

THE ISOLATION AND LOCALIZATION OF ARBITRARY RESTRICTION
FRAGMENT LENGTH POLYMORPHISMS IN SOUTHERN AFRICAN
POPULATIONS

Vera Conn

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

Johannesburg, 1987

ABSTRACT

A human HaeIII/AluI DNA library in lambda bacteriophage Charon 4A, was screened for single-copy human DNA sequences. Twenty-three out of 85 potential human DNA sequences actually contained human DNA inserts, the remainder contained only Charon 4A sequences. These clones were used to screen seven random individuals for restriction fragment length polymorphisms (RFLPS) using fourteen different restriction enzymes.

Eight base pair substitution-type RFLPs and one rearrangement-type RFLP, involving 500 bp of DNA, were detected. Four of these RFLPs were found with the restriction enzyme TaqI and three with MspI, which both have CpG dinucleotides in their recognition sequences. The other RFLP was detected with AvaII. The di-allelic DNA rearrangement-type RFLP can be detected by many restriction enzymes. Each of two clones detected RFLPs with two restriction enzymes and another clone was found to detect three allele RFLP in the Caucasoid population. The PIC values obtained using these probes ranged between 0,20 and 0,43 in the different population groups studied.

These RFLP-detecting DNA sequences have been provisionally assigned to chromosomes 1, 2, 4, 6, 7 and 8 using a panel of somatic cell hybrid lines. Three of the RFLP-detecting sequences have been subcloned into plasmid vectors.

iii.

The frequencies of the alleles at each of the RFLP loci have been determined in the Caucasoid, Negroid and San population groups of southern Africa. One RFLP was found to be monomorphic in the Negroid and San populations. The frequencies of the rarer allele of all the other RFLPs was at least one percent in all the populations studied. A genetic distance analysis using the allele frequencies of the RFLPs in the different population groups studied showed the Caucasoid population to be separate from the Negroid and San populations which showed a closer relationship.

Six of the nine RFLPs have already been tested on some of the families from the Centre for the study of Human Polymorphism (C.E.P.H.) collaboration study. Two probes isolated in this study have been found to both map to chromosome one and to be linked (LOD score of 6.2, $\theta = 0.0$) and no recombinants have yet been observed.

The RFLPs detected in this study contribute to a much larger effort to map the human genome and in addition, they will contribute to our understanding of the origins and relationships of the diverse populations of southern Africa.

iv.

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 being submitted before for any degree or examination in any other university.



Ver: Conn

20th day of August, 1987.

v.

This work is dedicated to my husband Joseph Alexander who
has constantly supported me, my parents, and Lieb and
Kitty Verga for their encouragement and love.

Acknowledgments

I am very grateful to all the people who kindly participated in this study, including the many members of the different populations groups in South West Africa and Natal and the families from around Johannesburg.

I would like to thank my supervisor Professor Trefor Jenkins for his guidance and continuing support throughout this study. I would especially like to thank Dr. Michele Ramsay for her constant support, enthusiasm and encouragement.

Many thanks to my colleagues, Ruth Heim, Debbie Morris and Ananda Krause for their helpful discussions and the support that they gave me.

My thanks are extended to Dr. T. Maniatis for providing the genomic DNA library and Doctors U. Francke, T. Mohandas and A. Retief who helped to complete our panel of somatic cell hybrids by generously donating cell lines. A special thanks to the members of the Cytogenetics Unit who have worked very hard to provide us with somatic cell hybrid lines, especially Maria Dos Santos and Isabelle Marques.

I acknowledge the South African Institute for Medical Research, The South African Medical Research Council and the University of the Witwatersrand for financial assistance during the past three years. I would also like to thank Professor Jack Metz for the facilities which helped make such a study possible.

vii.

Thanks to Janice Gallant for assisting with the figures, Jo Waltham, Mel Anderson, Martha Ullrich and Gill Martin for all the photographic prints.

Finally, I would like to thank Joseph Alexander Verga for his never ending support and constant encouragement. I am also very grateful for his excellent assistance with all the computer work involved, especially during the printout stages.

ABBREVIATIONS

ACD	acid citrate dextrose
BSA	Bovine serum albumen
bp	base pairs
	degrees celcius
cm	centimeter
cpm	counts per minute
Ci	Curie
DNA	deoxyribonucleic acid
DNAse	deoxyribonuclease
DNA pol I	deoxyribonucleic acid polymerase I
dATP	deoxyadenylic acid
dCTP	deoxycytidylic acid
dGTP	deoxyguanylic acid
dTTP	deoxythymidylic acid
EDTA	ethylene diaminetetra-acetic acid
g	gram
Kb	kilobases
KCl	potassium chloride
l	litre
M	moles per litre
mA	milliamps
mg	milligrams
mM	millimolar / millimoles per litre

ix.

mV	millivolts
nm	nanometres
PVP	polyvinyl pyrrolidone
RNA	ribonucleic acid
RNAse	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
TCA	tri-chloro acetic acid
Tris	tris-(hydroxymethyl)-aminomethane
μ g	microgram
μ l	microlitre
UV	ultra violet

TABLE OF CONTENTS	PAGE
ABSTRACT.	ii
DECLARATION	iv
ACKNOWLEDGMENTS	vi
LIST OF FIGURES	xv
LIST OF TABLES.	xxii

CHAPTER ONE -INTRODUCTION

1.1 THE HUMAN GENOME	1
1.1.1 The Structure of the Human Genome	1
1.1.2 Human Variation.	5
1.1.2.1 Protein Polymorphism	9
1.1.2.2 DNA Polymorphism	12
1.1.3 The Isolation of Human DNA sequences	15
1.2 RESTRICTION FRAGMENT LENGTH POLYMORPHISMS	19
1.2.1 Discovery	19
1.2.2 Types of RFLPs found in the genome	22
1.2.3 Strategies to Isolate RFLPs	27
1.2.4 A Genetic Linkage Map using RFLPs	34
1.3 APPLICATIONS OF RFLPS	35
1.3.1 Medical applications	36
1.3.2 Anthropological	41
1.4 MAPPING THE HUMAN GENOME	43
1.4.1 Classical Genetics	43
1.4.2 Somatic Cell Genetics	44

PAGE

1.4.3	Molecular Genetics	53
1.4.4	Collaboration to Map the Human Genome	57
1.4.5	Conclusions	58
1.5	AIMS OF THIS STUDY	58
 CHAPTER TWO -- MATERIALS AND METHODS		
2.1	COLLECTION OF MATERIALS	60
2.2	ISOLATION OF SINGLE-COPY HUMAN PROBES	65
2.2.1	Screening a Human Genomic Library	65
2.2.2	Isolation of bacteriophage DNA	69
2.2.3	Restriction Enzyme analysis of Bacteriophage DNA	72
2.3	ISOLATION OF HIGH MOLECULAR WEIGHT HUMAN GENOMIC DNA	73
2.3.1	DNA Extraction	73
2.3.2	Analysis of Human DNA	75
2.4	RESTRICTION ENZYME ANALYSIS	76
2.5	GEL ELECTROPHORESIS	78
2.6	SOUTHERN BLOTTING TO SOLID SUPPORTS	81
2.6.1	The Southern Blotting Procedure	81
2.6.2	Modifications for Nylon Membranes	81

	PAGE
2.7 RADYOLABELLING OF DNA PROBES	83
2.7.1 Nick Translation	84
2.7.2 Monitoring the Labelling Reaction	85
2.7.3 Removal of the non-incorporated nucleotides	86
2.7.4 Oligolabelling of DNA fragments	87
2.8 DNA HYBRIDIZATION	89
2.9 AUTORADIOGRAPHY	90
2.10 STRIPPING AND REUSE OF MEMBRANES	91
2.11 DETERMINATION OF DNA FRAGMENT SIZES	92
2.12 LOCALIZATION USING SOMATIC CELL HYBRID LINES	93
2.13 LINKAGE STUDIES	96
2.14 RESTRICTION MAPPING	97
2.15 SUBCLONING OF BACTERIOPHAGE ISOLATES DETECTING RFLP.	99
2.15.1 Preparation of the vector DNA	100
2.15.2 Preparation of the human DNA fragments	104
2.15.3 Ligation and transformation	105
2.15.4 Recombinant Analysis	108
2.16 EVALUATION OF METHODS	111

PAGE

CHAPTER THREE - RESULTS

3.1 HUMAN GENOMIC PROBES	114
3.1.1 The Isolation of Human DNA Probes	114
3.1.2 Subcloning of certain Probes	120
3.2 RFLPs ISOLATED IN THIS STUDY	123
3.2.1 Restriction Fragment Length Polymorphisms	123
3.2.2 Localization of the RFLPs detecting DNA sequences.	151
3.2.3 Summary	165
3.3 CHARACTERIZATION OF UNUSUAL RFLP DETECTED BY VC85	166
3.4 POPULATION STUDIES	177
3.5 ESTIMATES OF GENETIC HETEROZYGOSITY	183
3.6 LINKAGE STUDY -PRELIMINARY RESULTS	185
3.7 SUMMARY	187

CHAPTER FOUR - DISCUSSION

4.1 THE POLYMORPHISM CONCEPT	188
4.2 THE ISOLATION OF SINGLE-COPY REGIONS IN THE HUMAN GENOME	193

PAGE

4.3 STRATEGIES USED TO SCREEN FOR RFLPs.	202
4.4 TYPES OF RFLPs ISOLATED	204
4.5 LOCALIZATION OF RFLP-DETECTING PROBES	210
4.6 RFLPs USED FOR POPULATION STUDIES	213
4.7 CONTRIBUTION TO MAPPING OF THE GENOME.	223
4.8 FUTURE STUDIES	228
APPENDIX A	231
APPENDIX B	235
APPENDIX C	242
REFERENCES	243

LIST OF FIGURES

PAGE

CHAPTER ONE

- 1.1. a) Schematic diagram of base pair substitution-type RFLP.
 b) Schematic diagram of DNA rearrangement t-type
 RFLP. 23
- 1.2. The hypervariable region 5' to the human insulin gene. 25

CHAPTER TWO

- 2.1. An autoradiograph of a nitrocellulose filter containing bacteriophage DNA and hybridized to radiolabelled total human DNA. 68
- 2.2. An ethidium bromide stained agarose gel containing human DNA, digested with the restriction enzyme AvaII and separated out by electrophoresis. 80
- 2.3. An agar plate of JM103 bacterial colonies after transformation with plasmid pUC19. 107

CHAPTER THREE

PAGE

- 3.1. a) An ethidium bromide stained agarose gel containing bacteriophage clones, digested with EcoRI. 116
- b). An autoradiograph of the Southern blot of the agarose gel in a) showing that some of the human inserts have hybridized to total human DNA. 116
- 3.2. An ethidium Bromide stained gel containing Lambda phage clones digested with EcoRI. 117
- 3.3. A schematic representation of pUC19 and the subclones pVC61, pVC63, pVC35. 121
- 3.4. An ethidium bromide stained gel containing pUC19, pVC85, Lambda HindIII marker, pVC63, pVC61, and the hybridization pattern of each of the fragments to its subclone. 122
- 3.5. An autoradiograph of Southern blot of an MspI digest of DNA from random Negroid individuals, hybridized to a) probe VC14, b.) the 11,0 MspI fragment of VC14 and c) the 8,5 kb MspI fragment of VC14. 125
- 3.6. Schematic representation of the MspI restriction pattern of probe VC14. 126

PAGE

- 3.7. An autoradiograph of a Southern blot of an MspI digest on DNA (C.E.P.H. no. 1424), hybridized to VC14 and b) the 11,0 kb MspI fragment. 127
- 3.8. An autoradiograph of a Southern blot of DNA from random Negroid individuals digested with MspI and hybridized to a) probe VC28 and b) the 3,6 kb MspI fragment of VC28. 129
- 3.9. An autoradiograph of the bacteriophage clone VC28 hybridized to a Southern blot containing DNA, digested with MspI, from a five member family. . . 131
- 3.10. An autoradiograph of a Southern blot of random Caucasoid DNA digested with TaqI and hybridized to a) probe VC28 and b) the 3,6 kb MspI fragment of VC28. 132
- 3.11. An autoradiograph of a Southern blot of TaqI digested DNA from three generation family (C.E.P.H. no. 13294) and hybridized to the 3,6 kb MspI fragment of probe VC28. 133
- 3.12. An autoradiograph of Southern blot of DNA from random Caucasoid individuals, digested with TaqI and hybridized to the 1,98 and 1,0 kb EcoRI fragments of probe VC61. 135

- 3.13. An autoradiograph of a Southern blot of a TaqI digest of a three generation family (C.E.P.H. no. 13294) hybridized subclone pVC61. 135
- 3.14. An autoradiograph of the probe VC63 hybridized to a Southern blot of DNA from random Caucasoid individuals, digested with TaqI. 137
- 3.15. An autoradiograph of a Southern blot of TaqI digests of DNA from some individuals of a three generation family (C.E.P.H. no. 1416), hybridized the subclone pVC63. 138
- 3.16. Schematic representation of the TaqI restriction pattern of probe VC63. 139
- 3.17. An autoradiograph of a Southern blot of DNA from random San individuals, digested with TaqI and hybridized to probe VC64. 141
- 3.18. Schematic representation of the TaqI polymorphism detected by VC64. 141
- 3.19. An autoradiograph of probe VC64 hybridized to a Southern blot of TaqI digests on DNA from a five child family. 142

xix.

PAGE

- 3.20. An autoradiograph of probe VC75 hybridized to a Southern Blot of AvaII digestions on DNA from random Caucasoid individuals. 144
- 3.21. An autoradiographs of Southern blots of DNA digested with AvaII from individuals of two families showing in a) the segregation of the 4,3 and 3,5 kb alleles and in b) the segregation of the 3,5 and 3,2 kb alleles in (C.E.P.H. no 1331) . 145
- 3.22. Schematic representation of the deletion/insertion-type RFLP detected by probe VC85. . . . 147
- 3.23. Autoradiographs of DNA from random individuals digested with four different restriction enzymes namely a) StuI b) PvuII c) BglIII d) BamHI, and hybridized to the probe VC85. 148
- 3.24. An autoradiograph of a Southern blot of DNA, from a family, digested with BamHI and hybridized to probe VC85. 150
- 3.25. An autoradiograph of probe VC14 hybridized to a Southern blot of somatic cell hybrid DNA digested with HindIII. 154
- 3.26. An autoradiograph of Southern blot of somatic cell hybrid DNA digested with HindIII and hybridized to probe VC61. 155

XX.

PAGE

3.27. An autoradiograph of Southern blot of DNA of somatic cell hybrid lines digested with <u>HindIII</u> and hybridized to probe VC64.	157
3.28. An autoradiograph of a Southern blot of somatic cell hybrid DNA digested with <u>HindIII</u> and hybridized to probe VC85.	158
3.29. An autoradiograph of probe VC28 hybridized to Southern blot of a <u>HindIII</u> digest of somatic cell hybrid DNA.	160
3.30. An autoradiograph of Southern blot of DNA from somatic cell lines digested with <u>HindIII</u> and hybridized to the 1,3 and 0,95 kb <u>EcoRI</u> fragments, of probe VC63.	162
3.31. An autoradiograph of Southern blot of DNA from somatic cell hybrid lines hybridized to probe VC75.	164
3.32. An ethidium bromide stained agarose gel of probe VC85 digested with different enzymes for restriction mapping purposes.	167
3.33. A partial restriction map of VC85	170
3.34. An autoradiograph of a Southern blot of DNA from random Caucasoid individuals, digested with <u>MspI</u> , and hybridized to probe VC85.	172

xxi.

PAGE

3.35. Restriction map showing <u>MspI</u> sites in the human genome detected by probe VC85.	173
3.36. An autoradiograph of a Southern blot of DNA from random Caucasoid individuals, digested with <u>TaqI</u> and hybridized to probe VC85.	174
3.37. Schematic representation of the <u>TaqI</u> restriction pattern of VC85 around the deletion/ insertion region.	176
3.38. Dendrogram indicating the genetic distance between four population groups studied.	182

LIST OF TABLES

PAGE

Chapter Two

2.1. The requirements of the restriction enzymes used in this study.	77

Chapter Three

3.1. The properties of human DNA inserts in Charon 4A.	118
3.2. Chromosomal content of somatic cell hybrid lines and corresponding hybridization pattern of DNA probes.	152
3.3. Summary of restriction fragment length polymorphisms detected and their chromosomal localizations.	165
3.4. Human fragment sizes (kb) detected by probe VC85 after Southern Blot Analysis.	168
3.5. DNA fragment sizes found in phage clone VC85.	168
3.6. Allele frequencies of DNA sequence polymorphisms in Populations studied	180
3.7. Statistical analysis of the genotype data between populations.	181
3.8. Genetic distance between populations.	182

xxiii.

- 3.9 Estimation of genetic heterogeneity at loci
identified by the probes isolated in this study. . 184
- 3.10. Summary of the results of the P.P.H. study. . . 186

CHAPTER ONE

INTRODUCTION

1 CHAPTER ONE INTRODUCTION

The unravelling of the human genome represents a challenge not only because it is so vast, but because it also organizes all the information coding for life, along the sequence of the DNA double helix.

Since it has become possible to study the human genome directly, using recombinant DNA technology, knowledge about the organization of the genome has rapidly increased. The number of functional DNA sequences is as yet unknown but it appears as though the entire genome is not required for normal function. The construction of a genetic linkage map, for each chromosome, with a specific set of evenly spaced polymorphic markers, will allow for the eventual fine mapping of the human genome. This study represents a contribution to the effort of finding some of these markers.

1.1 THE HUMAN GENOME

1.1.1 The Structure of the Human Genome

The human haploid genome has been estimated, on the basis of recombination fractions in males and females, to be contained within 33,3 Morgans of DNA (Renwick, 1971). The number of functional genes in the genome, based on an average gene length of 1 000 nucleotides and the amount of

DNA present, has been estimated to be about 50 000 (O'Brien, 1973; Bishop, 1974). These genes code for the basic amino-acid sequences of the polypeptide chains of proteins. From estimates of the number of genes in the genome, it has been concluded that more than 90% of the genome may be composed of non-functional, non-informational or "junk" DNA (Ohta and Kimura, 1971; O'Brien, 1973; Orgel and Crick, 1980).

Repeated DNA sequences have been found dispersed throughout eukaryotic genomes (Britten and Kohne, 1968). About 30% of the human genome consists of highly or mid-repetitive sequences and the rest contains single-copy regions, in which the sequences coding for functional genes are found (Schmid and Deininger, 1975). Highly repetitive regions in the human genome have been classified into long and short interspersed sequences known as LINES and SINES (Jelinek *et al.*, 1980; Singer, 1982; Schmid and Jelinek, 1982; Shafit-Zagardo *et al.*, 1982).

In the new recently-proposed terminology, LINES are called L1Hs, where L1 stands for the first long interspersed repeat and Hs for Homo sapiens, the species in which it was first isolated. L1Hs were first isolated from human DNA and were known as the KpnI family of repeats. They were found to be present in 15% of the plaques during screening of human genomic libraries (Shafit-Zagardo *et al.*, 1982). These repeat sequences varied in size between 1.2 and 1.9 kb and the sequences were resistant to digestion with the enzyme KpnI. The 3' end of the sequences are homologous to other repeat sequences. One KpnI member was shown to hybridize to

RNA extracted from a human and a monkey cell line, thus suggesting that some of these sequences could be transcribed (Lerman *et al.*, 1983). A new LINE sequence, L2Hs, which shows no cross-hybridization to the L1Hs family, has been found. It is of recent evolutionary development and is only found in gorilla and humans (Musich and Dykes, 1986).

