmid DNA pellet was mixed with 4g CsCl in 3ml TE, and 320µl 10mg/ml EtBr. The mixture was carefully transferred to a Beckman Quickseal polyallomer tube (13x51mm) suitable for centrifugation in a Beckman VT 65.1 rotor, and spun for at least 16 hrs at 45000 rpm in a Beckman L8-55 ultracentrifuge. The tubes were scanned under uv light, and two bands were visualized. The lower band, representing closed circular plasmid DNA, was removed through the side of the tube using a wide bore syringe. The upper band of nicked, linearized plasmid DNA and linear bacterial DNA was discarded. EtBr was removed from the CsCl plasmid solution by addition of an equal volume of isoamylalcohol,

inversion, and removal of the top layer. This step was repeated until the pink colour in the lower phase had disappeared. The plasmid DNA was dialyzed against TE pH 8 to remove the CsCl salt, using a Millipore 0.25μm type VS filter.

3.3.1.3 Preparation of Probe DNA Insert

In most instances the purified DNA insert was prepared from plasmid probes to improve hybridization results. The DNA insert was released from the probe by digestion with the appropriate enzyme(s) (see Table 3.1), and purified by phenol-freeze extraction.

Digestion of the plasmid DNA to release the human DNA insert followed manufacturer's specifications. The completion of digestion was checked by agarose gel electrophoresis of an aliquot of the digest. In the case of double digests, restrictions were done sequentially to monitor the digestion of each enzyme. The digested probe fragments were separated by electrophoresis in 1% low melting point agarose (Seakem), and the appropriate DNA band was cut out with a razor blade. A gel band of approximately 0.3ml was placed in an Eppendorf tube with an equal volume of phenol. The tube was vortexed well, frozen rapidly in an ethanol/ice/salt bath for 5 min in a microfuge. This procedure was repeated twice. The aqueous phase was

removed, and an equal volume of chloroform added. After centrifugation for 5 min in a microfuge, the aqueous phase was removed. The addition of 1/10 the volume 3M NaAcetate pH 4.6 and 2 volumes absolute ethanol followed. The DNA was precipitated at -70°C for 20 min or at -20°C for 2hrs to overnight. The solution was spun for 15 min in a microfuge to precipitate the DNA, the pellet was washed in 70% ethanol, and then resuspended in 100-250μl TE.

3.3.2 Genomic DNA Extraction and Digestion

Blood samples for DNA extraction were collected in vacutainer tubes with the anticoagulant acid citrate dextrose (ACD) or ethylenediaminetetra-acetic acid (EDTA). Samples were centrifuged to separate the red blood cells, buffy coat (containing white blood cells) and plasma. The packed cells, or buffy coat alone, were stored at -20°C until required for DNA extraction.

Total genomic DNA was extracted from stored samples by the methods of Sykes (1983). The sample was thawed, and mixed with an equal volume of 0.2% Triton X, 0.9% NaCl to lyse the red blood cells. The tube was centrifuged at 2700rpm for 15 min. The supernatant was discarded and the soft pellet resuspended in 30ml of the same saline solution by vigorous shaking. The centrifugation and washing steps were repeated until the pellet was no

longer red. The final pellet was dispersed with a flame-sterilized glass rod, and lysing buffer (7M urea, 0.3M NaCl, 10mM EDTA, 10mM Tris pH7.5) was added drop by drop to a total of 10ml, while continually working the pellet into solution. Then 2ml 10% SDS was added, and the tube incubated for 30 min at 37°C. The proteins were removed by extraction twice with 5ml chloroform and 10ml phenol, and once with 15ml chloroform. The DNA was precipitated by addition of 2-2½ volumes of icecold ethanol, and collected by spooling with a flame-sterilized glass rod, or by centrifugation in instances where DNA yield was low. DNA was dissolved in 1ml TE.

Restriction digestion followed followed manufacturer's specifications. Digests contained 5-10μg of human DNA, 4μl of the appropriate 10x digestion buffer (supplied by the enzyme manufacturer), 25 units of restriction endonuclease, 0.5μl 0.1M spermidine trihydrochloride, and distilled water to make the incubation mixture up to a final volume of 40μl. The mixture was incubated at the recommended temperature (65°C for *Taq*I, 37°C for all other enzymes) for 4hrs to overnight. An aliquot of 4μl was removed from each digest, and electrophoresed to determine if digestion was complete. Digested DNA samples were represented as an even smear on the gel. After digestion, 5μl loading buffer (0.1% bromophenol

blue, 0.1M EDTA, 50% glycerol) was added. This solution stops the reaction, and acts as a dense loading medium and tracking dye.

3.3.3 Agarose Gel Electrophoresis and Southern Blotting.

The DNA fragments resulting from restriction endonuclease digestion of high molecular weight DNA are size fractionated by electrophoresis through an agarose gel matrix, and transferred *in situ* to a solid support such as a nylon membrane by the method of Southern blotting (Southern 1975).

3.3.3.1 Agarose Gel Electrophoresis

Horizontal slab gels were prepared for gel electrophoresis. Agarose powder (Seakem HGT) was added to trisborate buffer (TBE) at a concentration of 0.8%, and dissolved by boiling. EtBr was added to a final concentration of 0.5μg/ml after the solution had cooled slightly. The gel mixture was poured into a gel mould (18x20cm) with a gel comb in place to form loading wells, and allowed to set at room temperature. Prior to electrophoresis, the gel was submerged in the buffer, and samples (containing loading dye) were loaded into individual gel slots. DNA markers (see Appendix One) were loaded into one or two slots for sizing of DNA

fragments. Samples were electrophoresed at 40V overnight. The conditions required for adequate separation of fragments differed according to the polymorphic system analyzed. Digests for the detection of polymorphism by random screening were electrophoresed until the blue dye had migrated to the end of the gel. Digests for the 49a RFLPs were separated as follows: *Taq*I digests were separated by electrophoresis until the bromophenol blue dye front had migrated 20-26cm from the origin. *Sst*I digests were separated until the 2kb fragment of lambda digested with *Hin*dII had migrated 18cm from the origin, while *Bgl*II, *Hin*dIII, *Pvu*II and *Pst*I digests were electrophoresed until the 3.53kb fragment of lambda digested with *Hin*dII and *Eco*RI had migrated 16cm, 15cm, 17.5cm and 18.5cm respectively, from the origin. DNA samples from certain individuals of known haplotype were included as standards on each *Sst*I, *Bgl*II, *Hin*dIII, *Pvu*II and *Pst*I gel. After electrophoresis was complete, the gel was photographed on Polaroid type 667 film over a transilluminator.

3.3.3.2 Southern Blotting

Southern blotting of DNA fragments in agarose gels to nylon membranes followed the standard procedures originally described (Southern 1975), but with modifications specified by the membrane manufacturer.

DNA fragments were transferred to one of two nylon membranes - Biodyne or Hybond-N. Details of the blotting solutions required for either membrane are given in Appendix One.

Gels were denatured for 30 mins, rinsed in distilled water, and neutralized for 30 mins. The gels were rinsed again before soaking in the transfer solution for 10 mins. Transfer took place on a glass plate supported on a trough filled with the transfer solution. A piece of Whatman 3MM filter paper was soaked in transfer solution, and draped over the glass plate with its ends in the transfer solution. The gel was placed on the wet filter paper, avoiding air bubbles, and the paper around the gel covered with clingwrap to ensure that capillary action took place through the gel only. The membrane, cut to size, was carefully placed on top of the gel, and rolled out with a sterile pipette to remove air bubbles. Three pieces of Whatman 3MM, soaked in 2xSSC, were placed on the membrane, followed by three dry pieces. A pile of dry absorbent paper, a glass plate and a 500g weight (usually a small telephone directory) were added. Transfer was left to proceed overnight. The well positions were then marked, and the membrane rinsed in 2xSSC to remove residual agarose. The blot was baked at 2hrs at 80°C, and stored in a sealed plastic bag until required.

3.3.4 Hybridization and Autoradiography

The DNA probe, or purified probe insert, was prepared for hybridization by incorporation of ^{32}P-dCTP by the random priming oligolabelling method (Feinberg and Vogelstein 1983). In this method, random hexanucleotides bind to complementary sequences on the denatured probe DNA, and act as primers for extension by the Klenow fragment of DNA Polymerase I. The mixture of nucleotides includes ^{32}P-dCTP, allowing for incorporation of radiolabelled nucleotides in the DNA probe. The multiprime kit prepared by Amersham was used, as specified by the manufacturer, for all oligolabelling reactions. The unincorporated ^{32}P-dCTP was separated from the labelled probe by passage through a sephadex G-50 column (Maniatis *et al* 1987). Just prior to hybridization, the probe was denatured by boiling for 2 min, and placed on ice.

Baked filters were prehybridized and hybridized according to manufacturer's specifications. All hybridizations took place at 42°C in the presence of 50% deionized formamide. Filters were placed in a sealed plastic bag, or in a sealable plastic tub, with 10ml of prehybridization solution per blot (18x20cm). Prehybridization occurred for 1hr in the case of Hybond-N membranes, and for 6hrs to overnight for Biodyne membranes. The labelled DNA probe was added to

the hybridization mix in the same bag/tub, resealing took place, and hybridization was allowed to proceed for 24-48hrs. Hybridized filters were then washed to remove excess probe, and non-specifically hybridized DNA. Low stringency washes were used to reveal the polymorphisms detected by 49a/f. Biodyne filters were washed twice in 2xSSC at room temperature (15 min each), and twice in 2xSSC, 0.1%SDS at 42°C (30 min each). Hybond-N filters were washed in 2xSSPE, 0.1%SDS: twice at room temperature (15 min each), once at 42°C (30 min), and once at 65°C (10 min). All other post-hybridization washes followed manufacturer's specifications. Biodyne filters were rinsed twice at room temperature in 2xSSC (15 min each), washed once in 0.1xSSC, 0.1%SDS at 42°C (30 min), and in the same solution at 65°C (60 min). Hybond-N filters were rinsed twice at room temperature in 2xSSPE, 0.1%SDS (10 min each), once in 1xSSPE, 0.1%SDS at 65°C (15 min), and once in 0.1xSSPE, 0.1%SDS at 65°C (10 min).

Fragments were visualized by autoradiography after 1-6 days exposure at -70°C with Kodak XAR film backed with Dupont Cronex intensifying screens.

After use, blots were prepared for reuse by stripping to remove bound DNA. Blots were soaked in denaturing solution (as used for Southern blotting) for 10 min, and then washed twice in a stripping solution, for 10

min each. All washes took place at room temperature. Stripped blots were stored in 2xSSC until required for hybridization.

3.3.5 PCR Detection of the XY275 Polymorphism

Sequence data kindly supplied by Nathan Ellis were used to design three primers for the amplification of XY275 and the X and Y boundary regions. The primer sequences from 5'to 3'were as follows:-
XY or pseudoautosomal 19mer: CTG AGA GTG GAA GTG TCG C
Y-specific 22mer: AGA AAA CTA GTA TTT TCC CCT C
X-specific 20mer: AAC AAG CTC ATC AGC GTG AC.

PCR reactions were carried out in a Perkin Elmer Cetus DNA Thermal Cycler. The final reaction volume was 25μl, and each reaction used 6.25pM of the appropriate primers (XY and X, or XY and Y), 6.25nM of each dNTP, 5-10μg of acetylated BSA, 2 units of Promega *Taq* polymerase, 2.5μl of 10x Promega *Taq* polymerase buffer, and approximately 0.5μg genomic DNA. The reaction was overlaid with 1 drop of mineral oil (Sigma), and incubated for 30 cycles (94°C,48s / 57°C,48s / 72°C,90s), with a 10 minute extension at 72°C.

Amplification of samples was verified by electrophoresis of a 5μl aliquot on 0.8% agarose. The remaining 20μl of amplified product was digested with 6

units of Promega or Amersham *Msp*I at 37°C for 2hrs, and analyzed by electrophoresis through 2% composite FMC agarose gels (3:1 Nusieve GTG:Seakem HGT). PCR amplification and digestion products are described in Figures 3.2 and 3.3.

3.3.6 Analysis of Results

Genetic distance measurements and cluster analysis of 49a/*Taq*I haplotype frequency data were computed using the methods of Harpending and Jenkins (1973), and the computer program Antana written, and kindly supplied, by Professor H.C.Harpending, Pennsylvania State University. The chi-square test was used to determine the level of significance of frequency differences. Nei's test of genetic diversity (Nei 1987) was used to determine the level of genetic diversity observed using different polymorphic systems.

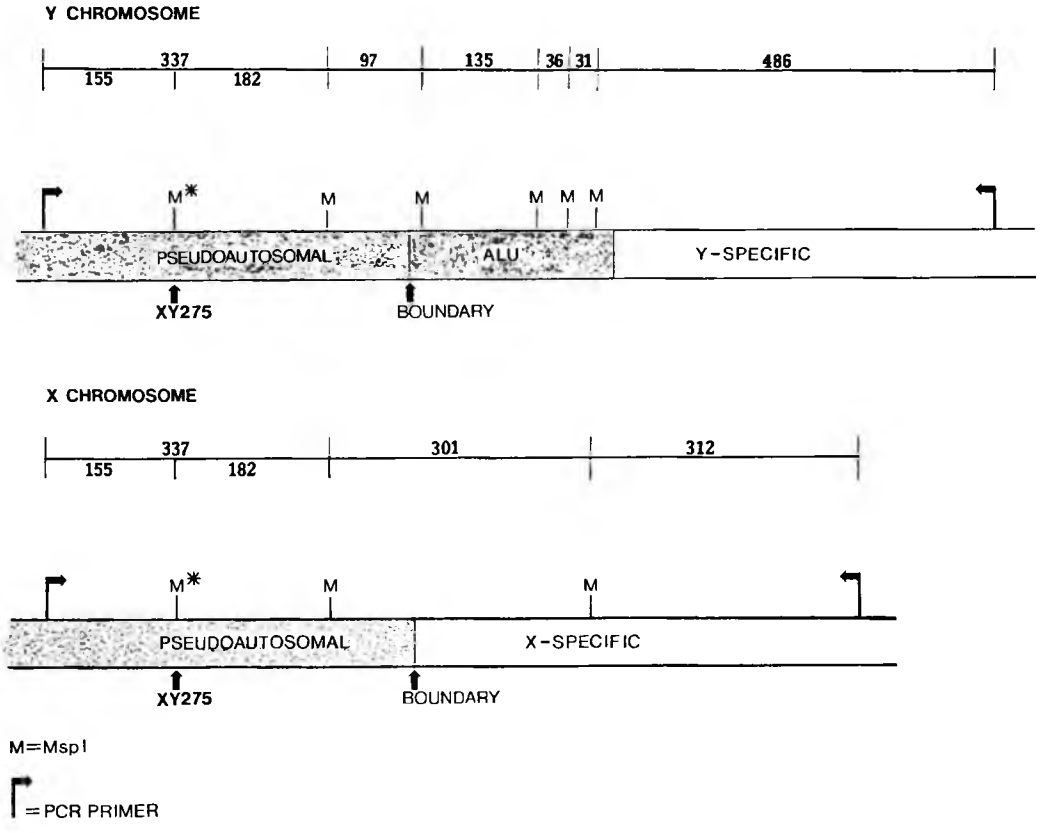

Figure 3.2. PCR amplification of the XY275 locus. Schematic representation of the XY275 PCR amplification and *MspI* digestion products expected using primers described in Section 3.3.5. An asterisk marks the XY275 polymorphism. Numbers above each chromosome represent fragment sizes in bp.

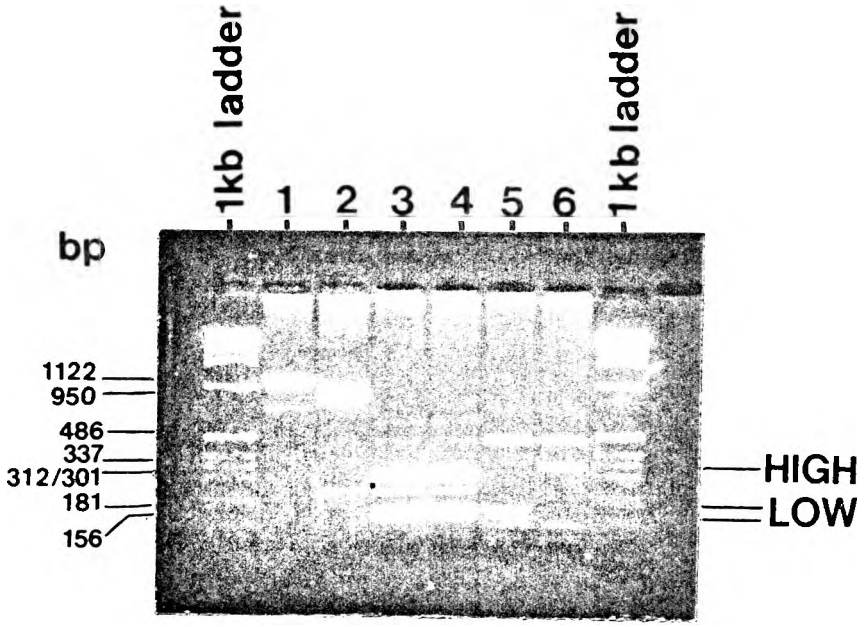

Figure 3.3. *Msp*I digestion products of PCR-amplified XY275 locus. Undigested and *Msp*I-digested X and Y boundary amplification products are separated on 2% composite gels (see Section 3.3.5). Appropriate fragment sizes are indicated in bp to the left of the photograph. Lane 1, undigested Y amplification product of 1122bp. Lane 2, undigested X amplification product of 950bp. Lanes 3 and 4, *Msp*I-digested X amplification products displaying *low* allele (181 and 156bp). Lanes 5 and 6, *Msp*I-digested Y amplification products displaying *low* (181 and 156bp) and *high* (337bp) alleles respectively. A spurious amplification product of approximately 250bp is observed in X amplification products, but does not interfere with interpretation of digestion results.

CHAPTER FOUR

THE SEARCH FOR Y-SPECIFIC POLYMORPHISM

4.1 INTRODUCTION

Many researchers have recognized the potential of Y chromosome studies in population and evolutionary genetics (Casanova *et al* 1985, Ngo *et al* 1986, Oakey and Tyler-Smith 1990), but progress in this area has been hindered by the scarcity of documented Y chromosome RFLPs. Several attempts have been made to systematically search for Y-specific RFLPs (Jakubiczka *et al* 1989, Malaspina *et al* 1990). These studies, which were confined to Caucasoid individuals, revealed a low level of Y chromosome polymorphism, ranging from <1/18000 nucleotides (Jakubiczka *et al* 1989) to <1/46515 nucleotides (Malaspina *et al* 1990). A number of probe/enzyme combinations exempt from preceding studies (Jakubiczka *et al* 1989, Malaspina *et al* 1990) were used to screen southern African Caucasoid and Negroid populations for Y-specific polymorphism in the present study.

4.2 RESULTS

A series of new polymorphisms were revealed by Y chromosome probes p49a and p49f. These complex RFLPs are believed to result, at least in part, from rearrangement processes, and are discussed in detail in Chapter Six. With these exceptions, no new Y-specific polymorphisms were detected. The pDP31 duplication polymorphism was revealed by the enzymes *Bgl*II, *Eco*RI and *Msp*I, as reported previously (Page *et al* 1982, Page *et al* 1984, Bowcock 1987). The enzyme *Pvu*II was also shown to reveal this rearrangement polymorphism. During screening for Y-specific polymorphism, an X-specific *Hin*dIII polymorphism was shown to be revealed in Negroids by pDP1007 (Spurdle and Jenkins 1990).

The number of basepairs screened was determined using the method of Cooper *et al* (1985). The total number of bp examined is conservatively estimated to be 20808. For the purpose of the calculation, probes detecting repeats or X-Y homologous regions were considered to detect only one Y-specific band, and 18 unrelated Y chromosomes were screened for each system (see section 3.2.1). Furthermore, the screening involving probes p49a and p49f was excluded from the analysis, since the complex nature of the polymorphism at this locus precluded a complete understanding of how the variation

is generated. At least some of the variation is believed to result from rearrangements, and not simple sequence polymorphism.

It must be acknowledged that those Y chromosome probes recognizing repeat sequences (pDP105, pDP97 and pY431-HinfA) may cross-react with sequences present elsewhere in the genome. However, the screening of father and son pairs did not reveal Y-specific Mendelian inheritance of any variant fragments, and limited family studies also excluded the possibility of autosomal inheritance of these spurious bands. Another problem to be considered is the fact that a single variant polymorphic site may be masked within multiple copies of an array, and also that variants may not be detected at the single copy level. Indeed, with the high stringency post-hybridization washes employed, no variation was detected in major hybridizing bands, although intensity differences in some minor hybridizing bands could not always be excluded. If these limitations are to be taken into account, and the data generated by those probes recognizing repeat sequences is excluded, the total number of basepairs screened is reduced to 11160.

It is interesting to note that similiar limitations have been observed for the p21A1/*Taq*I polymorphism. This RFLP was described as the interruption of a short

alternating repeat unit of 4kb and 7kb by the loss of one or several *TaqI* sites (Jakubicza *et al* 1989). The existence of several mutation events, either in time and/or position within the tandem repeat array, has since been verified by the comparative Y chromosome polymorphism study described in Chapter Nine. This finding correlates with observable differences in intensity of the 11kb fragment (unpublished data).

4.3 DISCUSSION

This methodical search for polymorphism on the Y chromosome of Caucasoids and Negroids has revealed a dearth of Y-specific RFLPs, as have two previous studies using Caucasoid samples and different probe/enzyme combinations. Collective results indicate that 290 probe/enzyme systems have been used to screen at least 85641 basepairs, revealing only one substitution polymorphism (Jakubiczka *et al* 1989) and one duplication polymorphism (Page *et al* 1982). This finding is at variance with the fact that all three studies employed the enzymes *TaqI* and *MspI*, which are known to exhibit a higher frequency of polymorphism in human DNA (Barker *et al* 1984). Furthermore, the enzyme choice included 5-9 of the 10 enzymes shown to be most

effective at revealing autosomal polymorphism (Wijsman 1984). Similiar random searches for autosomal RFLPs have been very successful (Feder *et al* 1985).

The lack of Y-specific variation is not confined to nucleotide changes occurring within restriction endonuclease sites. Studies on two Y-derived CA repeats have not revealed any polymorphism in 20 Caucasian/Asian males (Z.Ulinowski, D.Bailey, K.Taylor and J.Wolfe, personal communication). The repeats also appear to occur less frequently, with positive signals for 5% of Y cosmids, as opposed to 20% of cosmids derived from chromosome 9q. In addition, sequencing of the Y chromosome has revealed a very low level of nucleotide diversity, although somewhat higher than estimates based on RFLP studies (M.Hammer, personal communication). The implication is that many kilobases of each Y chromosome would need to be sequenced to match the information revealed by mtDNA D-loop sequences used in published evolutionary studies.

This dearth of polymorphism is unexpected in view of theories on the origin of the Y chromosome (Charlesworth 1991, Ellis 1991a). The proto-Y chromosome is believed to have evolved from a homologous proto-X chromosome, and with the exception of the genes involved in sex determination and the present region of homology between the sex chromosomes,

was prone to rearrangement processes and loss/inactivation of genes (Yen *et al* 1988, Lau *et al* 1989, Bardoni *et al* 1991, Legouis *et al* 1991, Nakahori *et al* 1991). The Y chromosome was presumably also free to accumulate satellite sequences and non-coding retrotransposed elements originating from autosomes, possibly as a response to selection for a chromosome size which would be appropriate for efficient meiotic segregation (Bardoni *et al* 1991). Despite this plethora of non-coding sequences, simple sequence polymorphism is rare, and most of the reported Y-specific polymorphism is generated, at least in part, by rearrangement processes (Casanova *et al* 1985, Oakey and Tyler-Smith 1990, Torroni *et al* 1990, Spurdle and Jenkins 1992).

Malaspina *et al* (1990) have proposed several explanations for the low frequency of polymorphism on the Y chromosome. One suggestion is that the mechanisms protecting the Y chromosome from recombination may simultaneously protect it from mutagenesis, but the role of recombination in maintaining genetic variation is considered more important (Malaspina *et al* 1990). The latter explanation arises from the postulate by Clark (1987) that Y chromosome polymorphism is unlikely to be maintained by natural selection. An additional explanation for the observed phenomenon may be inferred from sequencing studies of the pseudoautosomal boundary

region (Ellis *et al* 1990a, Ellis *et al* 1990b). The Y boundary was shown to exhibit less sequence variation than the X boundary, suggesting that the Y boundary evolved more recently (Ellis *et al* 1990a, Ellis *et al* 1990b). It is possible that insufficient time has elapsed since the origin of the present-day Y chromosome to allow for the accumulation of significant polymorphism. However, a survey for long range Y-specific polymorphism using pulsed-field gel electrophoresis has identified several novel hypervariable blocks (M.Jobling, personal communication), contrasting with the results from RFLP screening. These findings suggest that such large scale polymorphism is generated at a greater rate than simple sequence polymorphism.

4.4 SUMMARY

DNA samples of Negroid and Caucasoid origin were screened for Y-specific RFLPs. A total of 7 Y chromosome probes and 13 restriction enzyme digests were used to examine a conservative estimate of 20000bp, and no new Y-specific polymorphisms were revealed by these systems. The paucity of polymorphism on the Y appears to be unrelated to possible bottleneck effects during raciation.

CHAPTER FIVE

THE 49a/*Taq*I POLYMORPHISM IN POPULATION STUDIES

5.1 INTRODUCTION

The ability of Y chromosome probes p49f and p49a to detect multiple Y-specific haplotypes (Ngo *et al* 1986) has made it a prime candidate for a study of gene flow in southern African populations. The two probes are different subclones of cosmid 49 (Bishop *et al* 1983), and both reveal about fifteen Y-specific *Taq*I bands corresponding to a low copy number sequence (Ngo *et al* 1986). Five of these bands (A,C,D,F and I) were originally shown to be present, absent, or variable in length (Ngo *et al* 1986), and are said to constitute independent allelic series. Guerin *et al* (1988b) subsequently reported the absence of band B as a variant, Lucotte *et al* (1990b) recently reported absence of band G in Asiatic Indian populations, and band H was found to be missing in Italian Caucasians (Torroni *et al* 1990). Absence of band O was also reported in Italian Caucasians (Torroni *et al* 1990).

Since the Y-specific *Taq*I fragments are syntenic and span a region of at least 80kb, each combination of fragments is said to define a haplotype of the human Y chromosome (Ngo *et al* 1986). Forty-five such 49a/f

haplotypes have been reported (Ngo *et al* 1986, Breuil *et al* 1987, Lucotte *et al* 1989, Lucotte *et al* 1990b, Torroni *et al* 1990). The nomenclature for the first 24 49a/f/*Taq*I haplotypes is straightforward, but the recent publication of Torroni *et al* (1990) excludes haplotypes 20-24 of Lucotte *et al* (1990b). These authors have named their 22 new haplotypes from Ht20 to Ht41, thus creating a discrepancy in haplotype nomenclature. Ht24 of Lucotte *et al* (1990b) and Torroni *et al* (1990) is fortuitously the same. Seven of the new haplotypes reported by Torroni *et al* (1990) were observed in the present study.

The 49a/f probes have been used in a regional population study (Breuil *et al* 1987), and to establish a possible genealogy of the Y chromosome in human chromosomes (Hazout and Lucotte 1987, Lucotte *et al* 1989, Lucotte *et al* 1990b). The Y-specific 49a haplotypes were determined for a number of different southern African populations in the present study. The resulting frequency data have been used to assess the affinities of the populations, and to elucidate the nature of the historical interactions between them.

5.2 RESULTS

Probes p49f and p49a were initially used consecutively to determine the complete p49/*Taq*I haplotype. Consequently it was discovered that the purified *Xba*I-*Bam*HI p49a fragment revealed all the Y-specific fragments detected by p49f, with the exception of fragment N (see Figure 5.1). All five allelic series reported by Ngo *et al* (1986) could thus be detected by hybridization with p49a alone.

Screening of southern African populations with p49a/f has revealed a number of new *Taq*I allelic variants (Table 5.1). The variants depicted as apparently co-inherited A alleles are not believed to be due to partial digestions, since the fragment sizes remain unchanged when digested with increasing concentrations of *Taq*I (data not shown). In all instances where DNA from other appropriate family members was available, strict Mendelian inheritance of new allelic variants was successfully demonstrated *(A3;A2, G0,* and *H0)*.

The new variants were detected with both p49a and p49f. Suggestions as to the origin of the new variants are detailed in the discussion. The frequencies of the alleles (or allelic variants) in the four major population groups are shown in Table 5.2.

Each new allelic variant results, of necessity, in the creation of a new haplotype. Certain of the new variants are found in more than one haplotype, and new combinations of previously reported alleles have also been revealed (Table 5.3). The new haplotypes observed in this study (including the 8 new haplotypes of disputable nomenclature reported by Torroni *et al*, 1990) are shown in Table 5.3. Haplotypes have been numbered 25-62, following on from Lucotte's haplotypes 20-24 (Lucotte *et al* 1990b). The hybridization patterns for most of the new haplotypes are displayed in Figure 5.1.

Haplotype frequencies were calculated for each of the 23 populations, and also for larger groupings based on linguistic classification. The 49a/*Taq*I haplotype frequencies are shown in Appendix Two. The haplotype frequencies were used to calculate the genetic distances between the populations, and these distances were then subjected to cluster analysis. The clustering as revealed by analysis of populations grouped according to their broad classification is shown in Figure 5.2.

FIGURE 5.1 The 49a/TaqI hybridization patterns for new haplotypes detected in southern African populations. Haplotype numbers are indicated at the top of each sample lane, and the position of bands/alleles is indicated to the left of each autoradiograph. Haplotypes 31, 44, 51, 61 and 63 are not represented in this figure.

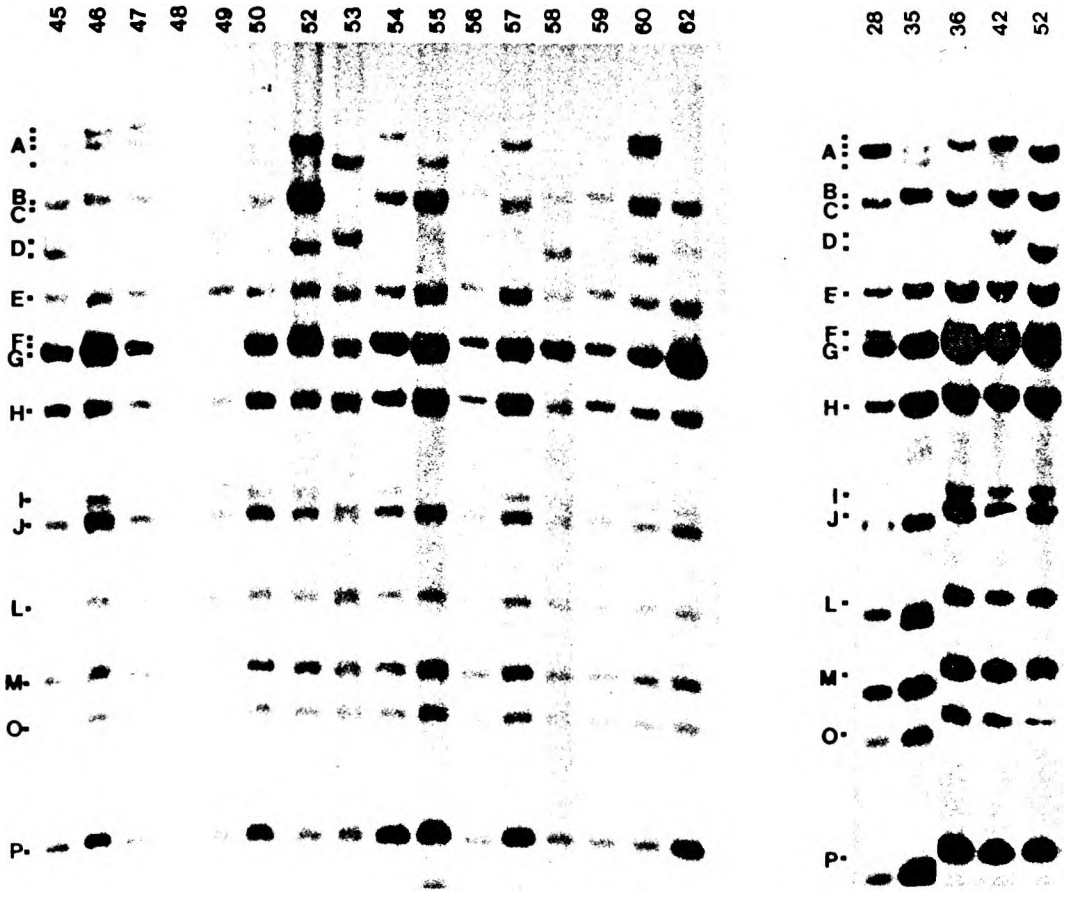

FIGURE 5.1 continued. Hybridization patterns for 49a/TaqI haplotypes detected in southern African populations.