The repeat sequence known as Alu is a member of the SINE repeat family and is found to be highly repeated in human and non-human primate DNA. Based on renaturation studies, it has been estimated that there are between 300 000 and 500 000 copies per human haploid genome, thus accounting for 3 to 5% of the genome. The average Alu sequence is about 300 bp long and consists of an internal imperfect duplication of a shorter sequence (Schmid and Jelinek, 1982). If these sequences were randomly distributed, they would be present once in every 2 200 bp in the genome, but the distances between Alu sequences have been shown to be variable. In the 50 kb of DNA containing the β -globin-like gene cluster, only eight Alu sequences have been found and five have been found within a 26 kb region containing the human α -like globin gene cluster (Fritsch *et al.*, 1980). The human insulin gene locus is flanked by two Alu sequences, which are separated by 21 kb (Bell *et al.*, 1980). Ninety-four percent of the clones of a human genomic library were found to be positive for Alu sequences (Crampton *et al.*, 1981; Schmid and Jelinek, 1982). Approximately 60% of these sequences can be cleaved at a common site using the restriction enzyme AluI (Houck *et al.*, 1979; Schmid and Jelinek, 1982).

Other repetitive DNA families have been isolated and characterized from the human genome including alphoid sequences (Willard, 1984), a Sau3A family of repeats (Kiyama et al., 1986) and four types of satellite DNA (Gosden et al., 1975; Mitchell et al., 1979; Frommer et al., 1982).

Some of the mid-repetitive sequences found in the human genome have been shown to have a coding function. Examples include the genes coding for the 45S precursor of the 18S and 28S ribosomal RNA (rRNA), which have a structural function in the assembly of the ribosome; they are repeated to the extent of 150-300 copies per haploid genome (Jeanteur and Attardi, 1969; Wellauer and Dawid, 1979). The genes coding for 5S rDNA and transfer RNA (tRNA) have also been found to be repeated in the human genome (Hatlen and Attardi, 1971). The human histone genes coding for the H2A, H2B, H3 and H4 proteins are each repeated 30 to 40 times (Wilson and Melli, 1977).

Single-copy regions of the human genome contain the unique coding sequences for the polypeptide chains of proteins and there are also other sequences presently of unknown function. These single-copy regions are interspersed with repetitive sequences (Schmid and Deininger, 1975; Strayer et al., 1983).

Methylation of sequences in the human genome

DNA sequences have been shown not to be equally mutable and those that are highly mutable are termed "hotspots" (Benzer, 1961). Hotspots have been found in the lacZ gene of Escherichia coli (E.coli) and have been shown to be due to a spontaneous base pair substitution, occurring frequently at 5-methylcytosine residues (Coulondre et al., 1978). These hotspots disappeared if the cytosine (C) residues were not methylated. It was proposed that the spontaneous deamination of the 5-methylcytosine resulted in the formation of thymine (T) which then would not be excised by the repair enzyme (Coulondre et al., 1978; Duncan and Miller, 1980). Methylation at CpG sequences has also been reported in the human genome (Van der Ploeg and Flavell, 1980; Cooper, 1983). A relative deficiency of CpG sequences has been found in mammals in addition to a heavily methylated genome thus supporting the suggestion that 5-methylcytosine tends to mutate abnormally frequently to thymine (Bird, 1980). In eukaryotic genomes repetitive sequences also tend to be hypermethylated (Ehrlich et al., 1982).

1.1.2 Human Variation

Inherited variation in the human genome has been studied for two main reasons. Firstly, because of its relevance to inherited disorders, where changes in genes cause

alterations in the protein products and result in phenotypic defects. Secondly, to increase the understanding of normal variation and the role that it plays in evolution.

Variation at the DNA level that results in a protein or enzyme product which is structurally or functionally impaired could lead to an inherited disorder. The seriousness of the disorder will depend on the effects of the mutant allele in the heterozygous or homozygous form. DNA changes which give rise to inherited diseases can therefore be regarded as extreme examples of the allelic variation which is ubiquitous throughout the human population (Harris, 1980).

A genetic polymorphism was 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 polymorphism was further defined by Bodmer and Cavalli-Sforza (1976) as "the occurrence of two or more alleles at a particular locus where the rarer allele is present at a frequency of at least one percent".

Polymorphism is, therefore, a term reserved for those loci which have alleles at significant frequencies in that population. Alleles present at a frequency of less than one percent are termed variants (Meisler, 1983).

How a mutation is maintained and reaches polymorphic frequencies has been the subject of discussion by two main opposing schools of thought represented by the selectionists and the neutralists.

The " new mutation, random genetic drift " hypothesis of Kimura (Kimura, 1968a, 1968b; Kimura and Ohta, 1971; Crow, 1973) proposes that variation is present in the population due to random genetic drift or chance. Sampling processes will determine which of the gametes, produced by an individual of one generation, will give rise to new individuals in the next. If this mutation is not deleterious, it has a chance of being passed onto half of the individuals in the next generation and its chances of persisting decrease with each new generation. The chances of a new mutation surviving were discussed by Fisher over 50 years ago :

" In a reasonably large population, where each pair of parents is replaced on average by two children who survive to have children, the probability that a new mutant will still be present after, say, 15 generations is only about 1 in 9." (Fisher, 1930 cited in Harris, 1980, p380).

The chances of a new mutation surviving are greater if a population is expanding than if it is declining. At any one time the number of neutral alleles in a population will

represent a balance between the generation of new alleles by mutation and the elimination of others by random genetic drift.

The selectionists argue that natural selection gives rise to genetic diversity. Selection will act against those alleles which decrease the biological fitness and in favour of those which increase fitness, where fitness is a measure of the relative contribution which individuals of different genotypes make to the next generation. Selection could, of course, have acted in the past but may no longer be operating in the present environment and since the allele frequency changes are slow, over many generations, the polymorphism still persists (Fisher, 1930; Haldane, 1962; and Wright, 1966).

Linkage disequilibrium may also be responsible for the existence of genetic polymorphisms. When certain alleles at two closely linked loci occur together more often than would be expected by chance, i.e. they are in linkage disequilibrium, one of the alleles could be maintained as a result of selection operating on the nearby gene. This effect is known as "hitch-hiking" (Thompson, 1977; Bodmer and Thompson, 1977). The HpaI site 3' to the human β -globin gene is associated with the sickle cell mutation in some populations and occurs at a higher frequency in that population due to linkage disequilibrium and selection for the sickle cell gene in a hyperendemic malaria area (Kan and Dozy, 1978a). It appears as though most DNA sequence

polymorphisms are neutral and are present in the populations at frequencies which have been reached due to random genetic drift (Cooper et al., 1985).

1.1.2.1 Protein Polymorphism

The first examples of polymorphism to be found in man included the ABO blood group system discovered by Landsteiner at the beginning of this century, and red-green colourblindness found in certain male individuals who were not able to distinguish between certain colours (Kalmus, 1965). The phenylthiourea taste polymorphism was discovered when certain individuals were found to be unable to taste the extreme bitterness of the substance phenylthiourea (Fox et al., 1932; Harris and Kalmus, 1949). The glucose-6-phosphate-dehydrogenase polymorphism in Negroids was recognized because certain individuals developed a haemolytic anaemia when primaquine was used as an anti-malarial drug (Hockwald et al., 1952). This was later shown to be due to a deficiency of this X-linked enzyme (Carson et al., 1956).

When techniques such as starch gel electrophoresis were developed, it became possible to identify polymorphisms in proteins and a host of enzymes. A polymorphism in human haptoglobin was one of the first to be recognized by starch gel electrophoresis, when serum from different individuals was shown to have distinct characteristic patterns (Smithies, 1955). Further technical alterations, including the addition of denaturing or reducing agents to the gel

matrix allowed for greater efficiency in the separation of polypeptide chains of proteins. In this way subtypes of haptoglobin, were demonstrated and three common alleles were identified (Connell et al., 1962; Smithies et al., 1962).

In the 1960s, these findings led to the independent studies, by Harris and Hopkinson in humans and by Lewontin and Hubby in *Drosophila pseudoobscura*, which demonstrated that variation was a common normal phenomenon. They used electrophoretic techniques which were expected to detect only that variation which led to a change in the charge of a protein. Both groups showed that enzyme polymorphism was ubiquitous in naturally occurring populations (Harris, 1966; Harris and Hopkinson, 1972; Lewontin and Hubby, 1966). By 1977, 104 different human enzyme loci had been screened for electrophoretic variation and 33 of these loci showed variation in at least one of the major population groups (Caucasoid, Negroids, Asiatic Indians and Orientals). Some of the electrophoretic polymorphisms were limited specifically to one population group (Hopkinson et al., 1976).

Proteins of blood plasma were also screened for variation using electrophoresis. Haptoglobin could be identified by its property of binding to hemoglobin (Smithies, 1955) and transferrin by its binding to iron (Smithies and Hiller, 1959). Further protein polymorphisms were identified with the development of isoelectric focusing in which a protein migrates to a position in a pH gradient gel corresponding to its isoelectric point (Svensson, 1962). A two-dimensional electrophoretic technique allows for the identification of

tissue protein polymorphisms. In the first dimension denatured proteins are separated using isoelectric focusing and the second dimension separates proteins according to their molecular weight by electrophoresis in SDS (O'Farrell, 1975).

The ABO blood group polymorphism was the first to be identified using immunological techniques and naturally occurring antibodies (Landsteiner, 1901). Today, an extensive series of antisera against other red cell surface antigens has been described (Race and Sanger, 1975). Variation has been found on the surface of almost all nucleated cells due to the presence of cell surface antigens which are encoded by the Major Histocompatibility locus (HLA). Four loci (A, B, C and DR) have been defined by the use of carefully absorbed antisera in cytotoxicity tests on peripheral blood lymphocytes (A, B and C) and B lymphocytes (DR). These antigenic differences are due to variation at four closely linked loci, on chromosome 6, each having a large number of alleles (reviewed by Amos and Kostyu, 1980).

Results of protein polymorphisms were used to estimate that any individual was likely to be heterozygous at about 6% of his loci (Harris and Hopkinson, 1972). This was an underestimate since only those changes in the coding regions, which altered the electrophoretic mobility of the enzyme, were detected. Neel (1984) considered the impact of a number of recent studies on protein variation on the index of heterozygosity using techniques of isoelectric focusing, thermostability studies, and 1D and 2D polyacrylamide electrophoretic techniques. Studies on enzymes had allowed

for the subtyping of electromorphs and resulted in a higher estimation of heterozygosity. The heterozygosity estimate from enzymes studies was calculated to be 17-18% , using data on electrophoretic variants, thermostability variants, and an estimate of non-charge change amino acid substitutions, which were not detected by electrophoresis or thermostability studies. The new estimate of heterozygosity from studies on soluble proteins using 1D electrophoresis and 2D polyacrylamide electrophoretic techniques was calculated to be 9%. Therefore, the overall estimate of heterozygosity using the new data from soluble proteins and enzymes was 12-13% (Neel, 1984).

1.1.2.2 DNA Polymorphism

Protein polymorphisms reflect a change in the DNA sequence coding for specific polypeptide chains. The development of recombinant DNA technology has allowed for the direct analysis of variation at the DNA level. DNA sequence polymorphisms were detected using restriction enzymes and a cloned DNA sequence as a probe. This type of variation has been termed a restriction fragment length polymorphism (RFLP) and will be addressed in section (1.2).

DNA sequence polymorphisms are mostly due to nucleotide substitutions which could be now be detected around genes, as well as within genes. These sequence changes would not have previously been detected at the protein level unless an altered product had been produced due to a mutation which had disrupted the normal splicing out of the intervening

sequences or sequences 5' and 3' to the coding regions which are important for normal transcriptional and translational processing of the mRNA.

Human heterozygosity has further been estimated by studying DNA sequence variation. DNA heterozygosity was initially estimated from data on the β -globin gene cluster region, where 1 variable site per 100 nucleotides was found to exist (Jeffreys, 1979). This estimate was revised to 1 in 200 nucleotides using population genetics theory (Ewens et al., 1981) and further data on the β -globin cluster region changed the estimate to 1 in 500 bp (Kazazian et al., 1983). The incidence of RFLPs near the human albumin gene suggested that 1 in 95 bp varied (Murray et al., 1983).

All the available data on DNA sequence variation, detected using cloned coding and non-coding regions, were used to calculate human heterozygosity based on the formula developed by Nei (1975). The formula was based on the total number of base pairs screened and the polymorphic sites identified (Cooper and Schmidtke, 1984). The heterozygosity rate of the human genome was estimated to be 0.00376 for the autosomes and 0.00115 for the human X chromosome. Schmidtke and Cooper (1984) criticized estimates of heterozygosity which were derived from the β -globin study of Jeffreys (1979) because they were based on a single region containing several coding sequences, and recalculated the results at the albumin locus to be 0.0025, rather than 0.011 as calculated by Murray et al., (1983).

Random single-copy probes provided the first unbiased estimates of DNA heterozygosity since the coding potential of the DNA sequences within these clones was unknown. The average heterozygosity was calculated to be 0,003. Therefore estimates derived from the β -globin locus (Jeffreys, 1979) appear to be representative of the rest of the genome (Cooper et al., 1985). Data on the X-chromosome (Murray et al., 1982; Bakker et al., 1983; Aldridge et al., 1984) indicated that heterozygosity of this chromosome was not significantly different from the rest of the genome (Cooper et al., 1985). Ohno's law (1967) has predicted the conservation of the X chromosome because the localization of identical coding sequences on the X chromosome is conserved between species. Heterozygosity of the X chromosomal sequences were therefore expected to reflect the same constraint in non-coding regions (reviewed by Cooper et al., 1985).

Highly polymorphic regions have been found near the human insulin gene, the α -globin gene cluster and the heavy chain of the immunoglobulin complex. Because not many such regions have been found, it was suggested that a non-uniform distribution of variation over the human genome existed (Bell et al., 1981; Higgs et al., 1981; Mignone et al., 1983). Results from base pair substitution type RFLPs have also indicated this non-uniform distribution and clustering in the non-coding introns or intergenic regions. The RFLPs observed within the human β -like globin gene loci were within introns and flanking regions of the genes (Jeffreys 1979; Poncz et al., 1983), and only 1 out of the 12 RFLPs detected was found within an exon (Kazazian et al., 1983).

Forty-three polymorphisms were observed in 11 cloned alcohol dehydrogenase genes, 2,7 kb in length, which were isolated from five populations of *Drosophila* and only 1 amino acid substitution was observed (Kreitman, 1983).

From the above studies it has been concluded that most polymorphic variation occurred outside the coding regions of the DNA, at a frequency of about 1 in every 200 to 300 bp. When the results of DNA sequence variation studies were compared with previous estimates from protein data, DNA heterozygosity was found to be between 6 and 10 times higher. The uneven distribution pattern of RFLPs throughout the human genome reflects the sites of mutational processes and the action of selection. If mutations are more or less randomly distributed along the DNA, the distribution of RFLPs outside of the coding regions, may reflect the action of selection (Neel, 1984). These findings agreed with the predictions of the neutral theory of molecular evolution of Kimura (1983,p238) where it was predicted that the human genome should contain $1,4 \times 10^7$ heterozygous nucleotides due to neutral mutation, (cited in Cooper et al., 1985).

Heterozygosity of the human genome is useful for the study of genetic variability in mutation rates, natural selection and genetic drift. It is important for genomic mapping and is presently being used in constructing a linkage map. Variation is exploited in the preclinical and prenatal diagnosis of human genetic disorders, the detection of carriers of genetic disease, and is also useful in medico-legal work, particularly for disputed paternity

investigations. The applications of DNA variation between normal individuals will be discussed in section 1.3.

1.1.3 The Isolation of Human DNA sequences

The isolation and analysis of coding and non-coding regions of human DNA sequences was the result of major technical advances followed by a number of important discoveries.

Restriction endonucleases which could cleave DNA at specific recognition sites were discovered and became important tools to divide up the human genome into unique fragments (Smith and Wilcox, 1970). The first recombinant DNA hybrid molecules were constructed between eukaryotic (Drosophila melanogaster) and bacteriophage DNA (Thomas, Cameron and Davies, 1974). The construction of recombinant DNA molecules requires a number of essential steps for the insertion of DNA fragments from any source into a viral or plasmid replicon. These steps include the specific cleavage of both donor and recipient DNA, the covalent rejoining of the two molecules, the transformation or infection of the host bacterial strain and the recovery of the recombinant DNA by selection systems (Rodríguez et al., 1977).

By 1977 suitable plasmid and bacteriophage vectors had been constructed. Plasmids are small circular DNA molecules which are found independently replicating in certain bacterial strains. They often contain genes which confer resistance to certain antibiotics and hence their host bacteria have an advantage in certain environments

(Rodriguez et al., 1977). The Charon series of bacteriophage Lambda were also suitable vectors because the two internal EcoRI DNA fragments could be removed and replaced with eukaryotic DNA, without the loss of the ability to grow lytically. DNA fragments ranging in size from 7 to 20 kb can be inserted into these vectors (Blattner et al., 1977). In vitro packaging of recombinant phage DNA into phage protein heads, which is an essential step for their recovery, was developed by Hohn and Murray (1977).

In order to identify and isolate a gene or region of DNA of interest, a number of strategies were needed. Firstly, two screening methods were developed, one is used to screen bacterial colonies : the presence of hybrid plasmids and the other is used to screen bacteriophage plaques for foreign DNA inserts (Grunstein and Hogness, 1975; Cami and Kourilsky, 1978; Benton and Davies, 1977).

Secondly, the DNA sequence which is to be used as a probe needs to be labelled. Nucleic acid hybridization is a very powerful technique used to detect and quantitate specific DNA and RNA sequences. One of the DNA strands in the reaction can be labelled, usually with an isotope. The ability of *E. coli* DNA polymerase I to catalyze a nick translation reaction was recognized by Kelly et al., (1970) and a procedure to radiolabel DNA sequences was refined by Rigby et al., (1977).

These developments increased the speed of screening large numbers of bacterial colonies or plaques. The potential of this technique in the isolation of unique human genes rather than the physical purification of the gene from mRNA, was realized (Benton and Davies, 1977).

The Southern Transfer technique was developed to analyze specific DNA fragments generated by restriction enzymes and separated by electrophoresis in agarose gels. A technique to transfer the DNA directly from an agarose gel to a solid support (nitrocellulose) was developed by Southern (1975). These filters could then be hybridized to radiolabelled DNA sequences. After autoradiography the sequence of interest is revealed as a series of dark bands on an X-ray plate. This method is still widely applied in the field of human genetics.

Human genomic and cDNA libraries

The above technology led to the construction of human complimentary DNA (cDNA) and genomic libraries which allowed for the isolation of many human genes and other single-copy DNA sequences.

In the construction of eukaryotic cDNA libraries, messenger RNA (mRNA) was isolated from a cell, transcribed into complimentary DNA (cDNA) using the enzyme reverse transcriptase and cloned into suitable plasmid vectors (Higuchi *et al.*, 1976). A cDNA library was, therefore, a representation of the DNA sequences which are expressed in the cell type from which the RNA was first isolated. The

first human cDNA clone to be isolated was human chorionic somatomammotropin from a cDNA library constructed from placental RNA (Shine et al., 1977).