TABLE 5.1

New Allelic Variants Detected in Southern African Populations

Allelic Variant	Fragment Size(s)(kb)[a]	Population(s)
A5[b]	~28.0[c]	SA Indian, SA Jewish
A5;A3	~28.0;18.0	SA European, Richtersveld "Coloured"
A4;A3	20.0;18.0	SA Indian, Richtersveld and JHB "Coloured", Tswana
A4;A3;A2	20.0;18.0;17.0	SA European
A3;A2	18.0;17.0	SA European, SA Jewish, SA Indian, San, Swazi, Dama, JHB "Coloured"
D3	6.6	JHB "Coloured"
D2;D1	7.5;8.0	SA European
F2;F1[d]	5.9;5.5	Richtersveld "Coloured"
G0	absent	SA Indian, Pedi, Tsonga, Tswana
H0[e]	absent	Tsumkwe San

[a]Fragment sizes are consistently larger than those reported by Ngo et al (1986).

[b]This A5 fragment probably represents the same A5 fragment reported by Torroni et al (1990), but fragment sizes have not yet been compared directly.

[c]Size out of range with DNA size markers used.

[d]Fragment *F2* may represent the 5.2kb fragment reported by Torroni et al (1990), but fragment sizes have not yet been compared directly.

[e]Torroni et al (1990) report the absence of H to be associated with the absence of bands P and R. The techniques of electrophoretic separation employed in this study mostly excluded identification of 49a/TaqI fragments smaller than the O fragment. However, in at least one example, H0 was shown to be associated with the absence of band P.

TABLE 5.2

Frequency of 49a/*Taq*I Alleles in Four Major Population Groups

Allelic Variant or Allele	Caucasoid (152)	Negroid[a] (452)	Khoisan (117)	JHB "Coloured" (66)
A5;A3	0.020	0.000	0.000	0.000
A4;A3	0.007	0.002	0.000	0.000
A4;A3;A2	0.007	0.000	0.000	0.000
A3;A2	0.039	0.004	0.231	0.045
A5	0.020	0.000	0.000	0.000
A4	0.046	0.007	0.017	0.015
A3	0.466	0.124	0.291	0.515
A2	0.349	0.133	0.205	0.243
A1	0.000	0.710	0.231	0.152
A0	0.046	0.020	0.025	0.030
B1	0.960	1.000	0.974	0.985
B0	0.040	0.000	0.026	0.015
C1	0.329	0.011	0.009	0.303
C0	0.671	0.989	0.991	0.697

TABLE 5.2 continued.

D1;D2	0.007	0.000	0.000	0.000
D1	0.513	0.013	0.009	0.259
D2	0.243	0.015	0.017	0.259
D3	0.000	0.000	0.000	0.015
D0	0.237	0.972	0.974	0.467
F2;F1[b]	0.000	0.000	0.000	0.000
F1	0.947	0.962	0.667	0.970
F0	0.053	0.038	0.333	0.030
G1	0.993	0.991	1.000	1.000
G0	0.007	0.009	0.000	0.000
H1	1.000	1.000	0.974	1.000
H0	0.000	0.000	0.026	0.000
I1	0.789	0.863	0.786	0.833
I0	0.211	0.137	0.214	0.167

[a]Only Bantu-speaking Negroids included in this category.

[b]Allelic variant *F2;F1* has been observed only in the Richtersveld population.

TABLE 5.3

New Y-specific 49a Haplotypes Reported in This Study

Taq49a Haplotype[a]	Alleles							
	A	B	C	D	F	G	H	I
25	A3/A2	B1	C1	D0	F1	G1	H1	I1
26	A3	B1	C0	D3	F1	G1	H1	I0
27[b]	A0	B1	C0	D0	F0	G1	H1	I1
28	A3	B0	C1	D0	F1	G1	H1	I0
29	A3	B1	C0	D0	F0	G1	H1	I0
30	A3/A2	B1	C0	D0	F1	G1	H1	I1
31	A3/A2	B1	C0	D0	F1	G1	H1	I0
32	A3	B1	C0	D0	F0	G1	H1	I1
33[c]	A2	B1	C0	D0	F0	G1	H1	I1
34[d]	A1	B1	C0	D0	F0	G1	H1	I1
35	A3/A2	B1	C0	D0	F0	G1	H1	I0
36	A4	B1	C0	D0	F1	G1	H1	I1
37	A4/A3	B1	C0	D2	F1	G1	H1	I1
38	A3/A2	B0	C0	D0	F1	G1	H0	I0
39	A0	B1	C0	D0	F0	G1	H1	I0
40	A4/A3	B1	C0	D0	F1	G1	H1	I1
41[e]	A1	B1	C0	D0	F1	G0	H1	I0
42	A5	B1	C0	D1	F1	G1	H1	I1
43	A4	B1	C0	D0	F1	G1	H1	I0
44	A4/A3	B1	C0	D0	F0	G1	H1	I1

TABLE 5.3 continued.

45	A4/A3/A2	B1	C1	D1/D2	F1	G1	H1	I1
46	A5/A3	B1	C0	D0	F1	G1	H1	I1
47	A5/A3	B1	C1	D2	F1	G1	H1	I1
48	A3	B0	C0	D1	F1	G1	H1	I0
49	A0	B0	C0	D1	F1	G0	H1	I0
50[f]	A3/A2	B1	C1	D2	F1	G1	H1	I1
51	A2	B0	C0	D1	F1	G1	H1	I0
52	A3	B1	C0	D2	F1	G1	H1	I1
53	A2	B0	C0	D1	F0	G1	H1	I0
54	A4	B1	C0	D0	F0	G1	H1	I1
55	A2	B1	C0	D0	F0	G1	H1	I0
56[g]	A2	B1	C0	D0	F1	G1	H1	I0
57	A3	B1	C1	D0	F0	G1	H1	I1
58	A0	B1	C0	D2	F0	G1	H1	I1
59	A4	B1	C0	D0	F0	G1	H1	I0
60[h]	A4/A3	B1	C1	D2	F1	G1	H1	I1
61	A4	B1	C1	D2	F1	G1	H1	I1
62	A0	B1	C0	D1	F1/F2	G1	H1	I1
63[i]	A0	B1	C0	D0	F1	G1	H1	I0

[a]The Roman numeral system of naming haplotypes (Ngo et al 1986) has been replaced with the arabic system in this study because of the large number of haplotypes recorded.

[b]Ht40 of Torroni et al (1990). [c]Ht41 of Torroni et al (1990). [d]Ht23 of Torroni et al (1990). [e]Ht38 of Torroni et al (1990). [f]Ht34 of Torroni et al (1990). [g]Ht27 of Torroni et al (1990). [h]Ht37 of Torroni et al (1990). [i]Ht22 of Torroni et al (1990).

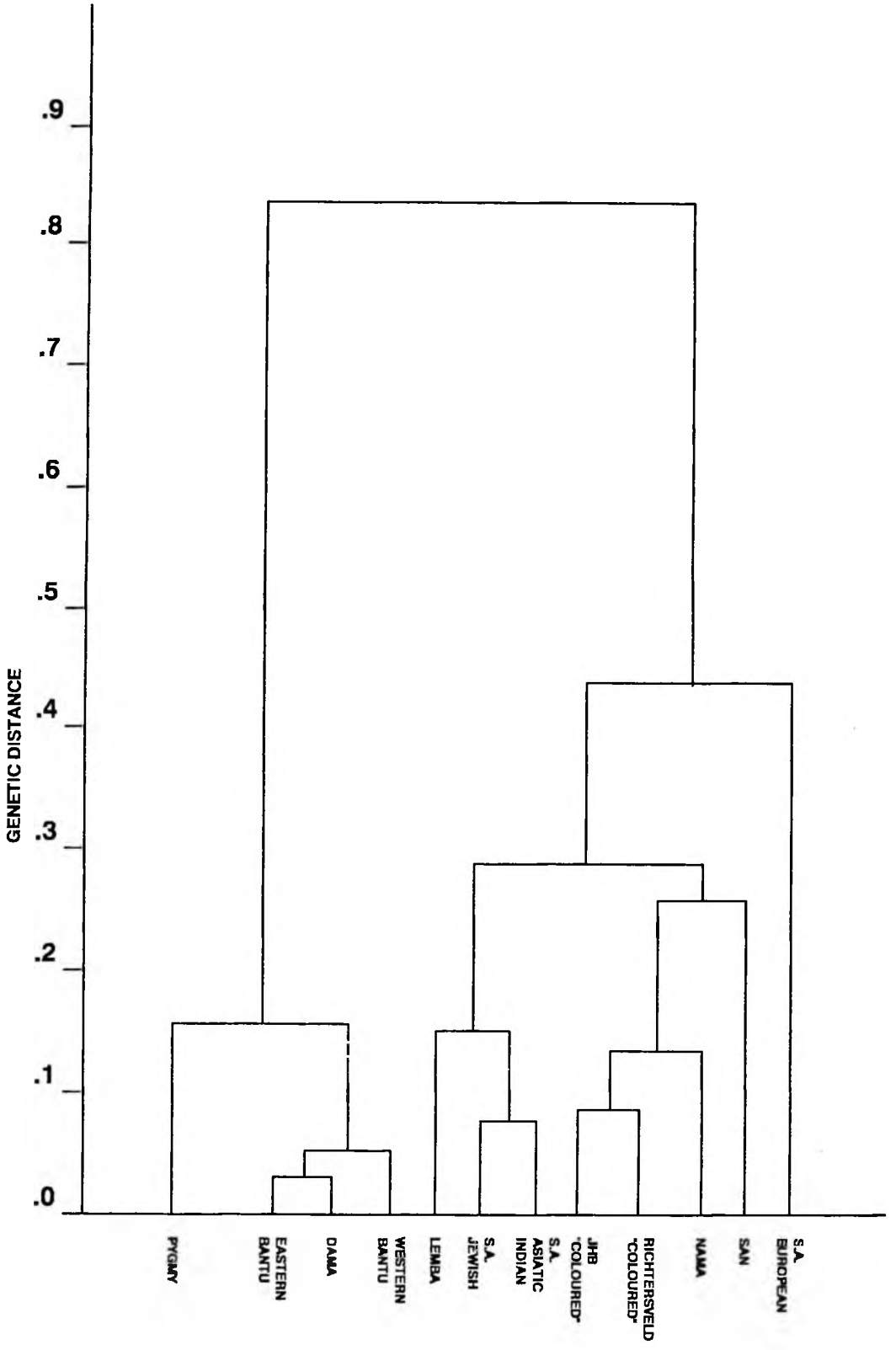

FIGURE 5.2 Clustering implied by 49a/TaqI haplotype frequency data. The dendogram is derived from the genetic distances displayed in Table 5.4, which are themselves based on 49a/TaqI haplotype frequency data.

TABLE 5.4

Genetic Distances Between Larger Population Groupings.

	E.Bantu	W.Bantu	Lemba	Dama	Nama	San	Jewish	Indian	Eur.	Col.	Richt.	Pygmy
E.Bantu		.053	.317	.026	.163	.347	.505	.489	.641	.281	.292	.079
W.Bantu			.476	.043	.380	.502	.722	.700	.838	.474	.496	.154
Lemba				.285	.291	.289	.127	.147	.440	.170	.207	.246
Dama					.227	.308	.454	.446	.577	.267	.291	.063
Nama						.262	.278	.277	.400	.109	.124	.182
San							.267	.239	.430	.216	.178	.198
Jewish								.078	.310	.145	.169	.357
Indian									.346	.128	.149	.334
European										.130	.304	.507
JHB Col.											.086	.208
Richters												.205
Pygmy												

5.3 DISCUSSION

5.3.1 Origin of the 49a/*Taq*I Polymorphism

An understanding of the origin and nature of the polymorphisms detected by probes 49f and 49a is crucial to the understanding of the origin of new allelic variants detected in this study. These Y-specific DNA probes are said to detect a family of moderately repeated sequences (Ngo *et al* 1986). The repeat sequences are present in restriction fragments of different sizes, corresponding to the Y-specific bands detected under non-stringent conditions. Certain groups of Y-specific bands are considered independent "allelic series" or "loci" since they can be absent, present or variable in length (Ngo *et al* 1986). The apparent co-existence of alleles of the same allelic series has since been reported (Spurdle *et al* 1989, Torroni *et al* 1990).

The generation of the *Taq*I RFLPs detected at each locus has been proposed to be due to point mutations within *Taq*I restriction sites (Ngo *et al* 1986), arising in, or leading to, fragments that could co-migrate with other major non-polymorphic bands like E, N, and O, and thus remain undetected. This behaviour was illustrated by the apparent co-migration of a putative *D0* fragment with B (Ngo *et al* 1986), but haplotype data presented

by Torroni *et al* (1990), as well as in this study (Table 5.3) indicate that this co-migration is at the very least inconsistent since haplotypes (Ht) 30 and 33 of Torroni *et al* (1990), and Ht28 and Ht38 (this study) each include both alleles *B0* and *D0*, suggesting that *D0* may co-migrate with a smaller non-polymorphic band.

The increased polymorphism frequency of *Taq*I sites in general has been ascribed to the presence of the CpG dinucleotide in the recognition sequence (Barker *et al* 1984), a hotspot for point mutations. The possibility of insertion/deletion or rearrangement processes as a mechanism for generating the observed polymorphism were excluded on the premise that variations of hybridization patterns were not observed with other restriction digests (Ngo *et al* 1986).

The co-existence of "alleles" at several loci (Spurdle *et al* 1989, Torroni *et al* 1990, this study) has led to a reassessment of the nature of the polymorphism detected by 49a. If the *A* locus is considered as an example of co-existing "alleles", there are several possible explanations. The co-existing bands may represent the original A fragments, and these fragments would thus not be allelic to each other. However, the presence of three A bands in one individual (Table 5.1 and Figure 5.1) suggests that duplication processes, followed by differential CpG mutation of the duplicated

loci, may be responsible for the observed co-existence. This explanation for co-existing alleles is also that favoured by Torroni *et al* (1990). Thus each combination would represent one allele in itself. Densitometric studies could be used to quantify the number of *A* loci present in a given A band and to thus distinguish between single and muliple copy alleles. Unfortunately, such studies have proved unreliable due to the large range of hybridization intensities present in any given 49a/*Taq*I haplotype, as well as the large area over which hybridization conditions have to be uniform (>23kb to <2kb).

Such a theory of duplication followed by differential CpG mutation, would also explain the co-inheritance of D1 and D2. It is of interest that allelic variant *D1/D2* is found together with the allelic variant *A4;A3;A2* (Ht45, Table 5.3), and furthermore, that the D2 band is more intense than the D1 band (Figure 5.1), suggesting that there are more copies of D2 than D1. This would imply that there are at least three copies of the *A* and *D* loci, and also that the *A* and *D* loci are probably contiguous and were duplicated as a unit. In support of this hypothesis, co-existing A and D alleles occurring in a single Italian Caucasoid individual has also been reported by Torroni *et al* (1990).

Co-existence of the F1 and F2 bands could be explained in a similiar way. In this case the allelic variant is not associated with co-inheritance of bands at another locus (Ht62, Table 5.3), and does not give any clues as to the molecular arrangement of loci. All other allelic variants (Table 5.1 and Figure 5.1) can be explained in terms of CpG mutation in *Taq*I sites resulting in a change in fragment length. Fragment A5 has probably resulted from the loss of a *Taq*I site at the extremity of an A4 fragment, since no fragment of size intermediate between A4 and A5 has been observed. For this reason, it is also likely that the A5 fragment reported in this study represents that reported by Torroni *et al* (1990). D3 is probably a smaller precursor fragment to D2, and loss of a *Taq*I site at the extremity of the D3 fragment results in D2. Conversely, D3 may be the product of a newly created site within D2, although restoration of *Taq*I sites is considered to be a rare event (Bird 1980). *G0* and *H0* are probably best explained as alleles at the *G* and *H* loci respectively that co-migrate with non-polymorphic bands. No particular allele at any other locus is strictly associated with *G0*. *H0* is associated with the presence of a more intense F1 band (Spurdle *et al* 1989). Torroni *et al* (1990) concluded that the absence of bands H-P-R together with B in a single individual was due to a deletion. However, the fragment sizes of H and P add up to a fragment size very similiar to that

of F1. Although the absence of band R has not been confirmed for those individuals with *H0* reported in this study, it is not inconceivable that fragments H and P, with or without the small (<1kb) fragment R, are masked as fragment F1 in these individuals.

Thus it is possible that loss of *Taq*I sites by CpG mutation alone, or a combination of duplication followed by differential CpG mutation, can explain the origin of the new allelic variants detected in the present study (Table 2). Ngo *et al* (1986) excluded duplication as a mechanism for generation of the 49a polymorphism because they did not detect polymorphism with other enzymes. However 49a has since been shown to detect polymorphism with several other enzymes (Chapter Six). The 49a haplotypes revealed by the different enzymes do not correspond strictly with one another, which is not incomprehensible since all these polymorphisms may reflect restriction site changes in addition to deletion/insertion changes.

The origin and nature of the repeat sequences detected by 49a need to be discussed with regard to understanding how duplications can occur on the haploid Y chromosome. Probes 49a and 49f also detect complementary sequences on an autosome, and a human testis cDNA transcript corresponding to 49f has been mapped to chromosome 3 (Cohen-Hagenhauer *et al* 1987).

It has been suggested that the Y sequences represent a pseudogene, and may have occurred on the Y as a result of duplication or retrotranscription (Leroy *et al* 1987). Lucotte *et al* (1990a) conclude, from comparative hybridization studies in primates, that the autosomal sequence was retrotranscribed onto the Y chromosome during primate evolution. RNA-mediated transposition or retrotranscription is the most likely event to explain the transfer of an autosomal gene to the Y chromosome. This event is described as the insertion of double-stranded DNA copies of an RNA transcript at random positions in the genome (Maeda and Smithies 1986). Since these RNA transcripts lack regulatory sequences, the pseudogenes formed from such transcripts are mostly inactive.

Only one and two autosomal *Taq*I fragments of apparent uniform intensity are detected by 49a and 49f respectively, as opposed to a number of varying Y-specific fragments. This implies that the copy number increase detected on the Y chromosome occurred only after the transfer from autosome to the Y. The most likely explanation for the duplication of the 49a repeat is gene amplification. This is described as differential replication of particular sequences as a result of multiple initiation events at a single origin of replication (Maeda and Smithies 1986). The "onion skin" forms of chromosome generated by such

re-initiations of replication are resolved into mitotically stable forms, giving rise to a cluster of multiple copies of the overreplicated gene. The individual genes are found to have orientation either the same or opposite to their neighbours, and the endpoints of the multiple copies are variable.

After gene duplication, the different copies diverge by accumulating base-pair substitutions, deletions, duplications or insertions. The decreased selection pressure on non-coding pseudogenes facilitates their degeneration, and such degeneration of duplicated 49a repeats could foreseeably result in the complex hybridization patterns observed. Divergence by duplication or deletion is possible due to polymerase slippage (Wolff *et al* 1988) and could explain the origin of some of the new allelic variants (Table 5.1), although only very short repeats are generally believed to have originated in this way (Morral *et al* 1991, Schlotterer and Tautz 1992).

5.3.2 Population Frequencies of 49a/*Taq*I Alleles

The new allelic variants described here (Table 5.1) occur at low frequencies in the populations studied (Table 5.2), with the exception of variant *A3;A2*, which occurs at a frequency of 0.231 in the Khoisan group. This same variant was found at a frequency of 0.039 in

S.A. Caucasoids, lower than the frequency of 0.08 reported in Italian Caucasoids (Torroni *et al* 1990). The considerable number of new allelic variants observed in this study can be attributed to the large number and variety of new populations screened.

The frequency of previously reported alleles in the four major population groups studied reveals major differences between the groups (Table 5.2). Allele *A1* is by far the most common allele of the *A* series in Bantu-speaking Negroids of southern Africa, occurring at a frequency of 0.710. This is in direct contrast to data reported by Lucotte *et al* (1989), where *A1* occurs at a frequency of only 0.036 in a sample of 56 Negroids from Bangui (Central African Republic), and *A3* is the most common allele at a frequency of 0.804. However, Torroni *et al* (1990) reported a high frequency of *A1* (approximately 0.80) in Africans. Alleles *A3* and *A2* are the most common alleles in the Caucasoid population, as well as in the Johannesburg "Coloured" population. The similarity in *A* allele frequency between the Caucasoid and JHB "Coloured" populations correlates with the cluster analyses generated from 49a/*Taq*I haplotype genetic distance data (Figure 5.2), in that both results lend support to the theory that Caucasoid gene flow into the hybrid groups of southern Africa was male-specific. The Khoisan group has allele frequencies spread fairly evenly between *A3;A2*, and *A3*, *A2* and *A1*.

A0 occurs at frequencies lower than 0.05 in all population groups. New variants *A5;A3, A4;A3, A4;A3;A2* and *A5* are found at low frequencies, and mostly in the Caucasoid population alone. Similarly, variants *A5;A2, A4;A3* and *A3;A2* described by Torroni *et al* (1990) were observed in Caucasoids.

The *B* series is represented mostly as *B1*, in agreement with the findings of Guerin *et al* (1988) and Torroni *et al* (1990). *C1* occurs at low frequency in the Negroid and Khoisan groups, and rises to frequencies of 0.329 and 0.303 in the Caucasoid and Johannesburg "Coloured" groups, respectively. In agreement with these findings, Torroni *et al* (1990) reported *C1* to be absent in Africans, and at a frequency of 0.53 in Italian Caucasoids. In contrast, *C1* was the most common allele (frequency 0.411) in the Bangui sample of Lucotte *et al* (1989). The *D* series is strikingly represented by *D1* in the Caucasoid group, while *D0* occurs at very high frequency in the Negroid and Khoisan groups, and at relatively high frequency in the JHB "Coloured" group. Again, these frequencies seem to be at variance with the situation in the Bangui group, where *D0* occurs at a frequency of only 0.124, and *D1* at 0.609 (Lucotte *et al* 1989). However, Torroni *et al* (1990) found D to be absent in Africans. Allele *F1* generally represents the *F* series, with the marked exception of the Khoisan group, where *F0* is observed at a frequency of 0.333.

This is in contrast to the previous reports of low heterozygosity at the *F* locus (Ngo *et al* 1986, Lucotte *et al* 1989, Torroni *et al* 1990). *I1* represents the *I* locus at a high frequency in all population groups, which corresponds with the frequency of 0.714 in the Bangui Bantu population (Lucotte *et al* 1989), 0.674 in a mixed Asiatic Indian group (Lucotte *et al* 1990b), and 0.776 and 0.885 in the Italian Caucasoid and African samples studied by Torroni *et al* (1990).

The major differences in *A* and *D* allele frequencies between the Bangui Bantu group (Lucotte *et al* 1989) and the southern African Bantu-speaking Negroid group (this report) imply that these two Negroid groups are genetically distant. This cannot be easily explained, since all the Bantu-speakers, it is proposed, have originated from the same proto-Bantu stock, probably as recently as 2000-3000 years before present (Oliver and Fagan 1975), with the possible exception of the north-western Bantu. The latter languages have been shown to be more distant from each other than those of the rest of the Bantu sphere (Oliver and Fagan 1975), which has led to the suggestion that the north-western Bantu people may have diverged from the proto-Bantu-speaking population before they became a single coherent population (Oliver and Fagan 1975). However, the Bangui Bantu-speaking population of the Central African Republic is likely to be classified as

Equatorial Bantu (Murdock 1959), and would thus have originated from the same proto-Bantu stock as the southern African Bantu-speakers. Knowledge about the possible operation of random genetic drift in the Bangui group is lacking, as is the extent of admixture. This population is, however, represented by an urban sample (Lucotte, personal communication), so the latter possibility cannot be excluded.

5.3.3 Haplotype and Cluster Analysis

The allele frequencies (Table 5.2) discussed in the preceding section have been included mainly for comparison with previous reports. Although explanations may have been proposed for the generation of the new variants detected, none has yet been proven. Thus it would not be prudent to regard the alleles and their frequencies as reliable markers. However, when combined in the form of a haplotype, the 49a/*Taq*I band pattern represents the Y-specific genotype of an individual, and can thus be employed in population and gene flow studies.

A total of 31 new haplotypes were identified in the 23 populations studied (Table 5.3), in addition to 8 haplotypes of disputable nomenclature reported by Torroni *et al* (1990). Of these 31 haplotypes, 21 were observed at low frequency (0.02-0.12) in only one

population (Appendix Two). The remaining new haplotypes were observed in at least two populations, generally at low frequency, although Ht29 occurs at a frequency of 0.31 in the Tsumkwe San, and Ht30 reaches frequencies of 0.29 and 0.26 in the Tsumkwe and Sekele San respectively, (Appendix Two). Most of the new haplotypes occur in the Caucasoid populations (or in the hybrid groups believed to have arisen from male-specific Caucasoid admixture with indigenous populations), while the Khoisan and Negroids possess relatively few new haplotypes for their sample size. However, these findings do not correlate directly with gene diversity values (Appendix Two), because although Caucasoids exhibit great diversity, so do the Khoisan. The large gene diversity of Caucasoids can be related to the numerous haplotypes they possess, but the increased gene diversity of the Khoisan over the Negroids may be attributed to a more even spread of haplotype gene frequency in the former.

The previously described haplotypes Ht6, Ht14, Ht16, Ht17, Ht20, Ht21 and Ht23 were not observed in the populations screened, as well as all but 9 of the new haplotypes described by Torroni *et al* (1990). Eight of these are identified in Table 5.3, while the remaining haplotype, Ht24, was first described by Lucotte *et al* (1990b), also as Ht24. Ht6, Ht14 and Ht16 were reported in the initial sample of 44 Parisians of various ethnic

origins (Ngo *et al* 1986), as well as in the Bangui sample (Lucotte *et al* 1989). Ht17 is Papua-specific (Breuil *et al* 1987), and Ht20, Ht21 and Ht23 were reported in a mixed Asiatic Indian group (Lucotte *et al* 1990b). Interestingly, Ht18 and Ht19 were reported to be Pygmy-specific (Lucotte *et al* 1989). However, Ht18 was observed at low frequency in the Asiatic Indian, Tswana and Richtersveld "Coloured" groups, and at a considerable frequency in the Lemba population, while Ht19 was observed in the Johannesburg "Coloured" population at a frequency of 0.02 (Appendix Two).

Ht4 is the most common haplotype in all Negroid groups studied, and also occurs at substantial frequencies in the Nama and the Sekele San populations. In agreement with this data, Torroni *et al* (1990) found Ht4 to be the most common in Africans, at a frequency of 0.686. Surprisingly, this haplotype occurs infrequently in the Tsumkwe San (Appendix Two), and is absent from the Bangui and Pygmy samples studied by Lucotte *et al* (1989). Ht15 is most representative of the S.A. European population, a finding reported by Torroni *et al* (1990) with regard to an Italian Caucasoid sample. However, the S.A. Jewish population is distinctive due to the substantial frequencies of Ht7 and Ht8. The presence of Ht8 in the Nama population may lend support to the theory of the semitic origin of this group

(Jeffreys 1968), but Ht8 is also found at relatively high frequency (0.17) in the S.A. Asiatic Indian population.

Interestingly, Ht18 and Ht19 were reported to be Pygmy-specific (Lucotte *et al* 1989), but were both absent from the Pygmy sample studied here. Furthermore, Ht18 was observed at low frequency in the Asiatic Indian, Tswana and Richtersveld "Coloured" groups, and at a considerable frequency in the Lemba population, while Ht19 was observed in the Johannesburg "Coloured" population at a frequency of 0.02 (Appendix Two). It is of some concern that not one of the haplotypes reported to occur in the Aka Pygmy sample of 37 individuals from the Central African Republic (Lucotte *et al* 1989) have been observed in the Pygmy group presented in this study (Appendix Two), which includes individuals of the Aka tribe (see section 3.1.7). Admixture of the Aka Pygmy population studied by Lucotte *et al* (1989) with local Negroid farmers should be an implausible explanation for this disparity, because Ht4, the common Negroid haplotype, is absent from these Pygmies. However, as mentioned previously during the discussion of 49a/*Taq*I allele frequencies, the central African Bangui Bantu sample studied by the same group (Lucotte *et al* 1989) appears to be distinct from other Bantu-speakers (Table 5.2), and the analysis of 49a/*Taq*I haplotype frequencies in Africans from Senegal and Cameroon (Torroni *et al* 1990) and southern African

Negroids (Spurdle *et al* 1992, this study) have also shown the Bangui Bantu to be genetically discordant. If the Bantu-speaking Negroid Bangui population is trully different from other Negroids, it is possible that gene flow from this disparate Negroid group could have resulted in the Aka Pygmy gene frequencies reported by Lucotte *et al* (1989). However, this hypothesis does not explain the absence of common haplotypes in the two different Pygmy samples apparently drawn from the same parent population. An alternative explanation is that the allele typing of the Bangui Bantu and Aka Pygmy populations, presented in a single study (Lucotte *et al* 1989), was not consistent with preceding and subsequent studies using the 49a/*Taq*I polymorphism.

The cluster analysis of the 23 different population groups proved to be sensitive to small changes in haplotype frequency, suggesting that the data were not very robust at this level of analysis. The different Negroid tribal groups did not cluster in the cohesive groups as expected from their linguistic classification, although there was a clear split into Eastern and Western Bantu groups in the Negroid cluster. However, studies of mitochondrial (mt) DNA polymorphisms in southern African populations have also not delineated populations strictly according to linguistic classification (H.Soodyall, personal communication), despite the increased mutation rate of

mtDNA, and thus the expected increase in sensitivity of mtDNA polymorphism over nuclear DNA polymorphisms. Another problem to be considered in such haplotype analysis of populations is the classification of the groups themselves. Harding and Sokal (1988) reported that genetic distances between European language families did not reflect their accepted linguistic relationships. Languages evolve more rapidly than genes, and can undergo rapid replacement, even by an invading minority (Cavalli-Sforza *et al* 1988), and thus the present language classification of a group may not reflect its genetic past.

The cluster analysis of larger population groupings proved to be less sensitive to minor frequency changes, and revealed some interesting associations (Figure 5.2). Firstly, all the Negroid groups form a distinct cluster separate from the Khoisan, Caucasoids and hybrid groups. The clustering of the enigmatic Khoisan-speaking Dama with the Eastern Bantu suggests that this population may originate from a Bantu-speaking group that followed the eastern pattern of migration from west-central Africa (Huffman 1980). The Pygmy population is placed in the Negroid cluster, although it is slightly distanced from the Negroid groups. This finding corroborates evidence that the Pygmy groups have typical African genetic profiles, and

correlates with suggestions that there has been substantial gene flow from Negroid farmers to Pygmy populations (Cavalli-Sforza 1986).