A human genomic DNA library was prepared from a foetal liver (Lawn et al., 1978). The DNA was partially digested using restriction endonucleases, such that any sequence would then be completely represented in the library, and cloned into bacteriophage Charon 4A. Many other human DNA libraries have since been constructed (Blattner et al., 1978; Williamson et al., 1981).

The isolation of cDNA and genomic clones for the mouse β -globin genes led to the discovery that eukaryotic genes were not contiguous but interrupted by stretches of DNA called introns or intervening sequences (Tilghman et al., 1978). Intervening sequences were then shown to be present in the mouse immunoglobulin and human β -globin genes (Tonegawa et al., 1978; Lawn et al., 1978).

1.2 RESTRICTION FRAGMENT LENGTH POLYMORPHISMS

1.2.1 Discovery

A restriction fragment length polymorphism (RFLP) represents DNA sequence variation detected by using a restriction enzyme, where loss or gain of a restriction site leads to an alteration in the length of the DNA fragment generated. The rarer variant must have attained a frequency of at least one percent.

There are three classes of restriction enzymes with type II being the most useful in molecular biology. Restriction fragment length differences were first recognized after the discovery and isolation of type II restriction endonucleases (Smith and Wilcox, 1970). The first type II, site-specific endonuclease was purified from Haemophilus influenzae and was shown to degrade phage T7 DNA into 40 unique fragments. The specific recognition site of the enzyme was found to be 5-6 bp long (Smith and Wilcox, 1970; Kelly and Smith, 1970). More than 300 type II enzymes are now known (reviewed by Kessler, Neumaier and Wolfe, 1985).

RFLPs were first used as a tool for genetic analysis to map the temperature sensitive mutations of Adenoviruses. Specific restriction fragments carrying the mutation were identified and the physical and genetic maps of the mutations were deduced (Grodzicker et al., 1974). The maternal inheritance of mitochondrial DNA in mammals was demonstrated using RFLPs. A distinct pattern of mitochondrial restriction fragments was observed in the mitochondrial DNA of a mule. This pattern corresponded to the restriction fragments of the maternal and not the paternal mitochondria (Hutchison III et al., 1974). The first human restriction fragment length differences between individuals were also discovered in mitochondrial DNA (Potter et al., 1975).

Genomic RFLPs were detected after the cloning and isolation of specific human genes. A PstI restriction difference between the two δ - β -globin gene clones indicated that the foetus, from whom the cDNA library was constructed, was a

heterozygote for this PstI site in the intervening sequence of the δ -globin gene. This PstI difference was confirmed by a Southern blot analysis of the foetal liver DNA (Lawn et al., 1978) and was later shown to be present in one other individual and his mother which indicated that it was stably inherited (Jeffreys, 1979). This variation has not been found in a significant frequency and hence has been termed a variant.

The first human genomic RFLP was found due to variation in some Negroid individuals, at a site 3' to the β -globin gene and it was detected with the restriction enzyme HpaI (Kan and Dozy, 1978a). In Caucasoids, the β -globin gene was found on a 7,6 kb HpaI fragment but fragment length variation was found to be 7,6, 7,0 or 13,0 kb, in American Negroids and Asians (Kan and Dozy, 1978a). The 13,0 kb fragment was shown to be associated with the sickle gene in some Negroid populations. This polymorphism made possible the first prenatal diagnosis of a foetus at risk for Sickle Cell Anaemia, using recombinant DNA technology (Kan and Dozy, 1978b).

1.2.2 Types of RFLPs found in the genome

Different types of DNA sequence changes can result in RFLPs.

Base pair substitution

Mutations lead to base pair changes in the genome, which alter the restriction enzyme-recognition sequences. These mutations are either transitions or transversions where a transition is the mutation of a purine to a purine or a pyrimidine to a pyrimidine and a transversion is the change from a purine to pyrimidine or a pyrimidine to a purine. The restriction fragment length generated is then increased or decreased, depending on whether a recognition site was created or lost by the mutation (Figure 1.1a). Base pair substitutions are usually only recognized by one restriction enzyme.

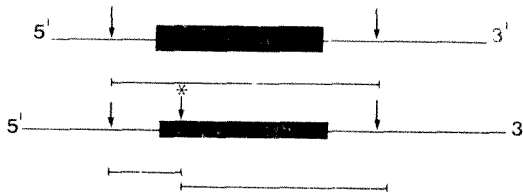
DNA rearrangements

DNA rearrangements alter the sequence of the bases along a region of DNA, thereby altering the positions of the recognition sites of a number of restriction enzymes. Rearrangements can occur due to deletions or insertions or even inversions of various DNA sequences.

Deletions / Duplications / Insertions

The duplication or insertion of a DNA sequence will lead to a constant increase in the DNA fragment lengths generated by any enzymes which cut the DNA 5' and 3' to the rearrangement. Deletions will follow the same pattern except that the differences in the fragment lengths generated will be reduced. This fragment length variation will be equal to the number of bases added or deleted as

a)



b)

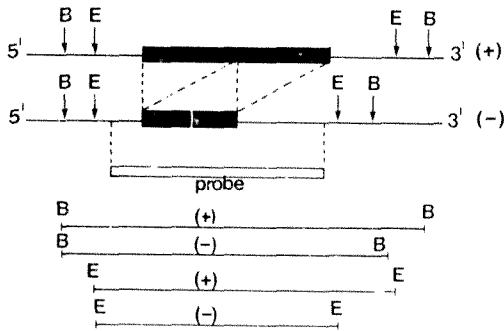


Figure 1.1. a) Schematic diagram of base pair substitution-type RFLP where $\blacksquare$ indicates the region cloned and the (*) indicates the polymorphic site. The human DNA fragments generated are shown below the diagram. b) Schematic diagram of DNA rearrangement-type RFLP where $\blacksquare$ indicates the region which is rearranged. The human fragments generated are indicated below when two enzymes (B and E) are used. A (+) indicates that the region is present and (-) indicates that the region is absent on the chromosome, generating DNA fragments which show a constant decrease in fragment length.

illustrated in Figure 1.1b. If an enzyme has a recognition site within the inserted or deleted region, the DNA fragments generated will behave as base pair substitution-type RFLPs, where a site has been created within a larger fragment generating two smaller fragments.

Hypervariable regions

This type of DNA rearrangement is due to the presence of a small sequence of DNA which is tandemly repeated along a length of DNA. The fragment length will vary according to the number of times this sequence is repeated. This type of RFLP was found 5' to the human insulin gene. A 14 bp consensus sequence was tandemly repeated and many alleles, of different sizes, can be defined at this locus (Bell *et al.*, 1981; Bell *et al.*, 1982). These alleles have been classified into three classes (Figure 1.2) (Bell *et al.*, 1984). This type of polymorphic region has been found elsewhere in the human genome such as 3' to the α -globin gene and will be discussed later (Higgs *et al.*, 1981).

The majority of DNA sequence polymorphisms that have been reported were due to single nucleotide changes, which altered the recognition site of a restriction enzyme (Skolnick *et al.*, 1984). Those DNA rearrangements however, which involved small regions of DNA would only have been detected if a restriction site was present within the rearranged region and these would have been grouped together with the base pair substitution-type RFLPs.

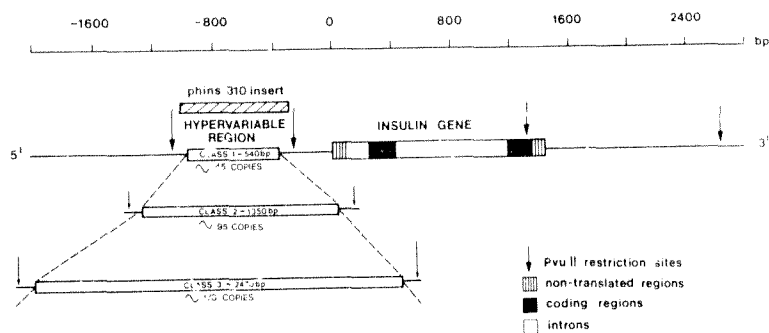


Figure 1.2. The hypervariable region 5' to the human insulin gene showing three classes of alleles generated by different copy numbers of the HVR consensus sequence, when hybridized to the probe pHINS310. (Adapted from Bell *et al.*, 1984)

The first arbitrary RFLP -pAW101

Most RFLPs were initially detected using cloned sequences of DNA which represented genes. The first arbitrary RFLP was detected using a 16 kb single-copy DNA sequence (pAW101) and more than eight alleles were defined.

This region was proposed to be highly polymorphic and a number of enzymes could be used to reveal differences in fragment sizes. The fragment length variation was demonstrated to be inherited in a co-dominant Mendelian fashion in a three generation family (Wyman and White, 1980).

This region was localized to chromosome 14q21-ter, by using somatic cell hybrids and in situ hybridization techniques (DeMartinville et al., 1982; Donlon et al., 1983) and it was renamed D14S1 since it was the first single-copy DNA region to be assigned to chromosome 14 (Human Gene Mapping Workshop, 1983). It is interesting that D14S1 has been localized proximally to the human immunoglobulin heavy chain genes, which are another highly variable region in the human genome (Balazs et al., 1982; Mignone et al., 1983).

The region causing the polymorphism was later found to be outside of the sequence which was initially cloned. The flanking sequences had resisted cloning and this led to the discovery that certain sequences, containing inverted repeats, are not stable in *recA+* hosts. It was only possible to isolate this sequence when a *recB* or *recC* host strain was used (Leach and Stahl, 1981; Wyman et al.,

1984). The D14S1 region was shown to consist of a 60 mer repeat sequence, which also contained inverted repeats. Differences in the copy number of this sequence are responsible for the polymorphic variation reported at this locus (Wyman et al., 1985). Furthermore, these results indicated that the human genomic libraries prepared using recA⁺ hosts may not be complete since such unstable sequences would have been lost during library amplification.

1.2.3 Strategies to Isolate RFLPs

In any study to screen for RFLPs a number of parameters must be evaluated. Firstly, the type of DNA sequence to be used as a probe; secondly, the restriction enzymes to be used; and thirdly the number of individuals to be screened.

Sources and isolation of DNA probes

A large number of potential sources exist for DNA probes. Firstly, human genomic libraries which contain large or small DNA inserts can be used. These libraries will contain a high frequency of repetitive human DNA sequences and therefore must be screened for single-copy regions by hybridization to radiolabelled total human genomic DNA (Botstein et al., 1980; Kao et al., 1982). The efficiency in isolation of single-copy sequences was found to be greater using libraries with smaller rather than larger inserts (Pearson et al., 1981).

Secondly, complimentary DNA (cDNA) libraries have also been used as a source of probes to screen for RFLPs. Such clones are relatively free of repetitive DNA and were initially used to show polymorphisms around the human insulin, and the α - and β -globin genes. In a study to evaluate the usefulness of cDNA clones compared with genomic clones in the search for RFLPs, the results demonstrated that single-copy cDNA clones did not reveal more RFLPs (Helentjaris and Gesteland, 1983).

Thirdly, the combination of somatic cell genetics and recombinant DNA technology led to the construction of autosomal and X chromosome-specific DNA libraries. Chromosome-specific single-copy clones could be isolated from these libraries and used as DNA probes (Schneckpeper *et al.*, 1979; Gusella *et al.*, 1980). When techniques of flow cytometry were applied to the sorting of human chromosomes (Young *et al.*, 1981), purified individual chromosomes could be used to construct autosomal and X chromosome-specific DNA libraries (Davies *et al.*, 1981; Krumlauf *et al.*, 1982; Kunkel *et al.*, 1982). The latter DNA libraries were free from any rodent DNA sequences. Chromosome-specific genomic probes have been a very valuable source of probes to detect DNA variation, especially when there has been interest in a specific region of a chromosome.

Fourthly, cloned genes have been a valuable source of probes in the search for RFLPs as demonstrated by the large number of RFLPs detected using the α - and β -globin genes (reviewed by Antonarakis, 1985a) and those reported at human gene mapping (HGM 6, (1981), HGM 7 (1983) and HGM 8 (1985).

Such RFLPs exist within and around genes and are important for linkage studies of these regions especially if the gene is known to be involved in disorders of medical relevance.

Once potential single-copy clones have been isolated they must be analyzed by restriction digests to ascertain whether they contain human DNA. These human inserts can again be checked for repetitive DNA by Southern blot analysis against total human genomic DNA (Kao *et al.*, 1982).

Evaluation of Restriction enzymes used to screen for RFLPs

In order to detect DNA variation, Southern blots, prepared from the DNA of random individuals digested with a selection of restriction enzymes, must be screened using a radiolabelled clone. A number of studies have evaluated the usefulness of certain restriction enzymes to detect DNA variation (Bastie-Sigeac and Lucotte, 1983; Bishop *et al.*, 1983; Wjisman, 1984).

It has been reported that enzymes, *MspI* and *TaqI*, which both have a CpG dinucleotide in their recognition sequence, show an increased detection rate of RFLPs (Schafer and White, 1982; Skolnick and White, 1982). The CpG detection rate for RFLPs was reported to be 0.23 compared to 0.13 for other enzymes (Barker *et al.*, 1984). *MspI* and *TaqI* cleave the sequences CCGG and TCGA respectively. These CpG dinucleotide sequences are frequently methylated and the spontaneous mutation of the methylated C residue could alter the recognition site of these two enzymes. The restriction

enzyme BglII also has been shown to be useful in the search for RFLPs using smaller genomic probes (Pearson et al., 1981)

Number of subjects

In screening studies to detect RFLPs a small number of individuals (6 to 9) are usually tested because it has been shown theoretically that only the low frequency, rare variants would be missed and an increased efficiency in the isolation of high frequency polymorphisms would be gained (Skolnick and White, 1982). An approach for the isolation and detection of high frequency RFLPs, using large genomic probes combined with certain restriction enzymes, has been proposed (Feder et al., 1985).

When variation among individuals was detected it was investigated further by means of family studies to confirm that the different alleles were inherited in a co-dominant Mendelian fashion. Those variants in the population having the rarer allele present at a frequency of greater than 1% constitute a restriction fragment length polymorphism as defined above.

New Strategies used to isolate RFLPs

The number of arbitrary RFLPs reported on the human gene mapping conferences increased from 18 in 1981, to 95 in 1983 and to 245 in 1985 (Willard et al., 1985). This number of reported RFLPs was an underestimate since many researchers had not contributed their data to this collection (Schumm et

al., 1985). Ninety percent of the reported RFLPs were due to single base pair changes, 40% of the RFLPs had rare allele frequencies less than 0.2 and the polymorphic information content (PIC) values of 90% of the RFLPs were less than 0.4 (Skolnick et al., 1984). The PIC value is the probability that a given offspring of a parent carrying the rare allele, at the index locus, will allow for the deduction of the parental genotype at the marker locus (Botstein et al., 1980).

The initial limiting factor in the search for RFLPs was the extent of repetitive sequences found in the human genome. Cosmid vectors developed by Hohn and Collins (1980) contained up to 45 kb of genomic DNA, were not used as probes unless the internal single-copy human fragments isolated and subcloned.

A technique to use cosmids as RFLP-detecting probes has been developed by Litt and White, (1985). The repetitive DNA sequences are blocked out prior to hybridization by incubation of the radiolabelled cosmid clone with a vast excess of sheared total human genomic DNA.

Pre-reassociation of repetitive sequences occurs but the single-copy regions remain single-stranded (Litt and White, 1985; Sealey et al., 1985). When cosmid clones are used a larger genomic region is screened for variation and hence more polymorphic sites are defined. The alleles at the loci often show linkage disequilibrium between each other and therefore haplotype studies are needed. Two highly polymorphic regions have been defined on chromosomes 2q and 12q using this method (Litt and White, 1985; Buraker et al.,

1986). This method has been compared to previous strategies and has been found to be as successful in the detection of single-copy RFLPs (Barker et al., 1985).

A number of highly polymorphic regions have been found in the human genome and these have been termed "hypervariable regions" (HVR), and they are all due to short tandem repeats. Examples are found 5' to the human insulin gene (Bell et al., 1982), as discussed above, and around the human α -like globin gene loci. These regions have been found 3' to the α -globin gene (Higgs et al., 1981), within the larger intron of the pseudozeta and zeta globin genes (Proudfoot et al., 1982) and between the pseudozeta and zeta globin genes (Goodbourn et al., 1983). The HVR 3' to the α -globin gene has been found to be due to differences in the copy number of a 17 bp sequence (Jarman et al., 1987). The repeat sequence within the intron of the zeta globin gene is 14 bp long and is similar to the human insulin gene repeat (Proudfoot et al., 1982). Other similar sequences have been reported for the c-HaRas-1 oncogene and the human μ light chain of the immunoglobulin complex (Goldfarb et al., 1982; Mignone et al., 1981). The human myoglobin gene contains a 33 bp repeat sequence, similar to the sequences described above, which was flanked by an 8 bp direct repeat characteristic of a target site duplication generated by transposable elements (Jeffreys et al., 1980a).

Strategies have now been developed to search for similar hypervariable regions elsewhere in the human genome. The core sequences of the HVR's have been shown to detect many other autosomal loci simultaneously. The discovery of

minisatellite sequences using the 33 bp myoglobin core sequence as a probe, has revealed that a unique pattern of fragments is generated for every individual resulting in a "DNA fingerprint" (Jeffreys *et al.*, 1985a). Synthetic oligonucleotide sequences corresponding to the core sequences of HVRs of the insulin, myoglobin, zeta and α -globin and other genes have been prepared and these oligomers are being used as probes to search for other highly polymorphic regions in the human genome. A number of highly polymorphic loci, which are due to the variation in the number of tandem repeats, have been found (personal communication Dr. R. L. White, 1986; Nakamura *et al.*, 1987).

1.2.4 A Genetic Linkage Map using RFLPs

Prior to 1980, the RFLPs which had been found were all associated with cloned genes. The frequency of RFLPs at these loci was rather low due to conservation of such regions. Botstein *et al.*, (1980) proposed that a search for RFLPs, using cloned single-copy human DNA regions as probes, combined with Southern blot analysis of the DNA of random individuals, digested with different restriction endonucleases, should be conducted. These RFLPs which are found in the single-copy regions of the genome could, by linkage analysis, be shown to be associated with genes. This meant that the location of a gene could be found without its prior isolation. The identification of a sufficient number of RFLPs, at even intervals throughout the human genome, could be used to construct a genetic linkage

map of the human genome. These single-copy regions could be localized using the techniques of somatic cell hybridization and in situ hybridization to human chromosomes. Such regions could then be used in linkage analysis to other known markers as well as between each other.

It was proposed that only about 150 evenly-spaced highly polymorphic DNA markers would be needed to span the human genome (Botstein et al., 1980). This number was based on estimates that the autosomal length of the human genome was 33 Morgans (Renwick, 1971). These RFLPs need to be found at intervals of 20 cM and to have PIC values of at least 0,5. Many more markers first need to be tested before this task will be completed. The estimated number of RFLPs required to map the human genome was revised to 1500, by Lange and Boehnke (1982), as these markers would be random with respect to each other and would fall into clusters resulting in sizeable gaps. PIC values for highly, reasonably and slightly informative markers are greater than 0,5, between 0,25 and 0,5 or less than 0,25 respectively. For two-allele systems the maximum possible PIC value is 0,38. As the number of alleles at a locus increases, the PIC value becomes closer to 1. Loci having high PIC values will be more favourable for mapping of the human genome but may not be very useful for population and evolution studies. The mapping of the human genome will be discussed further in section 1.4.