The presence of the Khoisan populations in the "Caucasoid" cluster was at first perplexing, since, like the Pygmy, the Khoisan are firmly believed to have an African genetic profile (Tobias 1974, Cavalli-Sforza 1986). An earlier analysis of a subset of this frequency data (Spurdle and Jenkins 1992) showed the Sekele San (formerly called Omega San) to cluster with Negroids, and the Tsumkwe San to associate with Caucasoids. The latter finding was attributed largely to sample error, and recent Caucasoid admixture in the Tsumkwe San was considered unlikely (Spurdle and Jenkins 1992). However, an increase in sample size for the Tsumkwe San, and the removal of several individuals of doubtful genetic origin from the Sekele San sample, has resulted in a firm clustering of both of these groups with the Caucasoids (as seen in cluster analysis of the individual population groups). It is interesting to note that, in their study of human evolution, Cavalli-Sforza and coworkers (Cavalli-Sforza *et al* 1988, Cavalli-Sforza 1989) observed different clustering results for their San sample with the bootstrap method of statistical analysis. Although there was a basic African/non-African split, the African cluster was made up of a core of four

populations: Pygmies, west Africans, Bantus and Nilosaharans. Both the San and Ethiopians showed reciprocal affinity and joined the other Africans in only 54% of the bootstraps, otherwise tending to join the Caucasoids. This observation was taken to suggest that these two populations were intermediate between Africans and southwestern Asians (Cavalli-Sforza *et al* 1988, Cavalli-Sforza 1989). Although the result was not surprising for Ethiopians, who have had considerable contact with southern Arabia for at least 3000 years (Cavalli-Sforza 1989), it was unexpected in the case of the southern African Khoisan, whose distinctiveness has always been recognized (Nurse *et al* 1985). Cavalli-Sforza (1989) has suggested that ancient admixture with Caucasoids may have occurred 10000 or more years ago, when the Khoisan were presumably widely distributed and perhaps located in an area intermediate between Africa and south-west Asia. The date of 10000 years is approximate, and based on dates of the oldest fossil appearance in Egypt and Ethiopia (Cavalli-Sforza 1989).

The position of the Khoi Nama in the Caucasoid cluster may also be justified by the above explanation, although recent admixture of Nama males with the Richtersveld and Johannesburg "Coloured" groups is also possible. The presence of the "Coloured" groups in the "Caucasoid" cluster, away from the Negroid populations,

affirms the theory that males were responsible for the Caucasoid contribution to the admixture in the hybrid groups. The two hybrid groups cluster closely together on account of their similiar frequencies of Ht4 and Ht5. The S.A. Jewish and S.A. Asiatic Indian populations also cluster close together, due to similiar frequencies of Ht8, Ht11 and Ht24. The presence of the Lemba population in the Caucasoid cluster strengthens arguments that this Bantu-speaking group is of alien origin. The S.A. European population clusters relatively distantly from the "Coloured" and remaining Caucasoid groups, a phenomenon probably best explained by the heterogeneous origins of this population, consisting of historical and more recent immigrants from a number of European countries.

The cluster analysis of 49a/*Taq*I haplotype frequencies in southern African populations apparently fails to provide support for the African origin of modern man, since there is not a clear African/non-African split. However, only the presence of the Khoisan populations in the "Caucasoid" cluster requires explanation. Should theories put forward by Cavalli-Sforza (Cavalli-Sforza *et al* 1988, Cavalli-Sforza 1989) concerning ancient Caucasoid admixture in the Khoisan gain support, then the data are in agreement with evolutionary studies

based on nuclear and mitochondrial DNA markers (Wainscoat *et al* 1986, Cann *et al* 1987, Vigilant *et al* 1990).

5.4 SUMMARY

Y-specific 49a/*Taq*I haplotypes were determined for 933 individuals drawn from 23 different African populations. A total of 31 new haplotypes were observed, some of which contained new alleles or allelic variants. Duplication, in addition to C_pG mutation, is implicated in the generation of certain allelic variants. Cluster analysis of genetic distances between populations was calculated using the 49a/*Taq*I haplotype frequencies. Cluster analysis of the individual populations was sensitive to minor changes in population frequency, and did not delineate the Bantu-speaking Negroid populations strictly according to linguistic classification, although there was a clear split into Eastern and Western Bantu groups in the Negroid cluster. It is suggested that the Y-specific 49a/*Taq*I haplotype analysis of these populations is not sufficiently sensitive to accurately distinguish between the different Bantu-speaking Negroid tribal groups. Cluster analysis of larger groupings is more stable, and with the exception of the Khoisan, there was a basic split between African and non-African populations. The association of the Khoisan

with the Caucasoids is in agreement with the hypothesis put forward by Cavalli-Sforza that ancient admixture may have occurred between these two groups. Hybrid groups also cluster with the Caucasoid groups, indicating that male gene flow has occurred from the latter into the former, and the positioning of the Lemba population with the Caucasoids strengthens arguments that this alien group has Caucasoid origins.

CHAPTER SIX

THE *Bgl*II, *Hin*dIII, *Pst*I, *Pvu*II AND *Sst*I 49a RFLPS AND THEIR RELATIONSHIP TO THE *Taq*I POLYMORPHISM

6.1 INTRODUCTION

The generation of the 49a/*Taq*I RFLPs detected at each 49a "locus" was initially proposed to be due to point mutations within *Taq*I restriction sites (Ngo *et al* 1986), arising in, or leading to, fragments that could comigrate with other major non-polymorphic bands like E, N and O, and thus remain undetected. The possibility of insertion/deletion or rearrangement processes as a mechanism for generating the observed polymorphism was excluded on the premise that variations of hybridization patterns were not observed with other restriction digests (Ngo *et al* 1986). However, co-existence of "alleles" at the A, D and F loci (Spurdle *et al* 1989, Torroni *et al* 1990, Spurdle and Jenkins 1992, this study) has led to a reassessment of the nature of the polymorphism detected by 49a, since these co-existing "alleles" are believed to be due to duplication processes. It is thus of great interest that probes p49a and p49f have since been shown to

polymorphisms and their correlation with *Taq*I haplotypes is discussed, and their use in population studies is also indicated.

6.2 RESULTS

Y chromosome probe p49a was used to screen a range of restriction enzyme digests for RFLPs. Restriction endonucleases *Bam*HI, *Bgl*I, *Eco*RI, *Hae*III, *Msp*I, *Rsa*I and *Xba*I revealed no variation, whilst complex polymorphisms were detected with *Bgl*II, *Hin*dIII, P*st*I, *Pvu*II and *Sst*I. Y-specific Mendelian inheritance was observed for all the polymorphisms.

Each polymorphism consists of several variant fragments, ranging in number from 11 bands for *Sst*I, to 19 for *Bgl*II. The polymorphisms are shown in Figures 6.1-6.5, with details of fragment sizes indicated alongside. At least one constant Y-specific fragment is observed in the *Bgl*II, *Hin*dIII, *Pst*I and *Pvu*II digests. In the case of the *Pst*I polymorphism, the autosomal band is not detected under the conditions of electrophoresis (see section 3.3.3.1), since it is below the size range revealed (Figure 6.3). *Sst*I does not appear to reveal a constant band, although the presence of a constant band <1.2kb in size cannot be

(Cohenhagenhauer *et al* 1987), and an autosomally derived fragment is seen in the digests of female controls for *Bgl*II, *Hin*dIII, *Pst*I and *Pvu*II. The situation concerning autosomally-inherited fragments is very different for *Sst*I. Whereas the autosomal band is of constant size for *Bgl*II, *Hin*dIII, *Pst*I and *Pvu*II, at least two different autosomal fragments are revealed by *Sst*I. In most instances females possess band *e* only, but band *f* has also been observed in females. The two bands were not observed together in a sample of 20 Caucasoid and Negroid females, and screening also did not reveal any other possible autosomal fragments. However, certain males possess neither of the two autosomal bands (Figure 6.3), suggesting that an autosomal band may not always be detected. Furthermore, since fragments *e* and *f* occur together in several individuals, and band *e* is seen more commonly as the autosomal fragment in females (data not shown), the possibility that Y-specific fragments may mimic the autosomal fragment *f* cannot be excluded.

The frequency of the different Y-specific *Pvu*II fragments detected by 49a was calculated in order to determine the informativity of such data in population studies (Figure 6.6). The arrangement of polymorphic

for the five enzyme systems are defined in Tables 6.1-6.5, and their frequencies in the three major population groups studied are displayed in Figure 6.7. Each polymorphic system reveals differences in band intensity between haplotypes for a single fragment size (Figures 6.1-6.5), but the band intensity observed for a given haplotype appears to be consistent. In the case of *Sst*I, the haplotype included the "autosomal" fragments *e* and *f* because, due to the complications discussed above, it could not be specifically determined if they were autosomal or Y-specific. This means that the *Sst*I haplotype is not necessarily only Y-specific, but may represent autosomal variation as well. A profile of the different 49a RFLPs was determined for those individuals with results available for all the 49a systems (Table 6.6). Nei's measure of genetic diversity (Nei 1987) was used to determine the genetic variation displayed by the compound haplotype, and by the individual RFLPs. Genetic diversity was calculated for each population, and for the system overall (Table 6.7).

A comparative analysis of *Taq*I and *Pvu*II polymorphisms was also effected in two discrete population studies. The first study was undertaken to determine if a group

suspected from historical data (Bernstein *et al* 1986). The 8 inv(Y) Gujerati Muslim males possessed identical p49a/*Taq*I and p49a/*Pvu*II haplotypes (Table 6.8, Figure 6.8), as opposed to the 7 *Taq*I and 8 *Pvu*II haplotypes observed in 9 normal random Gujerati Muslim males. The second study involved an analysis of *Pvu*II haplotypes associated with the *Taq*I Hts 7 and 8 in the Lemba and Jewish populations mainly. These haplotypes are found commonly in South African Asiatic Indian and Jewish Caucasoids, and are also found in the enigmatic Bantu-speaking Lemba population. The latter group considers itself to be one of the lost Jewish tribes, but is believed by ethnographers to rather have Arabic origins (see section 3.1.4 for a full description). The sample of individuals possessing Ht7 or Ht8 also included a few Khoisan or Khoisan-speaking Negroid individuals, and an Asiatic Indian from the sample of random Caucasoids. It is apparent that the Lemba and Jewish samples share only a single *Pvu*II haplotype associated with *Taq*I Ht7 or Ht8 (Table 6.9), found in a single Jewish individual with Ht8.

FRAGMENT	SIZE(kb)	NATURE
a	>48.5*	P
b	>48.5	P
c	>48.5	P
d	>48.5	P
e	37.0	P
f	23.0	P
g	20.8	P
h	19.5	P
i	18.5	P
j	17.5	P
k	16.3	P
l	15.4	P
m	13.2	P
n	11.3	P
o	10.6	P
p	10.0	P
q	9.1	P
r	6.9	P
s	5.4	P
t	4.4	P
u	4.2	C
v	3.3	A

*Fragment migrates slower than undigested lambda.

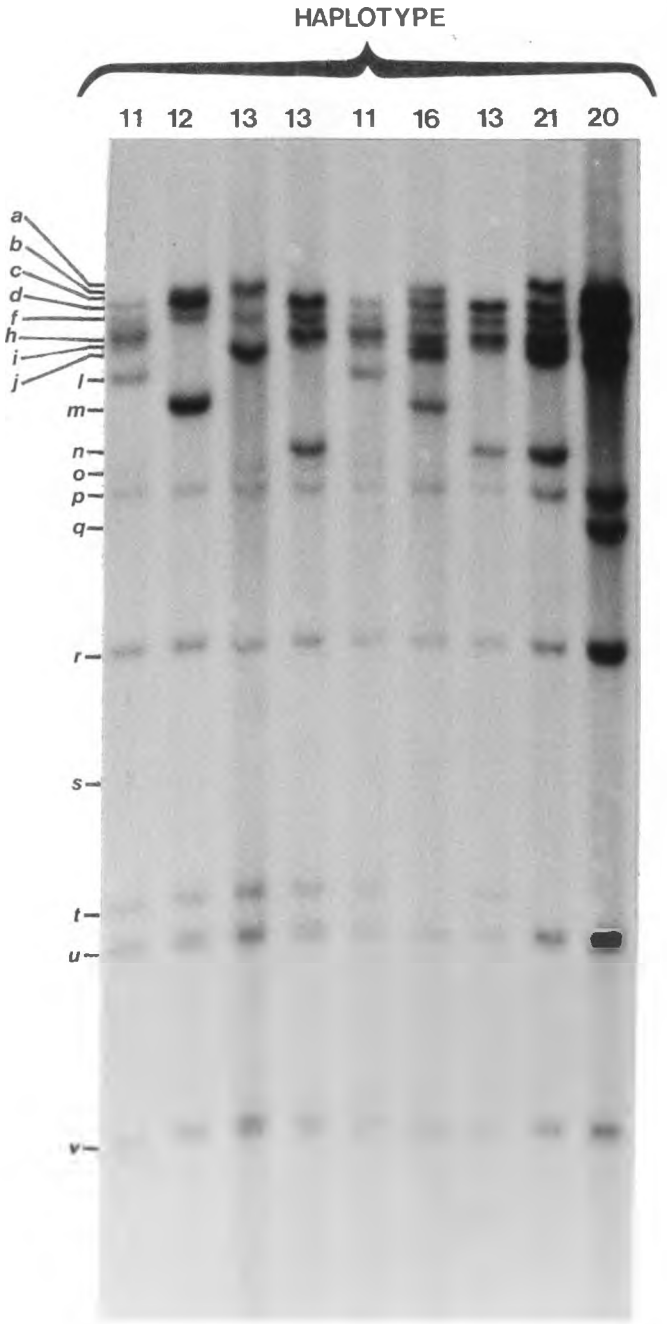

FIGURE 6.1 Hybridization pattern revealed by 49a with *BglII*. Fragment sizes are indicated on the left, with the nature of the fragments (P=Polymorphic, C=Constant, A=Autosomal).

FRAGMENT	SIZE(kb)	NATURE
a	>48.5*	P
b	>48.5	P
c	>48.5	P
d	25.0	P
e	21.4	P
f	19.5	P
g	17.0	P
h	12.3	P
i	11.8	P
j	9.3	P
k	8.5	P
l	7.7	A
m	6.0	P
n	4.4	P
o	3.5	C

*Fragment migrates slower than undigested lambda.

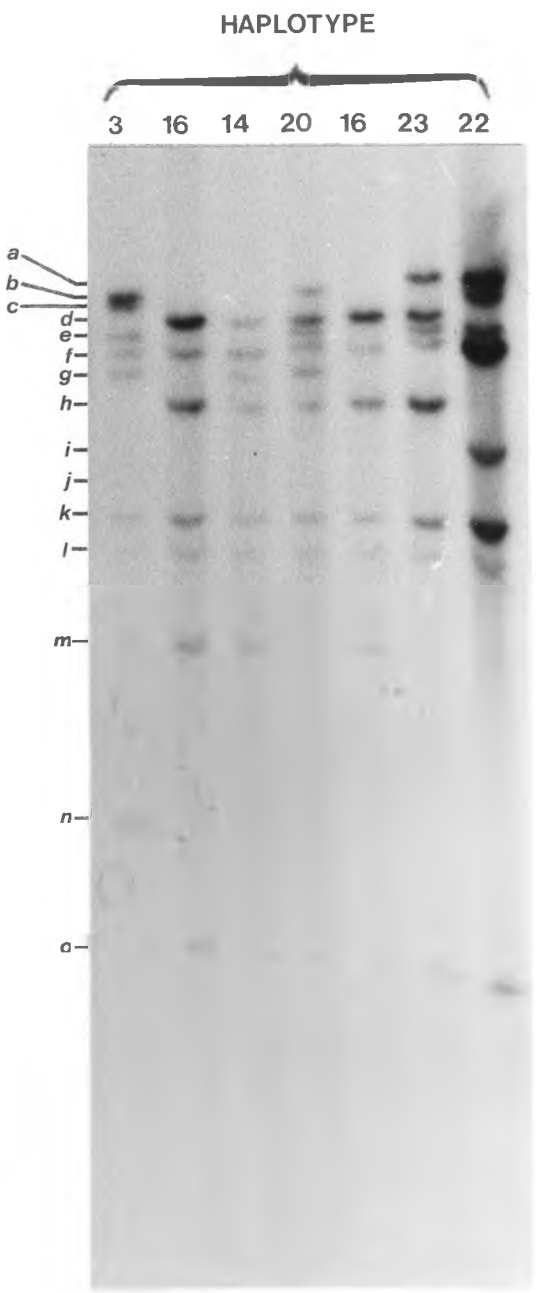

FIGURE 6.2 Hybridization pattern revealed by 49a with *Hind*III. Fragment sizes are indicated on the left, with the nature of the fragments (P=Polymorphic, C=Constant, A=Autosomal).

FRAGMENT	SIZE(kb)	NATURE
a	30.0	P
b	23.1	P
c	18.0	P
d	16.5	P
e	15.5	P
f	14.2	P
g	13.5	P
h	12.3	P
i	12.0	P
j	11.7	P
k	11.2	P
l	9.0	P
m	8.6	P
n	7.8	P
o	7.2	P
p	6.8	P
q	6.3	P
r	5.3	C
s	5.1	C
t	4.8	P
u	3.8	P
v	3.5	C

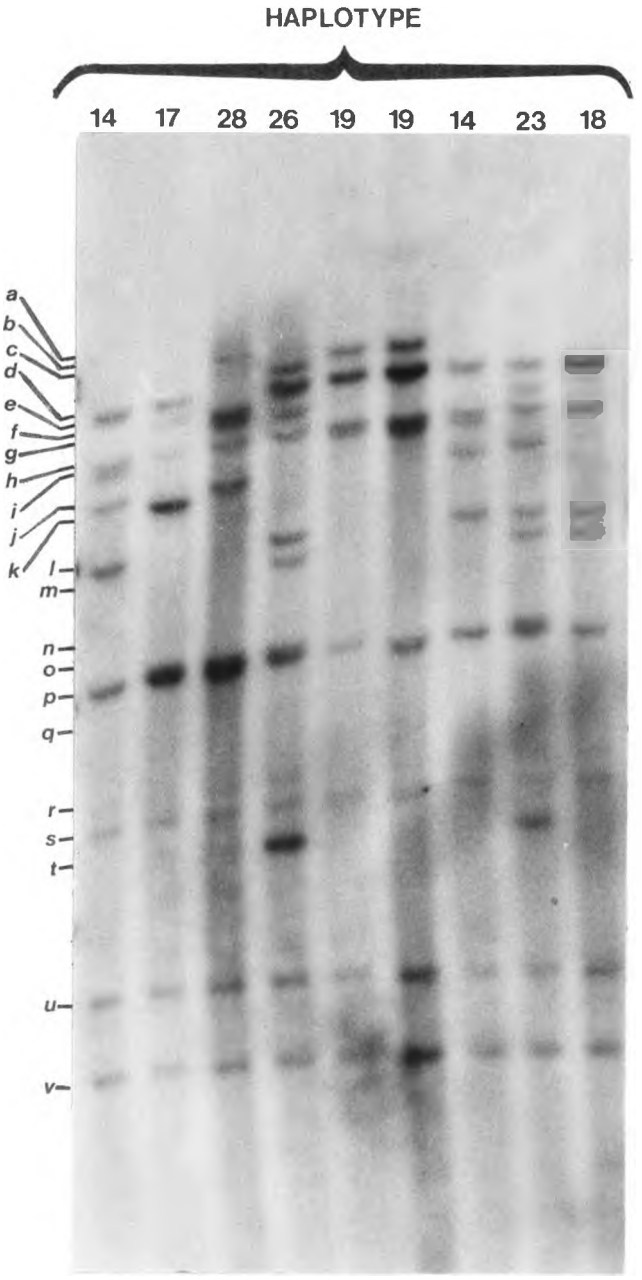

FIGURE 6.3 Hybridization pattern revealed by 49a with *Pst*I. Fragment sizes are indicated on the left, with the nature of the fragments (P=Polymorphic, C=Constant, A=Autosomal).

FRAGMENT	SIZE(kb)	NATURE
A[+]	>48.5[*]	P
A	>48.5	P
B	>48.5	P
B2	24.0	P
C	20.2	P
D	18.5	P
D2	16.0	P
E	15.0	P
E2	13.5	P
F	12.8	P
F2	11.0	P
G	10.5	P
G2	8.8	P
H	8.3	P
I	7.7	C
J	6.0	P
K	5.6	C
L	4.7	A
L2	4.6	P
M	3.6	P

[*]Fragment migrates slower than undigested lambda.

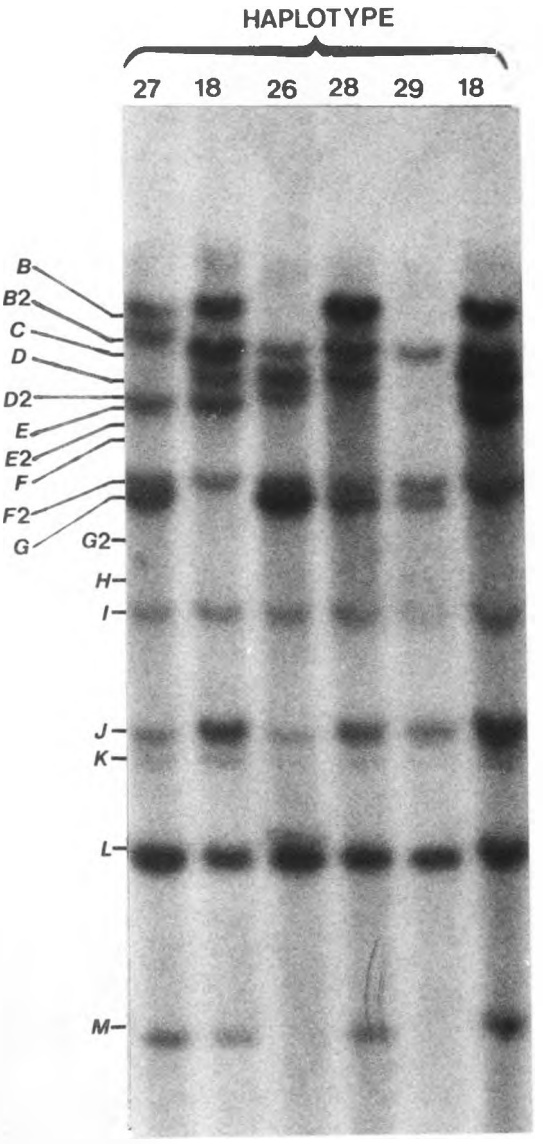

FIGURE 6.4 Hybridization pattern revealed by 49a with *PvuII*. Fragment sizes are indicated on the left, with the nature of the fragments (P=Polymorphic, C=Constant, A=Autosomal).

FRAGMENT	SIZE(kb)	NATURE
a	>48.5*	P
b	>48.5	P
c	>48.5	P
d	41.0	?A
e	27.0	?A
f	20.0	P
g	17.0	P
h	14.5	P
i	12.5	P
j	10.2	P
k	8.5	P

*Fragment migrates slower than undigested lambda.

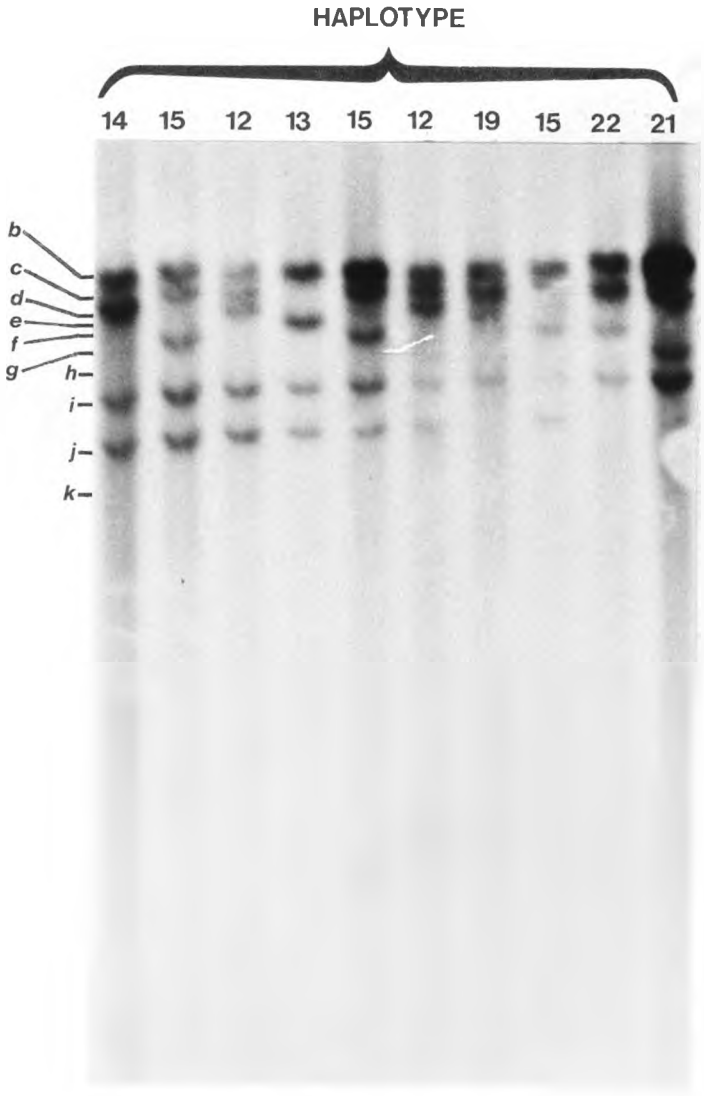

FIGURE 6.5 Hybridization pattern revealed by 49a with *Sst*I. Haplotype numbers are indicated to the left of each autoradiograph. Sizes of fragments (in kb) are indicated to the left of each autoradiograph, together with the nature of the fragment (P=Polymorphic, C=Constant, A=Autosomal).

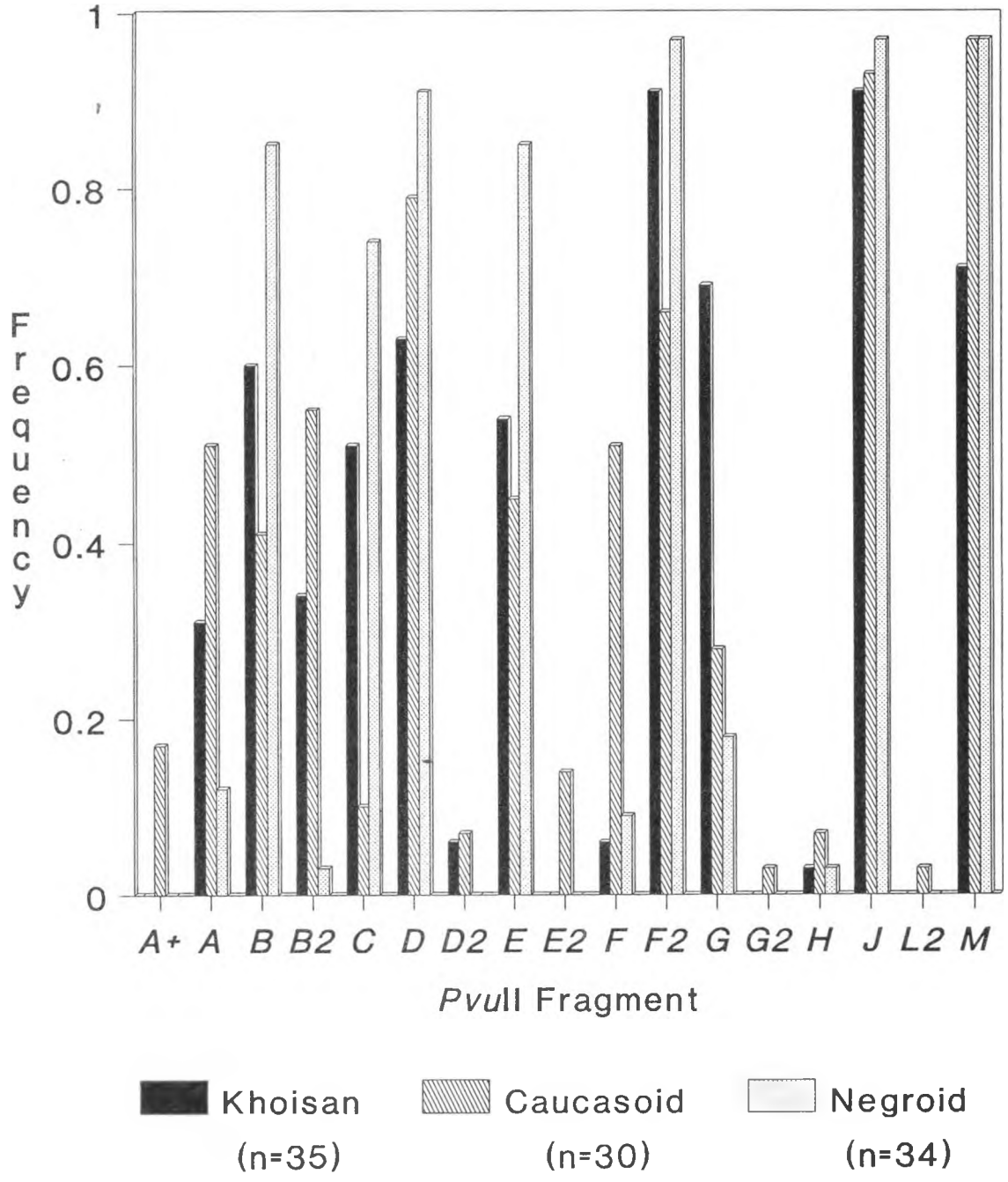

FIGURE 6.6 Frequency of the 49a/PvuII fragments in the Khoisan, Caucasoid and Negroid populations.