1.3 APPLICATIONS OF RFLPS

This section will review how RFLPs have already been successfully applied in human and population genetics and other field of medicine. An important application of RFLPs has been in prenatal diagnosis and carrier detection of individuals carrying the gene for a specific disorder. In the past six years since their discovery, RFLPs have been shown to be useful in many other aspects of science and medicine and some of these will be discussed.

1.3.1 Medical applications

Isolation of Genes

RFLPs have been the starting point for the isolation and identification of several genes responsible for major genetic disorders, for which the biochemical defect was unknown. The following strategy was generally used.

A set of RFLPs, localized to chromosomes, are used as known genetic markers to search for linkage to a genetic disorder of unknown location. All individuals in a family are typed using the RFLP and the segregation of the trait, for which the locus is unknown, is compared with the segregation of the alleles of the RFLP. Linkage studies are then performed on the data using tables compiled by Morton (1955a) or the computer programmes LIPED or LINKAGE (Ott, 1974; Lalouel and Lathrop, 1984a, 1984b). When positive LOD scores are found

linkage is suggested and steps to isolate the actual gene are taken. Finally, the cause of the genetic defect giving rise to the disease is sought.

Chromosome walking techniques are then used systematically to isolate the DNA sequences flanking the initial linked sequence. It involves the restriction mapping of this linked sequence and isolating the extreme 5' and 3' DNA fragments. The sequences are used to screen a human DNA library for sequences that correspond to these flanking sequences. These new sequences are then restriction mapped and orientated with respect to the original sequence. Bidirectional walking is needed because the orientation of the original sequence with regard to the disease locus is not known. This procedure was used to isolate the genes from the aca, ros, and the hithorax complex in Drosophila (Bender *et al.*, 1983). Two closely spaced markers, flanking the putative gene locus are needed for the eventual isolation of the locus. The distance between the two flanking markers can be estimated by observing the recombination between the RFLPs at a number of meiotic events.

One new technique, to facilitate walking, has been developed. In it cosmids are used as vectors because they can contain up to 40 kb of foreign DNA. Because conventional techniques of electrophoresis were limited to the resolution of 20 kb fragments of DNA, a new electrophoretic technique, known as pulse-field gel electrophoresis or orthogonal field alteration electrophoresis, was devised. This technique can be used to

separate large DNA molecules, of up to two million bp in size, which are generated using restriction enzymes having rare DNA recognition sites in the human genome (Swartz and Cantor, 1984; Carle *et al.*, 1986; Smith *et al.*, 1986). These electrophoretic techniques will facilitate the mapping of the human genome and the isolation of many genes (reviewed by Little, 1986; Shaw, 1986).

Chromosome jumping techniques have been developed to reduce the number of "walking steps" when large distances must be covered. It involves the ligation of a marker fragment to one end of the linear DNA fragment (up to 200 kb) under investigation. The fragment is then circularized and digested with an enzyme which does not have a recognition site within the marker fragment. This fragment would represent a marker for two DNA sequences which were normally separated by 200 kb. The left-hand marker can then be used to "jump" 200 kb downstream to the right-hand marker (Collins and Weissman, 1984).

Localization of genetic disorders

A number of genetic disorders have been localized to specific human chromosomes using the above described strategies. The gene causing Huntington's Disease was found to be linked to a random DNA segment, which was localized to chromosome 4pter (Gusella *et al.*, 1983; Gusella *et al.*, 1985). Cystic fibrosis was linked to the serogenetic enzyme marker, paraoxonase (PON) (Eiburg *et al.*, 1985) and tight linkage to DNA markers, which map to the long arm of chromosome 7, have been found (Knowlton *et al.*, 1985a;

White et al., 1985). A candidate for the potential gene causing cystic fibrosis has recently been isolated (Estivill et al., 1987). The gene for Duchenne Muscular Dystrophy was first shown to linked to two X-linked probes, RC8 and L1-23, which were localized to the short arm of the X-chromosome (Murray et al., 1982; Davies et al., 1983). A portion of the DMD gene has also recently been isolated after extensive research (Kunkel et al., 1985; Kunkel and co-workers, 1986; Monaco et al., 1986).

Highly polymorphic RFLPs have been very important in linkage studies where, because of their high PIC values, they are usually informative in all families (Donis-Keller et al., 1986; Higgs et al., 1986). The hypervariable region 3' to the human α -globin gene on chromosome 16 has been shown to be linked to Adult Polycystic Kidney Disease and also to phosphoglycolate phosphatase (Reeders et al., 1985; Reeders et al., 1986). A dominant gene which confers a strong predisposition to manic depressive disease in the Amish population, has been linked to the highly variable region at the HaRas gene locus, on chromosome 11p, (Egeland et al., 1987). No linkage was demonstrated in three Icelandic (Hodgkinson et al., 1987) and three North American kindreds (Detera-Wadleigh et al., 1987) thus indicating that this psychosis is heterogeneous.

The number of genes isolated through linkage to RFLP-detecting DNA markers is increasing and the goal of a complete genetic linkage map of the human genome is in

sight. Much work will however still be needed to find the loci responsible for many genetic disorders, even when a linkage map of the human genome is completed. (This will be discussed further in section 1.4). An extensive review of all genetic diseases which can be diagnosed using recombinant DNA technology has been published (Antonarakis et al., 1985a; Cooper and Schmidtke, 1986). The detection of markers for genetic diseases raises important medical, social, legal, and ethical issues and there are widespread implications for the future of genetic diagnosis.

Other Medical Applications

RFLPs have been applied to other medical fields. The clonal origin of tumours has been determined using X-linked RFLPs and this could have important clinical implications (Vogelstein et al., 1985). Highly informative RFLPs have been useful to evaluate graft rejection, graft-verses-host reactions and the recurrence of leukemia after bone marrow transplants (Knowlton et al., 1986). RFLPs, detected using the genes coding for class II HLA antigens, have been used to sub-divide the serologically detected class II antigens and an increased association of some of the DNA fragments with insulin-dependent diabetes-mellitus and Rheumatoid Arthritis has been shown (Festenstein et al., 1986).

RFLPs also are useful for medico-legal investigations including disputed paternity and forensic medical issues. Classical genetic markers including red cell enzymes, blood group antigens and HLA antigens have successfully been applied to paternity testing. These markers could be

replaced by DNA markers such as highly polymorphic RFLP loci (Kazazian et al., 1986; Smouse and Chakraborty, 1986). The minisatellite regions, detected using the myoglobin repeat sequence, have been used to determine the origin of a Ghanaian boy in an immigration test-case (Jeffreys et al., 1985b). These minisatellite fragments are stably inherited in a Mendelian fashion and the mutation rate is low (Jeffreys et al., 1985a). Minisatellite probes could have major implications in forensic medicine as it has been shown that results could be obtained from old blood stains (Gill et al., 1985).

1.3.2 Anthropological

As early as 1918, genetic markers were used for anthropological studies, when it was demonstrated that the ABO blood groups varied according to the ethnic origin of the population (Hirszfeld and Hirszfeld, 1919, cited in Cavalli-Sforza et al. 1986). In 1950, individuals were assigned to their ethnic group of origin by using ABO, MN and Rh blood group data (Boyd, 1950). However, it has been found that individual loci, are rarely sufficiently informative to draw conclusions about the ethnic origins of individuals. The accumulation of information from many genes should allow for the reconstruction of the evolutionary history of population groups.

In 1964, 5 loci were used to construct a tree of descent, of human ethnic groups, consisting of 15 populations (Cavalli-Sforza and Edwards, 1964). Such results gave

respect to their origin. Initially the accumulation of data for such studies has been slow, because the number of classical polymorphic markers were low. Some polymorphisms have been specifically found present in only one population group. For example, the Fy^a and Fy^b alleles at the Duffy blood group locus are not found in individuals of African origin, but in Caucasoids both Fy^a and Fy^b are found at significant frequencies. The presence of the Fy^a or Fy^b alleles or both in individuals of African origin have probably some degree of Caucasoid admixture. Gene flow between population groups can give rise to changes in the gene frequencies of the original parent population (reviewed in Bodmer and Cavalli-Sforza, 1976).

The discovery of RFLP which directly detect DNA sequence variation has allowed for these studies, on human populations, to proceed more rapidly. By analyzing the frequencies of the alleles in different populations it is possible to construct genetic distance maps between the populations (Cavalli-Sforza et al., 1986). Human racial groups have been studied at the nuclear level by analyzing differences in restriction sites at certain gene loci. The data from the human β -globin like gene loci show that there was an early split between the African and Caucasoid / Mongoloid populations and that a relatively small founder population immigrated out of Africa (Wainscoat et al., 1986; Jones and Rouhani, 1986). This interpretation has not been accepted by everyone (Giles and Ambrosia, 1986; Greenwood, 1986).

The same approach has been possible using mitochondrial DNA. The human mitochondrial genome is a circular molecule consisting 16,5 kb of DNA. It has been completely sequenced (Anderson et al., 1981) and shown to evolve 5-10 times more rapidly than human nuclear DNA. Mitochondria are maternally inherited and every individual therefore has only one type of mitochondria and no differences due to recombination are generated. Differences between mitochondrial DNA types in individuals can be determined by restriction mapping and high resolution mapping, using agarose and polyacrylamide gel electrophoresis. The analysis of 467 independent sites in the mitochondrial genome, using 12 different restriction enzymes, in 147 individuals, from 5 different populations revealed 133 different mitochondrial types (Cann et al., 1987). It has been estimated that a single ancestral mitochondrion has evolved to give rise to the variation present in man today. There are two primary lines of descent, the first being specific for Africans and the second for a few African and all other racial groups. This data was therefore consistent with nuclear data, from the β -globin gene sequences (Wainscoat et al., 1986).

The development of Y chromosomal probes, which detect RFLPs, will be useful to study the evolution of the paternal ancestor. A few polymorphic regions have been found on the Y chromosome and the variation appears to be present in all population groups studied (Page et al., 1982; Bishop et al., 1984; Casanova et al., 1985). Both mitochondrial and Y chromosome-specific RFLPs will be useful to study gene flow

between racial groups. The results will give information about the direction of gene flow and the common ancestors of the racial groups.

The origin of specific gene clusters and mutations can be analyzed using RFLPs. The 13 kb HpaI fragment associated with the $Hb\beta^S$ globin gene in some Negroid populations was also found to be associated with the $Hb\beta^C$ gene mutation. These results indicated that the mutation, which gave rise to the HpaI site, predated the origin of the sickling and the $Hb\beta^C$ mutation (Kan and Dozy, 1980). Further haplotype studies, using extensive restriction mapping, and 11 different polymorphic restriction sites around the β^S -globin gene, led to the conclusion that the mutation causing the sickle cell gene has arisen independently in three areas in Africa (Pagnier et al., 1984).

Type I Gaucher Disease, which is prevalent in the Ashkenazi Jewish population, was proposed to have originated in a single extended kindred. However, restriction mapping of the defective gene has demonstrated heterogeneity, thus indicating that the mutation causing Gaucher Disease has arisen more than once, even in this small population group (Sorge et al., 1985). Haplotype studies around the Pi^Z gene causing α_1 -antitrypsin deficiency has indicated that the mutant allele was always associated with an identical haplotype, in all populations groups, thus suggesting there is probably a single origin of this mutation, which has predated the divergence of the populations in which it occurs (Cox et al., 1985).

Gene clusters, such as the α - and μ -like globin gene clusters and the human growth hormone somatomammotrophin gene clusters, are said to have originated by gene duplication and translocation events to the same or different chromosomes. These events can be analyzed by using RFLPs at and around these loci (reviewed by Jeffreys, 1981; Chakravarti et al., 1984).

1.4 MAPPING THE HUMAN GENOME

The time-span during which the human genome has been mapped can be divided into a number of periods. Classical Genetics was first used to map genes by analysis of the segregation patterns of traits in families. Somatic Cell Genetics was used initially to map human genes to specific chromosomes using cell culture techniques. Finally molecular genetic techniques have resulted in the isolation of human genes from human DNA libraries and, by exploiting somatic cell hybrids or by means of in situ hybridization to an ever increasing number of genes being mapped to specific chromosomes every year. (Extensive reviews are available by Renwick, 1971; Ruddle and Craagan, 1975; McKusick and Ruddle, 1977; McKusick, 1979; Shows et al., 1982; Ruddle, 1984).

The construction of a complete human genetic linkage map is now possible due to the wealth of techniques and material which have become available over the past few years. This

section will discuss the developments and technical breakthroughs which have played an important role in mapping of the human genome.

Historical background

In the early 1900's, Sutton and Boveri theorized, that the "factors" proposed by Mendel, which were later called genes by Johannsen (1909), were carried on chromosomes. Hunt, Morgan and Sturtevant (1911), demonstrated that in *Drosophila* genes were linearly arranged on chromosomes and that the distance between two gene loci on a chromosome could be inferred, according to the frequency of recombination between these traits among offspring of particular parental pairs. The gene causing colour blindness was the first to be localized and it was assigned to the X chromosome by Wilson (1911), (reviewed by Ruddle and McKusick, 1977).

Human linkage was first demonstrated between the genes for haemophilia and colour blindness (Bell and Haldane, 1937) and further X-linkage was shown between the loci for glucose-6-phosphate dehydrogenase and colour blindness in American Negroes (Porter et al., 1962).

The first autosomal linkage was found between the Lutheran blood group and the Secretor trait (Mohr, 1951) but the actual autosome involved was only identified as chromosome 19 by Eiberg et al., in 1983. In 1968, the first

chromosomal assignment of an autosomal linkage group, the ABO blood group/ Nail-Patella syndrome/ adenylate kinase, was shown to be on chromosome 9 (Rapley et al., 1968).

No genes were initially assigned to the human Y chromosome even though Painter had shown the presence of a Y chromosome in all males in 1920 and had proposed (1950) that the Y chromosome carried a testis-determining factor (reviewed by Ruddle and McKusick, 1977). In 1956 the correct number of human chromosomes was shown to be 46, with each person having 22 pairs of autosomes and one pair of sex chromosomes. Females were shown to have two X chromosomes and males to have an X and a Y chromosome (Tjio and Levan, 1956, Ford and Hamerton, 1956).

1.4.3 Classical Genetics

Morgan (1911) suggested, from studies on X-linked mutants in *Drosophila*, that genes are linked as result of being carried on the same chromosome, and that genes which were closer together would therefore be found to be coupled more frequently than distant genes. Sturtevant (1913) demonstrated that, in *Drosophila melanogaster*, it was possible to map the genes on the X chromosome linearly, using the frequency of recombination as an index of the relative distances between the loci on the chromosome (reviewed by Conneally and Rivas, 1980). Loci which are on the same chromosome were termed syntenic by Renwick (1969). Linkage has been defined as, "the occurrence of two loci which are sufficiently close together on a chromosome such

that their assortment is recognized as being non-independent". Linked loci are syntenic but the converse is not necessarily true (Renwick, 1969). The classical method of localization of human genes proceeded slowly by means of linkage analysis in family studies, which was made by careful observation of the segregation of traits within families.

There were two approaches to the localization of genes. Firstly, linkage analysis was used to construct a genetic map of the loci, depending on the rate of occurrence of crossovers between the loci. Secondly, the approach was the actual physical assignment of a gene to a chromosomal locus by somatic cell genetics and molecular hybridization studies (see below).

Linkage Analysis

The recombination fraction (θ) was defined as the proportion of strands that experience an odd number of crossover events between two loci. It is the unit of measurement of map distance expressed as a centimorgan (cM) or a map unit. The relationship between recombination frequency and physical length of DNA, between the loci, is that one centimorgan (cM) is roughly equal to 1 000 kb. The maximum recombination frequency between two loci is 50% , which indicates that they are not linked. For short intervals between loci the frequency of recombination will be directly proportional to the size of the map interval, but as the

distance increases the relationship becomes more non-linear due to multiple crossovers and interference (Renwick, 1971).

The LOD score method is a quantitative expression of the likelihood of linkage. The result gives the likelihood that such a combination would have occurred if the two loci examined were linked at a given recombination fraction (θ). The odds in the LOD is a likelihood ratio:

$$\text{LOD} = \log_{10} \frac{\text{Likelihood of observed pedigree assuming } \theta=0}{\text{Likelihood of observed pedigree assuming } \theta=0,5}$$

LOD scores of different families can be added together. The recombination fraction giving the highest LOD score is taken as the maximum likelihood estimate. A LOD score of +3,0 indicate that the likelihood of linkage is 1000 : 1 and have been taken on recommendation of Morton to mean that "there is very strong evidence of linkage". A LOD score of - 2,0 or less at $\theta=0,2$ is taken as "strong evidence against linkage" (Morton, 1955a). Linkage analysis between two loci, using the classical method of LOD scores, was initially restricted to simple Mendelian traits in nuclear families, however a number of linkage groups were successfully determined.

Genes on the Sex Chromosomes

The observation of two X chromosomes in females and one X chromosome in males allows the genes to be assigned to the X chromosome with relative ease. The first gene assignments

distance increases the relationship becomes more non-linear due to multiple crossovers and interference (Renwick, 1971).

The LOD score method is a quantitative expression of the likelihood of linkage. The result gives the likelihood that such a combination would have occurred if the two loci examined were linked at a given recombination fraction (θ). The odds in the lod is a likelihood ratio:

$$\text{LOD} = \log_{10} \frac{\text{Likelihood of observed pedigree assuming } \theta=0}{\text{Likelihood of observed pedigree assuming } \theta=0,5}$$

LOD scores of different families can be added together. The recombination fraction giving the highest LOD score is taken as the maximum likelihood estimate. A LOD score of +3,0 indicate that the likelihood of linkage is 1000 : 1 and have been taken on recommendation of Morton to mean that "there is very strong evidence of linkage". A LOD score of - 2,0 or less at $\theta=0,2$ is taken as "strong evidence against linkage" (Morton, 1955a). Linkage analysis between two loci, using the classical method of LOD scores, was initially restricted to simple Mendelian traits in nuclear families, however a number of linkage groups were successfully determined.

Genes on the Sex Chromosomes

The observation of two X chromosomes in females and one X chromosome in males allows the genes to be assigned to the X chromosome with relative ease. The first gene assignments

namely colour blindness and haemophilia were both to the X chromosome. The segregation of traits from mother to son and from father to daughter indicated X-linkage. The phenotypes of the male offspring give definitive information on the phase of the chromosome. The genes for maleness were proposed to be on the Y chromosome, when it was observed that the Y chromosome was essential to the development of testis regardless of the number of X chromosomes present as was shown in males having Klinefelters Syndrome (Jacobs and Strong, 1959). The initial identification of the H-Y antigen in the presence of the Y chromosome, due to skin graft rejection between male and female mice, led to the proposal in 1955 that this gene was on the Y chromosome (Eichwald and Silmsker, 1955). This controversy has not yet been resolved.