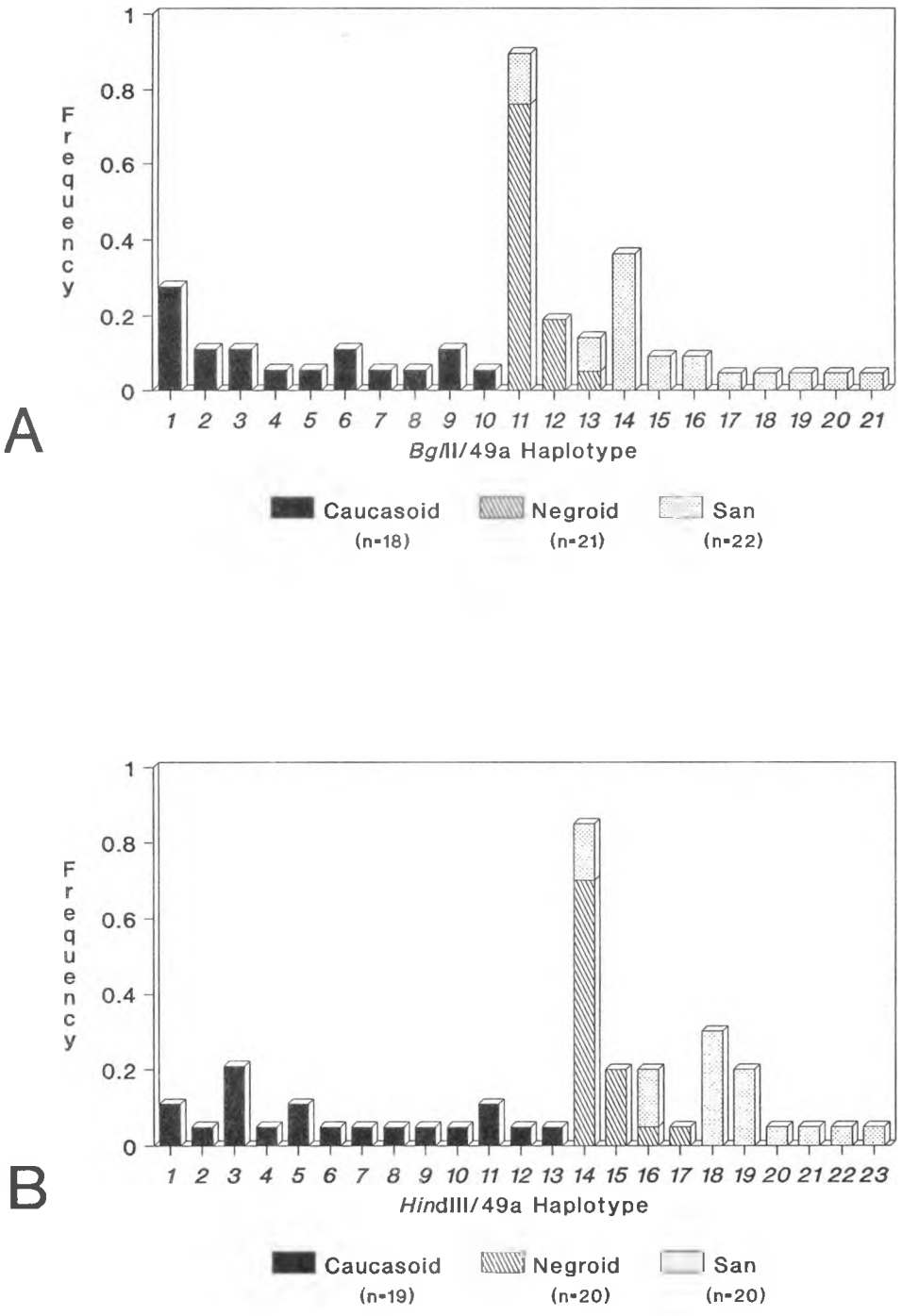

FIGURE 6.7 Haplotype frequencies for 49a polymorphisms in the Caucasoid, Negroid and Khoisan populations. A) *Bgl*II B) *Hind*III

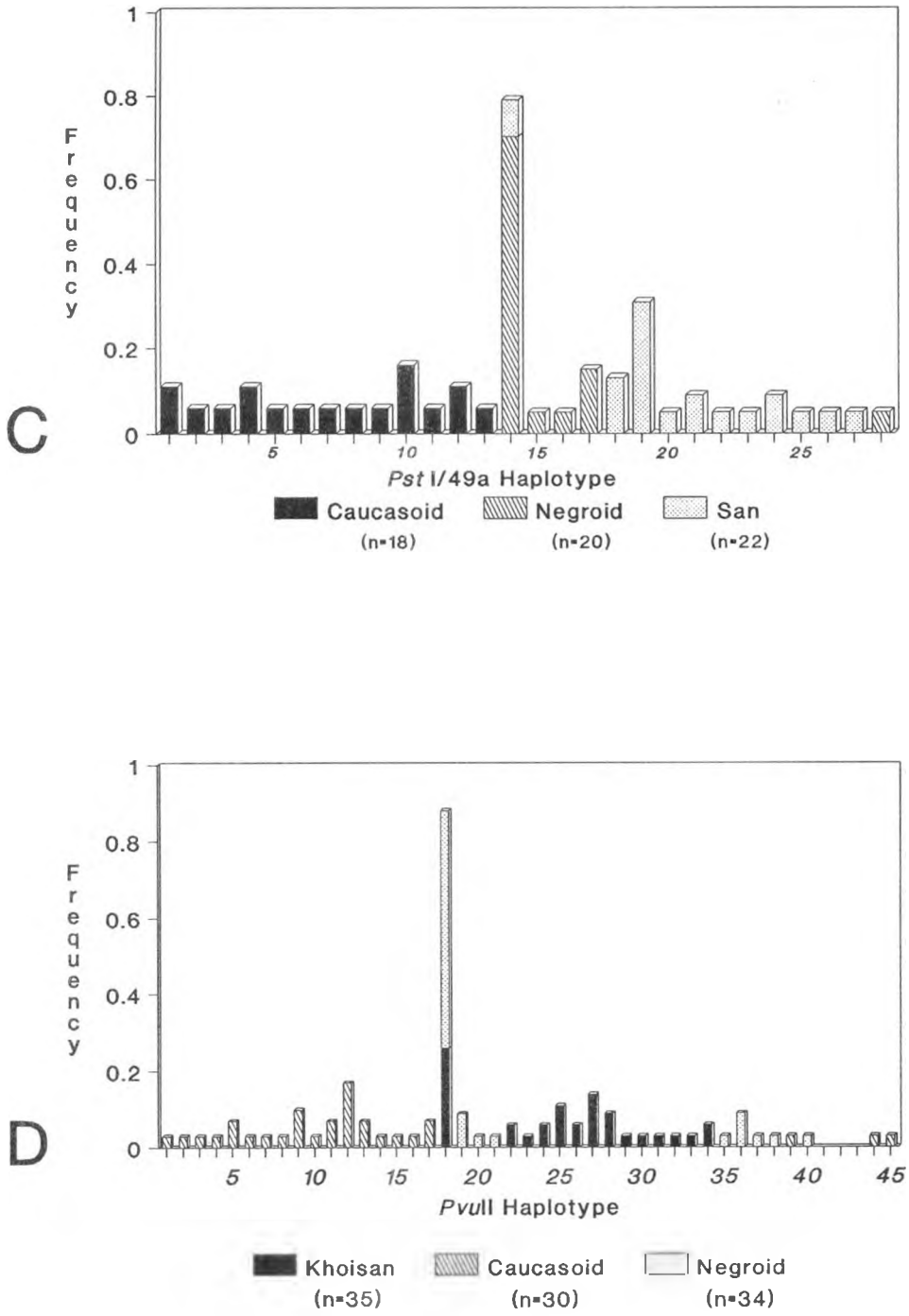

FIGURE 6.7 continued. Haplotype frequencies for 49a polymorphisms. C) *Pst*I D) *Pvu*II

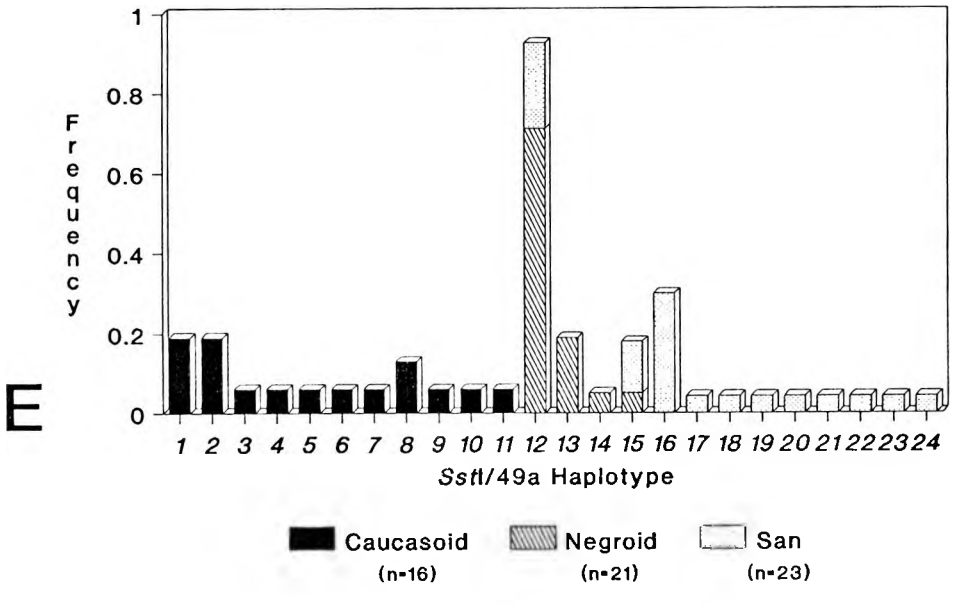

FIGURE 6.7 continued. Haplotype frequencies for 49a polymorphisms. E) *Sst*I.

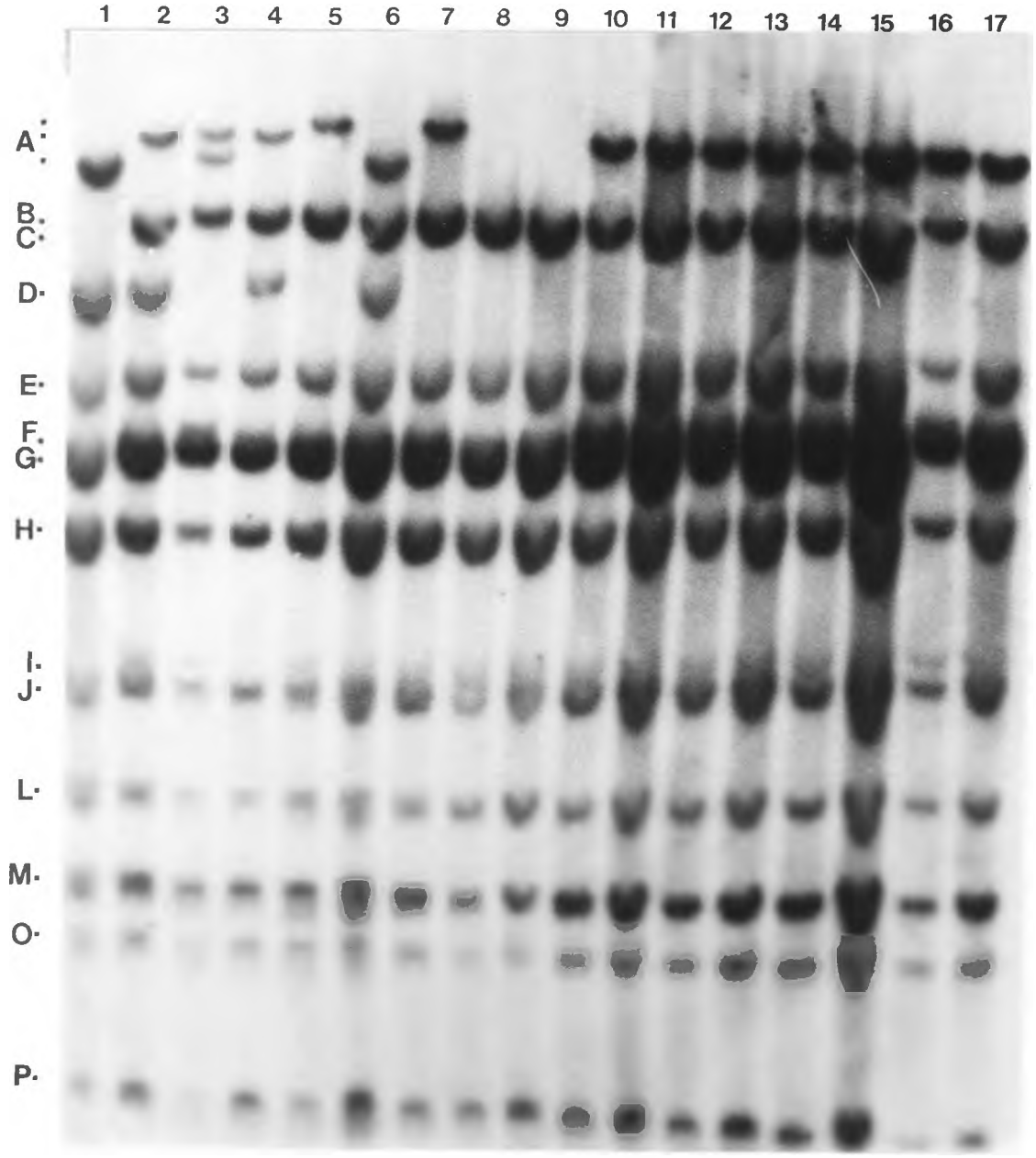

FIGURE 6.8 Y-specific 49a/*Taq*I and 49a/*Pvu*II hybridization patterns detected in inv(Y) and normal Gujerati Muslim males. Fragments are indicated to the left of the autoradiograph, and individuals are numbered at the top of each lane. A) *Taq*I hybridization patterns.

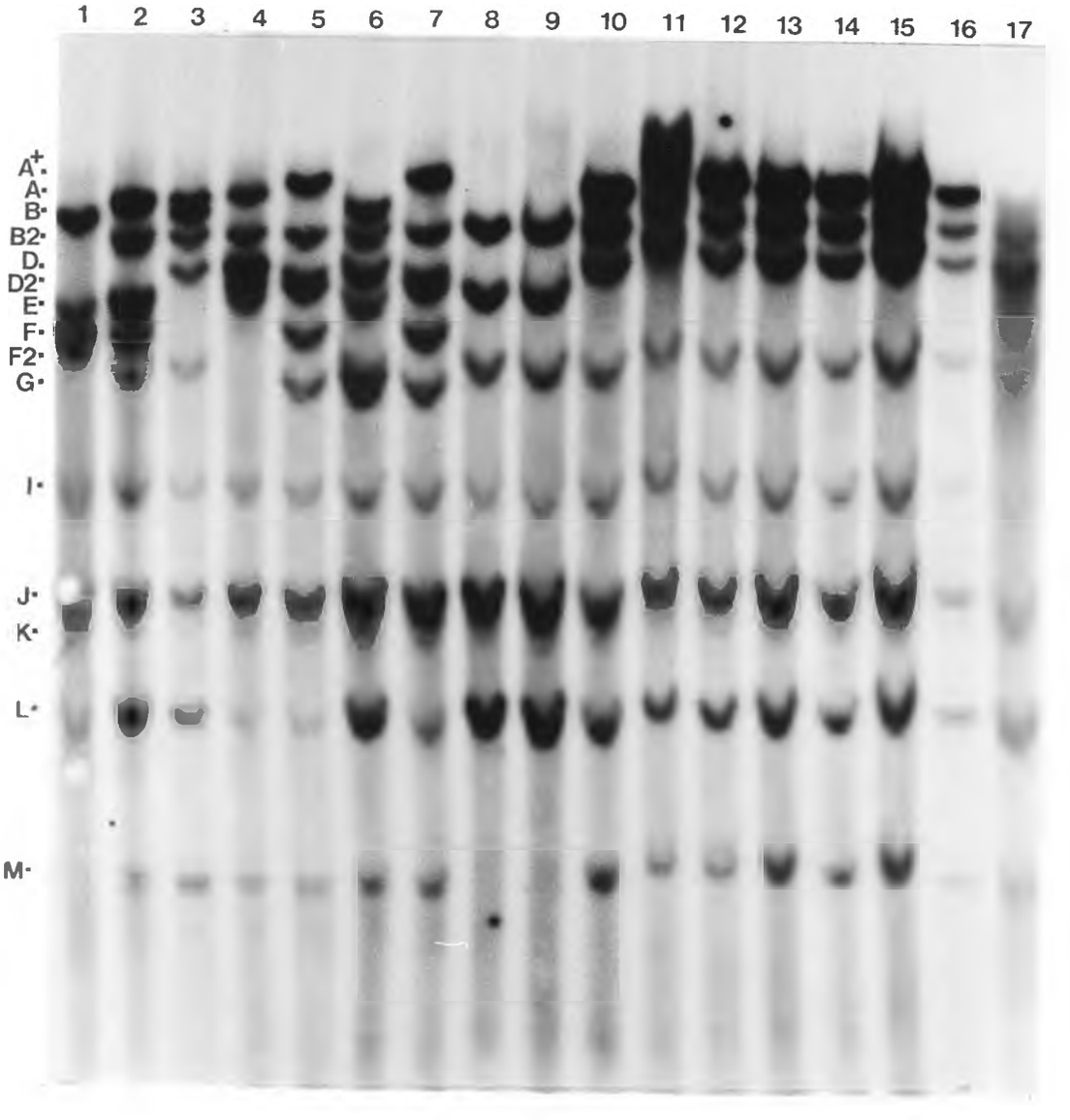

FIGURE 6.8 continued. Y-specific 49a/*Taq*I and 49a/*Pvu*II hybridization patterns detected in inv(Y) and normal Gujerati Muslim males. B) *Pvu*II hybridization patterns.

TABLE 6.1

*Bgl*II/49a Haplotypes (Ht)

	Polymorphic fragments[a]																	
Ht	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	q	r	s	t
1	*	*	c	*	*	f	*	*	*	j	*	*	m	*	o	*	r	*	t
2	*	b	*	*	*	f	*	*	*	j	*	*	m	*	o	*	r	*	t
3	*	b	*	d	*	f	*	*	*	*	*	l	m	*	o	*	r	*	t
4	a	*	c	*	*	f	*	*	*	j	*	*	*	*	o	*	r	*	t
5	*	*	*	d	*	f	*	*	*	j	*	*	m	*	o	*	r	*	t
6	*	b	*	*	*	*	*	*	*	j	*	*	m	*	o	*	r	*	t
7	*	*	c	*	*	f	*	*	*	j	k	*	*	*	o	*	r	*	t
8	*	*	c	*	e	f	*	*	*	j	k	*	*	*	*	*	r	s	t
9	*	*	*	d	*	f	*	h	*	j	k	*	*	*	o	q	*	*	t
10	*	*	c	*	*	f	*	*	*	j	*	*	*	*	o	*	r	*	t
11	*	*	*	d	*	f	*	h	i	*	*	l	*	*	o	*	r	*	t
12	*	*	c	d	*	f	*	*	*	*	*	*	m	*	*	*	r	*	t
13	*	b	*	*	*	f	*	*	*	j	*	*	*	*	o	*	r	*	t
14	*	*	c	d	*	f	*	*	*	*	*	l	*	*	*	*	r	*	t
15	*	*	*	d	*	f	g	*	*	j	*	l	*	*	o	*	r	*	t
16	*	b	*	d	*	f	*	h	i	j	*	*	m	*	*	*	r	*	*
17	*	*	c	*	*	f	*	h	i	j	*	*	*	n	o	*	r	*	*
18	*	*	*	d	*	f	*	h	i	*	*	*	*	n	o	*	r	*	t
19	*	*	*	d	*	f	*	*	*	j	*	l	*	*	o	*	r	*	t
20	*	b	*	*	*	f	*	*	*	j	*	*	*	*	*	q	r	*	*
21	a	*	*	d	*	f	*	h	i	j	*	*	*	n	*	*	r	*	*

[a]Absence of bands is indicated by an asterisk (*).

TABLE 6.2

*Hind*III/49a Haplotypes

	Polymorphic fragments[a]												
Ht	a	b	c	d	e	f	g	h	i	j	k	m	n
1	*	b	c	*	e	f	g	*	*	*	k	m	*
2	*	b	c	d	e	f	g	*	*	*	k	m	*
3	*	b	c	*	e	f	g	*	*	*	k	m	n
4	*	*	c	*	e	f	g	*	*	*	k	m	*
5	*	b	*	d	*	f	g	*	*	*	k	m	*
6	a	*	c	*	e	f	*	*	*	*	k	m	*
7	*	*	c	d	e	f	g	*	*	*	k	m	*
8	*	b	*	*	e	*	g	*	*	*	k	m	*
9	*	*	c	*	e	f	*	*	i	*	k	m	*
10	*	b	*	*	e	f	*	*	i	j	k	m	*
11	*	*	*	d	e	f	*	*	i	*	*	m	*
12	*	*	*	d	*	*	*	*	i	*	*	m	*
13	*	*	c	*	e	f	*	*	*	*	k	m	*
14	*	*	*	d	*	f	g	h	*	*	k	m	*
15	*	*	c	d	*	f	g	*	*	*	k	m	*

TABLE 6.2 continued

16	*	*	*	d	*	f	*	h	*	*	k	m	*
17	*	b	c	*	e	f	*	*	*	*	k	m	*
18	*	*	c	d	*	f	*	*	*	*	k	m	*
19	*	*	*	d	e	f	g	h	*	*	k	m	*
20	*	b	*	d	e	f	g	h	*	*	k	*	*
21	*	*	c	d	e	f	*	h	*	*	k	*	n
22	a	*	c	*	e	f	*	*	i	*	k	m	*
23	a	*	*	d	e	f	*	h	*	*	k	*	*

[a]Absence of bands is indicated by an asterisk (*).

TABLE 6.3

PstI/49a Haplotypes

	Polymorphic fragments[a]																
Ht	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	q	t	u
1	*	b	*	*	*	*	g	h	*	j	k	*	*	*	*	*	*	u
2	a	*	*	*	*	f	g	h	*	j	k	*	*	n	*	*	*	u
3	*	*	*	*	*	f	g	h	*	*	k	*	*	*	*	*	*	u
4	*	b	*	*	*	f	*	h	i	j	k	*	*	*	*	*	*	u
5	*	b	*	*	*	f	g	h	*	*	*	*	*	*	*	*	*	u
6	*	*	*	*	*	*	g	h	*	j	k	*	*	*	*	*	*	u
7	*	*	c	*	*	f	g	*	*	j	k	*	*	*	*	*	*	u
8	*	b	*	*	*	f	*	h	*	j	k	*	*	*	*	*	*	u
9	*	*	*	*	*	f	g	h	*	j	*	*	*	*	*	*	*	u
10	*	b	*	*	*	f	g	h	*	j	k	*	*	*	*	*	*	u
11	*	b	*	*	*	*	g	h	*	j	*	*	*	*	o	*	*	u
12	*	*	*	*	e	*	g	h	*	j	*	*	*	*	*	q	*	*
13	*	*	*	*	*	f	g	h	*	*	*	*	*	*	*	*	*	u
14	*	*	*	d	*	*	*	h	i	j	*	l	*	*	*	*	*	u
15	*	b	*	d	*	f	*	h	*	*	k	*	*	*	*	*	*	u

TABLE 6.3 continued

16	*	*	*	*	e	*	*	h	i	j	*	l	*	*	*	*	u
17	*	*	*	d	*	f	*	h	*	*	k	*	*	*	*	*	u
18	*	*	*	d	*	*	*	h	*	*	*	l	m	*	*	*	u
19	*	b	*	d	*	*	*	h	i	*	*	*	*	*	*	*	u
20	*	*	*	d	*	*	g	h	i	j	*	*	m	*	*	*	u
21	*	*	*	d	*	f	g	h	*	*	k	l	*	*	*	t	u
22	*	*	*	d	*	*	*	h	*	j	*	l	*	*	*	*	u
23	*	*	*	d	*	*	g	h	*	j	*	l	m	*	*	t	u
24	*	*	*	d	*	*	g	h	i	j	*	l	*	*	*	*	u
25	*	*	*	*	*	f	g	h	*	*	*	*	*	*	q	*	u
26	*	*	c	*	e	*	g	h	*	*	*	l	m	*	*	t	u
27	*	*	c	d	*	*	g	h	*	*	*	l	m	*	*	t	*
28	a	*	*	*	*	f	g	h	*	j	*	*	*	*	*	*	u

[a]Absence of bands is indicated by an asterisk (*).

TABLE 6.4

PvuII/49a Haplotypes

	Polymorphic Fragments[a]															
Ht	A+	A	B	B2	C	D	D2	E	E2	F	F2	G	G2	H	J	L2	M
1	*	*	B	*	*	*	*	E	*	F	*	*	*	*	J	*	M
2	*	A	*	B2	*	*	*	E	*	F	F2	*	*	*	J	*	M
3	*	A	B	B2	*	D	*	*	*	*	F2	*	*	*	J	*	M
4	*	A	*	B2	*	D	*	E	*	*	*	*	*	*	J	*	M
5	A^+	*	*	B2	*	D	D2	*	*	F	*	G	*	*	J	*	M
6	*	*	*	B2	*	D	*	E	*	*	F2	G	*	*	J	*	M
7	*	*	*	B2	*	*	*	E	*	*	F2	*	*	*	J	*	M
8	*	*	*	B2	*	*	*	E	*	*	F2	*	*	*	J	*	*
9	*	A	*	B2	*	D	*	*	*	*	F2	*	*	*	J	*	M
10	A^+	*	*	*	C	D	*	*	*	*	*	*	*	*	J	*	M
11	A^+	*	B	*	*	*	*	E	*	F	F2	*	*	*	J	*	M
12	*	A	*	*	*	D	*	*	*	F	F2	G	*	*	J	*	M
13	*	A	B	*	*	D	*	*	*	F	F2	*	*	*	J	*	M
14	*	*	B	B2	*	D	*	*	*	F	F2	*	*	*	J	*	M
15	*	A	*	B2	*	D	*	E	*	*	F2	*	*	*	J	*	M
16	*	*	B	B2	*	D	*	E	E2	*	*	*	*	*	J	L2	M
17	*	*	B	B2	C	D	*	E	E2	*	*	*	*	H	J	*	M
18	*	*	B	*	C	D	*	E	*	*	F2	*	*	*	J	*	M

TABLE 6.4 continued

19	*	A	B	*	*	*	*	*	*	F	F2	*	*	*	J	*	M
20	*	*	*	*	C	D	*	E	*	*	F2	*	*	*	J	*	M
21	*	A	*	*	*	D	*	*	*	*	F2	*	*	*	J	*	M
22	*	A	*	B2	*	*	*	*	*	*	*	*	*	*	*	*	M
23	*	*	*	*	*	D	*	*	*	*	F2	G	*	*	J	*	*
24	*	A	B	*	*	D	*	*	*	*	F2	G	*	*	J	*	*
25	*	A	*	B2	*	*	*	E	*	*	F2	G	*	*	J	*	M
26	*	*	*	*	C	D	D2	*	*	*	F2	G	*	*	J	*	*
27	*	*	B	B2	*	*	*	E	*	*	F2	G	*	*	J	*	M
28	*	*	B	*	C	D	*	*	*	*	F2	G	*	*	J	*	M
29	*	*	*	*	C	*	*	*	*	*	F2	G	*	*	J	*	*
30	*	A	*	*	*	*	*	E	*	*	F2	G	*	*	J	*	M
31	*	A	B	*	*	D	*	*	*	*	*	G	*	*	*	*	M
32	*	A	*	B2	*	D	*	*	*	*	F2	G	*	H	J	*	*
33	*	*	B	*	C	D	*	*	*	*	F2	G	*	*	J	*	*
34	*	*	*	*	C	D	*	*	*	F	F2	G	*	*	J	*	*
35	*	*	B	B2	C	D	*	E	*	*	F2	*	*	*	J	*	M
36	*	*	B	*	*	D	*	E	*	*	F2	G	*	*	J	*	M
37	*	*	*	*	C	D	*	E	*	*	F2	G	*	*	*	*	*
38	*	*	*	*	C	D	*	E	*	*	*	*	*	*	J	*	M
39	*	A	B	*	*	D	*	*	E2	F	*	*	G2	*	J	*	M
40	*	*	*	*	*	D	*	*	*	*	F2	G	*	H	J	*	M
41	*	A	B	*	C	*	*	E	*	*	F2	G	*	*	J	*	M
42	*	A	B	*	C	D	*	*	*	*	*	*	*	*	J	*	M
43	*	A	*	B2	*	D	*	E	*	*	*	*	*	H	J	*	M

TABLE 6.4 continued

44	*	*	B	*	*	D	*	E	*	*	*	*	*	*	J	*	M
45	A[+]	*	*	B2	C	D	*	E	*	*	F2	*	*	*	J	*	M
46	*	*	B	B2	*	D	*	E	*	*	F2	G	*	*	J	*	M
47	*	*	B	B2	*	D	*	E	*	*	F2	*	*	*	J	*	M
48	*	*	B	B2	*	D	*	E	*	*	*	*	*	*	J	*	M
49	*	*	*	B2	*	D	D2	E	*	F	*	*	*	*	J	*	M
50	*	*	B	B2	*	D	D2	E	*	*	*	*	*	*	J	*	M
51	*	*	B	B2	*	D	*	*	*	F	*	*	*	*	J	*	*
52	*	*	B	B2	*	D	*	*	*	*	F2	*	*	*	J	*	M
53	*	*	B	B2	*	D	D2	E	*	F	*	*	*	*	J	*	M

[a]Absence of bands is indicated by an asterisk (*).

TABLE 6.5

*Sst*I/49a Haplotypes

	Polymorphic fragments[a]										
Ht	a	b	c	d	e	f	g	h	i	j	k
1	a	*	*	d	*	f	*	*	i	j	*
2	a	b	*	d	*	f	*	*	i	j	*
3	a	*	*	*	e	f	*	*	*	*	*
4	a	b	*	d	*	*	*	*	i	j	*
5	a	b	*	*	e	f	*	*	i	j	*
6	a	b	*	d	*	*	*	*	i	j	k
7	a	b	c	d	e	*	*	h	*	j	*
8	*	b	*	d	*	f	*	*	i	j	*
9	*	b	*	d	*	*	*	h	*	j	*
10	a	*	*	d	*	*	*	*	*	j	*
11	a	b	*	d	e	*	*	*	i	j	*
12	*	b	c	*	e	*	*	*	i	j	*
13	*	b	*	*	*	f	*	*	i	j	*
14	*	b	*	d	*	*	*	*	i	j	*
15	*	b	c	*	*	*	g	*	i	j	*

TABLE 6.5 continued

16	*	b	*	*	e	*	*	*	i	j	*
17	*	b	*	d	e	*	*	*	i	j	*
18	a	*	c	*	*	f	*	*	i	*	*
19	*	b	c	*	*	f	*	*	i	*	*
20	*	*	*	*	*	*	g	*	i	*	*
21	*	b	*	d	*	*	*	h	i	*	*
22	*	b	c	d	*	*	g	*	i	*	*
23	a	*	*	*	e	*	*	*	i	j	*
24	*	b	*	*	*	*	*	*	*	j	*

[a]Absence of bands is indicated by an asterisk (*).

TABLE 6.6

Compound 49a Haplotypes

	Haplotypes							
Compound	*Taq*I	*Pvu*II	*Bgl*I	*Hind*III	*Sst*I	*Pst*I	Frequency	Population
1	15	13	1	1	1	1	.02	Caucasoid
2	50	13	1	2	1	1	.02	Caucasoid
3	15	12	2	3	2	2	.02	Caucasoid
4	11	9	1	4	2	3	.02	Caucasoid
5	50	11	3	5	3	4	.02	Caucasoid
6	46	10	4	6	4	5	.02	Caucasoid
7	15	12	2	1	2	8	.02	Caucasoid
8	50	11	3	5	5	4	.02	Caucasoid
9	15	12	6	3	1	10	.02	Caucasoid
10	2	16	8	10	6	11	.02	Caucasoid
11	12	17	9	11	7	12	.04	Caucasoid
12	15	12	1	3	8	10	.04	Caucasoid
13	11	9	10	13	11	13	.02	Caucasoid
14	4	18	11	14	12	14	.17	Negroid/San
15	1	18	11	16	12	14	.02	Negroid
16	15	21	13	17	14	28	.02	Negroid
17	4	18	11	14	12	16	.02	Negroid
18	10	19	12	15	13	17	.06	Negroid
19	3	18	11	14	12	14	.04	Negroid/San

TABLE 6.6 continued.

20	4	20	11	14	12	14	.02	Negroid
21	11	28	13	16	15	18	.04	San
22	30	27	14	18	16	19	.08	San
23	4	18	15	19	17	20	.02	San
24	30	25	14	18	16	19	.04	San
25	4	18	11	14	12	22	.02	San
26	33	26	17	21	20	23	.02	San
27	11	28	18	16	15	18	.02	San
28	4	18	19	19	12	24	.02	San
29	4	18	15	19	12	24	.02	San
30	35	32	20	22	21	25	.02	San
31	54	33	21	23	22	26	.02	San
32	32	34	16	20	19	21	.04	San

TABLE 6.7

Genetic Diversity Values for the 49a Polymorphisms[a]

	Genetic Diversity (n)				Total Number of
Enzyme	Caucasoid	Negroid	San	Overall	Haplotypes Observed
*Bgl*II	0.883(18)	0.393(21)	0.833(22)	0.871(61)	21
*Hin*dIII	0.921(19)	0.476(20)	0.835(20)	0.888(59)	23
*Pst*I	0.932(18)	0.493(19)	0.867(22)	0.904(60)	28
*Pvu*II	0.945(30)	0.603(34)	0.885(35)	0.892(99)	42
*Sst*I	0.911(16)	0.466(21)	0.852(23)	0.863(60)	24
*Taq*I	0.844(19)	0.655(35)	0.843(35)	0.875(89)	24
Compound	0.947(15)	0.717(16)	0.928(20)	0.952(51)	32

[a]Values were determined for *Taq*I using only data from individuals screened for all five other 49a polymorphisms.