Genes on Autosomes

The first autosomal assignment of the gene for the Duffy blood group, (Fy), to chromosome 1, was based on the co-segregation of the Fy gene with inherited structural variants of the heterochromatic region on the long arm of chromosome 1 (Donahue *et al.*, 1968). The location of autosomal genes was more difficult by family studies, however some linkage groups were found, as has been previously mentioned and others include the Rh blood group and the elliptocytosis locus (Lawler and Sandler, 1954) and the ABO blood group and Nail-patella syndrome (Renwick and Lawler, 1955). These linkage groups were assigned to chromosome 1 and 9 respectively when chromosome mapping

techniques became defined (Ruddle et al., 1972 and Rapley et al., 1968).

1.4.2 Somatic Cell Genetics

A number of technical discoveries were responsible for the development of somatic cell genetics. Somatic cell hybrids are the result of parasexual events, that is union of two cells that are not gametes, in culture. The nuclei of two different cells were fused forming an interspecific hybrid cell which had double the number of chromosomes (Barski et al., 1961). The potential use of somatic cell hybrids to study cell physiology and function was recognized. Hybrid cells were shown to express the HIA antigens of both parental cell lines (Gershon and Sachs, 1963). In 1965, the first man-mouse hybrids between HeLa cells and mouse tumour cells were formed (Harris and Watkins, 1965; Ephrussi and Weiss, 1965). The observation that there was a selective loss of the human chromosomes in the mouse-human hybrid lines was the landmark in the development of a very powerful tool for chromosome-gene mapping (Weiss and Green, 1967).

The human phenotypes of the hybrid cells could be related to those human chromosomes that were retained within the cell line (reviewed by Renwick, 1971, Ruddle and Creagan 1975, McKusick, 1975). The identification of a specific human gene product, correlated with the presence or absence of a specific human chromosome in the hybrid cell, was initially used to map the gene for thymidine kinase to chromosome number 17 (Miller et al., 1971).

Human protein products and cell surface antigens, that were constitutively expressed in the hybrid cell lines, could be initially mapped because they could be identified (Ruddle and Creagan, 1975). Selection systems were developed to recover hybrid lines containing specific chromosomes efficiently, such as the HAT growth medium (Littlefield, 1964). In man/rodent hybrid lines, the rodent cell lines were selected for mutations of either the hypoxanthine guanine phosphoribosyl transferase (HGPRT) or the thymidine kinase (TK) gene by resistance to certain drugs. After fusion the lines were grown in the HAT medium, containing hypoxanthine, aminopterin and thymine. The aminopterin blocks the de novo pathway of purine biosynthesis and the hybrid cell is dependent on the activity of human HGPRT and TK genes to survive. The surviving hybrid lines will contain at least the human X chromosome and chromosome number 17.

Evolutionary Divergence between man and rodents has resulted in aminoacid changes in homologous enzymes. These enzymes are functionally identical but structurally altered. It is possible to identify the presence of a human gene in a hybrid clone by distinguishing between the human and rodent gene product due to a differences in the electrophoretic mobility of the product (Ruddle, 1973).

The ability to identify specific human chromosomes has also had a major impact on human gene mapping. Quinacrine and giemsa were shown to bind differentially to specific regions of the human chromosomes such that each chromosome displayed a unique pattern of bands and these techniques became known

as Q-banding and G-banding respectively (Cassperson et al., 1970; Cassperson et al., 1971; Sumner et al., 1971).

Further techniques to distinguish between rodent and human chromosomes in culture have also been developed (Bobrow and Cross, 1974).

1.4.3 Molecular Genetics

Molecular genetics has had a major influence on the field of human gene mapping and has been combined with cytogenetic techniques, such as somatic cell genetics and in situ hybridization. This has resulted in the localization of many human DNA sequences to specific human chromosomes. This technology is presently being used to construct a human linkage map to cover the human genome.

Molecular hybridization studies using somatic cell hybrids

Initially, hybridization studies were performed in solution using the human α - and β -globin cDNA clones and somatic cell hybrid clones containing different human chromosomes using Cot analysis. The α - and β -globin genes complexes were localized to chromosomes 16 and 11, respectively (Deisseroth et al., 1977, Deisseroth et al., 1978).

These studies were not entirely satisfactory therefore DNA filter hybridization studies were done using a panel of somatic cell hybrid lines and Southern blot analysis. Southern blots were prepared from a number of somatic hybrid lines each containing a different complement of human

chromosomes and these were used to localize cloned DNA sequences. The DNA sequence was radiolabelled and hybridized to the somatic cell DNA immobilized on the filter. On the basis of elimination, due to presence and absence of individual chromosomes, a sequence could be tentatively assigned to a specific human chromosome (Kamarck et al., 1984).

The mapping of chromosomes has been facilitated by the use of structural chromosomal aberrations such as translocations, deletions and inversions (Ferguson-Smith and Aitken, 1982). Somatic cell hybrid lines were constructed using these structural aberrations and helped in the regional assignment of genes. The Human Genetic Mutant Cell Repository was established in 1972 in Camden, New Jersey. It is an important source of cell cultures for genetic research especially the regional mapping of genes (Greene, 1978). The American Tissue Culture Centre also aims at the collection and distribution of important material for genetic research (Benarde and Nierman, 1985).

Chromosome specific DNA libraries

Human mitotic metaphase chromosomes can be sorted into size groups using a fluorescence particle sorter (Carrano et al., 1979). This has been applied to chromosome sorting of cell lines containing translocations, deletions and rearrangements of genetic material. These chromosomes can then be sorted individually due to an alteration in size and such chromosomes can be used to prepare somatic cell hybrids lines (Krumlauf et al., 1982; Kunkel et al., 1982). The

human G-gamma A-gamma-6- β globin genes were mapped to the short arm of chromosome 11 using somatic cell hybrids, constructed from flow sorted chromosomes which had a translocation between the human chromosome 11 and the X chromosome (Lebo et al., 1979).

Chromosome-specific DNA libraries were constructed using the process of chromosome sorting combined with recombinant DNA technology (Schmeckpeper et al., 1979; Gusella et al., 1980) or using somatic cell hybrid lines containing a single human chromosome. Recombinants containing human sequences were isolated by hybridization to radiolabelled total human genomic DNA and then single-copy human sequences present within the clones were isolated. These libraries have been an important source of chromosome-specific DNA probes. Many chromosome-specific DNA libraries have been constructed and are available from various laboratories (Willard et al., 1985).

In situ Hybridization to Metaphase Chromosomes

In situ hybridization has been described as the marriage between molecular biology and cytology (Hsu, 1979). The hybridization of radiolabelled rRNA in solution to denatured DNA of intact cytological preparations, and the subsequent in situ hybridization of radiolabelled LNA to DNA of cytological preparations, was demonstrated by Pardue and Gall (1969a, 1969b). Radiolabelled mouse satellite DNA was localized to the centromeric heterochromatic regions of mouse chromosomes (Pardue and Gall, 1970). This technique was applied to the localization of human DNA sequences and

initially, only repetitive genes were localized to human metaphase chromosomes. Human repetitive and satellite DNA was mapped to the pericentric regions of chromosomes (Saunders et al., 1972; Jones and Prosser, 1973; Gosden et al., 1975). Those human genes found as mid-repetitive sequences in the genome, such as the genes for 5S ribosomal RNA and the genes for 18S and 28S rRNA were localized to chromosome 1 region q42-43 and to the secondary constriction, not the satellite region, of the short arm of the human acrocentric chromosomes respectively (Steffensen et al., 1974; Evans et al., 1974; Chandler and Yunis 1978). The first human structural genes to be localized using this method were the human histone genes, which are tandemly repeated forty fold in the human genome, and were localized to chromosome 7q (Wilson and Milli, 1977).

Harper and Saunders (1981a) demonstrated that it was also possible to localize single-copy regions of DNA to metaphase chromosomes. This was due to the discovery that Dextran Sulphate accelerates the rate of hybridization, one-hundred fold, of randomly cleaved single-stranded DNA probes to immobilized DNA on filters (Wahl et al., 1979). A random single-copy DNA sequence was the first to be localized using this technique. This technique was extended to the localization of other single-copy genes such as human β -globin, insulin, α globin and placental lactogen (Gerhard et al., 1981; Harper et al., 1981; Barton et al., 1982; Harper et al., 1982). The first random RFLP-detecting DNA probe to be localized was D14S1 (pAW161) to 14q321-322 (Donlon et al., 1983). This *in situ* localization confirmed previous somatic cell hybrid studies

(DeMartinville et al., 1982, Balazs et al., 1983). Many single-copy random RFLP-detecting sequences have now been localized by this method (Human Gene-Mapping workshops 7 and 8, 1983 and 1985).

1.4.4 Collaboration to Map the Human Genome

Collaboration between laboratories has also been set up in order to map the human genome more efficiently. A major collaborative project was initiated by Professor J. Dausett in order to pool all available RFLP data at the Centre d'Etude du Polymorphisme Humain (C.E.P.H.) in Paris in 1984. Cell cultures from forty large families were established and DNA from them has been made available to a number of interested researchers. These families have been typed for a wide range of serogenetic markers and the commitment of each collaborator is to type all informative families with their RFLP-detecting probes. A database combining all this information has been set up and made available to collaborating laboratories. Linkage analysis will be performed between each RFLP detecting DNA sequence and other markers. This project is in principle an extension of the proposal of Botstein et al., (1980) to construct a genetic linkage map of the human genome using all known human polymorphic markers. Furthermore, the use of large multigenerational families for linkage analysis increases the probability of finding linkage between genetic markers (Skolnick et al., 1984).

Eight Human Gene Mapping Workshops have already been held where all interested researchers meet and update all available data on the human gene map. Human gene mapping workshop numbers are given to all arbitrary DNA sequences, of unknown function, once they have been assigned. A human gene mapping database has been set up at Yale University and all information regarding mapping of the human genome is placed into it (Miller et al., 1984, Miller et al., 1985).

1.4.5 Conclusions

The mapping of the human genome is an enormous task. After a slow start, gene mapping is proceeding rapidly as new developments have occurred. The discovery of RFLPs started a new era in the mapping of the human genome, because these DNA sequences could potentially be found throughout the genome. In these cases, the importance of the DNA sequence was its function as a genetic marker on a chromosome. An important achievement in this field is the collaboration between researchers, which eliminates the duplication of work and promotes the sharing of material. This is exemplified best by the Centre for the Study of Human Polymorphism. A complete human linkage map within a few years is now a realistic goal.

1.5 AIMS OF THIS STUDY

The construction of a human genetic linkage map using highly informative DNA markers which are equally spread across the human genome was proposed by Botstein et al (1980). This linkage map would need approximately 150 to 1500 DNA markers with reasonable PIC values. However, before this could be realized, a much larger number of RFLP-detecting sequences needed to be isolated and localized to the human chromosomes. The markers which will ultimately make up the map will be drawn from those that are found to be the most useful in this pool. The main aim of my study was to contribute to the mapping of the human genome by searching for and characterizing a number of RFLPs in the human genome.

The more specific aims of this study were:

1. To isolate single-copy human DNA sequences from a human genomic library.
2. To use these single-copy sequences as DNA probes to search for polymorphic variation among Caucasoid individuals.
3. To show by means of family studies that the RFLPs were inherited in a co-dominant Mendelian fashion.
4. To determine the population frequencies of these RFLPs in Southern African Populations, namely the Bantu-speaking Negroids and the San.

59.

5. To assign these RFLP-detecting DNA sequences to human chromosomes using somatic cell hybrid lines.

CHAPTER TWO

MATERIALS AND METHODS

2 CHAPTER TWO -- MATERIALS AND METHODS

This Chapter discusses the materials and methods which were used in this study. All of the recombinant DNA experiments were carried out in a laboratory classified as P2, according to the guidelines of SAGENE (The South African Committee for Genetic Experimentation). The composition of many of the solutions mentioned in this Chapter are given in appendix B.

2.1 COLLECTION OF MATERIALS

Screening studies

Blood samples were taken from consenting random Caucasoid individuals at the S.A.I.M.R. so that a repeat sample could be obtained if any interesting polymorphisms were found. Fifty ml of blood was collected into ACD or EDTA sample tubes.

Five Caucasoid families having at least four children were obtained. After informed consent, 50 ml of blood was drawn into 1 clotted sample tube and 10 ACD tubes. Serogenetic markers were studied on these samples in order to exclude any families where there was non-paternity.

Population studies

Three population groups were used in this study, namely the Caucasoids, who are recent immigrants to this country, and two indigenous groups namely the Negroids and the San (reviewed by Nurse et al., 1985)

i. The Caucasoid population

English and Afrikaans-speaking white individuals, who lived in Johannesburg, were used for screening studies. The Afrikaaner-speaking population is a more biological distinct group of people compared with the English, and accounts for 60% of the white population in South Africa. A number of the English speaking people had Afrikaaner and British ancestors. These individuals are grouped together as the Caucasoid population.

In addition, a large number of DNA samples were received from members of large three-generational Caucasoid families as part of the C.E.P.H. programme. All the data from the non-related parents of these families was also used for population studies. These individuals were from Utah, in North America and from France. The Venezuelan and Algerian families were both excluded.

ii. The Southern African Negroids

Many different tribal groups of Bantu-speaking Negroids are found in Southern Africa, but due to urbanization inter-tribal relationships have increased and they are now no longer isolated groups of people. In this study individuals from different tribal groups were obtained, including the Nguni-speaking people, namely Zulu, Xhosa and Swazi people, the Sotho-speaking people, namely the Northern Sotho or Pedi, Southern Sotho and Swazi and some individuals from the Venda tribe.

The Zulu belong to the Natal division of the Nguni Population. They are descendants of a number of independent Nguni Chiefdoms and live in an area which is encompassed by Natal and Kwazulu. The Xhosa people are found in the western Cape and Swazi people in the independent country of Swaziland. Both groups of people are similar to the Zulu tribe in terms of language and custom.

The Sotho group of Bantu-speaking people consists of three groups namely the Pedi or Northern Sotho, the Southern Sotho and the Tswana. The Pedi live in the Northern Transvaal regions, the Tswana are from a wide area extending beyond the northern Cape province and Transvaal into the Republic of Botswana and the Southern Sotho inhabit regions in the Southern and Eastern Transvaal.

The Venda are still a distinct group of people found in the Northern Transvaal. They are distinguished by their language and have remained an isolated group mainly due to their geographical location.

Morphology of the Negroids

The African Negroes are of medium height, and are characterized by a brown to dark brown skin colour. They have dark brown to black woolly hair. They are generally of muscular build and have rounded limbs.

Sources of Samples

A number of samples of blood were obtained by informed consent from a group of Zulu school children who lived in Ubombo, Kwazulu in 1983. All other samples were obtained from individuals who lived and worked in the urban areas surrounding Johannesburg.

iii. The San or Bushman Population

These individuals are the best known examples of the Hunter-gatherer peoples of Southern Africa. They are indigenous to the whole of south and south-eastern Africa and have probably occupied the region for many centuries, if not millennia. Today these people have been confined to the Kalahari desert, an area which largely precludes agriculture. During the past decade many of the San individuals have been resident in settlement at Tsumkwe in Namibia. The San have been divided into three linguistic

groups, a Northern, a Central and a Southern group, but the cultural differences among these people are minimal. The majority of San individuals are found in Northern Namibia and consist of !Kung individuals, who belong to the Northern Bushman language group. The San were, until recently, a relatively isolated group of people and therefore it is possible that mutant genes and polymorphic loci could have assumed high frequencies in this population due to random genetic drift.

Morphological features of the San

The San people can be characterized by a small stature with a light yellowish brown skin colour. Their hair consists of spirally tufts (peppercorns). They are usually of slender build, with thin not heavily muscled limbs. They often have slanting eyes with epicanthic folds, lobeless ears and other features of the head and skull.

Collection of samples

Blood samples from random San individuals were obtained during departmental field studies on the population of Tsumkwe in 1981 and 1982. Due to their geographical location, it was not always possible to follow up individuals to obtain repeat samples when interesting results were found.

2.2 ISOLATION OF SINGLE-COPY HUMAN PROBES

2.2.1 Screening a Human Genomic Library

A human genomic DNA library was a gift of Dr. T. Maniatis 1978. The library was constructed by digesting human DNA with HaeIII and AluI, adding synthetic linkers and then cloning these fragments into the EcoRI sites, at positions 19890 and 34990, of the Charon 4A bacteriophage (Lawn et al., 1978; Maniatis et al., 1978). The DNA region between these two EcoRI sites, of Charon 4A, codes for genes which are not essential for the replication and bacterial chromosomal integration of the phage. This region could be removed and replaced with human DNA sequences.

The host cells, EK2 bacterial strain LE392, were prepared by growing an aliquot to an overnight density of 0,2 at OD(600). (EK2 is a biological safety rating for containment, Blattner et al., 1977). The cells were pelleted at 3 700 rpm for 20 minutes and the pellet was resuspended in 0,1 M magnesium sulphate to an OD(600) of 0,15. A 10 μ l aliquot of the library, was diluted using SM buffer to $1,0 \times 10^7$, $1,0 \times 10^8$ and $1,0 \times 10^9$, mixed with the host cells (at a ratio of 1 phage in 100 000 host cells). The phage and LE392 dilutions were mixed with 5 ml of top layer NZCYM (0,6% Bacto-agar), at 55°C and plated onto agar plates prepared with NZCYM containing 1,2% Bacto-agar. When the top layer had set the plates were incubated overnight at 37°C. These results determined which of the phage dilutions

gave non-overlapping plaques. All of the previous days dilutions were then adjusted to this dilution factor, replated out with LE392 and incubated overnight at 37°C.

A large number of plates were seeded using LE392, (100 μ l host cells and 5 ml of top layer per plate) and allowed to dry for 15 minutes. Isolated plaques were picked with sterile toothpicks and placed according to a grid structure and allowed to grow overnight at 37°C.

Transfer of plaques to a membrane (Benton and Davies, 1977)

This method allows for the transfer of plaques in situ to a solid support, nitrocellulose discs (Schleicher and Schuell). Enough phage remains viable on the plates to be reisolated, after the phage which contained single-copy human DNA sequences had been detected. Nitrocellulose discs were placed onto a plate for a minute. While the transfer was proceeding the disc was numbered and corresponding identification marks were made on the filter and the plate. The filter was then placed, with the DNA side upwards, onto filter paper soaked in denaturing solution for a further minute. Then the filter was placed into the beaker of neutralizing solution. Any agar that was attached to the filter was gently rubbed off. This process was repeated until all the plates had been transferred to membranes. The filters were rinsed twice in 2 X SSC, allowed to dry and then baked under vacuum at 80°C, for 2 hours.

Hybridization of filters to radiolabelled genomic DNA

In order to isolate plaques containing single-copy human DNA sequences the filters were hybridized to radiolabelled total human genomic DNA. The filters were incubated in prehybridization solution containing 50% formamide at 42°C for 6-16 hours. The human DNA, labelled by nick translation (section 2.7.1) was denatured by boiling, mixed with hybridization solution which contains prehybridization solution plus 5% Dextran Sulphate and added to the filters. Hybridization was allowed to continue for 16-20 hours. The labelled DNA was removed and the filters were rinsed twice in 2X SSC, then with gentle shaking in 2X SSC for 30 minutes, finally at 55°C for 1 hour in post-hybridization wash (0.1% SSC, 0.1% SDS) with a change of solution after 30 minutes.

The filters were rinsed in 2X SSC, sealed in plastic and exposed to X-ray film (Trimax 3M XD) in cassettes with intensifying screens, at -70°C overnight. The X-ray film was developed in Kodak Developer and rapid fixer. The X-ray film was compared to the bacteriophage agar plates. Those phage which did not hybridize to total human genomic DNA potentially contained single-copy human DNA sequences (Figure 2.1).