TABLE 6.8

Y-specific p49a/*Taq*I and p49a/*Pvu*II Haplotypes Observed in Normal and Inv(Y) Gujerati Muslim males

Individual	Status	49a/*Taq*I Ht[a]	49a/*Pvu*II Ht
1	Normal	53	1
2	Normal	13	2
3	Normal	30	3
4	Normal	12	4
5	Normal	36	5
6	Normal	8	6
7	Normal	36	5
8	Normal	27	7
9	Normal	27	8
10-17	Inv(Y)	11	9

[a] *Taq*I haplotype numbers are those reported in this study, and include haplotype numbers 1 to 24 (Ngo *et al* 1986, Breuil *et al* 1987, Lucotte *et al* 1989, Lucotte *et al* 1990). Torroni *et al* (1990) did not include haplotypes 20 to 24 (Lucotte *et al* 1990b) in their list of haplotypes, thus the haplotype nomenclature is out of phase.

TABLE 6.9

PvuII Haplotypes Associated with *TaqI* Haplotypes Seven and Eight in the Jewish and Lemba Populations

TaqI Ht	*PvuII* Ht	Population (n)
7	47	Jewish (1)
	49	Jewish (1)
	50	Jewish (2)
	51	Jewish (1)
	52	Lemba (6)
	53	Dama (1)
8	6	Indian (1), Jewish (3)
	27	Lemba (3)
	41	Lemba (1), Dama (1)
	42	Lemba (1)
	43	Nama (1)
	46	Jewish (1)
	47	Lemba (4), Jewish (1)
	48	Jewish (1)

6.3 DISCUSSION

6.3.1 The Variability Revealed by 49a

The 49a polymorphisms revealed by *Bgl*II, *Hin*dIII, *Pst*I, *Pvu*II and *Sst*I show a marked similarity to each other. All of these RFLPs consist of a considerable number of fairly large fragments (some greater than 50kb) which can be present or absent in a variety of arrangements (Figures 6.1-6.5, Tables 6.1-6.5). The allelic nature of the fragments cannot be satisfactorily determined given the large number of fragments, the variability of the polymorphism, and the fact that fragments can also differ in intensity between different haplotypes. The latter finding suggests possible changes in fragment copy number, which would prohibit knowledge of the exact origin of each polymorphic fragment. Thus the overall pattern described by the haplotype is considered more reliable than frequencies of the individual fragments. Furthermore, studies of *Pvu*II fragment frequencies in the three major population groups indicate that *Pvu*II fragments are less useful than *Pvu*II haplotypes as population-specific markers. Most fragments are found in all three populations, and frequencies are often similar in the different groups (Figure 6.6). Rare fragments are, however, observed mostly in Caucasoids only, with the exception of fragments D2 and H.

The amount of variation revealed by 49a at the level of the haplotype is much greater. At least 21 haplotypes each were observed in the approximately 60 individuals studied for the *Bgl*II, *Hin*dIII, *Pst*I and *Sst*I RFLPs (Tables 6.1-6.3 and 6.5), while analysis of the *Pvu*II polymorphism in 127 individuals revealed 53 different haplotypes. Most haplotypes are observed in only a few individuals, and are population-specific (Figure 6.7). An exception to this statement is provided by the Negroid population, which is characterized by a high frequency (62-76%) of one particular haplotype, which is also present in the San population. One other haplotype is shared between the San and Negroid populations for all except the *Pst*I and *Pvu*II polymorphisms. These findings have implications for the origin of the Bantu-speaking Negroids, as discussed later.

A comparison of the genetic diversity values for each RFLP (Table 6.7) summarizes the descriptions of variability. In each case the Negroid sample has the lowest genetic diversity of the three populations studied, while the Caucasoid sample exhibits greatest genetic diversity. The overall genetic diversity is high for all the enzyme systems, but the compound haplotype, encompassing 32 haplotypes in 51

individuals, peaks at 0.952. Inferences concerning the nature of the 49a polymorphisms that may be drawn from this increased diversity are discussed below.

6.3.2 The Nature of the 49a Polymorphisms

The observation that 49a revealed polymorphisms with a number of different enzymes suggested that the 49a/*Taq*I polymorphism was not due to point mutations alone as originally described (Ngo *et al* 1986), but may also result from rearrangement processes. An analysis of the five different 49a polymorphisms occurring within a single individual, the compound haplotype (Table 6.6), suggests that there is poor correspondence between the 49a polymorphisms. A haplotype revealed by one enzyme is not necessarily always found with a partucular haplotype revealed by one of the other five enzymes, resulting in the observed increased genetic diversity of the compound haplotype over the individual RFLPs (Table 6.7). Each 49a polymorphism may thus be considered to be distinct and dynamic, and probably results from a combination of rearrangement events and restriction site mutations specific to that enzyme, further complicated by the presence of a variable number of similar repeat units.

6.3.3 Use of the 49a RFLPs in Population and Evolutionary Studies

The 49a RFLPs can clearly be used to distinguish the three major population groups studied here (Figure 6.7), and there is nothing to suggest that they should not exhibit similar variation in other population groups. Such variability suggests that the 49a polymorphisms might prove useful in paternity testing for male children in almost all population groups. The exception is in the Negroid population, where genetic diversity is reduced in comparison with the other population groups (Table 6.7).

The haplotype frequencies for the five enzyme systems (Figure 6.7) show clearly the presence of one haplotype in high frequency in the Negroid population. This phenomenon is restricted to the Negroids, and suggests a biological unity of the Bantu-speakers, in agreement with the proposed origin of all Bantu-speakers from a common proto-Bantu stock 2000-3000 years ago (Oliver and Fagan 1975). Furthermore, the existence of the same haplotype in the San, and also the sharing of one other haplotype for several of the enzyme systems, indicates that the Negroid and San populations may have arisen from a common ancestral group. Recent male-specific gene flow from Negroids to the isolated San is believed to be unlikely (Ramsay and Jenkins 1988), and is not

expected to account for the observed haplotype frequencies, although unidirectional gene flow from San females to certain Negroid chiefdoms has been reported (Jenkins 1970). Limited correlation between the different enzyme haplotypes occurs in the Negroids (Table 6.6), and may similarly be related to the proposition that Bantu-speakers are derived from a common stock. Thus genetic diversity values and correlation between haplotypes may be of importance in future population studies using the 49a RFLPs.

The studies using only *Taq*I and *Pvu*II polymorphisms have also proved useful in population studies. The common genetic origin of the inv(Y) Gujerati Muslim Indians is clearly indicated by their possession of identical 49a/*Taq*I and 49a/*Pvu*II profiles (Table 6.8, Figure 6.8). Since this finding supports historical data (Bernstein *et al* 1986), the corollary is that inferences drawn from correlations between haplotypes in the Bantu-speaking Negroid and Khoisan populations may hold greater weight. Results from the inv(Y) study also precipitated the analysis of *Pvu*II haplotypes in selected individuals possessing Ht7 or Ht8 (Table 6.9), in an attempt to reveal a common genetic origin of these *Taq*I haplotypes in the Lemba and Jewish populations and thus confirm theories concerning the Caucasoid origins of the Lemba population. Since the Lemba sample does not share a single *Pvu*II haplotype

with Jewish individuals, it is suggested that the Lemba and Jewish samples are genetically distinct. However, certain of the *Pvu*II haplotypes of one population are fairly similar to some of those seen in the other: Ht52 (found in the Lemba) differs from Ht47 (Jewish and Lemba) by one fragment only; Ht27 (Lemba) and Ht46 (Jewish) differ by one fragment; Ht47 (Lemba) differs from both Ht46 and Ht48 (Jewish) by one fragment. In contrast, Ht41 and Ht42, the two rare *Pvu*II haplotypes in the Lemba population, are quite dissimiliar to any of the *Pvu*II haplotypes found in the Jewish samples. The latter haplotypes probably have a different origin to Ht27, Ht47 and Ht52 in the history of the Lemba population. It should be noted that 13/15 Lemba individuals screened possess haplotypes containing the B2 fragment, which is rare in Negroids (occurring at a frequency of 0.05), but is fairly common in Caucasoid (frequency = 0.55). Overall,the *Pvu*II data do not exclude a distant Jewish origin of the Lemba *Taq*I Ht7 and Ht8. However, the Semitic influence (via Arab traders) in southern Africa began approximately 1000 years ago (Murdock 1959, Van Warmelo 1974), more recently than the time estimated for the origin and expansion of the Bantu-speaking Negroids. Thus, judging from the biological unity observed in the Bantu-speakers as indicated by the *Taq*I and *Pvu*II haplotypes, a strong correlation between the *Taq*I and *Pvu*II haplotypes would be expected in the Lemba and

their Semitic ancestors. Comparison of *Pvu*II haplotypes between the Lemba and Arab populations may be useful in clarifying the historical origin of the Lemba.

The role of the 49a polymorphisms in sophisticated evolutionary studies is limited by a poor understanding of the mechanism of generation of the variant fragments. Nucleotide diversity can be calculated for RFLPs resulting from restriction site mutations (Nei 1987), but complex polymorphisms involving rearrangements are not amenable to such analysis. However, simple comparisons of genetic diversity between populations have also been used in evolutionary studies, and it is generally considered that ancestral populations will exhibit the greatest variation. Using this approach, Africa has been pinpointed as the site of origin of modern man from mitochondrial (Cann *et al* 1987) and nuclear (Wainscoat *et al* 1986, Cavalli-Sforza *et al* 1988, Bowcock *et al* 1987, Bowcock *et al* 1991) DNA studies. However, Y chromosome studies using 49a do not follow this trend (Torroni *et al* 1990, Spurdle and Jenkins 1992, this study), and non-African populations show the greatest variation. It is thus not surprising that an extension of the 49a/*Taq*I study to include several new enzyme polymorphisms reveals decreased genetic diversity in Negroids. The male-specific inheritance pattern of the Y chromosome, together with cultural practices such as polygamy and extensive male

migration may be implicated in this reversal of expectations. Alternative approaches to the use of Y-specific polymorphisms (and 49a RFLPs in particular) in evolutionary studies need to be considered.

6.4 SUMMARY

Y chromosome probe 49a detects complex polymorphisms with *Bgl*II, *Hin*dIII, *Pst*I, *Pvu*II and *Sst*I, involving 19, 13, 18, 17 and 11 variant bands respectively. The haplotype, a description of the arrangement of fragments in an individual, exhibits great variation. Most haplotypes are population-specific and occur infrequently. In a study of approximately 60 individuals of Caucasoid, Negroid and San origin, 21, 23, 28 and 24 haplotypes were revealed by *Bgl*II, *Hin*dIII, *Pst*I and *Sst*I respectively, while a more extensive study of the 49a/*Pvu*II polymorphism in 127 individuals revealed 53 haplotypes. In general, no single haplotype for a given enzyme system strictly correlates with a haplotype from any of the other three systems, suggesting that each polymorphism results from a combination of restriction site mutations and rearrangement events. Accordingly, the compound haplotype, a combination haplotype composed of the six different 49a haplotypes, exhibits the most genetic diversity. Absolute correlation between *Taq*I and *Pvu*II haplotypes in a sample of inv(Y) Gujerati Muslim

Indians believed historically to be of common genetic origin has led to the belief that specific correlation between the different 49a polymorphisms indicates recent common genetic origin of individuals. In this way, the Bantu-speaking Negroid population appears to be genetically cohesive, and to have distant links with the Khoisan population. Analysis of *Pvu*II haplotypes associated with *Taq*I Ht7 or Ht8 in Lemba and Jewish individuals precludes a close genetic relationship between two groups. The reduced genetic diversity in the Bantu-speaking Negroid population may be ascribed to its proposed recent common origin, but the occurrence of decreased diversity in Negroids in general implicates genetic and cultural differences associated with a male-specific pattern of inheritance.

CHAPTER SEVEN

THE Y-LINKED XY275 POLYMORPHISM IN SOUTHERN AFRICAN POPULATIONS

7.1 INTRODUCTION

The pseudoautosomal region of the human sex chromosomes is defined as the region of strict homology on the distal tip of the short arms that pair at meiosis, and undergo homologous recombination (Weissenbach *et al* 1987). Recombination in the pseudoautosomal boundary region is limited at its proximal end by the pseudoautosomal boundary, the interface between pseudoautosomal and sex chromosome-specific regions. The pseudoautosomal boundary of the human Y chromosome is defined by the insertion of a 303bp *Alu* repeat sequence (Ellis *et al* 1989), an event thought to have created the boundary. However, recent studies on the pseudoautosomal boundaries of great apes and Old World monkeys indicate that the *Alu* element did not create the present day boundary, but only inserted at the pre-existing boundary after the great ape and Old World monkey lineages diverged (Ellis *et al* 1990c).

Polymorphisms of the pseudoautosomal boundary region have been studied in an attempt to formally define the boundary in terms of recombination (Ellis *et al* 1990b).

The X chromosomes were found to be polymorphic at five positions in a 300bp region, while all Y chromosomes were identical except for one distal polymorphism shared with the X chromosome (Ellis *et al* 1990b). This *Msp*I polymorphism was originally termed XY274 (Ellis *et al* 1990a), and was erroneously believed to result from a C to T transition 274bp distal to the *Alu* insertion site. The polymorphism has since been shown to be a G to T transversion located 275bp from the boundary, and has been renamed XY275 (Ellis 1991b). This correction does not significantly affect the interpretation of the primary data (Ellis 1991b), although it does exclude prediction of the ancestral form of the polymorphism. The two-allele polymorphism may be detected by hybridization of pseudoautosomal probe Hf0.2 to *Msp*I digested DNA, or by PCR amplification and *Msp*I digestion analysis of the boundary region (Ellis *et al* 1990a). The alleles are termed *high* for absence of the *Msp*I site, or *low* for presence of the site (Ellis *et al* 1990a).

The polymorphism was tested on a number of different populations, and a full range of *high* allele frequencies was observed on the X chromosome (Ellis *et al* 1990a). However, XY275 was shown to be fixed as the *high* allele on the Y chromosome in all but two populations - a Pygmy and a Kalahari San population (Ellis *et al* 1990a). The discovery of *(low,* PABY)

chromosomes in two African populations only, in conjunction with the strong linkage disequilibrium observed between XY275 and the Y boundary, led Ellis *et al* (1990a) to suggest that an ancestral *Homo sapiens* population possessing both *high* and *low* alleles was African in origin, and that a single class of Y chromosome with the *high* allele migrated out of Africa at the time of the African/non-African split about 90000 years ago. The present study was undertaken to determine the level of linkage disequilibrium between XY275 and the pseudoautosomal boundary in a number of different southern African populations, and to relate the results to theories put forward by Ellis and coworkers.

7.2 RESULTS

The frequencies of the XY275 *high* allele in the different populations studied are shown in Table 7.1. A wide range of P_X values (X-associated *high* allele frequencies) occur in these southern African groups, but the situation is different for the P_Y values (Y-associated *high* allele frequencies). As described by Ellis *et al* (1990a), strong disequilibrium between XY275 and the Y boundary is observed, with the *high* allele occurring almost exclusively in association with the Y chromosome. Exceptions to this fixation occur in the Khoisan, Negroid, Khoisan-speaking Negroid and hybrid groups, as well as in the Caucasoid Asiatic Indian population (Table 7.1).

TABLE 7.1

XY275 *High* Allele Frequency[a] in Southern African Populations

Population	N	P_X[b](S.E.)	N	P_Y[c](S.E.)
Negroid[d]				
Zulu	70	0.04(0.02)	39	0.97(0.03)
Xhosa	32	0.16(0.06)	10	1.00(0.00)
Ndebele	9	0.11(0.10)	9	1.00(0.00)
Swazi	71	0.13(0.04)	18	1.00(0.00)
S.Sotho	45	0.18(0.06)	27	0.85(0.07)
Pedi	40	0.10(0.05)	34	0.97(0.03)
Tswana	18	0.06(0.06)	25	0.92(0.05)
Venda	21	0.62(0.11)	21	1.00(0.00)
Tsonga	60	0.15(0.05)	23	1.00(0.00)
Herero	41	0.17(0.06)	37	1.00(0.00)
Himba	33	0.03(0.03)	35	1.00(0.00)
Ambo	65	0.14(0.04)	35	0.95(0.04)
Enigmatic Bantu-speakers				
Lemba	36	0.11(0.05)	40	1.00(0.00)

TABLE 7.1 continued.

Khoisan-speaking Negroid				
Dama	58	0.22(0.05)	26	0.92(0.05)
Khoisan				
Nama	73	0.10(0.04)	22	0.77(0.09)
Sekele San	52	0.25(0.06)	47	0.81(0.06)
Caucasoid				
S.A. European	53	0.58(0.07)	51	0.98(0.02)
S.A. Indian	67	0.22(0.05)	61	0.79(0.05)
S.A. Jewish	69	0.72(0.05)	33	1.00(0.00)
Hybrid				
"Coloured"	56	0.13(0.04)	56	0.98(0.02)
Richtersveld	44	0.20(0.06)	22	0.68(0.10)

[a]The XY275 polymorphism is represented by only two alleles, thus the frequency of the *low* allele is reciprocal to that of the *high* allele.

[b]Frequency of (*high*, PABX), the X chromosome-associated *high* allele.

[c]Frequency of (*high*, PABY), the Y chromosome-associated *high* allele.

[d]Broad population classification is indicated in **bold** type.

7.3 DISCUSSION

The wide range of P_X values is a phenomenon also described by Ellis *et al* (1990a) in their study of Caucasians, Oceanic populations, Amerindians and Africans. The P_X value of 0.58 (± 0.07) in the South African (S.A.) European population correlates well with that of 0.59 observed by Ellis *et al* (1990a) in Northern Europeans. The latter sample includes a few Ashkenazi individuals (N.Ellis, personal communication). The S.A. Ashkenazi Jewish population has a much higher P_X value of 0.72 (± 0.05), possibly due to genetic drift; when the approximately 40 000 Jewish immigrants came to South Africa in 1880-1910, they nearly all came from Lithuania, mostly in family groups and from only a few towns or villages (shtetls). The S.A. Asiatic Indian population is distinct from the other two Caucasoid groups with a P_X of only 0.22 (± 0.05). Likewise, the Negroid, Khoisan and hybrid groups have much lower P_X values, ranging from 0.03 to 0.25, with the exception of the Venda with a P_X of 0.62 (± 0.11). The Venda are relatively recent immigrants to southern Africa, and the unusually high P_X value may reflect frequencies found in central African Bantu-speaking populations, none of which has been screened to date. The Venda sample also represents individuals from a geographically localized area, and sampling error cannot be excluded as a possible

explanation for the high P_X value observed in this rather small sample. It is interesting that the P_X values of Caucasoids in general are much higher than those in African populations (Table 7.1), a phenomenon also observed in the populations screened by Ellis *et al* (1990a). These findings may be seen to lend support to the theory of an African/Non-African split, as suggested studies using nuclear polymorphisms (Wainscoat *et al* 1986), and mtDNA polymorphisms (Cann *et al* 1987, Vigilant *et al* 1991).

The P_Y values are mostly great (0.66-1.00). The *low* allele is found in association with the Y chromosome in the Sekele San population at the relatively high frequency of 0.19 (± 0.06). This is in agreement with the observation of the (*low*, PABY) haplotype in 6/14 San individuals from a distinct population from Tsumkwe by Ellis *et al* (1990a). (The Sekele and Tsumkwe populations are both !Kung or Northern Bush language speakers). The Khoi Nama also possess (*low*, PABY) chromosomes, at a frequency of 0.23 (± 0.09). The frequency of 0.32 (± 0.10) in the hybrid Richtersveld population, which is thought to have Nama origins (Nurse *et al* 1985), indicates that Nama gene flow into this population was not strictly maternal, although it may initially have been so. The (*low*, PABY) haplotype occurs sporadically in the different Bantu-speaking

Negroid groups, at frequencies ranging from 0.00 to 0.15, as well as in the Khoisan speaking Dama (0.08 ±0.05).

Ellis *et al* (1990a) reported fixation of the *high* allele on the Y chromosome in Caucasoids. However, (*low*, PABY) haplotypes have been observed in the S.A. European and Asiatic Indian populations sampled in this study (Table 7.1). Although the occurrence of a single (*low*, PABY) chromosome in the S.A. European population may be due to historical gene flow from other southern African groups, the substantial frequency of the (*low*, PABY) haplotype in S.A. Asiatic Indians (0.21 ± 0.05) is believed to be significant. It is possible that this rather high frequency is due to a sampling effect, and may not be representative of the Asiatic Indian population as a whole, since a relatively small group of Indians, from the provinces of Gujerat, Madras and Bengal gave rise to the S.A. Asiatic Indian population (Bernstein *et al* 1986). The (*low*, PABY) haplotype occurs at similiar frequencies in the Hindu and Muslim religious groups (5/26 and 3/14 respectively), and the differences are not significant (p=0.65). This is not totally unexpected since the spread of the Muslim religion to the Indian subcontinent, with its ancestral Hindu population, only occurred in the region of 1000 years ago.

The possibility that (*low*, PABY) chromosomes were generated by recent crossover between the X and Y in the Asiatic population cannot be excluded - the calculated rate of recombination between XY275 and the boundary is 6 x 10^{-5}/bivalent (N.Ellis, personal communication). Analysis of the XY275 polymorphism in populations of the Middle East would probably shed light on this question. However, one might have expected the notable presence of (*low*, PABY) chromosomes in other non-African populations, which has not been the case (Ellis *et al* 1990a; this study). It is especially surprising to note that the Rondonia Surui and Karitiana populations studied by Ellis *et al* (1990a) possess only (*low*, PABX) and (*high*, PABY) chromosomes (as calculated by maximum likelihood methods), and thus no crossover between X and Y appears to have taken place in these Amerindian groups.
The discovery of Y-associated *low* alleles in the Caucasoid S.A. Asiatic Indian population suggests that the conclusions drawn by Ellis *et al* (1990a) may have been premature. Although mitochondrial data (Cann *et al* 1987, Vigilant *et al* 1991) and autosomal nuclear data (Wainscoat *et al* 1986, Cavalli-Sforza *et al* 1988) both indicate an African origin of modern man, the XY275 data cannot be said to support the hypothesis. Neither do the data refute it. One must also acknowledge, however, that the population relationships inferred from Y-specific markers, or those in linkage

disequilibrium with the Y chromosome (as is the case with XY275), may not correspond to those obtained from studies with autosomal or mtDNA markers. If modern humans arose in Africa, then it is clear that more than one class of Y chromosome was already present and the frequency differences in present day populations are due to the operation of genetic drift after the populations migrated out of Africa.

7.4 SUMMARY

The level of linkage disequilibrium between the XY275 *Msp*I polymorphism and the X and Y boundaries was investigated in 21 different southern African populations. A full range of frequencies of the *high* allele was observed on the 1013 X chromosomes studied, in keeping with published data. In previous studies fixation of the *high* allele on the Y chromosome was observed in all but two groups - a Pygmy and a Tsumkwe San population. However, in this study of 673 Y chromosomes, the *low* allele was found to be associated with the Y chromosome in several different Bantu-speaking Negroid groups, the Khoisan-speaking Negroid Dama, the Khoisan, two groups of mixed ancestry and the South African Asiatic Indian population. The discovery of the *low* allele on Y chromosomes of Caucasoid individuals suggests that more than one class of Y chromosome gave rise to the present-day non-African population. The data also fail to provide support for the theory that Africa is the site of origin of *Homo sapiens*, but equally do not exclude it.

CHAPTER EIGHT

THE pDP31, p21A1 AND *ALU* INSERTION POLYMORPHISMS IN SOUTHERN AFRICAN POPULATIONS

8.1 INTRODUCTION

The potential of Y chromosome studies in population and evolutionary genetics, and the scarcity of Y chromosome polymorphisms available for such studies has been discussed at great length in previous chapters (see Chapter Two, Section 2.4, and Chapter Four). The utility of several known Y-specific polymorphisms in this field of study has, however, not been reported.

The Y-specific polymorphism detected by pDP31 (Page *et al* 1982) appears not to have been widely studied. It is reported to result from a duplication event (Page *et al* 1984), and is detected with several enzymes (Page *et al* 1982, Page *et al* 1984, Bowcock 1984). The duplication has been shown to occur in USA Caucasoid males at a frequency of 0.5, and was found to be absent in a sample of 10 Finnish individuals (D.Page, personal communication). The duplication was also shown to be present in southern African Negroid and Khoisan populations (Bowcock 1984), but sample sizes were inadequate for reliable estimation of population frequencies. Probe p21A1 reveals a simple *Taq*I point

mutation polymorphism within a repeat sequence at Yq11 (Jakubicza *et al* 1989). The level of polymorphism appears to be low, with the rarer allele reported to have a frequency of 0.05 in a sample of 74 German individuals (Jakubicza *et al* 1989). However, frequencies have not been determined in other populations, and the potential of this polymorphism in Y-specific population studies has not yet been ascertained. The Y-specific polymorphism YAP, resulting from the insertion of a 300bp *Alu* repeat into the Y chromosome at Yq11 has recently been discovered (M.Hammer, personal communication). The *Alu* element has been found present in a number of different populations, and sequence data indicate that the *Alu* element and surrounding DNA are identical in a number of individuals from several different populations (M.Hammer, personal communication). This finding suggests that the YAP *Alu* insertion results from a single insertion event (M.Hammer, personal communication).

A study of the pDP31, p21A1 and YAP Y-specific polymorphisms in southern African populations was undertaken in an attempt to determine the value of these polymorphisms in Y chromosome population and evolutionary studies.

8.2 RESULTS

The frequency of the pDP31/*Eco*RI duplication in southern African populations, grouped according to broad classification is shown in Table 8.1, and the frequencies of the p21A1/*Taq*I and Y *Alu* insertion polymorphisms in the 23 different populations studied are displayed in Tables 8.2 and 8.3 respectively.

TABLE 8.1

Frequency of the pDP31/*Eco*RI Duplication in Southern African Populations

Population	(n)	Frequency of Duplication	(S.E.)
Caucasoid	(30)	0.63	(0.09)
Hybrid	(79)	0.39	(0.05)
Negroid	(54)	0.04	(0.03)
Khoisan	(55)	0.07	(0.03)
Lemba	(18)	0.17	(0.09)

TABLE 8.2

Frequency of the p21A1/TaqI Polymorphism in African Populations[a]

Population	N	Frequency	(S.E.)
Caucasoid			
S African European	47	0.85	(0.05)
S African Asiatic Indian	57	0.82	(0.05)
S African Jewish	32	0.59	(0.41)
Total	136	0.78	(0.05)
Hybrid			
Johannesburg "Coloured"	66	0.74	(0.05)
Richtersveld "Coloured"	34	0.68	(0.08)
Total	100	0.72	(0.04)
Khoisan			
Nama	20	0.65	(0.10)
Tsumkwe San	33	0.61	(0.07)
Sekele San	46	0.61	(0.07)
Total	99	0.62	(0.05)

TABLE 8.2 continued.

Negroid			
Eastern Bantu			
Nguni: Zulu	53	0.83	(0.05)
Xhosa	23	1.00	(0.00)
Ndebele	14	0.79	(0.11)
Swazi	33	0.73	(0.08)
Sotho/Tswana: Southern Sotho	48	0.75	(0.06)
Pedi	52	0.85	(0.05)
Tswana	40	0.80	(0.06)
Tsonga	31	0.65	(0.09)
Venda	29	0.76	(0.08)
Western Bantu			
Ambo	42	0.90	(0.05)
Herero: Herero	45	0.98	(0.02)
Himba	35	0.89	(0.05)
Total	445	0.83	(0.02)

TABLE 8.2 continued.

Enigmatic Bantu-Speakers			
Lemba	48	0.73	(0.06)
Khoisan-speaking Negroid			
Dama	41	0.89	(0.05)
Pygmy			
Pygmy	32	0.94	(0.04)

[a]Frequency of 11kb fragment (absence of *Taq*I site).

TABLE 8.3

Frequency of the Y *Alu* Polymorphism[a]

Population	N	Frequency	(S.E.)
Caucasoid			
S African European	51	0.04	(0.03)
S African Asiatic Indian	63	0.00	(0.00)
S African Jewish	36	0.14	(0.06)
Total	150	0.05	(0.02)
Hybrid			
Johannesburg "Coloured"	66	0.24	(0.05)
Richtersveld "Coloured"	33	0.30	(0.08)
Total	99	0.26	(0.04)
Khoisan			
Nama	23	0.48	(0.10)
Tsumkwe San	38	0.11	(0.05)
Sekele San	45	0.44	(0.07)
Total (less Tsumkwe)	68	0.46	(0.06)

TABLE 8.3 continued.

Negroid			
Eastern Bantu			
Nguni: Zulu	47	0.81	(0.06)
Xhosa	22	0.95	(0.05)
Ndebele	14	0.79	(0.11)
Swazi	30	0.80	(0.07)
Sotho/Tswana: Southern Sotho	45	0.64	(0.07)
Pedi	52	0.79	(0.06)
Tswana	40	0.78	(0.07)
Tsonga	30	0.67	(0.09)
Venda	27	0.85	(0.07)
Western Bantu			
Ambo	38	0.92	(0.04)
Herero: Herero	46	0.61	(0.07)
Himba	35	0.89	(0.05)
Total	426	0.78	(0.02)

TABLE 8.3 continued.

Enigmatic Bantu-Speakers			
Lemba	47	0.30	(0.07)
Khoisan-speaking Negroid			
Dama	37	0.77	(0.07)
Pygmy			
Pygmy	32	0.76	(0.08)

[a]Frequency of chromosomes without the *Alu* insert is reciprocal to the value indicated in the table. Totals included values for all groups within the population only when the chi-square test indicated that no significant difference existed between them. The Tsumkwe San was excluded from the San total for this reason.

8.3 DISCUSSION

8.3.1 The pDP31 Duplication Polymorphism

The frequency of the pDP31 duplication (Table 8.1) is significantly higher in the Caucasoid population than in either of the large African samples ($p < 0.005$). The historical flow of male Caucasoid genes into the hybrid populations of southern Africa is thus validated by the conspicuous frequency of the duplication in the hybrid sample. In addition, the disparity in duplication frequency between the Negroid and hybrid populations is greater than between the Caucasoid and hybrid groups ($p<0.005$ for Negroids as opposed to $p<0.025$ for Caucasoids). The Lemba population is believed to have Semitic origins (Van Warmelo 1974), and accordingly has the duplication at a noteworthy frequency. This is in agreement with findings using the Y-specific 49a/*Taq*I polymorphism (see Chapter Five). The overall results would suggest that the pDP31/*Eco*RI duplication may prove to be a useful genetic marker in population studies, especially with regard to distinguishing between Caucasoid and Negroid gene flow.

8.3.2. The p21A1/*Taq*I Polymorphism

Results for the p21A1/*Taq*I polymorphism appear more complicated than the pDP31 data. It is immediately apparent that the rarer allele occurs at higher frequencies than originally reported in German individuals (Jakubicza *et al* 1989). Furthermore, the polymorphism exists in four major population groups, the Caucasoid, Negroid, Khoisan and Pygmy populations, as well as in the two groups of mixed ancestry. There do not appear to be any obvious trends in the population frequencies. The frequency of the *Taq*I site is significantly higher in the Khoisan than in the other groups, but differences between the Caucasoids and Negroids are not statistically significant. This is unexpected in light of results from studies using other Y chromosome polymorphisms in the same populations (see Chapters Five, Six and Seven), which always show the Negroids to be distinct from the Caucasoids. There is also great variation within the Negroid and Caucasoid groups, differences which prove to be significant ($p<0.005$ for Negroids and $p<0.025$ for Caucasoids). These findings suggest that p21A1/*Taq*I frequency data are of little value in population studies as a marker for Caucasoid or Negroid gene flow. This conclusion is emphasized by the results observed for the enigmatic Bantu-speaking Lemba. In accordance with theories of their Semitic origin (Van Warmelo 1974), this

population has been shown to have affinity with the Caucasoid populations using Y chromosome probes 49a (see Chapter Five) and pDP31 (see Section 8.3.1 above). However, the population frequencies for the p21A1/*TaqI* polymorphism, using the same sample of individuals, does not isolate the Lemba specifically from other Negroids, nor does it show any association with the individual Caucasoid groups. It might be suggested that the p21A1/*TaqI* polymorphism frequencies do not merely reflect genetic shifts and drift, but may result from repeated *TaqI* mutations at the locus. The *TaqI* recognition site is a hotspot for mutation (Barker *et al* 1984), and polymorphism is believed to result from loss of the site due to methylation of the CpG dinucleotide (Barker *et al* 1984). Thus the *TaqI* site detected by p21A1 may have been lost several times. Furthermore, p21A1 recognizes a repeat sequence, and the polymorphism was originally described as the interruption of a short alternating repeat unit of 4kb and 7kb by the loss of one or *several TaqI* sites, resulting in an 11kb fragment. Thus the existence of a single *TaqI* mutation event, in time and/or position, is unlikely.