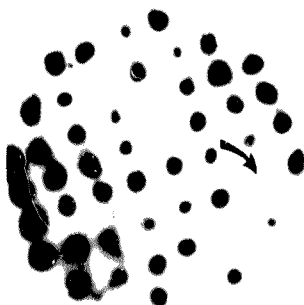


Figure 2.1. An autoradiograph of a nitrocellulose filter containing bacteriophage DNA and hybridized to radiolabelled total human DNA. The dark spots indicate the presence of repetitive human DNA contained in the bacteriophage and the arrow indicates a potential single-copy human DNA sequence contained in the phage, since there is no detectable hybridization to total human DNA.

Those phage were reisolated onto a freshly seeded plate and the screening procedure was repeated. Finally all the phage, which potentially contained single-copy human DNA, were picked and resuspended in eppendorf tubes containing 1 ml of SM buffer and a drop of chloroform. The tubes were sealed with parafilm and stored at 4°C.

2.2.2 Isolation of bacteriophage DNA

Large scale plate lysate isolation of Charon 4A DNA
(modified method of Maniatis et al., 1982)

Preparation of host bacteria

A fresh plate of the host bacteria (LE392) was prepared from a glycerol stock of the strain, stored at -20°C. The plate was incubated at 37°C overnight. An isolated colony was used to inoculate 50 ml of NZCYM in a 250 ml conical flask and allowed to grow overnight, with vigorous shaking, at 37°C to an OD(600) of 0,2. The cells were harvested and resuspended in 0,1 M magnesium sulphate to a final OD(600) of 0,15. These cells were stored at 4°C and could be used for up to 2 weeks.

Infection of LE392

The resuspended plaque was gently mixed and 10 µl was diluted 10X, 100X and 1000X with SM buffer. Each dilution was mixed with 100µl of the host cell preparation and incubated at 37°C for 10 minutes. The mixtures were added

to molten top agar (55°C) and plated onto NZCYM agar plates. Three control plates were prepared in order to check that the soft agar, plating cells and SM buffer were not contaminated. All plates were allowed to set for 15-30 minutes and then incubated overnight at 37°C.

Confluent lysis of host cells

Forty well isolated, evenly sized plaques were picked using a sterile Pasteur-pipette into 3 ml NZCYM broth containing 500 µl of host cells (LE392) and incubated for 30 minutes. Each sample was then mixed with 70 ml of molten top layer and poured onto 5 plates, 150 mm in diameter. When the top agar had set, the plates were incubated at 37°C until confluent lysis occurred (4-6 hours), that is when the top layer containing the bacteriophage clears in contrast to the control LE392 lawn. The plates were placed at 4°C for 30 minutes and then 15 ml of cold SM buffer was added to each plate.

Harvesting of the bacteriophage

The SM buffer was harvested using a 10 ml syringe, into sterile 150 ml bottles containing 2 ml of chloroform, which completed lysis. A 1 µg/ml final concentration of RNase and DNase was added and samples were left for an hour at room temperature. The lysate was then centrifuged at 9 000 rpm for 10 minutes at 4°C to remove the bacterial debris. The supernatant was centrifuged for 2½ hours at 17 000 rpm at

4°C to pellet the phage. The pellet containing the phage particles was dissolved in 2 ml of SM buffer and placed at 4°C overnight.

Cesium Chloride Step gradient to purify Bacteriophage particles

A CsCl step gradient was prepared to densities of 1,3; 1,5; 1,7 containing 9 g ;15 g; 24 g /20 ml of SM buffer respectively. Two ml of each CsCl solution was placed into 14 ml polyallomer centrifuge tubes (Beckmann). The lowest density (1,3) was added first then the 1,5 density was layered below it and finally the 1,7 was placed below both layers. The interface between each density was marked. The phage in SM buffer was gently layered on top of the gradient and the volume was made up using a low density mineral oil. The tubes were centrifuged at 35 000 rpm for two hours using Beckmann SW40Ti rotor, at 20°C.

After centrifugation a blue band was visible in the 1,5 density layer. The band was removed with a 19 gauge needle attached to a syringe and by puncturing the side of the tube below the band. The phage band was dialysed in treated dialysis tubing against 1000 X its volume in 1X dialysis buffer, for few hours at 4°C. (The dialysis tubing had been boiled in 2% sodium bicarbonate and 1 mM EDTA solution for 10 minutes and then in distilled water for 10 minutes. It was stored at 4°C in 50% ethanol and rinsed in distilled water prior to use).

Protein removal and bacteriophage DNA isolation

The phage were removed from the dialysis tubing and extracted twice with an equal volume of phenol (saturated thrice with 0,1 M Tris-HCl pH 8,0) and then three times with an equal volume of chloroform : isoamyl alcohol (24:1). The phage DNA was precipitated using 3 M sodium acetate; 0,5 M magnesium acetate pH 5,5 and 2 volumes of ice-cold ethanol. The samples were well mixed and placed at -20°C overnight or at -70°C for 15 minutes and then centrifuged for 30 minutes at 2 700 rpm. The DNA pellets were washed twice in 70% ethanol. The pellet was allowed to dry and resuspended in an appropriate volume of 1X TE buffer pH 8,0.

2.2.3 Restriction Enzyme analysis of Bacteriophage DNA

Agarose gel electrophoresis

Five μ l of the TE buffer containing the bacteriophage DNA was loaded onto a gel to check that it was of a high molecular weight and not degraded. The concentration of DNA was estimated by comparing it with a known quantity of DNA. Two μ g of the bacteriophage DNA and a sample of Charon 4A DNA was digested with EcoRI according to the instructions of the manufacturers. The samples were separated on an agarose gel by electrophoresis. Those phage DNA samples that contained human inserts could be identified by their restriction pattern. The agarose gel was transferred to nitrocellulose using the method of Southern (Section 2.6).

Hybridization to total genomic DNA

The nitrocellulose filters were hybridized to radiolabelled total human genomic DNA to check that none of the human fragments within the phage contained any repetitive DNA. If the phage hybridized very strongly to human DNA it was necessary to isolate the non-hybridizing fragments and to subclone them. These phage were not used in the initial screening experiments.

2.3 ISOLATION OF HIGH MOLECULAR WEIGHT HUMAN GENOMIC DNA

2.3.1 DNA Extraction

A number of DNA extractions have been used and the most successful, currently applied procedure is reported here (a method described by Sykes, 1983).

The blood samples which had been collected (section 2.1) were centrifuged at 2 700 rpm. The white blood cells (buffy coats), red blood cells and plasma were separated and frozen at -20°C. After thawing, the buffy coat was vigorously mixed with an equal volume of Triton X-100 and 0.9% NaCl and centrifuged at 2 700 rpm for 15 minutes to pellet the white blood cells. A red-brown supernatant was poured off, leaving a red-brown pellet. The pellet was resuspended in 30 ml of the Triton X 100, NaCl solution, mixed thoroughly and centrifuged again. The supernatant was discarded and if the pellet was not yet a white/pale pink colour the Triton X-100 : NaCl step was repeated.

In order to lyse the pellet it was resuspended, using a glass-rod, in a lysing buffer (7 M Urea, 0,01 M EDTA, 0,3 M NaCl 0,01 M Tris-HCl pH 7,5) added in a dropwise fashion until the volume was 10 ml and the pellet completely resuspended. Finally lysis was completed by adding 2 ml of 10% SDS and incubating at 37°C for 10 minutes.

The proteins were removed by first adding 10 ml of saturated phenol and 5 ml of the chloroform: isoamyl alcohol mix to the lysed cells. The samples were thoroughly mixed and centrifuged for 15 minutes at 2 700 rpm. The aqueous phase, which was separated from the organic phase, was collected and the phenol: chloroform: isoamyl alcohol step was repeated. The aqueous phase was then extracted once with 5 ml of chloroform:isoamyl alcohol and after centrifugation it was removed into a clean glass flask. The DNA was precipitated using 2½ volumes of ethanol, then spooled up using a sterile glass-rod, rinsed in 70% ethanol, dried and dissolved in at least 1 ml of TE buffer pH 8,0. The DNA solutions were stored at 4°C or frozen away at -70°C.

2.3.2 Analysis of Human DNA

Five µl of the DNA samples were mixed with one-tenth final volume of Ficoll Dye and the volume made up to 10 µl using 1X TE. They were loaded onto an agarose gel, containing ethidium bromide and electrophoresed in TEA buffer at 50 volts for 30-60 minutes. The sample was then placed on a UV-transilluminator (Chromato-Vue Transilluminator model 0-63) to check that it was high molecular weight DNA.

A 10 μ l aliquot of the sample was mixed with 990 μ l of 1X TE buffer and the absorbance was determined between 230 nm and 300 nm. DNA peaks at an absorbance of approximately 260 nm (One OD(260) is approximately equal to 50 μ g per ml of double stranded DNA). The ratio of absorbance at 260 to 280 is an indication of any protein contamination and should be greater than 1.8.

In order to test that the DNA sample was pure and would digest with a restriction enzyme, approximately 5 μ l of sample was incubated with a restriction endonuclease according to the instructions of the manufacturer and then electrophoresed for 30-60 minutes. The agarose gel was viewed using a UV light source to see that the DNA had been digested to completion and appeared as an even smear on a gel.

2.4 RESTRICTION ENZYME ANALYSIS

Restriction endonucleases are DNases that recognize specific oligonucleotide sequences, make double-stranded cleavages and generate unique, equal molar fragments of the DNA. They are very useful tools to study DNA at the molecular level because of their controllable, predictable and site-specific cleavage of DNA (Fuchs and Blakesley, 1983). Today, many restriction enzymes are commercially available and their optimum assay conditions have been given by their manufacturers. The restriction enzymes are kept in a storage buffer at -20°C .

Ten μg of total human genomic DNA was digested in the prescribed buffer for the restriction endonuclease. These buffers are now supplied with the restriction enzymes or can be prepared and stored at -20°C (Table 2.1). Two to three units of enzyme activity were used per μg of human genomic DNA. (One unit of enzyme activity is defined as the amount of enzyme required to digest to completion 1 μg of bacteriophage lambda DNA in 1 hour at the specified temperature). These restriction reactions were incubated at their specified temperatures (mostly 37°C) for 5 to 12 hours. One μl of 0,1 M spermidine trihydrochloride was added to the reactions. Spermidine trihydrochloride is a polyamine and causes a B to Z conformation change in the DNA (Feuerstein *et al.*, 1986). A final concentration of 0,01% of BSA was also added to stabilize the enzyme.

After the required incubation a one-tenth volume of the reaction was mixed with 5 μl of loading dye (Ficoll) and loaded onto an agarose gel. The sample was electrophoresed to check that the digestion was complete, then the activity of the enzyme was inhibited by adding a one-tenth volume of 0,5 M EDTA or by heating samples to 65°C for 10 minutes. The reactions were stored on ice. This procedure to reduce the reassociation of any sticky DNA ends and helps to remove any bound proteins. A one-tenth final volume of Ficoll dye was added to the samples prior to electrophoresis. Digested DNA samples need not be electrophoresed immediately and can be stored at 4°C in Ficoll Dye.

TABLE 2.1. THE REQUIREMENTS OF THE RESTRICTION ENZYMES USED IN THIS STUDY

ENZYME NAME	BACTERIA FROM WHICH ENZYME ISOLATED	DNA SEQUENCE CLEFT	REACTION BUFFER				
			Tris pH (mM)	KCl (mM)	MgCl ₂ (mM)	NaCl (mM)	DTT μ merc. (mM)
AluI	<i>Arthrobacter luteus</i>	AG ⁺ CT	6 7.6	-	6 50	10	-
AvaII	<i>Anabaena variabilis</i>	G ⁺ G(AT)CC	10 8.0	-	7 60	-	7
BamHI	<i>Bacillus amylotumefaciens</i>	C ⁺ GATCC	6 7.5	20	6	-	6
BglII	<i>Bacillus globigii</i>	A ⁺ GATCT	10 7.4	6 10	-	1	-
EcoRI	<i>Escherichia coli</i>	G ⁺ AATTC	50 7.5	-	10 50	1	-
HindIII	<i>Haemophilus influenzae</i>	GT(Py)(Pu)AC	10 7.6	-	10 50	15	-
HindIII	<i>Haemophilus influenzae</i>	A ⁺ AGCTT	10 7.6	-	10 50	14	-
MboI	<i>Moraxella bovis</i>	⁺ GATC	50 8.0	-	10 50	-	-
MspI	<i>Moraxella species</i>	CC ⁺ GG	50 8.0	-	10	-	1
PstI	<i>Providencia stuartii</i>	TCGGA ⁺ G	20 7.5	10	-	-	-
PvuII	<i>Proteus vulgaris</i>	CAG ⁺ CTG	10 7.5	-	10 50	-	7
RsaI	<i>Rhodospirillum sphaeroides</i>	GT ⁺ AC	10 7.9	-	6	-	0.5
SstI	<i>Streptomyces stanford</i>	GAGTC ⁺ TT	50 8.0	-	10 50	-	-
StuI	<i>Streptomyces tubercidicus</i>	AGG ⁺ CCT	6 7.9	-	6 50	-	7
TaqI	<i>Thermus aquaticus</i> Y11	TC ⁺ GA	6 7.4	-	6 100	-	6
XbaI	<i>Xanthomonas badrii</i>	T ⁺ CTAGA	6 7.4	-	6 100	-	-

Py-pyrimidine; T or C, Pu-purine G or A, T-thymine; G-guanine; A-adenine; and G-guanine.

 β -merc-2- β mercaptoethanol; DTT-dithiothreitol; Tris- Tris HCl

Other requirements: PstI-50 mM ammonium sulphate; StuI-0.07% TritonX-100.

2.5 GEL ELECTROPHORESIS

Agarose gel electrophoresis is a standard method used for separating DNA fragments according to their molecular weight, in the matrix created by the agarose, when a charge is applied across the gel. Negatively-charged DNA migrates towards the positive terminal. The DNA can be viewed directly due to the intercalation of ethidium bromide which causes DNA to fluoresce under UV-light (Sharp *et al.*, 1973). Varying concentrations of agarose are used to efficiently separate different molecular weight sizes of DNA, depending on the type of restriction enzyme used and the average DNA fragment length generated. Those restriction enzymes which have 6 to 8 base pair recognition sequences cut the human genome less often than those which have 4 base pair recognition sites. Higher agarose gel concentrations were used to separate smaller fragments generated by the frequent cutters.

The agarose gel was prepared by boiling agarose (type II high gelling temperature(HGT)) in running buffer (1X TEA buffer). After the gel had cooled to 55°C, ethidium bromide (0,5 to 1,0 µg/ml) was added and the gel poured onto a horizontal gel plate, 18 X 20 cm long, with a set of loading wells placed at one end. The gel solidified in 30-90 minutes and was ready for use. It was submerged in a trough filled with running buffer and attached to a power source.

Samples consisting of 5 to 10 µg of genomic DNA or 1 to 2 µg of bacteriophage DNA were mixed with a running dye (Ficoll Dye) and are loaded into slots and electrophoresed at a

constant voltage. Human genomic DNA is electrophoresed at 25-35 volts for 16-20 hours to separate fragments between 20 and 1 kb efficiently (Southern, 1983). When using Tris-Acetate (TEA) buffer it was necessary to mix of the buffer since the anode reservoir tends to become alkaline and the cathode reservoir acidic. The distance of migration of the smaller fragments (around 500 bp) corresponded to the dye front of the tracking dye, bromophenol blue. After electrophoresis the gel placed on a UV transilluminator and it was photographed using Polaroid type 667 film (Figure 2.2).

lanes 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

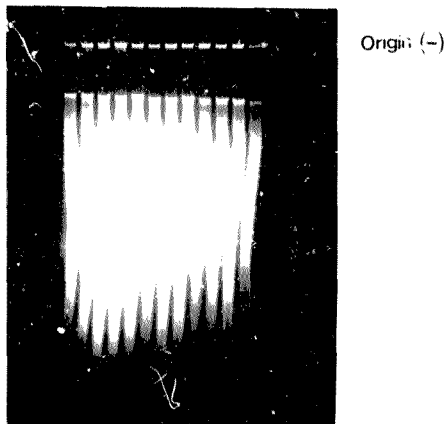


Figure 2.2. An ethidium bromide stained agarose gel containing human DNA, digested with the restriction enzyme *Avall* and separated out by electrophoresis lanes 4-15. Two Lambda bacteriophage molecular weight markers are shown in lanes 1 and 18. The origin where the samples were loaded is indicated.

2.6 SOUTHERN BLOTTING TO SOLID SUPPORTS

In 1975 Southern described a technique enabling one to transfer DNA in situ from an agarose gel onto a solid support, initially nitrocellulose. This preserved the relative positions of the DNA in the gel. The DNA on the filter can be hybridized to, using a radiolabelled probe, which is complementary to a certain sequence of DNA attached to the filter. After autoradiography of this filter a dark band will be visible on an X-ray film where the radiolabelled probe has hybridized to its complementary sequence. Nylon membranes are becoming more popular than nitrocellulose because they are more durable and can withstand repeated stripping and hybridization cycles. Nitrocellulose tends to disintegrate after about three successive hybridization cycles. (The different membrane solutions are given in Appendix B).

2.6.1 The Southern Blotting Procedure

The inefficient transfer of large DNA fragments, those greater than 4 kb, can be overcome since these fragments can be nicked using a UV light source for 10 minutes or by soaking the gel in 0,25 M HCl for 15 minutes. After photography of the gel the DNA in the gel was denatured, using an alkaline solution containing NaOH and NaCl for 30 minutes. The gel was rinsed twice in distilled water and then neutralized using either Tris-HCl buffer containing NaCl pH 7,0-7,5 or a sodium acetate buffer pH 5,5, for 20

minutes. This neutralization step was repeated if the pH of the gel was not yet 7.0. The gel was placed in 10 X SSC for 10 minutes.

The transfer membrane was cut to the exact size of the gel placed into water and then in 10 X SSC. Approximately 300 ml of 10 X SSC was placed into a plastic container. A glass plate was placed across the width of the container in order to support a sheet of Whatman 3MM paper, soaked in 10 X SSC. This formed a stack through which the salt solution could be drawn. The air bubbles must be removed from between each layer using a glass-rod. The gel was placed directly on top of the Whatman filter paper. The treated transfer membrane was placed onto the gel followed by two pieces of Whatman 3MM paper, prewet in 10X SSC. Finally a stack of paper towels, a glass plate and a weight of 1 to 1 kg were placed on top. Plastic was placed around the gel so that the solution only could be drawn directly through the gel into the paper towels. The transfer proceeded for 16-20 hours and paper towels were replaced if they became wet throughout. The filter was removed and baked at 80°C under vacuum for 2 hours between two sheets of filter paper and then stored at room temperature until used.

2.6.2 Modifications for Nylon Membranes

Zetaprobe (Biorad) and Zetabind (AMF Cuno)

The transfer of DNA onto these membranes is similar to nitrocellulose. It has been reported that UV nicking of DNA prior to transfer to nylon membranes can alter the binding ability of the DNA to the membrane and then inhibit the subsequent hybridisation to these membranes (Reed *et al.*, 1985). Acid depurination was recommended for Zetaprobe and Zetabind. These membranes must also be soaked for 10 minutes in 10 X SSC before setting up the transfer. When the transfer was complete the membranes were rinsed twice in 2 X SSC for 15 minutes with gentle agitation. These membranes require baking at 80°C for 3 hours.