8.3.3 The Y *Alu* Polymorphism

The frequency of the Y *Alu* insertion shows a clear trend in the populations studied (Table 8.3). The insert is largely absent in the Caucasoid population, occurs at slightly higher frequencies in the hybrid populations, at intermediate frequencies in the Khoisan, and at high frequencies in Bantu-speaking Negroids. The Tsumkwe San exhibits a frequency of the insert which is significantly lower than expected, and this "Caucasoid-like" phenotype has been observed in the same population with the 49a/*Taq*I polymorphism (see Chapter Five). The clustering of the Tsumkwe San sample with Caucasoid groups was first attributed to genetic drift of this isolated population, possibly accentuated by sample error (Spurdle and Jenkins 1992), but reanalysis of the 49a/*Taq*I data (see Chapter Five for details) has shown both the Sekele and Tsumkwe San to have affinity with Caucasoids, and ancient admixture of the San with Caucasoids is a possible explanation for the results (Cavalli-Sforza 1989). It would appear, from the results using the Y *Alu* polymorphism, that the ancestors of the Tsumkwe San were more exposed to Caucasoid admixture than were those of the Sekele San. The affinity between the enigmatic Bantu-speaking Lemba population and other Caucasoid groups strengthens

arguments, based on myth and cultural practices (Van Warmelo 1974), that this population may have Semitic origins.

The frequencies of the insert in the Khoisan-speaking Dama and the Pygmy sample are comparable with those found in the Bantu-speaking Negroid groups. This confirms the Negroid origins of the Dama (Nurse *et al* 1985), and also suggests that the *Alu* insertion became entrenched in Africa prior to raciation. In accordance with these findings, cluster analysis of the *Alu* insertion frequency data reveals a basic African/Non-African split (data not shown), as described by many researchers (Wainscoat *et al* 1986, Cann *et al* 1987, Cavalli-Sforza *et al* 1988, Vigilant *et al* 1991). The hybrid groups cluster together with the Caucasoid groups, affirming the theory that males were responsible for the Caucasoid contribution to the admixture in these groups of mixed ancestry (Nurse *et al* 1985).

No significant differences were observed (using the chi-squared test) between the various Nguni chiefdoms, or between the different Sotho/Tswana chiefdoms ($p=0.222$ and 0.244 respectively). This is not unexpected from previous results using the 49a/*Taq*I, 49a/*Pvu*II and XY274/*Msp*I polymorphisms (see Chapters Five, Six and Seven). In all these cases the

Bantu-speaking Negroid groups, believed to have originated from a proto-Bantu stock approximately 2000 years ago (Oliver and Fagan 1975), could not be distinguished strictly according to linguistic classification. The western Bantu Herero was shown to be significantly different from the Himba and Ambo western Bantu groups. This may be attributed to the fact that the Himba and Ambo samples were mostly localized and rural, whereas the Herero sample was collected in a more urban setting.

Since the *Alu* insert occurs in the major population groups - the Khoisan, Negroids and the Caucasoids - it almost certainly predates the diversification of man. The absence of the *Alu* insert in the Indian population may either be due to genetic drift after the proposed African/non-African split (Wainscoat *et al* 1986), or due to a bottleneck later on in the evolution of non-African populations.

8.4 SUMMARY

The pDP31 duplication polymorphism was studied in 241 individuals encompassing the southern African Khoisan, Negroid and Caucasoid populations. The duplication was present in all groups, but differences in frequencies were marked, with low frequencies of the duplication in African populations, and a high frequency in Caucasoids. Populations of mixed ancestry exhibited frequencies closer to Caucasoids than to Negroids, in accordance with their historical origins. The p21A1/*TaqI* polymorphism was studied in 904 individuals in 23 African population groups, encompassing the Caucasoid, Negroid, Khoisan and Pygmy populations. The frequency of the polymorphism was higher than previously reported, with the rarer allele occurring at frequencies ranging up to 0.41. No distinct trends in population frequency were observed, suggesting that the polymorphism is of little value in population studies. The existence of several *TaqI* mutations, in both time and position within the repeat unit, is indicated. The frequency of the Y *Alu* insertion was determined in 850 individuals from 23 different African population groups. A trend in frequency was observed, with the insert largely absent in Caucasoid populations, at intermediate frequency in the Khoisan, and at high frequency in Negroids. The insert predates

diversification of *Homo sapiens*, since it occurs in all groups, but frequency differences between populations suggest that it may be useful in population studies.

CHAPTER NINE

INTERRELATIONSHIPS BETWEEN THE DIFFERENT Y-LINKED POLYMORPHISMS

9.1 INTRODUCTION

Although a paucity of Y-specific polymorphism has been documented (Arnemann *et al* 1985, Jakubicza *et al* 1989, Malaspina *et al* 1990), a limited number of Y-linked polymorphisms has been reported. Unfortunately, the validity of results generated using these polymorphic systems has not been tested, and in some instances the data suggest that inconsistencies exist. Y chromosome probes 49f and 49a have proved to be popular in population studies because of their ability to detect multiple haplotypes (Ngo *et al* 1986), but the system is complex, and the mechanism of polymorphism is not fully understood. There is also some evidence to suggest that fragments of identical size have arisen more than once (Torroni *et al* 1990, Spurdle and Jenkins 1992, also see Chapter Five). Also of interest are population studies involving the XY275 *Msp*I polymorphism associated with the pseudoautosomal boundary (Ellis *et al* 1990a, Ellis 1991b). Theories concerning the origin and expansion of *Homo sapiens* have been constructed from the frequencies of the Y-associated alleles in the different populations (Ellis *et al* 1990, Spurdle *et al*, in press,

see Chapter Seven). However, the polymorphism is not actually Y-specific but is pseudoautosomal in nature, and despite its proximity to the boundary the possibility of recombination cannot be ignored. Other Y-specific polymorphisms include the point mutation detected by p21A1 with *Taq*I (Jakubicza *et al* 1989), the rearrangement polymorphism revealed by pDP31 (Page *et al* 1982), and the Y *Alu* polymorphism (see Chapter Eight for a more extensive description of these polymorphisms). The polymorphism revealed by p21A1 is described as the loss of a single *Taq*I site with a single copy of a repeat unit (Jakubicza *et al* 1989), but the existence of a single mutation event, either in time or in position, has not been proven. The pDP31 polymorphism is described as a duplication rearrangement (Bowcock 1984), however, it is not known if the duplication represents a single mutation event. The Y *Alu* polymorphism is a rearrangement polymorphism resulting from the insertion of an *Alu* element, and sequence data suggest that it has most likely originated only once (M.Hammer, personal communication). Furthermore it is not in close proximity to the pseudoautosomal boundary, and would not be subject to recombination. This simple polymorphism should thus provide a means to clarify the findings obtained using other Y-specific probes which may be less reliable. A study was undertaken to determine the relationship between the Y *Alu*

polymorphism and the four other Y-linked polymorphisms studied in the present sample of southern African populations.

9.2 RESULTS

The distribution of the Y-associated XY275 *high* and *low* alleles in relation to the Y *Alu* polymorphism is detailed in Table 9.1, and a schematic representation of the postulated events resulting in the observed associations is shown in Figure 9.1. The relationship between the 49a/*Taq*I allelic series and the *Alu* insertion is displayed in Table 9.2, and the association between the 49a/*Taq*I haplotype and YAP is shown graphically in Figure 9.2. The distribution of the p21A1/*Taq*I and pDP31/*Eco*RI alleles in relation to the Y *Alu* polymorphism is given in Table 9.3, while the Y chromosome combination haplotypes found in Caucasoids are listed in Table 9.4.

TABLE 9.1

Distribution of Y-associated XY275 *High* and *Low* Alleles in Relation to the Y *Alu* Insertion[a]

Population	Insert Present	Insert Absent
	High/Low	*High/Low*
Negroid		
Zulu	27/0	7/1
Xhosa	10/0	0/0
Ndebele	8/0	1/0
Swazi	14/0	1/0
Southern Sotho	11/1	12/2
Pedi	25/0	7/1
Tswana	18/0	5/1
Venda	14/0	3/0
Tsonga	17/0	6/0
Herero	23/0	14/0
Himba	29/0	4/0
Enigmatic Bantu-Speakers		
Lemba	10/0	22/0
Khoisan-Speaking Negroids		
Dama	10/0	3/1

TABLE 9.1 continued.

Khoisan		
Nama	8/0	4/3
Sekele San	21/2	15/7
Pygmy		
Pygmy	12/2	3/1
Caucasoid		
S.A. European	2/0	48/0
S.A. Asiatic Indian	0/0	45/13
S African Jewish	5/0	27/0
Hybrid		
Johannesburg "Coloured"	10/0	44/1
Richtersveld "Coloured"	4/0	7/6
	278/5	278/37

[a]Figures represent numbers of individuals.

TABLE 9.2

Relationship Between the 49a/*Taq*I Allelic Series and the *Alu* Insertion[a]

Allelic Series	Allele/Allelic Variant	Insert Absent	Insert Present
A	A0	20	9
	A1	15	376
	A2	175	10
	A3	150	66
	A3/A2	35	4
	A4	18	2
	A4/A3	6	1
	A4/A3/A2	1	0
	A5	3	0
	A5/A3	4	0
B	B0	10	0
	B1	417	468
C	C0	350	468
	C1	77	0

TABLE 9.2 continued.

D	D0	220	467
	D1	139	0
	D2	66	1
	D2/D1	1	0
	D3	1	0
F	F0	63	12
	F1	360	456
	F1/F2	4	0
G	G0	1	4
	G1	426	464
H	H0	3	0
	H1	427	468
I	I0	85	67
	I1	342	401

[a]Figures represent numbers of individuals.

TABLE 9.3

Association between the *Alu* Insertion and the p21A1/*Taq*I and pDP31/*Eco*RI Polymorphisms[a]

	Insert Present	Insert Absent
p21A1 (-)[b]	412 (N,K,P)	269 (C,N,K,P)
p21A1 (+)	49 (C,N,K)	138 (C,N,K,P)
pDP31 (-)[c]	72 (N,K)	90 (C,N,K)
pDP31 (+)	18 (C,N,K)	27 (C,N)

[a]Populations are indicated in parentheses. C=Caucasoid, N=Negroid, K=Khoisan, P=Pygmy.

[b](-) represents absence of *Taq*I site.

[c](-) represents absence of duplication.

TABLE 9.4

Caucasoid Y Chromosome Combination Haplotypes

49a B Series	49a C Series	*Alu* Insert	XY275	pDP31
B0	C0	absent	*low*	+
B0	C0	absent	*high*	-
B1	C0	present	*high*	+
B1	C0	absent	*high*	-
B1	C0	absent	*low*	+
B1	C1	absent	*high*	-
B1	C1	absent	*high*	+
B1	C1	absent	*low*	NR[a]

[a]No result available.

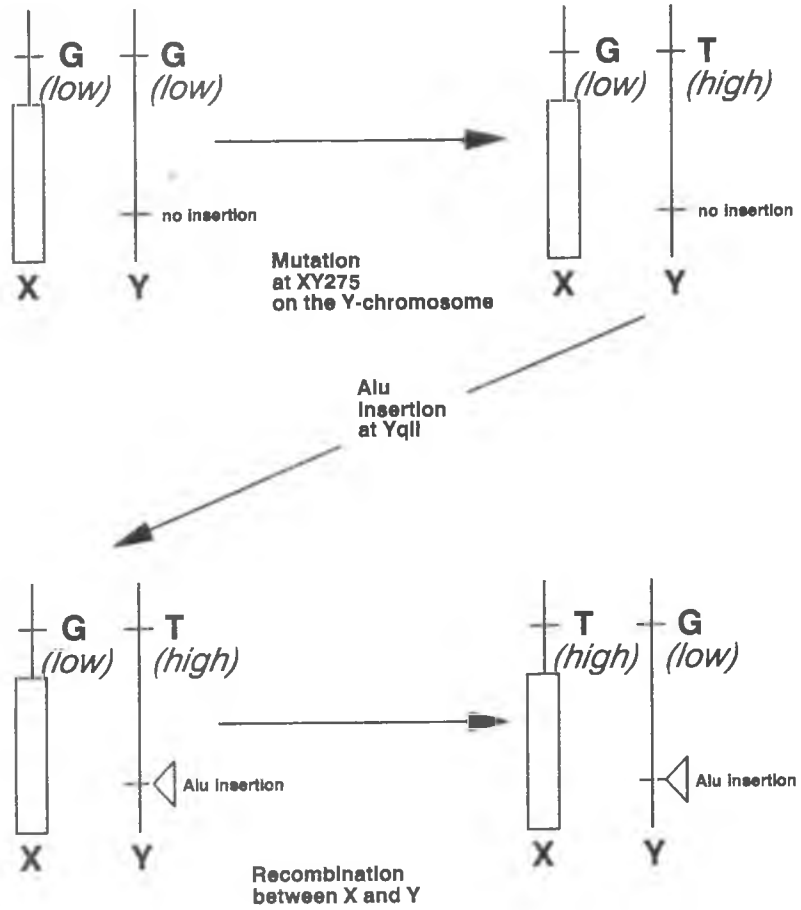

FIGURE 9.1 Schematic representation of postulated events resulting in the observed associations between the XY275 and Alu insertion loci.

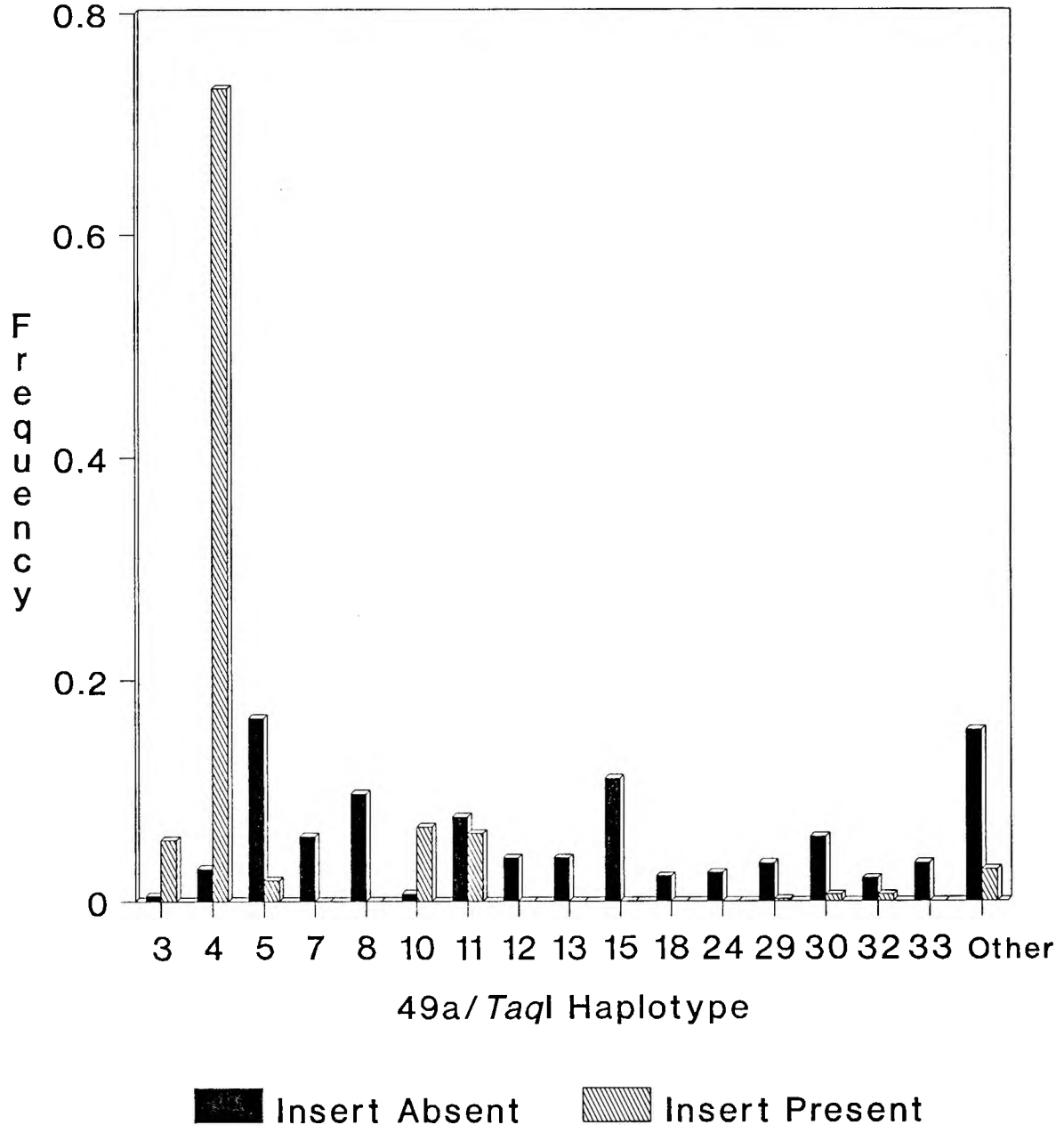

FIGURE 9.2 Association between the 49a/TaqI haplotype and Y *Alu* polymorphism. Haplotype numbers are those detailed in Chapter Five, and haplotypes occurring in both groups at a frequency less than 0.02 were pooled as "other". Sample size: n=468 for presence of insert; n=427 for absence of insert.

9.3 DISCUSSION

9.3.1 The Relationship between YAP and XY275

The XY275 *Msp*I polymorphism (Ellis *et al* 1990a) results from a G-»T transversion within the *Msp*I recognition site (Ellis 1991b). The XY275 polymorphism is subject to strong linkage disequilibrium on the Y chromosome, with the *high* allele having gone almost to fixation (Ellis *et al* 1990a). However, exceptions to this fixation (Ellis *et al* 1990, Spurdle *et al*, in press, Chapter Seven) indicate that the Y-associated XY275 polymorphism predates the diversification of *Homo sapiens*, and certainly the proposed African/non-African split dated to about 90 000 years ago (Cavalli-Sforza *et al* 1988). The *low* allele has been observed in an Asiatic Indian population (Spurdle *et al*, in press, Chapter Seven), implying that more than one Y chromosome migrated out of Africa in the event of an African origin of man. The XY275 polymorphism has also been studied in several chimpanzees and gorillas (N.Ellis, personal communication). All primates screened were shown to possess the XY275 *Msp*I cutting site, suggesting that the G nucleotide (*low* allele) is ancestral to the T nucleotide (*high* allele) at this locus.

The distribution of Y-associated XY275 alleles in Y chromosomes with and without the *Alu* insert was studied in an attempt to yield information about the time of insertion relative to the origin of the *Msp*I polymorphism. Data in Table 9.1 indicate that the *low* allele is absent in most chromosomes, as would be expected from XY275 allele frequency data (Ellis *et al* 1990, Spurdle *et al*, in press, Chapter Seven). However, the Y-associated XY275 *low* allele is found together with the absence of the *Alu* insert in 37 individuals from several different groups encompassing the Caucasoid, Negroid, Khoisan and Pygmy populations (Table 9.1). Furthermore, in the Sotho, Sekele San and Pygmy groups, 5 individuals were found to possess the Y-associated XY275 *low* allele in the presence of the *Alu* insert. Since the Alu insertion is believed to have occurred only once in the evolution of the human Y chromosome from sequencing data (M.Hammer, personal communication), these findings indicate that either recombination between the XY275 and YAP loci has occurred, or that the XY275 polymorphism has arisen more than once. Since the latter is unlikely to have occurred in the timespan for the origin of modern man (Ellis *et al* 1990a), the theory of recombination is favoured. In addition, it is suggested that the insertion took place on a Y chromosome with a *high* background (T nucleotide) after the XY275 polymorphism had arisen. Only one recombination event is required

for this scenario, as opposed to two events for the origin of the XY275 polymorphism on the X chromosome. A schematic representation of the possible set of events leading to the observed findings is shown in Figure 9.1.

9.3.2 The Correlation between YAP and the 49a/*Taq*I Polymorphism

The polymorphism revealed by 49a comprises up to 8 Y-specific bands which may be present, absent or vary in length (Ngo *et al* 1986, Guerin *et al* 1988, Spurdle *et al* 1989, Lucotte *et al* 1990, Torroni *et al* 1990, Spurdle and Jenkins 1992, Chapter Five). The bands were originally considered to represent independent loci, and variation was ascribed to the occurrence of *Taq*I point mutations (Ngo *et al* 1986). However, co-inheritance of "alleles" at several loci has led to the suggestion that duplication processes are also in existence (Torroni *et al* 1990, Spurdle and Jenkins 1992, Chapter Five). The correlation between the *Alu* insertion and the 49a/*Taq*I polymorphism was determined in an attempt to evaluate theories on the origin and evolution of the alleles, and their association into haplotypes.

A strong correlation between the presence of the *Alu* insert and the 49a/*Taq*I haplotype was observed in the 895 Y chromosomes typed for both these polymorphisms. The data are summarized in Figure 9.2, and indicate that the presence of the insert is significantly correlated with 49a/*Taq*I Ht4 ($p<10^{-6}$). In contrast, the distribution of Y chromosomes lacking the insert is more evenly shared between Ht5, Ht7, Ht8, Ht11, Ht15 and Ht30, amongst others. The associations between the *Alu* insert and the 49a/*Taq*I haplotype may be largely ascribed to a correlation between the insert and variation at the A series (Table 9.2). The correlation between the *Alu* insert and A series allelic variant *A1* is highly significant ($p<10^{-6}$), and the absence of the insert is associated mainly with A series allelic variants *A2* and *A3*. Analysis of the population frequencies of the 49a/*Taq*I haplotypes and A series allelic variants add dimension to these results. The frequency of the *Alu* insert is low in Caucasoids (Chapter Eight, Table 8.3), and accordingly these individuals account for the high proportion of Ht7, Ht8, Ht11, Ht12 and Ht15 in those chromosomes lacking the insert (Figure 9.2). The latter 49a/*Taq*I haplotypes possess A series allelic variants *A2* and *A3* between them, in accordance with results shown in Table 9.2. The frequency of the *Alu* insert is high in Negroids (Chapter Eight, Table 8.3), as is the frequency of 49a/*Taq*I Ht4, and the *A1* allele specifically (Torroni

et al 1990, Spurdle and Jenkins 1992, Chapter Five, Table 5.2), consistent with the observed data (Table 9.2).

These results also allow for conclusions to be drawn concerning the evolution of the A series "alleles". The different allele sizes were proposed to result from C to G point mutation within *Taq*I sites to create fragments of increasing size, whereas duplication events are indicated in the instances where "alleles" coexist (Torroni *et al* 1990, Spurdle and Jenkins 1992, Chapter Five). The occurrence of several alleles/allelic variants both with and without the insert indicates that the A series alleles have originated more than once during the origin and expansion of *Homo sapiens*. The data for the remaining allelic series (Table 9.2) can be interpreted in a similar fashion. It would appear that the B, C and H series arose only once, whilst the other allelic series must have arisen several times.

9.3.3 Other Y-Specific Polymorphisms and Their Association with the Alu Insertion

The relationship between the *Alu* insertion and the p21A1/*Taq*I and pDP31/*Eco*RI polymorphisms is summarized in Table 9.3. It is clear that if we assume that the *Alu* element was inserted once only, the p21A1 and pDP31

polymorphisms each arose more than once, or have reverted to the original state of the polymorphism. Neither theory is favoured *a priori* as an explanation for the pDP31 data. It is tempting to suggest, from population data in Table 9.3, that an initial pDP31 duplication event took place prior to raciation on a Y chromosome with the *Alu* insertion, and that a second event occurred early on in raciation in the absence of the *Alu* insertion. An alternative and equally plausible explanation is that the pDP31 duplication preceded the *Alu* insertion, and the duplication was lost later, with all events preceding raciation. The situation is different in the case of the p21A1/*Taq*I polymorphism. Most *Taq*I polymorphisms result from a C to T transition within the *Taq*I recognition site (Barker 1984), and the observed p21A1 results thus probably represent multiple unidirectional mutation events. The presence of the *Taq*I site would thus be considered the ancestral form of the polymorphism.

9.3.4 Ancestral Migration and Genetic Drift

The theory of the origin of modern man in Africa has gained much support from a variety of molecular studies in recent years (Wainscoat *et al* 1986, Cann *et al* 1987, Cavalli-Sforza *et al* 1988, Bowcock *et al* 1990, Vigilant *et al* 1991). The data have confirmed an African/non-African split, and revealed increased

genetic variation in the "ancestral" African populations. Y chromosome studies using p49a might also provide support for a Negroid/Eurasian split (Torroni *et al* 1990, Spurdle and Jenkins 1992, Chapter Five of this thesis), although in this case decreased heterogeneity has been observed in African populations (Torroni *et al* 1990, Spurdle and Jenkins 1992, Appendix Two of this thesis). The latter phenomenon has been ascribed to the unique male-specific inheritance of the Y chromosome, complicated by culturally-determined gender behaviour pattern differences.

The *Alu* insertion data suggest that at least two Y chromosomes migrated out of Africa, since Caucasoids possess chromosomes with and without the insertion (Table 8.1). A similar situation exists at the XY275 locus, with *high* and *low* alleles present in Caucasoids (Spurdle *et al*, in press, Chapter Seven). Since a recombination event is proposed to have occurred at the XY275 locus, it might be believed that the *low* allele in the Asiatic Indian sample arose in this way. However, it occurs in the absence of the *Alu* insertion, an association also found in the San, Pygmy and Negroids (Table 9.1). The *low* allele is only found in association with the insert in African populations, suggesting that it is this combination that results from the postulated recombination event. Furthermore,

all four possible combinations occur in at least the Negroid and San groups, intimating that the recombination preceded raciation.

Analysis of the association between the *Alu* insertion and the XY275 polymorphism also extends predictions of migration based on either system alone. Three of the four possible combinations exist in Caucasoids, indicating that three ancestral Y chromosomes existed in these populations. In addition, tremendous gene shifts are implicated by the observed population gene frequencies, with the *high* allele most common, and occurring in the absence of the insert in Caucasoids mainly, and in the presence of the insert mostly in African populations. Within the Caucasoids, differences exist between the Asiatic Indians and the other two groups, implying that genetic drift took place at this level as well.

The analysis was extended further by creating a combination haplotype including the 49a B and C series, the *Alu* insertion, and the XY275 and pDP31 loci. The 49a polymorphism was not used in its entirety, since most of the polymorphism revealed is believed to have arisen more than once. Only the B and C "loci" appear to have originated once, and are also polymorphic in Caucasoids. The p21A1/*Taq*I polymorphism was excluded from the combination haplotype since it appears that

multiple mutation events have also occurred at this locus (Table 9.3). The pDP31 duplication might be subject to such criticism as well, but this simple rearrangement polymorphism probably never arose as repeatedly as suspected for the 49a and p21A1 polymorphisms. Of the 32 possible arrangements for the combination haplotype, at least eight are detected in Caucasoids (Table 9.4), suggesting that at least eight ancestral Y chromosomes existed in these populations. However, even if the pDP31 data are excluded from the interpretation, it is impossible to create a flow chart of simple unidirectional mutation events to account for the observed haplotypes without implicating recurrent origin of polymorphism at the 49a B and C "loci" and/or the XY275 locus. Multiple mutations of the low copy repeat detected by 49a are favoured over repeated mutation or additional crossover at the XY275 locus as an explanation of the results, since the nature of the 49a polymorphism itself is suggestive of multiple rearrangement events, and, in addition, multiple mutation events appear to occur at all the other 49a "loci". In any event, it would appear that the combination haplotype is of little use in predicting migration and gene shifts, since the recurrent origin of polymorphism cannot be defined by locus, or in time.

9.4 SUMMARY

Several Y-specific polymorphisms were studied in up to 897 individuals belonging to a number of Southern African populations, and to a Pygmy sample drawn from two areas in central Africa (see Subjects and Methods). The Y *Alu* polymorphism is believed, from sequence data, to have originated once, but the *Msp*I XY275, 49a/*Taq*I, p21A1/*Taq*I and pDP31 duplication polymorphisms may have arisen more than once. The relationship between the *Alu* polymorphism and the other Y chromosome polymorphisms studied is not absolute, suggesting multiple origins of the latter RFLPs. A crossover is suggested to have occurred at the Y-linked XY275 locus, while the multiple origin of fragments of the same size by duplication is implicated in the case of the 49a/*Taq*I polymorphism. The p21A1/*Taq*I point mutation, situated within a repeat unit, is proposed to have taken place repeatedly in time and/or position, and the pDP31 duplication may either have arisen twice, or has reverted to the original allelic form. Thus the Y *Alu* polymorphism has proved useful in testing the validity of other Y chromosome polymorphisms.

CHAPTER TEN

GENERAL DISCUSSION

The role of Y chromosome polymorphisms in population and evolutionary genetics has not yet been clearly established, as a result of the singular lack of polymorphism discovered on this chromosome to date. The paucity of Y-specific polymorphism appears to be unrelated to possible bottleneck effects during human evolution, since it is characteristic of both Negroid and Caucasoid populations. These findings are especially troubling since interpopulation variation differences can only be truly evaluated in light of within-group variation, and it will be difficult to make progress and to draw accurate conclusions of an evolutionary nature as long as we are limited by the low levels of Y-specific variation.

The ability of 49a/f to detect numerous Y-specific haplotypes has lent itself to use in regional population studies (Breuil *et al* 1987, Lucotte *et al* 1989, Lucotte *et al* 1990, Torroni *et al* 1990, Spurdle and Jenkins 1992, and Chapter Five of this thesis). This study of southern African populations showed that the 49a/*Taq*I polymorphism was able to reveal differences between the Eastern and Western Bantu groups, but could not be used to accurately distinguish

between the linguistically affiliated tribal chiefdoms within these groups (Chapter Five, Figure 5.2). Similarities between the different Bantu-speaking groups, interpreted as evidence for a common proto-Bantu ancestral stock, were not only observed with the 49a/*Taq*I polymorphism, but also with the 49a/*Pvu*II, XY275 *Msp*I and Y *Alu* polymorphisms (Chapters Five, Six, Seven and Eight). The Negroid origins of the Khoisan-speaking Dama were confirmed by studies using the 49a/*Taq*I (Chapter Five) and the Y *Alu* insertion polymorphism (Chapter Eight), as suggested by research using serogenetic markers (Nurse *et al* 1985). The 49a/*Taq*I polymorphism showed the Dama to specifically associate with Eastern Bantu groups (Figure 5.2), in contrast to results from mitochondrial studies on the same populations, which grouped the Dama with the Western Bantu Herero population (H.Soodyall, personal communication). The exact origin of the Dama remains enigmatic, but one favoured theory is that ancestors of the Dama were historically enslaved by the Khoi, and modern Dama people represent the descendants of runaway slaves who escaped into the desert (Nurse *et al* 1985). The disparity between the results from Y chromosome and mitochondrial studies may perhaps be explained by the following scenario: If it is assumed that the Khoi may have originated in East Africa and only later migrated into southern Africa (Cooke 1965, Jenkins 1972),it is probable that the slaves they would have brought with

them, or captured en route, were Negroids of the Eastern Bantu culture. It is likely that the majority of such slaves would have been male, and on escaping would have been required to take wives from the area, probably of Western Bantu origin. Even if an alternative hypothesis is favoured for the origin of the Khoi, and they are considered to have evolved in the Kalahari (Elphick 1977), it is not impossible that they may have enslaved Bantu-speakers of the Eastern Bantu culture, since the ancestors of the Eastern Bantu Sotho/Tswana are believed to have stretched much farther to the west than they do now (Nurse *et al* 1985).

The hybrid groups of mixed racial ancestry have clearly been shown to have resulted from male-specific Caucasoid gene flow into Negroid groups, since they have greater affinity with Caucasoids as revealed by several Y-specific markers (Chapters Five and Eight). Accordingly, mitochondrial DNA markers show the hybrid populations to cluster with Negroids (H.Soodyall, personal communication). The Lemba population is postulated to have Semitic origins from anthropological data (Van Warmelo 1974), and the data from 49a/*Taq*I (Chapter Five), pDP31 and YAP (Chapter Eight) Y-specific genetic markers indicate that the Lemba group is distinctly associated with Caucasoids. The results suggest that the Caucasoid ancestors of this

population were male, and indeed, mitochondrial DNA markers (H.Soodyall, personal communication) and autosomal serogenetic markers (D.Dunn, personal communication) indicate that the Lemba are Negroid in origin. These data consolidate theories that the Lemba population stems from a group of Arab traders that took on Negroid wives when left stranded in southern Africa (Van Warmelo 1974).

The population studies reported in this thesis have not been able to conclusively determine the affinity between the Khoisan and Negroid populations, which are classically believed to have distant common origins (Tobias 1974, Nurse *et al* 1985). Consistent with such theories, the two groups seem to share a number of 49a haplotypes, as detected with *Taq*I (Chapter Five, Appendix Two), and *Bgl*II, *Hin*dIII, *Pst*I, *Pvu*II and *Sst*I (Chapter Six). Surprisingly, the 49a/*Taq*I cluster analysis (Figure 5.2) has shown the Khoisan groups to associate with Caucasoid populations. The possibility of ancient Caucasoid admixture, as postulated by Cavalli-Sforza (Cavalli-Sforza *et al* 1988, Cavalli-Sforza 1989), has been put forward as an eplanation for the observed result, and it is interesting to note that the Y *Alu* insertion occurs in the Khoisan at frequencies intermediate between those reported for the Caucasoids and Negroids (Table 8.3). In contrast, the pDP31 duplication results (Table 8.1)

suggest that the Khoisan have greater affinity with Negroids, but the sample size was considerably smaller in this instance, and did not include Tsumkwe San individuals. Another point to consider is that the San and Caucasoids actually share few 49a/*Taq*I haplotypes, and the similarity between them (as observed in Figure 5.2) appears to be largely due to a decreased level of Ht4 in these groups relative to the frequencies observed in Negroids (Appendix 2). Thus the results appear disparate, but it is possible that male-specific Caucasoid gene flow into ancient Khoisan stocks may have distorted the expected genetic cohesion of the Khoisan and Negroid populations.

Y-specific polymorphisms have not been used extensively in evolutionary studies. The 49a/*Taq*I polymorphism has been used to establish a genealogy of the Y chromosome, based on the successive loss of *Taq*I sites by CpG methylation (Hazout and Lucotte 1987, Lucotte *et al* 1989, Lucotte *et al* 1990). In this way, a haplotype found in the Pygmy population has been predicted to be ancestral. However, the proposed involvement of rearrangements in the generation of the 49a polymorphism (Torroni *et al* 1990, Spurdle and Jenkins 1992, and Chapter Five of this thesis) seriously questions the validity of such an approach, and the conclusions drawn therefrom, since polymorphic fragments of the same size may evolve in parallel.

Physical mapping of the area may lead to a better understanding of how the polymorphism is generated, and might allow for the development of PCR and sequencing systems to distinguish between haplotypes which are presently impossible to resolve. Without knowledge of the molecular basis of the 49a polymorphisms, future analyses using these systems will remain restricted. The presence of rearrangement events would nevertheless preclude evolutionary studies based on nucleotide substitution rates.

Studies using 49a/*Taq*I haplotype frequencies have, however, confirmed the Negroid/Eurasian dichotomy of extant human populations (Torroni *et al* 1990, Spurdle and Jenkins 1992, Chapter Five). If we take into consideration the hypothesis of ancient Caucasoid admixture in the Khoisan population (Cavalli-Sforza *et al* 1988, Cavalli-Sforza 1989), the cluster analysis portrayed in Figure 5.2 may be seen to represent an African/non-African split, as reported from mitochondrial DNA (Cann *et al* 1987, Vigilant *et al* 1991). This finding might be taken as support for the "Out of Africa" model of evolution of *Homo sapiens*. On the other hand, African populations have been shown to be far more homogeneous than Caucasoid populations, with *Taq*I Ht4 frequencies of 0.68 in western and central African Negroids (Torroni *et al* 1990), and 0.39-0.86 in southern African Negroids (Spurdle and

Jenkins 1992, Appendix Two). Accordingly, Negroid populations exhibit much lower genetic diversity (0.392-0.400) than Khoisan (0.808-0.840) and Caucasoid (0.737-0.900) populations, as determined from 49a/*Taq*I frequencies (Appendix 2). Studies using the other 49a RFLPs show the same pattern (Table 6.7). This contrasts with mtDNA studies where the greatest within-group variation occurred in African populations (Brega *et al* 1986, Cann *et al* 1987, Scozzari *et al* 1988, Vigilant *et al* 1991, Merriwether *et al* 1991). Further conflicting results are presented from a study using 30 autosomal DNA markers, where the mean heterozygosity was found to be lower in San than in Caucasoid and Negroid populations (Morris *et al* 1991). The latter observations were explained partly by the use of several DNA polymorphisms originally detected in Caucasoids, which might result *a priori* in a decreased heterozygosity in other populations (Morris *et al* 1991).

The discrepancy between Y-specific and other DNA studies may be partly explained by the common African practice of polygamy, the effect of which should not be underestimated, especially with regard to the genetic profile of a population. It has been claimed that polygamy is at least partly responsible for the high sickling rate in Africa (Konotey-Ahulu 1970), and the term "male procreative superiority index" (MPSI) has

been devised by Konotey-Ahulu (1980) to formally indicate the effects of polymgamy. The country-wide mean MPSI for 3095 fathers contacted in Ghana was calculated to be 2.03 (Konotey-Ahulu 1980), indicating that on average the Ghanaian father has twice as many children as the mother. It is postulated by Konotey-Ahulu (1980) that the high rates of common African genetic markers, such as extra digits, twinning and abnormal haemoglobins, cannot be dissociated from high values of MPSI. It is likely that the non-random matings resulting from polygamous marriages could cause extreme shifts in gene frequencies, and might have resulted in decreased Y-specific genetic diversity in Negroids, including the southern African Bantu-speaking Negroids. Polygamy is not thought to occur extensively in the San (Botha *et al* 1972), and this fact, together with the postulated ancient admixture with Caucasoids (Cavalli-Sforza *et al* 1988, Cavalli-Sforza 1989), could have resulted in an increased genetic diversity in the San relative to the Negroids. However, the long term effects of polygamy in all human populations should also be considered. The practice of polygamy is not restricted to African populations, and has been recorded in other populations as recently as biblical times. Such a phenomenon may have contributed to the decreased level of polymorphism observed on the Y chromosome in general.

The two allele deletion/insertion polymorphism detected by $p12f_2$ has also been shown to clearly distinguish Africans from non-Africans (Casanova *et al* 1985, Brega *et al* 1987), but has not been used in extensive evolutionary studies. Research on the pseudoautosomal boundary region and the *Msp*I polymorphism XY275 has proved to be more useful (Ellis *et al* 1990a, Ellis *et al* 1990b, Ellis 1991b). In the initial study, fixation of the Y-associated *high* allele occurred in all groups with the exception of two African populations, suggesting that a single class of Y chromosome migrated out of Africa (Ellis *et al* 1990a). However, the recent discovery of the "African-specific" Y-associated *low* allele in the Caucasoid Asiatic Indian population of South Africa intimates that more than one class of Y chromosome gave rise to the present-day non-African population (see Chapter Seven) .

There is also evidence from other Y polymorphism studies to suggest a division within the Caucasoid group. The 49a/*Taq*I haplotype frequency data reveal that while Europeans are distinctive due to a high frequency of Ht15 and Ht12 (Torroni *et al* 1990, Spurdle and Jenkins 1992, Appendix Two, A.Novelletto, personal communication), the situation is different for Indian and Jewish populations. Ht14 was found to be most common in Asiatic Indians from three locations in India (Lucotte *et al* 1990), however Ht7, Ht8, Ht11 and

sometimes Ht24 (but not Ht14) are most common in South African Asiatic Indian and Jewish populations (Spurdle and Jenkins 1992, Appendix Two), and in Ashkenazi and Sephardic Jews sampled in Israel and Tunisia (Santachiara Benerecetti *et al* 1992).

In addition, a study of three hypervariable loci detected by pulsed-field gel electrophoresis (Oakey and Tyler-Smith 1990) has led to the proposal that most European and Asian men are descended from two males. The haplotypes defined by the three loci encompassed 12 probe/enzyme combinations, including the *Hin*dIII/pYα1 polymorphism. Two groups were distinguished in this way - group 1 consisting of Caucasoids only, and group 2 containing Asians (including Orientals) and Caucasoids. The pYα1/*Hin*dIII "+" allele occurred infrequently (0.05) in group 1, but in all members of group 2, occurring at an overall frequency of 0.41 in the two combined groups. The latter polymorphism has also been studied in Egyptians, Italians and Sardinians, as well as in a British sample (A.Novelletto, personal communication). The "+" allele occurs at much higher frequencies (0.75-0.87) in all but the British sample (0.33), suggesting that the Egyptian, Italian and Sardinian populations are largely comprised of "group 2" individuals. Furthermore, screening for the XY275 polymorphism in these populations showed the Y-associated *low* allele to be present in the Egyptian

sample only, at a frequency of 0.09 (A.Novelletto, personal communication). This implies that at least three ancestral Y chromosomes existed in these populations. The hypervariable polymorphisms reported by Oakey and Tyler-Smith (1990) have not yet been studied in Negroid populations, and research in this area may prove to be of significance in Y-specific human evolutionary studies.

Evidence from the comparative studies of Y chromosome polymorphisms presented in this thesis (Chapter Nine) indicate that there are at least three combinations of the alleles of the Y *Alu* insertion and XY275 polymorphisms on Caucasoid Y chromosomes (Table 9.1). These are also found in Khoisan and Negroid individuals, albeit at vastly different frequencies. The fourth combination occurs only in three African populations, and is believed to have resulted from a crossover event (Figure 9.1) which appears to have preceded raciation. If a single crossover event is assumed to have occurred, then African populations could be said to exhibit greater variation than non-African populations, and the results may be taken as support for the "Out of Africa" hypothesis.

It would not be inappropriate to state that the paucity of conventional polymorphism on the Y chromosome has hindered the use of the Y chromosome in the study of

human evolution. It appears that a concerted collaborative effort will be required if this chromosome is to contribute to our understanding of human origins and evolution to anything like the same extent that nuclear and mitochondrial DNA variation have done. To this end, a Y chromosome consortium has been established by N.Ellis, M.Hammer and this author. The aim is to allow collaborative Y chromosome research on common cell lines representing populations from all continents of the world. It must be acknowledged, however, that Y chromosome polymorphisms reveal evolutionary patterns strikingly different from those demonstrated by nuclear and mitochondrial DNA studies. Such differences may be due to the unique pattern of paternal gene flow, complicated by cultural practices accentuating male/female behavioural differences. Polygamy and geographical mobility are two obvious factors enhancing this phenomenon. The logic of approaches to evolutionary studies employing autosomal traits and mitochondrial DNA markers would thus be invalid for corresponding, and possibly complementary, Y chromosomal studies.

REFERENCES

ALLISON AC (1954). Protection afforded by sickle-cell trait against subtertian malarial infection. *Br Med J* 1: 290-294.

ARNEMANN J, COOKE HJ, JAKUBICZA S, SCHMIDTKE J (1985). Human Y-chromosome derived cloned DNA sequences. *Cytogenet Cell Genet* 40: 571.

BARBUJANI G, JAQUEZ GM, LIGI L (1990). Diversity of some gene frequencies in European and Asian populations. V. Steep multilocus clines. *Am J Hum Genet* 47: 867-875.

BARDONI B, ZUFFARDI O, GUILO S, BALLABIO A, SIMI P, CAVALLI P, GRIMOLDI MG, FRACCARO M, CAMERINO G (1991). A deletion map of the human Yq11 region: implications for the evolution of the Y chromosome and tentative mapping of a locus involved in spermatogenesis. *Genomics* 11: 443-451.

BARKER D, SCHAFER M, WHITE R (1984). Restriction sites containing CpG show a higher frequency of polymorphism in human DNA. *Cell* 36: 131-138.

BERNSTEIN R, WADEE A, ROSENDORFF J, WESSELS A, JENKINS T (1986). Inverted Y chromosome polymorphism in the Gujerati Muslim Indian population of South Africa. *Hum Genet* 74: 223-229.

BERTRANPETIT J, CAVALLI-SFORZA LL (1991). A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann Hum Genet* 55: 51-67.

BISHOP JO (1974). The gene numbers game. *Cell* 2: 81-86.

BISHOP CE, GUELLAEN G, GELDWERTH D, VOSS R, FELLOUS M, WEISSENBACH J (1983). Single copy DNA sequences specific for the human Y chromosome. *Nature* 303: 831-832.

BLANCHE H, ZOGHBI HY, WANG JABS E, DE GOUYON B, ZUNEC R, DAUSSET J, CANN HM (1991). A centromere-based genetic map of the short arm of human chromosome 6. *Genomics* 9:429-428.

BODMER WF (1981). Gene clusters, genome organization, and complex phenotypes. When the sequence is known, what will it mean? *Am J Hum Genet* 33: 664-682.

BODMER WF, CAVALLI-SFORZA LL (1976). *Genetics, Evolution and Man.* WH Freeman, San Francisco.

BOTHA MC, DU TOIT ED, JENKINS T, VAN LEEUWEN A, D'AMARO J, MEERA KHAN P, VAN DER STEEN G, VAN ROOD JJ, VAN DER DOES JA (1972). The HL-A system in Bushmen (San) and Hottentot (Khoikhoi) populations of South West Africa. In: Dausset J, Colombani J (eds) *Histocompatibility Testing*. Munksgaard, Copenhagen, pp 421-432.

BOTSTEIN D, WHITE RL, SKOLNICK M, DAVIS RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32: 314-331.

BOWCOCK A (1984). *Identification of Restriction Fragment Length Polymorphisms in Southern African Human Populations*. PhD thesis, University of the Witwatersrand, Johannesburg.

BOWCOCK AM, BUCCI C, HEBERT JM, KIDD JR, KIDD KK, FRIDLANDER JS, CAVALLI-SFORZA LL (1987). Study of 47 DNA markers in five populations fron four continents. *Gene Geogr* 1: 47-64.

BOWCOCK A, KIDD J, MOUNTAIN J, HEBERT J, CAROTENUTO L, KIDD K, CAVALLI-SFORZA L (1991). Drift, admixture, and selection in human evolution: a study with DNA polymorphisms. *Proc Natl Acad Sci USA* 88: 839-843.

BREGA A, SCOZZARI R, MACCIONE L, IODICE C, WALLACE DC, BIANCO I, CAO A, SANTACHIARA BENERECETTI AS (1986). Mitochondrial DNA polymorphisms in Italy. I. Population data from Sardinia and Rome. *Ann Hum Genet* 50: 327-338.

BREGA A, TORRONI A, SEMINO O, MACCIONE L, CASANOVA M, SCOZZARI R, FELLOUS M, SANTACHIARA BENERECETTI AS (1987). The $p12f_2$/TaqI Y-specific polymorphism in three groups of Italians and in a sample of Senegalese. *Gene Geogr* 1: 201-206.

BREUIL S, HALLE L, RUFFIE J, LUCOTTE G (1987). Polymorphisme d'une sonde Y-specifique nommee p49 chez les Papous Baruyas de Nouvelle Guinee. *Ann Genet* 30: 209-212.

CANN RL, BROWN WM, WILSON AC (1984). Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* 106: 479-499.

CANN RL, STONEKING M, WILSON AC (1987). Mitochondrial DNA and human evolution. *Nature* 325: 31-36.

CASANOVA M, LEROY P, BOUCEKKINE C, WEISSENBACH J, BISHOP C, FELLOUS M, PURRELO M, FIORI G, SINISCALCO M (1985). A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* 230: 1403-1406.

CAVALLI-SFORZA LL (1974). The genetics of human populations. In: Katz S (ed) *Biological Anthropology. Readings from Scientific American.* WH Freeman, San Francisco, pp 265-273.

CAVALLI-SFORZA LL (1986). African Pygmies: An evaluation of the state of research. In: Cavalli-Sforza LL (ed) *African Pygmies*. Academic Press, Orlando, pp 361-426.

CAVALLI-SFORZA LL (1989). The last 100,000 years of human evolution: The vantage points of genetics and archaelogy. In: Giacobini G (ed) *HOMINIDAE. Proceedings of the 2nd International Congress of Human Paleontology. Turin, 28 September - 3 October, 1987.* Jaca, Milan, pp 401-413.

CAVALLI-SFORZA LL, BODMER WF (1971). *The Genetics of Human Populations*. WH Freeman, San Francisco.

CAVALLI-SFORZA LL, EDWARDS AWF (1967). Phylogenetic analysis - models and estimation procedures. *Am J Hum Genet* 19: 233-257.

CAVALLI-SFORZA LL, KIDD JR, KIDD KK, BUCCI D, BOWCOCK AM, HEWLETT BS, FRIEDLANDER J (1986). DNA markers and genetic variation in the human species. *Cold Spring Harbor Symp Quant Biol* LI: 411-417.

CAVALLI-SFORZA LL, PIAZZA A, MENOZZI P, MOUNTAIN J (1988). Reconstruction of human evolution: bringing together genetic, archaelogical and linguistic data. *Proc Natl Acad Sci USA* 85: 6002-6006.

CHARLESWORTH B (1991). The evolution of sex chromosomes. *Science* 251: 1030-1033.

CLARK AG (1987). Natural selection and Y-linked polymorphism. *Genetics* 115: 596-577.

COHEN-HAGENHAUER O, SEBOUN E, NGUYEN VAN CONG, DE TAND MF, BISHOP C, FELLOUS M, FREZAL J (1987). Mapping of a cDNA probe selected from a human testis cDNA library by a chromosome Y derived genomic probe, to human chromosome 3. *Cytogenet Cell Genet* 46: 596.

COOKE CK (1965). Evidence of human migrations from the rock art of Southern Rhodesia. *Africa* 35: 263-85.

COOPER DN, KRAWCZAK M (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum Genet* 83: 181-189.

COOPER DN, KRAWCZAK M (1990). The mutational spectrum of single base-pair substitutions causing human genetic disease: patterns and predictions. *Hum Genet* 85: 55-74.

COOPER DN, SCHMIDTKE J (1984). DNA restriction fragment length polymorphisms and heterozygosity in the human genome. *Hum Genet* 66: 1-16.

COOPER DN, SCHMIDTKE J (1989). Diagnosis of genetic disease using recombinant DNA. Second edition. *Hum Genet* 83: 307-334.

COOPER DN, SMITH BA, COOKE HJ, NIEMANN S, SCHMIDTKE J (1985). An estimate of unique DNA sequence heterozygosity in the human genome. *Hum Genet* 69: 201-205.

COOPER DN, YOUSSOUFIAN H (1988). The CpG dinucleotide and human genetic disease. *Hum Genet* 78: 151-155.

COTTON RGH, RODRIGUES NR, CANPBELL DR (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc Natl Acad Sci USA* 85: 4397-4401.

DAUSSET J, CANN H, COHEN D, LATHROP M, LALOUEL J-M, WHITE R (1990). Centre d'Etude du Polymorphisme Humain (CEPH): Collaborative genetic mapping of the human genome. *Genomics* 6: 575-577.

DE BENEDICTUS G, ROSE G, PASSARINO G, QUAGLIARIELLO C (1989). Restriction fragment length polymorphism of human mitochondrial DNA in a sample population from Apulia (Southern Italy). *Ann Human Genet* 53: 311-318.

DOBYNS WB, CURRY CJR, HOYME HE, TURLINGTON L, LEDBETTER DH (1991). Clinical and molecular diagnosis of Miller-Dieker syndrome. *Am J Hum Genet* 48: 584-594.

DRACOPOLI NC, O'CONNELL P, ELSNER TI, LALOUEL J-M, WHITE RL, BUETOW KH, NISHIMURA DY, MURRAY JC, HELMS C, MISHRA SK, DONIS-KELLER H, HALL JM, LEE MK, KING M-C, ATTWOOD J, MORTON NE, ROBSON EB, MAHTANI M, WILLARD HF, ROYLE NJ, PATEL I, JEFFREYS AJ, VERGA V, JENKINS T, WEBER JL, MITCHELL AL, BALE AE (1991). The CEPH consortium linkage map of human chromosome 1. *Genomics* 9: 686-700.

EHRET C (1973). Patterns of Bantu and central Sudanic settlement in central and southern Africa. *Transafr J Hist* 3: 1-71.

ELLIS NA (1991a). The human Y chromosome. *Sem Dev Biol* 2: 2331-240.

ELLIS NA (1991b) Erratum. *Am J Hum Genet* 49: 908.

ELLIS NA, GOODFELLOW PJ, PYM B, SMITH M, PALMER M, FRISCHAUF A-M, GOODFELLOW PN (1989). The pseudoautosomal boundary in man is defined by an *Alu* repeat sequence inserted on the Y chromosome. *Nature* 337:81-84.

ELLIS NA, KIDD J, GOODFELLOW PJ, KIDD K, GOODFELLOW PN (1990a). Strong linkage disequilibrium between the XY274 polymorphism and the pseudoautosomal boundary. *Am J Hum Genet* 46: 950-955.

ELLIS NA, TAYLOR A, BENGTSSON BO, KIDD J, ROGERS J, GOODFELLOW P (1990b). Population structure of the human pseudoautosomal boundary. *Nature* 344: 663-665.

ELLIS NA, YEN P, NEISWANGER K, SHAPIRO LJ, GOODFELLOW PN (1990c). Evolution of the pseudoautosomal boundary in Old World monkeys and great apes. *Cell* 63: 977-986.

ELPHICK R (1977). *Kraal and Castle: Khoikhoi and the founding of White South Africa.* Yale University Press, New Haven and London.

EWENS WJ, SPIELMAN RS, HARRIS H (1981). Estimation of genetic variation at the DNA level from restriction endonuclease data.
Proc Natl Acad Sci USA 78: 3748-3750.

EXCOFFIER L, PELLEGRINI B, SANCHEZ-MAZAS, SIMON C, LANGANEY A (1987). Genetic and history of sub-Saharan Africa. *Yearbook of Phys Anthropol* 30: 151-194.

FEDER J, YEN L, WIJSMAN E, WANG L, WILKINS L, SCHRODER J, SPURR N, CANN H, BLUMENBERG M, CAVALLI-SFORZA LL (1985). A systematic approach for detecting high-frequency restriction fragment length polymorphisms using large genomic probes. *Am J Hum Genet* 37: 635-649.

FEINBERG AP, VOGELSTEIN B (1983). A technique for radiolabelling DNA restriction fragments to high specific activity. *Ann Biochem* 137: 266-267.

FEY MF, WAINSCOAT JS, MUKWALA EC, FALUSI AG, VULLIAMY TJ, LUZZATTO L (1990). A *Pvu*II restriction fragment length polymorphism of the glucose-6 phosphate dehydrogenase gene is an African-specific marker. *Hum Genet* 84: 471-472.

FISHER RA (1930). *The Genetical Theory of Natural Selection.* Clarendon Press, Oxford.

GILL P, JEFFREYS AJ, WERRETT DJ (1985). Forensic application of DNA 'fingerprints'. *Nature* 318: 577-579.

GILLIAM TC, TANZI RE, HAINES JL, BONNER TI, FARYNIARZ AG, HOBBS WJ, MACDONALD ME, CHENG SV, FOLSTEIN SE, CONNEALLY PM, WEXLER NS, GUSELLA JF (1987). Localization of the Huntington's disease gene to a small segment of chromosome 4 flanked by D4S10 and the telomere. *Cell* 50: 565-571.

GOODFELLOW P, DARLING G, WOLFE J (1985). The human Y chromosome. *J Med Genet* 22: 329-344.

GREENBERG JH (1972). Linguistic evidence regarding Bantu origins. *J Afr Hist* 13: 189-216.

GROOT PC, MAGER WH, HENRIQUEZ NV, PRONK JC, ARWERT F, PLANTA RJ, ERIKSSON AW, FRANTS RR (1990). Evolution of the human α-amylase multigene family through unequal, homologous, and inter- and intrachromosomal crossovers. *Genomics* 8: 97-105.

GUERIN P, ROUGER P, LUCOTTE G (1988). A new *Taq*I variant detected with the p49 probe on the human Y chromosome. *Nucl Acids Res* 156: 7759.

GUERIN P, BERRICHE S, INTRATOR S, ROUGER P, SALMON D, SALMON C, LUCOTTE G (1988a). Les techniques de polymorphismse de l'A.D.N. de sondes du chromosome Y appliquees au probleme de filiations pere-fils dans les analyses de paternite. *Revue Francaise de Transfusion et Immuno-hematologie* 31: 513-520.

GUERIN P, ROUGER P, LUCOTTE G (1988b). A new *Taq*I variant detected with the p49 probe on the human Y chromosome. *Nucleic Acids Res* 156: 7759.

GUSELLA JF, WEXLER NS, CONNEALLY PM, NAYLOR SL, ANDERSON MA, TANZI RE, WATKINS PC, OTTINA K, WALLACE MR, SAKAGUCHI AY, YOUNG AB, SHOULSON I, BONILLA E, MARTIN JB (1983). A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306: 234-238.

HARDING RM, SOKAL RR (1988). Classification of the European language families by genetic distance. *Proc Natl Acad Sci USA* 85: 9370-9372.

HARPENDING HC, JENKINS T (1973). Genetic distance among Southern African populations. In: Crawford MH, Workman PC (eds) *Methods and theories of anthropological genetics*. University of New Mexico press, Albuquerque, pp177-199.

HARRIS H (1980). *The Principles of Human Biochemical Genetics*. Third Edition, Elsevier/North Holland, Amsterdam.

HARRIS H, HOPKINSON DA (1972). Average heterozygosity per locus in man: an estimate based on the incidence of enzyme polymorphisms. *Ann Hum Genet* 36: 9-20.

HARTL DL, CLARK AG (1989). *Principles of population genetics*. Second Edition, Sinauer Associates, Sunderland, Massachusetts.

HAZOUT S, LUCOTTE G (1987). Vers une genealogie du chromosome Y. *Ann Genet* 29: 246-252.

HIGGS DR, GOODBOURN SEY, WAINSCOAT JS, CLEGG JB, WEATHERALL DJ (1981). Highly variable regions of DNA flank the human α-globin genes. *Nucleic Acids Res* 17: 4213-4224.

HUFFMAN TN (1982). Archaeology and Ethnohistory of the African Iron Age. *Ann Rev Anthropol* 11: 133-150.

JAKUBICZA S, ARNEMANN J, COOKE H, KRAWCZAK M, SCHMIDTKE J (1989). A search for restriction fragment length polymorphism on the human Y chromosome. *Hum Genet* 84: 86-88.

JANSSEN LAJ, SANDKUYL LA, MERKENS EC, MAAT-KIEVIT JA, SAMPSON JR, FLEURY P, HENNEKAM RCM, GROSVELD GC, LINDHOUT D, HALLEY DJJ (1990). Genetic heterogeneity in tuberous sclerosis. *Genomics* 8: 237-242.

JEFFREYS AJ (1979). DNA sequence variants in the gamma-A, gamma-G, δ- and ß-globin genes of man. *Cell* 18: 1-10.

JEFFREYS AJ, TURNER M, DEBENHAM P (1991). The efficiency of multilocus DNA fingerprint probes for individualization and establishment of family relationships, determined from extensive casework. *Am J Hum Genet* 48: 824-840.

JEFFREYS AJ, WILSON V, THEIN SL (1985a). Hypervariable 'minisatellite' regions in human DNA. *Nature* 314: 67-73.

JEFFREYS AJ, WILSON V, THEIN SL (1985b). Individual-specific 'fingerprints' of human DNA. *Nature* 316: 76-79.

JEFFREYS MDW (1968). *Some semitic influences in Hottentot culture*. Fourth Raymond Dart Lecture. Witwatersrand University Press, Johannesburg.

JENKINS T (1972). *Genetic Polymorphisms of Man in Southern Africa.* M.D. Thesis, University of London.

JENKINS T (1982). Human evolution in Southern Africa. In: Bonne-Tamir B (ed) *Human genetics, Part A. The unfolding genome.* Alan R Liss, New York, pp 227-253.

JENKINS T, ZOUTENDYK A, STEINBERG AG (1970). Gammaglobulin groups (Gm and Inv) of various Southern African populations. *Am J Phys Anthropol* 32: 197-218.

KAN YW, DOZY AM (1978). Polymorphism of DNA sequence adjacent to the ß-globin structural gene: relationship to sickle mutation. *Proc Natl Acad Sci USA* 75: 5631-5635.

KASPERCZYK A, DIMARTINO NA, KRONTIRIS TG (1990). Minisatellite allele diversification: the origin of rare alleles at the HRAS1 locus. *Am J Hum Genet* 47: 854-859.

KIMURA M (1968a). Evolutionary rate at the molecular level. *Nature* 217: 624-626.

KIMURA M (1968b). Genetic variability maintained in a finite population due to mutational production of neutral and nearby neutral isoalleles. *Genet Res* 11: 247.

KONOTEY-AHULU FID (1970). Maintenance of high sickling rate in Africa - role of polygamy. *J Trop Med Hyg* 73: 19-21.

KONOTEY-AHULU FID (1980). Male procreative superiority index (MPSI): the missing coefficient in African anthropogenetics. *Br Med J* 281: 1700-1702.

KUNKEL TA, SONI A (1988). Mutagenesis by transient misalignment. *J Biol Chem* 263: 14784-14789.

LATHROP GM, O'CONNELL P, LEPPERT M, NAKAMURA Y, FARRALL M, TSUI L-C, LALOUEL J-M, WHITE R (1989). Twenty-five loci form a continuous linkage map of markers for human chromosome 7. *Genomics* 5: 866-873.

LAU EC, MOHANDAS TK, SHAPIRO LJ, SLAVKIN HC, SNEAD ML (1989). Human and mouse amelogenin gene loci are on the sex chromosomes. *Genomics* 4: 162-168.

LEGOUIS R, HARDELIN J-P, LEVILLIERS J, CLAVERIE J-M, COMPAIN S, WUNDERLE V, MILLASSEAU P, LE PASLIER D, COHEN D, CATERINA D, BOUGUELERET L, DELEMARRE-VAN DE WAAL H, LUTFALLA G, WEIISENBACH, J, PETIT C (1991). The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell* 67: 423-435.

LEROY P, SEBOUN E, MATTEI MG, BISHOP CE, FELLOUS M (1987). Isolation of autosomally located testis specific transcripts with human Y derived DNA probes. *Cytogenet Cell Genet* 46: 649.

LUCOTTE G, GUERIN P, HALLE L, LOIRAT F, HAZOUT S (1989). Y chromosome DNA polymorphisms in two African populations. *Am J Hum Genet* 45: 16-20.

LUCOTTE G, BARRIEL V, GUERIN P, ABBAS N, RUFFIE J (1990a). Retro-transposition de la sequence humain homologue a la p49f sur le chromosome Y au cours de l'evolution des singes anthropoides. *Biochem Syst and Ecol* 18: 199-204.

LUCOTTE G, SRINIVA F, LOIRAT F, HAZOUT S, RUFFIE J (1990b). The p49/*Taq*I Y-specific polymorphisms in three groups of Indians. *Gene Geogr* 4: 21-28.