Biodyne (Pall)

Biodyne need not be wet before being placed onto the gel and it also does not require any post-transfer washes. We preferred to prewet the membrane prior to transfer. Biodyne must be baked at 80°C for 1 hour, but a vacuum is not necessarily.

2.7 RADIOLABELLING OF DNA PROBES

DNA can be labelled using radio isotopes ^{32}P , ^{35}S , ^{125}I and ^3H . DNA Polymerase I isolated from *E. coli* has three enzymatic activities although the enzyme consists of only a single polypeptide chain. These are 5'-3' Polymerase

activity, 5'-3' exonuclease activity and 3'-5' exonuclease activity. DNA Polymerase I adds nucleotide residues to the 3'-hydroxyl terminus of a nicked double-stranded DNA molecule, and also removes the nucleotide sequences 5' to the nick. This sequential removal and addition of nucleotides is the basis of the nick translation reaction (Rigby *et al.*, 1977). Some of the nucleotides which are removed are replaced with radiolabelled nucleotides, resulting in a DNA molecule which has radioactive residues. The Klenow Fragment of the Polymerase I enzyme has been cloned and has 5' to 3' polymerase activity and no endonuclease activity. This enzyme is used to label DNA fragments to high specific activities using the primer extension technique (Feinberg and Vogelstein, 1983).

2.7.1 Nick Translation

A nick translation reaction consists of 1/2 to 1 μ g of DNA, usually a vector such as a plasmid or bacteriophage containing a human DNA insert, nick translation buffer containing 100 mM dNTPs, magnesium chloride ions and Tris-HCl pH 7.5, ^{32} P radiolabelled nucleotides, a low concentration of DNase (1.0×10^{-5}) and 5 units of DNA polymerase I. The reaction was incubated at 15°C for an hour. The DNase is important since it introduces single-stranded nicks into the DNA. Nick translation kits are now commercially available.

2.7.2 Monitoring the Labelling Reaction

This method gives a rough estimate of how the reaction is proceeding. A 1 μ l aliquot of the reaction was placed onto a glass microfibre filter (Whatman). The total number of counts per minute on the filter were estimated using a Geiger Counter (Weil). The filter was placed onto a vacuum filter and was washed with approximately 15 ml of 10% trichloroacetic acid (TCA) containing 0,2 M PPi, followed by three washes with 15 ml of 5% TCA. The non-incorporated nucleotides were washed through the filter and the DNA, containing the incorporated nucleotides, remains on the filter. The activity of the filter was again counted. The percentage incorporation was calculated as follows:

$$\frac{\text{Acid precipitable counts}}{\text{Total counts}} \times 100$$

When the reaction had reached the required incorporation of 10-20% for DNA fragments or greater than 50% for plasmids and bacteriophage, the reaction was stopped by adding a one-tenth volume of 0,5 M EDTA and it was placed on ice.

Spermidine trihydrochloride was found to increase the incorporation of nucleotides. One μ l of a 0,01 M solution was added to the reaction when the incorporation was low. The mechanism is not known however spermidine has been reported to stimulate the incorporation of gamma-³²P ATP in end-labelling reactions by inhibiting nucleases which may be

present in some preparations of polynucleotide kinase (Maniatis et al., 1982).

2.7.3 Removal of the non-incorporated nucleotides

The reaction volume was made up to 200 μ l using TE buffer containing 100 μ g/ml of denatured sonicated salmon sperm DNA. An equal volume of saturated phenol was added, mixed well and centrifuged to separate the two phases. The aqueous phase was collected. The phenol-phase was again extracted using 200 μ l of TE buffer containing salmon sperm DNA and the aqueous phases were pooled. Sephadex G-50 was swollen in 1X TE buffer pH 8.0 for at least 3 hours prior to the preparation of the column. A 10 ml glass column was rinsed thoroughly in distilled water then in TE buffer. The swollen Sephadex G-50 was poured into the column avoiding air bubbles, and allowed to settle.

A small amount of orange G dye was added to the prepared probe which was then loaded onto the top of the column and TE buffer was used to elute the DNA, containing incorporated nucleotides. The non-incorporated nucleotides move slightly ahead of the orange G dye and the larger DNA molecules move first through the column. Three drop fractions were collected and monitored using a Geiger counter. The first peak, representing the labelled DNA, was pooled and placed on ice. A 10 μ l aliquot was used to determine the total specific activity of the DNA using a Liquid Scintillation Spectrometer (Packard Tri-carb). Specific activities of $5,0 \times 10^7$ - $1,0 \times 10^8$ per μ g DNA were routinely obtained and were

sufficient for hybridization. The labelled DNA fraction was boiled for 10 minutes and stored at -20°C prior to hybridization.

2.7.4 Oligolabelling of DNA fragments

Preparation of DNA fragments

The DNA fragment to be used as a probe was first isolated by digestion of the vector with the appropriate restriction enzyme and by gel electrophoresis in a purer low gelling temperature (LGT) agarose. After electrophoresis the gel was soaked in water containing ethidium bromide ($0.2\mu\text{g/ml}$) for 30 minutes. This removed the TEA buffer which could inhibit further enzyme reactions with the DNA. The correct DNA fragment was excised, removing as much of the agarose as possible, and then placed into a preweighed eppendorf tube. Three ml of water was added per gram of gel. The tube was placed into a boiling water bath for 7 minutes, vortexed and then briefly placed onto ice before being stored at -20°C . This step denatured the DNA fragment and dissolved the gel completely.

The oligolabelling reaction

The DNA fragment was thawed, boiled for 3 minutes and then placed at 37°C for up to 1 hour prior to the labelling reaction. The reaction was prepared using DNA ($10-64\text{ ng}$), oligolabelling buffer, $2\text{ }\mu\text{l}$ of 10 mg/ml gelatin or BSA, two units of Klenow Fragment, $50\text{ }\mu\text{Ci}$ of the radioactive

nucleotide and water to make the volume up to 50 μ l. Oligolabelling buffer was prepared from solution A, B and C in the ratio of 100:250:150. Solution A contained 1,25 M Tris-HCl pH 8,0, 0,125 M magnesium chloride, 250 mM β -mercaptoethanol and 0,5 M dNTPs. Solution B consisted of 2 M Hepes pH 6,6. Solution C contained the random hexanucleotides, p(dN)6, dissolved to 90 OD units per ml of water. The reaction was allowed to proceed at room temperature overnight. The percentage incorporation was determined as described in section 2.7.2. and found to be between 15 and 70%. The non-incorporated nucleotides were removed using a G-50 sephadex column as described above (section 2.7.3). It was not necessary to remove the incorporated nucleotides, the reaction was stopped and phenol extracted as described, and the aqueous phase was boiled and ready to use as a probe.

2.8 DNA HYBRIDIZATION

Prehybridization

Some of the membranes used required treatment prior to prehybridization. Zetaprobe (Biorad) and Zetabind (AMF Cuno) both required a 65°C wash in 0,5% SDS; 0,1X SSC; 0,1 M Tris-HCl pH 7,5 for at least an hour. During the prehybridization step the pores on the membrane were blocked using a non-homologous DNA (usually salmon sperm DNA) which was contained in the prehybridization solution and helps to avoid non-specific background. The prehybridization solution containing 50% deionised formamide, was preheated

to 42°C. The filter was prewet in water and then placed inside a plastic bag. Approximately 10 ml of prehybridization solution was added for a 18 X 19 cm membrane and up to five 18 X 19 cm blots could be placed into 1 plastic bag. All the air bubbles were removed, the bag was heat-sealed and placed into a 42°C incubator, with gentle agitation for 4-16 hours.

Hybridization

The denatured DNA probe was mixed with the hybridization solution (prehybridization solution containing 5% Dextran Sulphate). The prehybridization solution was removed and the hybridization solution was added to the membranes. Any air bubbles were removed and the bag was heat-sealed. Hybridization was allowed to continue at 42°C for 16-20 hours, with gentle agitation. The hybridization solution was removed and this could be reused on other pretreated DNA blots if the specific activity was still high (and the probe was not degraded due to the radioactivity). Before reusing the probe, it must be heated at 70°C for 10-30 minutes.

Post-hybridization washing

This was an important step and varying conditions were used depending on the type of membrane and probe. The membranes were placed into a container and rinsed 3 times with 200 ml of 2X SSC, then placed on a shaker in 2X SSC for 30 minutes, at room temperature. The membranes were washed in 0.1X SSC; 0.1% SDS at 42°C, for 30 minutes and then at 55-65°C for a further 1 hour. The decrease in radioactivity on the

membrane was monitored using a geiger counter until background activity was reached. The membranes were finally rinsed in 0,1X SSC and then sealed in plastic bags.

2.9 AUTORADIOGRAPHY

The membranes were exposed to Kodak X⁷PS or Trimax 3M XD film in cassettes with tungsten/phosphate calcium intensifying screens (Du Pont), at -70°C for 1 to 7 days (depending on the specific activity obtained for the probe). The low temperature enhances the signal emitted and therefore the detection of the probe. The film was developed in Kodak GBX developer in a dark room for 3-4 minutes, rinsed in running water for 30 seconds and then fixed in Kodak Rapid Fixer for a further 3-5 minutes. If the background was high the filters were rewashed at 65-70°C in 0,05% SDS; 0,05 X SSC for a further hour and re-exposed to X-ray film.

2.10 STRIPPING AND REUSE OF MEMBRANES

Membranes could be reused and the successful stripping and hybridization depended on the type of membrane used. The basic principle involved using a high alkaline solution, containing NaOH. This caused the DNA to become single stranded and therefore the probe becomes dissociated from the membrane. The blots are then treated with a neutralizing solution to return the membranes to pH of 7,0. Finally the

membranes are rinsed in a 2X SSC solution. The membranes can be stored in 2X SSC at room temperature or at 4°C prior to reuse.

It has been well documented that nitrocellulose is in general much less hardy and the membrane usually disintegrates after three hybridization cycles. Nylon membranes such as Zetaprobe (Biorad), Zetapore (AMF Cuno) and Biodyne (Pall) are able to withstand this treatment and are far less brittle. Ten successful hybridization cycles with Zetaprobe and thirteen with Biodyne have been achieved. Biodyne was usually stripped using 50% formamide and sodium phosphate pH 6.4, however it has successfully been stripped using the protocol for Zetaprobe. Zetaprobe and Zetabind require 30 minutes at 42°C in 0.4 M NaOH followed by 30 minutes at 42°C in Tris-HCl pH 7.5, 0.5% SDS, 0.1X SSC. The potential to strip and reuse these membranes was a major time-saver in this study.

2.11 DETERMINATION OF DNA FRAGMENT SIZES

Preparation of molecular weight markers

Molecular weight markers were prepared by digesting bacteriophage lambda DNA using restriction enzymes EcoRI and HindIII in the appropriate buffers. A double digest using both enzymes was also prepared, where the enzyme which required the lowest salt buffer was first used. The reactions were analysed by gel electrophoresis to ascertain that they were complete. The digested DNA was then diluted

a 1000 times, mixed with denatured, sonicated salmon sperm DNA (the final concentration of DNA in the molecular weight marker was 10 μ g), Ficoll loading dye and TE buffer to make up to 1 ml. The exact size of these fragments was available from sequencing data and therefore they can effectively be used as molecular weight standards (Appendix C).

Graphical Analysis

The distance that the molecular weight standard had migrated from the origin was measured. The known molecular weight of the DNA fragments was plotted against the distance that each fragment had migrated, from the origin, using semi-logarithmic paper. The size of the unknown fragment was determined by measuring the distance that it had migrated from the origin and then reading off the appropriate molecular weight size from the graph.

2.1.1 LOCALIZATION USING SOMATIC CELL HYBRID LINES

Somatic Cell Hybrids were constructed and also maintained by the Cytogenetics Unit, S.A.I.M.R. as part of an on going research project of this department (Dos Santos, 1986). Other cell lines were gifts from Dr. U. Francke, Dr. D. Retief and Dr. T. Mohandas. These somatic cell lines were cultured and analysed for any chromosome changes prior to use. DNA was extracted from these lines according to the following two methods

The first method was obtained from Dr. C. Boehm, The John Hopkins Hospital, Baltimore.

The cells were harvested from the culture flasks using a rubber scraper and rinsed twice in saline (0,9% NaCl). The cells were then pelleted and frozen at -70°C . A pellet of cells was thawed and resuspended in amniocyte lysis buffer. A small amount of powdered Proteinase K was added and the samples were incubated at 55°C for 4 hours. The cells were extracted three times with equal volumes of phenol: chloroform: isoamyl alcohol (25:24:1) and then once only with chloroform: isoamyl alcohol (24:1). The DNA was precipitated using 3 M sodium acetate, 0,1 M magnesium acetate pH 5,5 and twice the volume of 95% ethanol, at -70°C for 30 minutes. The samples were then centrifuged at 2 700 rpm for 20 minutes and the DNA pellet was rinsed twice with 70% ethanol, dried and dissolved in an appropriate volume of TE buffer, pH 8,0. The DNA was placed on a gel and the concentrations were estimated.

The second method was obtained from the Cytogenetics Unit, Tygerberg Hospital, Stellenbosch.

This method does not require the collection of the cells instead lysis of the cells occurs directly in the flask. Ten confluent flasks of cells per hybrid line were obtained. These cells were washed with 10 ml of PVS, a saline phosphate buffer.

A lysis buffer containing 10 mM Tris-HCl, 0.1% SDS, 1 mM EDTA was prepared. Proteinase K [final concentration of 50 µg/ml] was added to 15 ml of lysis buffer per line extracted. Five ml of lysis buffer was added to 3 out of 10 flasks. These flasks were incubated at 45°C for 5-10 minutes. Lysis of the cells occurred and the lysis buffer became thick and very sticky. The 5 ml of lysate from each flask was then poured into a further 3 flasks. The incubation was repeated and the lysate buffer was again poured into a further 3 flasks. Finally the 15 ml of lysate containing all the lysed cells from the first 9 flasks was poured into the tenth flask.

After the cells in this flask had been lysed the cell lysate was placed at 68°C for 30 minutes, then at 45°C for 90 minutes. This lysate was extracted 5 times with equal volumes of chloroform: isoamyl alcohol (24:1) with 10 minute centrifugations at 2700 rpm, between each step. Finally, the cell DNA was precipitated by adding 5 M NaCl to a final concentration of 0.3 M and twice the volume of 95% ethanol. The DNA was gently spooled up using a sterile glass-rod, rinsed in 70% ethanol and finally redissolved in 1X TE buffer.

The yield from the second method was far superior to that from the first. It overcame degrading problems experienced previously, which could be due to the way in which the cells were collected. It also cut out the time-consuming step of cell collection.

Determination of chromosomal location of probes

A panel of somatic cell DNA was set up such that every chromosome was uniquely represented with a combination of hybrid lines. Twenty μ g of DNA of each line was digested with an appropriate restriction enzyme. Three controls, DNA from mouse, Chinese hamster and usually a human male individual were also prepared. The samples were electrophoresed in a 1% agarose gel and transferred to a nylon membrane, Biodyne (Pall). In order to localize a probe to a specific chromosome, it was radiolabelled by nick translation and hybridized to the membrane as previously discussed in sections 2.7, 2.8 and 2.9.

From the pattern of hybridization to different somatic cell lines the probe was tentatively assigned to that chromosome which was only present in all the lines that showed positive hybridization. The controls were important to show whether there was any cross-homology of the probe to the mouse or Chinese hamster DNA, which would also explain any anonymous bands that were sometimes present on the blot.

2.13 LINKAGE STUDIES

The department joined the CEPH collaboration project in July 1986. Samples of DNA from parents and a few families were initially obtained. The strategy to type these families was firstly to screen all the parents of the families in order to determine which families were informative for the particular RFLPs. (An informative family for linkage

studies is when at least one parent is a heterozygote). Secondly, family Southern blots were prepared, the probes were hybridized to these blots and the genotypes of all the individuals in the family were determined. Finally this data was entered into a computer database for linkage studies using the programme LINKAGE to all other DNA and protein markers which have also been placed into the data base.

These blots were all prepared using nylon membranes so that all RFLPs detected with the same enzyme could be reused. The blots were stored in 2X SSC between hybridizations. DNA from the study which was not available will to be typed as soon as it is obtained, in order to fulfil the commitment of the Department to the C.E.P.H. programme.

2.14 RESTRICTION MAPPING

Restriction mapping is the ordering of restriction sites along a sequence of DNA. It involves the digestion of the DNA fragment of interest with a series of enzymes, calculating the sizes of each fragment and then linearly arranging the fragments in their order along the DNA sequence. Double digestions which involved the digestion of the fragment with two or more enzymes are also needed. These were done in increasing order of the salt concentration requirements of the restriction enzymes.

A number of approaches to restriction mapping are possible

and the following approach was adopted :

- i. The single and double digestions of the bacteriophage clone with the required enzymes were done. The DNA fragments were then separated using electrophoresis in agarose gels with molecular weight standards. Charon 4A DNA without human DNA inserts was also prepared with the same enzymes and this helped to determine which fragments in the clone were of phage origin. The gel was photographed and then transferred to nylon membranes as discussed (section 2.6).
- ii. The human genomic fragment sizes were obtained from Southern blots of random individuals which had been hybridized to this DNA sequence.
- iii. A Southern blot was prepared using the DNA of an individual digested with many enzymes.

These blots were then hybridized to radiolabelled DNA from different sources.

- i. The clone containing the human DNA sequence to be restriction mapped was hybridized to blots ii. and iii.
- ii. Total human genomic DNA was hybridized to blot i, to determine whether any of the human fragments contained repetitive DNA sequences.

- iii. Charon 4A DNA was hybridized to blot i. to determine which human fragments were the flanking sequences of the clone and hence were also attached to phage arms when restricted with all enzymes besides EcoRI.
- iv. Different human EcoRI fragments of the clone to be mapped were separated by digestion with EcoRI, electrophoresed in LGT agarose and excised from the gel. These fragments could be labelled using primer-extension techniques and were hybridized to all the blots described above.

With all of the above results, it was possible to determine each different DNA fragment length and the linear order of the fragments along the sequence of DNA.

2.15 SUBCLONING OF BACTERIOPHAGE ISOLATES DETECTING RFLPS

Subcloning of DNA fragments, from a 15-20 kb human DNA insert of a bacteriophage Lambda, is a way to eliminate repetitive sequences and to reduce the size of the probe. This also allows for higher specific labelling and clearer hybridization results. The DNA sequences that were subcloned have all been placed into the C.E.P.H. programme and will therefore be made available to other collaborators. A suitable vector is necessary for the distribution of any RFLP-detecting sequences. For subcloning criteria for suitable vectors include selectable markers to determine which clones contain inserts, the maximum size of the insert that the vector can take up and

the size of the vector itself. Furthermore the newly, constructed recombinant must be stable and easily isolated. The vectors chosen were pUC18 and pUC19, which had been developed at the University of California (Vieira and Messing, 1982).