MAEDA N, SMITHIES O (1986). The evolution of multigene families: human haptoglobin gene. *Ann Rev Genet* 20: 81-108.

MALAN JS (1980). *Peoples of South West Africa/Namibia.* Haum Publishers, Pretoria.

MALASPINA P, PERSICHETTI F, NOVELLETTO A, IODICE C, TERRENATO L, WOLFE J, FERRARO M, PRANTERA G (1990). The human Y chromosome shows a low level of DNA polymorphism. *Ann Hum Genet* 54:297-305.

MANIATIS T, FRITSCH EF, SAMBROOK J (1982). *Molecular cloning - a laboratory manual.* Cold Spring Harbor Laboratory, New York.

MAYNARD SMITH J, HAIGH J (1974). The hitch-hiking effect of a favourable gene. *Genet Res* 23: 23-35.

MEISLER MH (1983). DNA polymorphisms. *Nature* 303: 108.

MERRIWETHER DA, CLARK AG, BALLINGER SW, SCHURR TG, SOODYALL H, JENKINS T, SHERRY ST, WALLACE DC (1991). The structure of human mitochondrial DNA variation. *J Mol Evol* 33: 543-555.

MIGNONE N, FEDER J, CANN H, VAN WEST B, HWANG J, TAKAHASHI N, PIAZZA A, CAVALLI-SFORZA LL (1983). Multiple DNA fragment polymorphisms associated with immunoglobulin μ chain switch-like regions in man. *Proc Natl Acad Sci USA* 80: 467-471.

MORRAL N, NUNES V, CASALS T, ESTIVILL X (1991). CA/GT microsatellite alleles within the cystic fibrosis transmembrane conductance regulator (CFTR) gene are not generated by unequal crossingover. *Genomics* 10: 692-698.

MORRIS DJ, HEIM RA, VERGA V, DENTER M, DUNN DS, JENKINS T (1991). Study of 30 DNA markers in three Southern African populations. *Gene Geogr* 5: 1-12.

MOURANT AE, KOPEC AC, DOMANIEWSKA-SOBCZAK K (1976). *The distribution of the human blood groups and other polymorphisms (Second ed)*. Oxford University Press, London.

MULLEY JC, HAAN EA, GEDEON AK, SUTHERLAND GR (1990). Predictive diagnosis for polycystic kidney disease using DNA markers. *Med J Aust* 152: 287-289.

MURDOCK G (1959). *Africa - Its peoples and their culture history*. McGraw-Hill, New York.

MYERS RM, LARIN Z, MANIATIS T (1985). Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* 230: 1242-1246.

NAKAHORI Y, TAKENAKA O, NAKAGOME Y (1991). A human X-Y homologous region encodes "amelogenin". *Genomics* 9: 264-269.

NAKAMURA Y, LEPPERT M, O'CONNELL P, WOLFF R, HOLM T, CULVER M, MARTIN C, FUJIMOTO E, HOFF M, KIUMLIN E, WHITE R (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235: 1616-1622.

NEEL J (1984). A revised estimate of genetic variation in human proteins: implications for the distribution of DNA polymorphisms. *Am J Hum Genet* 36: 1135-1138.

NEI M (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NEWTON CR, GRAHAM A, HEPTINSTALL LE, POWELL SJ, SUMMERS C, KALSHEKER N, SMITH JC, MARKHAM AF (1989). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res* 17: 2503-2516.

NGO KY, VERGNAUD G, JOHNSSON C, LUCOTTE G, WEISSENBACH J (1986). A DNA probe detecting multiple haplotypes of the human Y chromosome. *Am J Hum Genet* 38: 407-418.

NOLL WW, COLLINS M (1987). Detection of human DNA polymorphisms with a simplified denaturing gradient gel electrophoresis technique. *Proc Natl Acad Sci USA* 84: 3339-3343.

NURSE GT, JENKINS T (1977). *Health and the hunter-gatherer*. Karger, Basel.

NURSE GT, WEINER JS, JENKINS T (1985). *The peoples of Southern Africa and their affinities*. Clarendon press, Oxford.

O'BRIEN SJ (1991). Ghetto legacy. *Curr Biol* 1: 209-211.

OAKEY R, TYLER-SMITH C (1990). Y chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. *Genomics* 7: 325-330.

OLIVER R, FAGAN B (1975). *Africa in the Iron Age, c. 500 BC to AD 1400*. Cambridge University Press, Cambridge.

ORITA M, SUZUKI Y, SEKIYA T, HAYASHI K (1989). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5: 874-879.

PAGE D, DE MARTINVILLE B, BARKER D, WYMAN A, WHITE R, FRANCKE U, BOTSTEIN D (1982). Single-copy sequence hybridizes to polymorphic and homologous loci on human X and Y chromosomes. *Proc Natl Acad Sci USA* 79: 5352-5356.

PAGE D, HARPER M, LOVE J, BOTSTEIN D (1984). Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* 311: 119-123.

PHILLIPSON DW (1977). The spread of the Bantu language. *Sci Am* 236: 106-114.

PIAZZA A, CAPPELLO N, OLIVETTI E, RENDINE S (1988). A genetic history of Italy. *Ann Hum Genet* 52: 203-213.

PIAZZA A, SGARAMELLA-ZONTA L, GLUCKMAN P, CAVALLI-SFORZA LL (1975). The fifth histocompatability workshop gene frequency data: a phylogenetic analysis. *Tissue Antigens* 5: 445-463.

PONCZ M, SCHWARTZ E, BALLANTINE M, SURREY S (1983). Nucleotide sequence analysis of the δß-globin gene regions in humans. *J Biol Chem* 258: 11599-11609.

RAMSAY M, JENKINS T (1988). Alpha-globin gene cluster haplotypes in the Kalahari San and Southern African Bantu-speaking blacks. *Am J Hum Genet* 43: 527-533.

RICHARDS B, HORN GT, MERRILL JJ, KLINGER KW (1991). Characterization and rapid analysis of the highly polymorphic VNTR locus D4S125 (YNZ32), closely linked to the Huntington disease gene. *Genomics* 9: 235-240.

RUANO G, KIDD KK (1991). Coupled amplification and sequencing of genomic DNA. *Proc Natl Acad Sci USA* 88: 2815-2819.

SANTACHIARA BENERECETTI AS, SEMINO O, PASSARINO G, MORPUGO GP, FELLOUS M, MODIANO G (1992). DNA polymorphisms in Ashkenazi and Sephardic Jews. In: Bonne-Tamir B, Adam A (eds) *Genetic Diversity among Jews: Diseases and Markers at the DNA Level.* Oxford University Press, New York, pp 45-50.

SCHUGHART K, KAPPEN C, RUDDLE FH (1989). Duplication of large genomic regions during the evolution of vertebrate homeobox genes. *Proc Natl Acad Sci USA* 86: 7067-7071.

SCHLOTTERER C, TAUTZ D (1992). Slippage synthesis of simple sequence DNA. *Nucl Acids Res* 20: 211-215.

SCHUMM JW, KNOWLTON RG, BRAMAN JC, BARKER DF, BOTSTEIN D, AKOTS G, BROWN VA, GRAVIUS TC, HELMS C, HSIAO K, REDIKER K, THURSTON JG, DONIS-KELLER H (1988). Identification of more than 500 RFLPs by screening random genomic clones. *Am J Hum Genet* 42: 143-159.

SCOZZARI R, TORRONI A, SEMINO O, SIRUGO G, BREGA A, WALLACE DC (1988). Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms. *Am J Hum Genet* 43: 534-544.

SEMINO O, TORRONI A, SCOZZARI R, BREGA A, DE BENEDICTUS G, SANTACHIARA-BENERECETTI AS (1989). Mitochondrial DNA polymorphisms in Italy III Population data from Sicily: a possible quantitation of maternal African ancestry. *Ann Human Genet* 53:193-202.

SKRAASTAD MI, VERWEST A, BAKKER E, VEGTER-VAN DER VLIS M, VAN LEEUWEN-CORNELISSE I, ROOS RAC, PEARSON PL, VAN OMMEN G-JB (1991). Presymptomatic, prenatal, and exclusion testing for Huntington disease using seven closely linked DNA markers. *Am J Med Genet* 38: 217-222.

SLIGHTOM JL, KOOP BF, XU P, GOODMAN M (1988). Rhesus fetal globin genes - concerted gene evolution in the descent of higher primates. *J Biol Chem* 263: 12427-12438.

SMITH B, SKARECKY D, BENGTSSON U, MAGENIS RE, CARPENTER N, WASMUTH JJ (1988). Isolation of DNA markers in the direction of the Huntington disease gene from the G8 locus. *Am J Hum Genet* 42: 335-344.

SOUTHERN EM (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Evol* 19: 503-527.

SPURDLE AB, JENKINS T (1990). Candidate TDF probe pDP1007 detects an X polymorphism with *Hind*III at ZFX. *Nucl Acids Res* 18: 3430.

SPURDLE AB, JENKINS T (1992). Y chromosome probe p49a detects complex *Pvu*II haplotypes and many new *Taq*I haplotypes in southern African populations. *Am J Hum Genet* 51: 107-125.

SPURDLE AB, MORRIS D, JENKINS T (1989). Y-chromosome probe p49f detects new *Taq*I variants and complex *Pvu*II haplotypes. *Cytogenet Cell Genet* 51: 1084.

SPURDLE A, RAMSAY M, JENKINS T. The Y-associated XY275 *low* allele is not restricted to indigenous African peoples. *Am J Hum Genet* 1992, in press.

STARK GR, WAHL GM (1984). Gene amplification. *Ann Rev Biochem* 53: 447-491.

STEINBERG AG, JENKINS T, NURSE GT, HARPENDING H (1975). Gammaglobulin groups of the Khoisan peoples of southern Africa: evidence for polymorphism for a *Gm* ^{1,5,13,14,21} haplotype among the San. *Am J Hum Genet* 27: 528-542.

STRINGER CB, ANDREWS P (1988). Genetic and fossil evidence for the origin of modern humans. *Science* 239: 1263-1268.

SUMMERS KM (1987). Restriction fragment length polymorphisms on the X chromosome in populations from Papua New Guinea, Thailand, NorthWestern Australia and China. *Gene Geogr* 1: 207-217.

SYKES BC (1983). DNA in heritable disease. *Lancet* ii: 787-788.

TOBIAS V (1974). The biology of the southern African Negro. In: Hammond-Tooke WD (ed) *The Bantu-speaking peoples of southern Africa*. Routledge and Kegan Paul, London, pp3-45.

TORRONI A, SEMINO O, SCOZZARI R, SIRUGO G, SPEDINI G, ABBAS N, FELLOUS M, SANTACHIARA BENERECETTI AS (1990). Y chromosome DNA polymorphisms in human populations: differences between Caucasoids and Africans detected by 49a and 49f probes. *Ann Hum Genet* 54: 287-296.

TUDDENHAM EGD, COOPER DN, GITSCHIER J, HIGUCHI M, HOYER LW, YOSHIOKA A, PEAKE IR, SCHWAAB R, OLEK K, KAZAZIAN HH, LAVERGNE J-M, GIANELLI F, ANTONARAKIS SE (1991). Haemophilia A: database of nucleotide substitutions, deletions, insertions and rearrangements of the factor VIII gene. *Nucleic Acids Res* 19: 4821-4833.

VAN WARMELO NJ (1974). The classification of cultural groups. In: Hammond-Tooke WD (ed) *The Bantu-speaking peoples of southern Africa*. Routledge and Kegan Paul, London, pp56-84.

VANSINA J (1984). Western Bantu expansion. *J Afr Hist* 25: 129-145.

VIGILANT L, STONEKING M, HARPENDING H, HAWKES K, WILSON AC (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253: 1503-1507.

VOGEL F, KOPUN M (1977). Higher frequencies of transitions among point mutations. *J Mol Evol* 9: 159-180.

WAINSCOAT JS, HILL AVS, BOYCE AL, FLINT J, HERNANDEZ M, THEIN SL, OLD JM, LYNCH JR, FALUSI AG, WEATHERALL DJ, CLEGG JB (1986). Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* 319: 491-493.

WEIFFENBACH B, FALLS K, BRICKER A, HALL L, MCMAHON J, WASMUTH J, FUNANAGE V, DONIS-KELLER H (1991). A genetic linkage map of human chromosome 5 with 60 RFLP loci. *Genomics* 10: 173-185.

WHITE MB, CARVALHO M, DERSE D, O'BRIEN SJ, DEAN M (1992). Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* 12: 301-306.

WOLFF RK, NAKAMURA Y, WHITE R (1988). Molecular characterization of a spontaneously generated new allele at a VNTR locus: no exchange of flanking DNA sequence. *Genomics* 3: 347-351.

WOLFF RK, PLAETKE R, JEFFREYS AJ, WHITE R (1989). Unequal crossingover between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. *Genomics* 5: 382-384.

WOLPOFF MH (1989). Multiregional evolution: the fossil alternative to Eden. In: Mellars PA, Stringer CB (eds) *The Human Revolution*. University of Edinburgh Press, Edinburgh, pp62-108.

WONG Z, WILSON V, PATEL I, POVEY S, JEFFREYS AJ (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann Hum Genet* 51: 269-288.

WIJSMAN EM (1984). Optimizing selection of restriction enzymes in the search for DNA variants. *Nucleic Acids Res* 12: 9209-9226.

YEN PH, MARSH B, ALLEN E, TSAI SP, ELLISON J, CONNOLLY L, NEISWANGER K, SHAPIRO LJ (1988). The human X-linked steroid sulphatase gene and a Y-encoded pseudogene: evidence for an inversion of the Y chromosome during primate evolution. *Cell* 55: 1123-1135.

ZOUTENDYK A, KOPEC AC, MOURANT AE (1955). The blood groups of the Hottentots. *Am J Phys Anthropol* 13: 691-698.

APPENDIX ONE

MEDIA AND SOLUTIONS

Distilled water was used to make up all solutions.

Chloroform

Chloroform as used in protein extraction by organic solvents is 24:1 chloroform:isoamylalcohol. It is reported that this mixture produces better results than chloroform alone (Maniatis *et al* 1987).

Deionized Formamide

formamide	50ml
Amberlite resin	5g

Mix together, stir for 30 min at room temperature. Filter twice through Whatman No.1 filter paper. Store in aliquots at -20°C.

Denaturing Solution

The same solution is used for Biodyne and Hybond-N membranes, and also for stripping of blots.

1.5M NaCl
0.5M NaOH

DNA Markers

Commercial preparations of DNA markers were used undiluted as visible DNA markers. Lambda digested with *Hin*dIII and *Eco*RI, or digested with *Hin*dIII alone (prepared by Boeringer Mannheim) and a 1kb ladder (prepared by Bethesda Research Laboratories) were commonly used for the sizing of DNA fragments.

Invisible DNA markers used in Southern blotting were prepared as follows:

100μl ficoll loading dye
1μg sheared herring sperm DNA
16.7ng commercial lambda DNA molecular weight marker
distilled water to make up to a volume of 1ml

This mixture gives a final concentration of 10μg of sheared DNA in the 10μl aliquot loaded, and ensures that marker DNA will migrate at the same rate as 10μg of digested human genomic DNA. The lambda marker DNA is visualized by hybridization with ^{32}P-dCTP-labelled lambda DNA concurrent with normal hybridization.

Hybridization Solution

Biodyne

0.2% SDS
6xSSC
50mM Hepes pH 7.0
5xDenhardt's solution
0.02mg/ml poly A
0.5mg/ml herringsperm DNA
0.2g/ml dextran sulphate

Freeze in aliquots at -20°C. Dilute with deionized formamide just prior to use.

Hybond-N

Hybridization solution is as for prehybridization solution. Denatured radiolabelled DNA probe is added directly to the prehybridized blot.

LB (Luria Bertani) Medium

Bacto-Tryptone	10g
Bacto Yeast Extract	5g
NaCl	10g
Agar (for LB Agar only)	12g

Adjust pH to 7.5 with NaOH. Make up to a final volume of 1l with distilled water. Sterilize by autoclaving.

Lysozyme Solution for Plasmid DNA Extraction

25mM Tris-HCl pH 7.5

10mM EDTA

15% Sucrose

Make up to volume with distilled water, and sterilize by autoclaving. Add 2mg/ml lysozyme just before use.

Phenol

Melt phenol. Extract several times with an equal volume of 0.1M Tris pH 8.0. Add 8-hydroxyquinoline to a final concentration of 0.1%. Extract once more or until the pH of the aqueous phase reaches 8.0.

Post Hybridization Washing Solutions

Washing solutions were prepared from the stock solutions of 20xSSC or 20xSSPE, and 10% SDS.

Prehybridization Solution

Biodyne

0.2% SDS
6xSSC
50mM Hepes pH 7.0
5xDenhardt's solution
0.02mg/ml poly A
0.5mg/ml herringsperm DNA

Freeze in aliquots at -20°C. Dilute with deionized formamide just prior to use.

Hybond-N

10xSSPE
10xDenhardt's solution
1% SDS
0.25mg/ml herringsperm DNA

Freeze in aliquots at -20°C. Dilute with deionized formamide just prior to use.

Neutralizing Solution

Biodyne

3M NaAcetate pH 5.5

Hybond-N

1.5M NaCl

0.5M Tris pH 7.2

0.001M EDTA

20xSSC

3M NaCl

0.3M NaCitrate

20xSSPE

3.6M NaCl

0.2M $NaH_2PO_4.H_2O$

0.02M EDTA pH 7.7

Stripping Solution

3M NaCl

0.5M Tris pH 7.0

10xTBE

Tris Base 108g

Boric Acid 55g

Na_2EDTA 9.3g

Make up to 1l, and sterilize by autoclaving. Dilute tenfold with distilled water before use.

TE

10mM Tris

1mM EDTA

Bring pH to 8.0 with HCl. Make up to volume with distilled water.

Transfer Solution

Biodyne

10xSSC

Hybond-N

20xSSC

APPENDIX ONE

Frequencies of 49a/*Taq*I Haplotypes in Southern African Populations[a]

Population	N	H^D	Ht1	Ht2	Ht3	Ht4	Ht5
E.Bantu	325	.392	.01 (.01)	.00 (.00)	.06 (.01)	.60 (.03)	.16 (.02)
Zulu	53	.615	.02 (.02)	.00 (.00)	.02 (.02)	.58 (.07)	.13 (.05)
Xhosa	23	.761	.00 (.00)	.00 (.00)	.22 (.09)	.39 (.10)	.00 (.00)
Ndebele	14	.344	.00 (.00)	.00 (.00)	.00 (.00)	.79 (.11)	.21 (.11)
Swazi	33	.559	.00 (.00)	.00 (.00)	.03 (.03)	.64 (.08)	.18 (.07)
Sotho	48	.587	.00 (.00)	.00 (.00)	.02 (.02)	.60 (.07)	.23 (.06)
Pedi	53	.543	.00 (.00)	.00 (.00)	.04 (.03)	.66 (.07)	.13 (.05)
Tswana	41	.599	.00 (.00)	.00 (.00)	.05 (.03)	.61 (.08)	.17 (.06)
Tsonga	31	.666	.03 (.03)	.00 (.00)	.03 (.03)	.52 (.09)	.26 (.08)
Venda	29	.597	.03 (.03)	.00 (.00)	.21 (.08)	.59 (.09)	.14 (.06)
W.Bantu	127	.400	.00 (.00)	.00 (.00)	.03 (.02)	.77 (.04)	.02 (.01)
Ambo	44	.426	.00 (.00)	.00 (.00)	.09 (.04)	.75 (.07)	.00 (.00)
Himba	37	.259	.00 (.00)	.00 (.00)	.00 (.00)	.86 (.06)	.05 (.04)
Herero	46	.476	.00 (.00)	.00 (.00)	.00 (.00)	.72 (.07)	.02 (.02)
Dama	41	.639	.00 (.00)	.00 (.00)	.05 (.03)	.59 (.08)	.02 (.02)
Lemba	49	.807	.00 (.00)	.00 (.00)	.02 (.02)	.20 (.06)	.06 (.03)
Nama	23	.808	.00 (.00)	.00 (.00)	.00 (.00)	.26 (.09)	.35 (.10)
San	84	.840	.00 (.00)	.00 (.00)	.02 (.02)	.17 (.04)	.00 (.00)
Omega San	46	.825	.00 (.00)	.00 (.00)	.02 (.02)	.28 (.07)	.00 (.00)
Tsumkwe San	38	.786	.00 (.00)	.00 (.00)	.03 (.03)	.03 (.03)	.00 (.00)
S.A.Jewish	36	.871	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)
S.A.Indian	63	.900	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.European	53	.737	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)	.04 (.03)
JHB "Coloured"	66	.910	.00 (.00)	.00 (.00)	.02 (.02)	.13 (.04)	.13 (.04)
Richtersveld	34	.915	.09 (.05)	.00 (.00)	.00 (.00)	.12 (.06)	.12 (.06)
Pygmy	32	.759	.00 (.00)	.00 (.00)	.06 (.04)	.44 (.09)	.03 (.03)

Population	Ht6	Ht7	Ht8	Ht9	Ht10	Ht11
E.Bantu	.00 (.00)	.01 (.01)	.00 (.00)	.00 (.00)	.06 (.01)	.04 (.01)
Zulu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.19 (.05)	.00 (.00)
Xhosa	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.22 (.09)	.04 (.04)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.06 (.04)	.03 (.03)
Sotho	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.04 (.03)
Pedi	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)	.07 (.04)
Tswana	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.05 (.03)
Tsonga	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)
W.Bantu	.00 (.00)	.01 (.01)	.00 (.00)	.00 (.00)	.03 (.02)	.03 (.02)
Ambo	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.05 (.03)	.07 (.04)
Himba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)	.00 (.00)
Herero	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)	.02 (.02)	.02 (.02)
Dama	.00 (.00)	.02 (.02)	.02 (.02)	.00 (.00)	.10 (.05)	.05 (.03)
Lemba	.02 (.02)	.12 (.05)	.35 (.07)	.00 (.00)	.04 (.03)	.06 (.03)
Nama	.00 (.00)	.00 (.00)	.04 (.04)	.00 (.00)	.04 (.04)	.09 (.06)
San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.07 (.03)
Omega San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.09 (.04)
Tsumkwe San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.05 (.04)
S.A.Jewish	.00 (.00)	.22 (.07)	.19 (.06)	.00 (.00)	.03 (.03)	.17 (.06)
S.A.Indian	.00 (.00)	.08 (.03)	.17 (.05)	.00 (.00)	.00 (.00)	.13 (.04)
S.A.European	.00 (.00)	.04 (.03)	.02 (.02)	.02 (.02)	.00 (.00)	.05 (.03)
JHB "Coloured"	.00 (.00)	.02 (.02)	.08 (.03)	.00 (.00)	.03 (.02)	.10 (.04)
Richtersveld	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)	.17 (.06)
Pygmy	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.09 (.05)	.06 (.04)

Population	Ht12	Ht13	Ht15	Ht18	Ht19	Ht22
E.Bantu	.00 (.00)	.00 (.00)	.01 (.01)	.01 (.01)	.00 (.00)	.00 (.00)
Zulu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Xhosa	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Sotho	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pedi	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)
Tswana	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)
Tsonga	.00 (.00)	.00 (.00)	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
W.Bantu	.02 (.01)	.00 (.00)	.02 (.01)	.00 (.00)	.00 (.00)	.00 (.00)
Ambo	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Himba	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Herero	.04 (.03)	.00 (.00)	.07 (.04)	.00 (.00)	.00 (.00)	.00 (.00)
Dama	.00 (.00)	.00 (.00)	.05 (.03)	.00 (.00)	.00 (.00)	.00 (.00)
Lemba	.00 (.00)	.00 (.00)	.00 (.00)	.12 (.05)	.00 (.00)	.00 (.00)
Nama	.00 (.00)	.00 (.00)	.04 (.04)	.00 (.00)	.00 (.00)	.00 (.00)
San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Omega San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsumkwe San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Jewish	.00 (.00)	.00 (.00)	.06 (.04)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Indian	.03 (.02)	.19 (.05)	.00 (.00)	.03 (.02)	.00 (.00)	.02 (.02)
S.A.European	.13 (.05)	.00 (.00)	.49 (.07)	.00 (.00)	.00 (.00)	.00 (.00)
JHB "Coloured"	.06 (.03)	.07 (.03)	.21 (.05)	.00 (.00)	.02 (.02)	.00 (.00)
Richtersveld	.03 (.03)	.03 (.03)	.03 (.03)	.03 (.03)	.00 (.00)	.00 (.00)
Pygmy	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)

Population	Ht24	Ht25	Ht26	Ht27	Ht28	Ht29
E.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.01 (.01)	.00 (.00)	.00 (.00)
Zulu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Xhosa	.00 (.00)	.00 (.00)	.00 (.00)	.04 (.04)	.00 (.00)	.00 (.00)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Sotho	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)
Pedi	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)
Tswana	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsonga	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)	.00 (.00)	.00 (.00)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
W.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ambo	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Himba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Herero	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Dama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Lemba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Nama	.00 (.00)	.00 (.00)	.00 (.00)	.04 (.04)	.00 (.00)	.00 (.00)
San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.17 (.04)
Omega San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.04 (.03)
Tsumkwe San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.31 (.08)
S.A.Jewish	.11 (.05)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Indian	.09 (.04)	.00 (.00)	.00 (.00)	.06 (.03)	.00 (.00)	.00 (.00)
S.A.European	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
JHB "Coloured"	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.02 (.02)	.00 (.00)
Richtersveld	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pygmy	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)	.00 (.00)	.16 (.06)

Population	Ht30	Ht31	Ht32	Ht33	Ht34	Ht35
E.Bantu	.01 (.01)	.01 (.01)	.01 (.01)	.00 (.00)	.00 (.00)	.00 (.00)
Zulu	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)
Xhosa	.00 (.00)	.00 (.00)	.04 (.04)	.00 (.00)	.00 (.00)	.00 (.00)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.03 (.03)	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Sotho	.00 (.00)	.00 (.00)	.06 (.03)	.00 (.00)	.00 (.00)	.00 (.00)
Pedi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tswana	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsonga	.00 (.00)	.00 (.00)	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
W.Bantu	.00 (.00)	.00 (.00)	.01 (.01)	.00 (.00)	.02 (.01)	.00 (.00)
Ambo	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)
Himba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Herero	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.04 (.03)	.00 (.00)
Dama	.02 (.02)	.00 (.00)	.02 (.02)	.00 (.00)	.02 (.02)	.00 (.00)
Lemba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Mawa	.00 (.00)	.00 (.00)	.00 (.00)	.04 (.04)	.00 (.00)	.00 (.00)
San	.27 (.05)	.00 (.00)	.05 (.02)	.16 (.04)	.00 (.00)	.01 (.01)
Omega San	.26 (.06)	.00 (.00)	.09 (.04)	.13 (.05)	.00 (.00)	.02 (.02)
Tsumkwe San	.29 (.07)	.00 (.00)	.00 (.00)	.18 (.06)	.00 (.00)	.00 (.00)
S.A.Jewish	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Indian	.03 (.02)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.European	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
JHB "Coloured"	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Richtersveld	.00 (.00)	.00 (.00)	.03 (.03)	.05 (.04)	.00 (.00)	.00 (.00)
Pygmy	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)

Population	Ht36	Ht37	Ht38	Ht39	Ht40	Ht41
E.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.01 (.01)
Zulu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Xhosa	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Sotho	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pedi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.04 (.03)
Tswana	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)
Tsonga	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
W.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.01 (.01)	.00 (.00)	.00 (.00)
Ambo	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Himba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Herero	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)
Dama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Lemba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Nama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
San	.00 (.00)	.00 (.00)	.04 (.02)	.02 (.02)	.00 (.00)	.00 (.00)
Cmega San	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)
Tsumkwe San	.00 (.00)	.00 (.00)	.08 (.04)	.03 (.03)	.00 (.00)	.00 (.00)
S.A.Jewish	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Indian	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)
S.A.European	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
JHB "Coloured"	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Richtersveld	.00 (.00)	.09 (.05)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pygmy	.00 (.00)	.00 (.00)	.00 (.00)	.09 (.05)	.00 (.00)	.00 (.00)

Population	Ht42	Ht43	Ht44	Ht45	Ht46	Ht47
E.Bantu	.01 (.01)	.01 (.01)	.01 (.01)	.00 (.00)	.00 (.00)	.00 (.00)
Zulu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Xhosa	.00 (.00)	.04 (.04)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Sotho	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pedi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tswana	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)
Tsonga	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
W.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ambo	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Himba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Herero	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Dama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Lemba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Nama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Omega San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsumkwe San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Jewish	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Indian	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.European	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.02 (.02)	.04 (.03)
JHB "Coloured"	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Richtersveld	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.03 (.03)	.00 (.00)
Pygmy	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)

Population	Ht48	Ht49	Ht50	Ht51	Ht52	Ht53
E.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.01 (.01)	.00 (.00)
Zulu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)
Xhosa	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Sotho	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pedi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tswana	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)
Tsonga	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
W.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ambo	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Himba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Herero	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Dama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Lemba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Mama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.04 (.04)	.00 (.00)
San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Omega San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsumkwe San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Jewish	.00 (.00)	.00 (.00)	.00 (.00)	.06 (.04)	.00 (.00)	.00 (.00)
S.A.Indian	.02 (.02)	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)
S.A.European	.00 (.00)	.00 (.00)	.05 (.03)	.02 (.02)	.02 (.02)	.00 (.00)
JHB "Coloured"	.00 (.00)	.00 (.00)	.03 (.02)	.00 (.00)	.00 (.00)	.00 (.00)
Richtersveld	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pygmy	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)

Population	Ht54	Ht55	Ht56	Ht57	Ht58	Ht59
E.Bantu	.00 (.00)	.01 (.01)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Zulu	.00 (.00)	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Xhosa	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Sotho	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pedi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tswana	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsonga	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
W.Bantu	.00 (.00)	.00 (.00)	.01 (.01)	.00 (.00)	.00 (.00)	.01 (.01)
Ambo	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Himba	.00 (.00)	.00 (.00)	.03 (.03)	.00 (.00)	.00 (.00)	.00 (.00)
Herero	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)
Dama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)
Lemba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Nama	.00 (.00)	.00 (.00)	.04 (.04)	.00 (.00)	.00 (.00)	.00 (.00)
San	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Omega San	.04 (.03)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsumkwe San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Jewish	.00 (.00)	.00 (.00)	.00 (.00)	.06 (.04)	.00 (.00)	.00 (.00)
S.A.Indian	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.European	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)	.00 (.00)
JHB "Coloured"	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Richtersveld	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pygmy	.00 (.00)	.09 (.05)	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)

Population	Ht60	Ht61	Ht62	Ht63
E.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Zulu	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Xhosa	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Ndebele	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Swazi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Sotho	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Pedi	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tswana	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsonga	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Venda	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
W.Bantu	.00 (.00)	.00 (.00)	.00 (.00)	.01 (.01)
Ambo	.00 (.00)	.00 (.00)	.00 (.00)	.02 (.02)
Himba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Herero	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Dama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Lemba	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Nama	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Omega San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
Tsumkwe San	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Jewish	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
S.A.Indian	.00 (.00)	.03 (.02)	.00 (.00)	.00 (.00)
S.A.European	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)
JHB "Coloured"	.02 (.02)	.00 (.00)	.00 (.00)	.00 (.00)
Richtersveld	.03 (.03)	.00 (.00)	.12 (.06)	.00 (.00)
Pygmy	.00 (.00)	.00 (.00)	.00 (.00)	.00 (.00)

[a]Haplotypes 1-24 are those reported by Ngo *et al* (1986), Lucotte *et al* 1989 and Lucotte *et al* (1990b). Haplotypes 25-63 are those reported in this study and displayed in Table 5.3. Standard errors are presented in parentheses after the haplotype frequencies frequencies. Population frequencies denoted in boldface type are those used for the cluster analysis (see Figure 5.2).

[b]Nei's measure of gene diversity (Nei 1987).