These plasmids, in host strain HB101, were received from Dr. D. Rawlings, University of Capetown and are also commercially available (B.R.L. LTU). They contain the PvuII/EcoRI fragment of pBR322 which carries the ampicillin resistance gene, β -Lactamase, and the origin of replication. The multiple cloning site of the M13mp sequencing vectors and part of the Lac-Z (β -Galactosidase) gene have been added. DNA fragments may be cloned into the unique restriction sites in the multiple cloning region. Selection for recombinants is relatively easy in suitable host strains. Insertional inactivation of the Lac-Z gene results in white colonies when plated on 5-bromo-4-chloro-3-indolyl- β -galactoside (X-Gal) plates in the presence of isopropyl- β -D- thiogalactopyranoside (IPTG). Only those host cells which contain a plasmid expressing the β -Lactamase gene will survive on ampicillin. Non-recombinants form blue colonies when they are transformed into a suitable host strain, e.g. JM103.

2.15.1 Preparation of the vector DNA

In order to isolate plasmid DNA, the size and the configuration differences between the DNA of plasmids and E.coli chromosome are used. Plasmid DNA is much smaller and

remains in covalently closed circular form, while the E.coli chromosomal DNA is broken down into linear molecules during the extraction procedure. Most of this linear DNA is pelleted with the cell remnants.

Plasmid DNA extraction

(SDS Lysis Method, Maniatis et al., 1982)

An isolated colony containing the plasmid was taken from a LA ampicillin [50 µg/ml] plate and sequentially amplified in 10 ml, 25 ml and finally 500 ml of Luria Broth, each containing 50 µg/ml of ampicillin at 37°C with vigorous shaking. Each culture was allowed to grow until the OD(600) was close to 0.6, representing late log phase of bacterial growth and the 500 ml culture was grown for 2½ hours. chloramphenicol [170 µg/ml], an antibiotic which inhibits bacterial protein synthesis by binding to the 50S subunit of the ribosome, was added. Bacterial DNA synthesis is inhibited but the independently replicating plasmid continues to amplify during overnight incubation at 37°C.

The bacterial cultures were harvested by centrifugation at 4 000 rpm for 10 minutes at 4°C. The supernatant was removed and autoclaved. Ice-cold STE was used to resuspend the bacterial pellet and the centrifugation step was repeated. The pellet was resuspended in 10% sucrose, 50 mM Tris-HCl, pH 8.0. Fresh lysozyme [2 mg/ml] in 0.25 M Tris-HCl, pH 8.0 was added with 0.25 M EDTA, mixed well then placed on ice for 10 minutes. Six ml of 10% SDS solution was added and gently dispersed using a sterile glass-rod to

avoid shearing the DNA as the bacterial cell walls broke down. Eight ml of 1.0 M NaCl was added and mixed evenly into the sample which was then left on ice for an hour. The high molecular weight bacterial chromosomal DNA and the bacterial debris were removed by centrifuging the sample at 17 000 rpm for 1 hour at 4°C. The plasmid DNA and low molecular weight DNA remained suspended in the supernatant and the rest formed a pellet which was discarded.

In order to remove the proteins the supernatant was extracted twice using an equal volume of phenol: chloroform: isoamyl alcohol mix (25:24:1), and then once using only chloroform: isoamyl alcohol (24:1). The DNA was precipitated using twice the volume of ethanol and placed at -70°C for 15 minutes then centrifuged for 15 minutes at 3 000 rpm at 4°C. The DNA pellet was washed twice in 70% ethanol and then dissolved in a small volume of TE buffer, pH 8.0.

The final step was the purification of closed circular plasmid DNA by centrifugation in CsCl. A CsCl gradient with an intercalating dye, ethidium bromide, was used in order to separate plasmid and bacterial DNA. Ethidium bromide binds to covalently closed circular plasmid DNA and linear DNA and after ultracentrifugation in CsCl the plasmid DNA bands at a higher density (Radloff et al., 1967). For every 1 ml of TE buffer, 1 gram of solid CsCl and 80 µg of ethidium bromide were added. Purple fuzzy aggregates representing bacterial protein-ethidium bromide complexes were observed.

The solution was kept covered and placed into Beckmann polycarbonate tubes and centrifuged in a Ti65 fixed angle rotor, at 45 000 rpm for 60 hours at 20°C.

The tubes were observed using a UV light and two bands were visible. The upper band represented nicked linearized plasmid DNA and linear bacterial DNA while the lower band closed circular plasmid DNA. The upper layer was removed and the lower band was collected with a 2 ml syringe.

An equal volume of isoamyl alcohol was added to the plasmid to remove the ethidium bromide and it was gently inverted several times. The ethidium bromide was absorbed by the isoamyl alcohol and remains in the top layer which was removed. This step was repeated until the pink colour in the lower phase had completely disappeared. In order to remove the CsCl the lower phase, containing the plasmid DNA was dialyzed in pretreated dialysis tubing (section 2.2.2.) at 4°C against TE buffer, pH 8.0 for several days. The plasmid DNA was removed from the tubing and stored at 4°C. The plasmid DNA concentration was determined as described (section 2.3.2).

This plasmid to be used as a vector was digested using a suitable restriction enzyme, (*EcoRI* for this study). A small aliquot was loaded onto an agarose gel to check that digestion was complete. The digested vector was prepared for ligation by extraction twice with equal volumes of phenol to remove the enzyme and other proteins, extracted once with chloroform: isoamyl alcohol and finally twice with water saturated ether. The ether was removed by incubation

in a 68°C water bath for 10-30 minutes. The DNA was precipitated using 3 M sodium acetate 0,1 M magnesium acetate, pH 5,5 and 95% ethanol, washed twice with 70% ethanol then resuspended in a small amount of 1X TE buffer. The DNA concentration was estimated on a gel.

2.15.2 Preparation of the human DNA fragments

The human EcoRI DNA fragment which detected the RFLP had been identified by oligolabelling and hybridization procedures (section 2.7.4) to the correct restriction digests of human DNA. Subcloning of the correct fragment eliminated the continuous steps involved in the preparation of fragments for oligolabelling from the bacteriophage clone.

Thirty µg of the bacteriophage clone containing the human DNA sequence was digested for 2-3 hours with EcoRI. A small aliquot was tested for complete digestion on a trial gel. The rest of the sample was then loaded onto a 1% agarose gel (low gelling temperature), containing 0,5-1 µg/ml of ethidium bromide. The samples were separated overnight at 25 volts with visible lambda molecular weight markers. The correct human EcoRI fragment was excised using a handheld UV light and placed into pretreated dialysis tubing with a small amount of 0,05 X TEA buffer. The fragment was eluted into the dialysis bag overnight at 10 mA in 0,05 X TEA.

The fragment was then collected and purified using an Elutip-d column (Schleicher and Schuell). The column was pretreated with a high salt buffer (1 M NaCl, 20 mM Tris-HCl pH 7,3-7,5, 1 mM EDTA) then washed through with low salt buffer (0,2 M NaCl, 20 mM Tris-HCl pH 7,3-7,5, 1 mM EDTA). The fragment was then loaded into the column and eluted with the high salt buffer, precipitated with 95% ethanol and washed twice with 70% ethanol. After the ethanol was completely removed and the pellet was resuspended in water or TE buffer, pH 8,0.

2.15.3 Ligation and transformation

The Ligation Reaction

The ligation reaction in this study was a recombination reaction for limited polymerization with circle closure. It was recommended that 5 pM/ml of DNA was used for such ligations. [One pM = (MwtDNA) μ g or (0,66 x size kb) μ g] (method of Rawlings, University of Capetown). The DNA ligation reaction consisted of appropriate concentrations of vector and insert DNA, 1 mM ATP, a ligation buffer (10 X contains 0,66 M Tris-HCl, pH 7,6, 0,006 M magnesium chloride and 0,1 M DTT) and T4 DNA ligase (Amersham). The ligation reaction was set up containing the correct concentrations of DNA, ligation buffer, ATP and 0,25 units of T4 DNA ligase for "sticky-ended" ligations. The final volume of the reaction was between 20 and 50 μ l. The reaction was incubated at 15°C for 18-24 hours.

Preparation of Competent Cells

(modified method of Maniatis *et al.*, (1982) and Dr. D. Rawlings pers. communication)

The suitable host, JM103, [Δ (*lacpro*), *hsdR*⁻, *supE*, *shcB*, *endA*, *thi*, *strA*, Lambda-/ P'traD36, *prqAB*, *lacI*, Z ϕ M15] was grown overnight in 10 ml of Luria broth. One hundred μ l of the overnight culture was placed into 100 ml of Luria Broth and incubated with vigorous shaking for 90-120 minutes until the OD(600) was between 2 and 3. The cells must still be in the logarithmic phase of growth. The cells were harvested by centrifugation at 2 700 rpm at 4°C and resuspended in 50 ml 50 mM Calcium Chloride, 10 mM Tris-HCl pH 8.0, left on ice for 30 minutes before being centrifuged at 2 700 rpm for 10 minutes to pellet the cells. The cells were finally resuspended in 1.5 ml of the calcium chloride solution and left on ice for at least 4 hours. This step increases the frequency of transformation four to sixfold.

Transformation of the Host Cells

The DNA to be transformed must not exceed 100 ng. The concentrations of the DNA in ligations were estimated and the volume containing approximately 50 ng was mixed with 1 X TE buffer pH 8.0 to a final volume of 100 μ l. One hundred μ l of the competent cells was added to the DNA ligation mix. The reaction was left on ice for 30 minutes, heat shocked in a 42°C waterbath for 2 minutes, then placed on ice again for 5 minutes. One ml of Luria Broth was added to

the reactions and they were then incubated at 37°C for 1 hour. This allowed for the expression of the β -Lactamase gene.

To analyze the recombinants plates were prepared using LA (1.2%) cooled to 55°C and ampicillin [final concentration 100 μ g/ml] was added. X-gal was prepared by dissolving 25 mg in 1.25 ml of dimethylformamide. The final concentration of X-gal is 50 mg/l Luria Agar. IPTG was dissolved in water and 500 μ l was added to 1.0 l of Luria Agar [final concentration 100 mM]. The plates were poured, allowed to set at room temperature for at least an hour and then dried at 37°C for 10-15 minutes. The transformed cells were gently mixed and different volumes were plated out. A flat sterile glass-rod was used to spread the colonies. When the liquid was absorbed into the agar the plates were incubated at 37°C for 12-18 hours.

2.15.4 Recombinant Analysis

The plates were examined for the presence of white and blue colonies (figure 2.3). The white colonies indicated successful recombinants which had lost β -galactosidase activity. Control plates containing the vector ligated to itself should have no white colonies. If present it could indicate that there was a mutation in the Lac-Z gene during ligation of the plasmid or else there was conferrence of ampicillin resistance to the host cells, JM103 from the

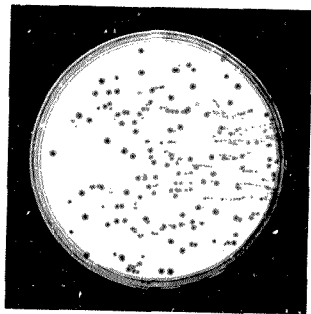


Figure 2.3. An agar plate of JM103 bacterial colonies after transformation with plasmid pUC19. The blue colonies contain non-recombinant plasmids and the few white colonies contain recombinant plasmids. The very small white colonies represent non-transformed JM103 bacteria, which have started to grow on the plate due to conference of resistance from the ampicillin resistant colonies.

ampicillin resistant surrounding cells. Any potentially positive recombinants were again plated out on Amp/X-gal/IPTG plates with overnight incubation.

A one-day plasmid preparation was used to analyze positive clones using a modified method of Promega Biotec. One isolated colony was used to inoculate 100 ml Luria broth containing 100 µg/ml of ampicillin. The cultures were incubated overnight with vigorous shaking and harvested by centrifugation at 2 700 rpm for 30 minutes. The cells were lysed by resuspension in 5 ml of lysozyme solution (25 mM Tris-HCl pH 7.5, 15% sucrose, 10 mM EDTA and 2 mg/ml of lysozyme) and placed in ice water for 20 minutes.

Twelve ml of 1% SDS, 0.4 N NaOH was added and the reaction was gently inverted and placed on ice for 10 minutes. Finally 3 M sodium acetate was added, to precipitate out the proteins and chromosomal DNA, evenly mixed by inversion and placed on ice for a further 20 minutes. The bacterial debris and chromosomal DNA were precipitated by centrifugation at 15 000 rpm for 10 minutes. The supernatant was collected and RNase was added to a final concentration of 1 µg/ml and incubated at 37°C for 30 minutes. The supernatant was extracted twice with an equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) and once only with chloroform: isoamyl alcohol (24:1). The plasmid DNA was precipitated using twice the volume of ethanol and then washed twice with 70% ethanol. The pellet was dried and resuspended in 1X TE buffer.

In order to release the insert the plasmid DNA was digested using the enzyme, EcoRI, which had been used to subclone the fragment. The reaction was electrophoresed and the fragment sizes were determined. Multiple transfers of the gel to nylon membranes were performed. This involved setting up of a normal Southern transfer but changing the membrane increasing the length of transfer for each successive membrane. These were hybridized to the DNA fragments from the original recombinant phage. Finally the new recombinants were radiolabelled by nick-translation and hybridized to a Southern blot containing human DNA digested with the enzyme that detected the RFLP.

Those recombinants which satisfied all of the above criteria were then replated and grown overnight in 5 ml of Luria Broth. One ml of the culture was mixed with glycerol [final concentration of 35%] and stored at -70°C.

2.16 EVALUATION OF METHODS

The methods used in this study were mostly standard techniques discussed at length by Maniatis et al., (1982) and those which have given the most reliable results have been reported. A number of modifications to the basic techniques were made when it was deemed necessary.

A number human genomic DNA extraction techniques have been used and the technique which gave the highest yield is reported here. This technique was advantageous in that it took only 3½ hours to do. It has also been found that an

extra phenol extraction in the protein removal stage is helpful in yielding DNA of better quality. The extraction of DNA from bacteriophage usually yielded pure quantities of DNA and hence other shorter methods were not explored. The one-day mini-plasmid preparation used to analyse the recombinants after subcloning yielded very large amounts of good quality DNA. It is therefore suggested that this one-day procedure combined with a CsCl gradient step would be suitable for use if there was a problem with contaminating chromosomal DNA.

A number of problems have been experienced when digesting DNA with restriction enzymes such as MspI and EcoRI. The DNA was pure enough to be digested with all other enzymes and so the problem was specific for these two enzymes. Sometimes, an alteration in the quantity of enzyme used or extra phenol extractions on the DNA allowed for digestion of the samples. Variation in the quality of enzyme obtained from the supplier also could be a contributing factor. Both enzymes could be giving problem due to partial methylation at these sites. Spermidine trihydrochloride was found to facilitate the digestion of human DNA but care was taken not to use large amounts or the DNA precipitated out of solution.

The most important modification to Southern blotting was the change over from nitrocellulose membranes to nylon membranes. The cost of nylon membranes was double that of nitrocellulose but the ability to reuse these membranes resulted in an enormous saving of time and hence was cost-effective. Our group was most satisfied with Biotyne

(Pall) which was durable and easy to handle. DNA transfer was successful in 10 x SSC and a stripping procedure without formamide was used.

Our experience demonstrated that Biotrans was less sensitive than nitrocellulose after hybridization. This has recently been shown to be due to baking of the membrane and that by exposing the membrane to a specific UV light after transfer instead of baking, sensitivity after hybridization is enhanced (Khandjian, 1987). The addition of 5% Dextran Sulphate to the hybridization solution of Biotrans was also found to increase the sensitivity of detection using this membrane. It was also found that 0.1% SDS helped reduce background caused by non-specific hybridization. The hybridization solution given in Appendix B was a general solution used for all nylon membranes. Plastic bags replaced perspex hybridization boxes and up to five large membranes were placed together in one plastic bag during hybridization. The amount of hybridization solution needed was not five times that specified per square cm on each blot but 20 ml was sufficient for 5 large (20 X 18 cm) membranes. If the hybridization mixture was used over more than two blots, results were obtained using Kodak XAR5 film and exposing the membranes for up to seven days.

Oligolabelling of specific human DNA fragments was preferred to nick-translation of the whole bacteriophage clone. The specific activities obtained were much higher and the results were often clearer. Column separation of the non-incorporated nucleotides from the oligolabelled DNA fragments was necessary. However, to prevent any

unnecessary exposure to radioactive nucleotides, the non-incorporated nucleotides were usually removed using Sephadex G-50 columns. It was found that 50 μ Ci of ^{32} P was sufficient to label 64 ng of DNA for results on five large Southern blots.

Two different somatic cell hybrid DNA extraction protocols were used. The yield of DNA from the second method was far superior to of the first. The second method also cuts out the time-consuming step of cell collection. In this method somatic hybrid cells are lysed in their culture flasks and larger DNA yields were obtained. This suggested that the low yield from the first method could have been due to lysis of the cells during collection and then the DNA was susceptible to any endonucleases that were present. This would account for the degrading problems that were experienced. The disadvantage of the second method was that an overall cytogenetic analysis could not be done on all the culture flasks. This could lead to inaccurate results if some cells containing certain chromosomes were misrepresented in those flasks which were analysed.

A further development in the field was the substitution of Biotin labelled nucleotides for ^{32} P labelled nucleotides (Leary *et al.*, 1983). This technique of labelling DNA probes is however not favourable to a study such as this because the need to reuse Southern blots was an important advantage. Subsequent hybridization studies have to be done using ^{32} P. Future developments in this field would be advantageous if radio-active nucleotides could be eliminated

113.

from hybridization procedures. Safety precautions were used at all times when using radio-active nucleotides until such developments are made in the field.

CHAPTER THREE

RESULTS

3 CHAPTER THREE - RESULTS

A human genomic DNA library was screened for single-copy sequences. These clones were used to screen random individuals, whose DNA was digested with a panel of restriction enzymes. A number of restriction fragment length polymorphisms were detected and these were confirmed to be inherited in a Mendelian fashion by means of family studies. Those DNA sequences which detected RFLPs were tentatively localized to human chromosomes by hybridization studies using a somatic cell hybrid panel. The allele frequencies of the various RFLPs were determined in three Southern African populations.

3.1 Human genomic probes

3.1.1 The Isolation of Human DNA Probes

Screening of the HaeIII/AluI human DNA library

The human DNA library was plated out and the titre (the number of plaque forming units per ml) of the library was determined. This had decreased from 1.0×10^{11} per ml in the previous screening (carried out in the Department, January, 1984) to 1.0×10^9 per ml. Approximately 1404 non-overlapping plaques were isolated and screened by the method of Benton and Davies (1977). Eighty-five plaques did not show any detectable hybridization to genomic DNA. An autoradiograph of radiolabelled total human genomic DNA hybridized to a plate containing phage with and without

repetitive human sequences is shown in figure 2.1 (page 68). The "gaps" represent plaques of bacteriophage which contain single-copy human DNA inserts or lack human inserts altogether.

Isolation of bacteriophage DNA containing human inserts

DNA from each bacteriophage clone was digested with the restriction endonuclease EcoRI to indicate which of the clones contained human DNA sequences. Twenty-three of the 85 clones had restriction patterns which indicated that they contained human DNA sequences. Some of those which detected RFLPs can be seen in figure 3.1a and 3.2. The Lambda Charon 4A DNA has EcoRI fragments of 19,9; 10,9; 7,8 and 6,6 kb in length. The joining of the left and right phage arms at their cohesive ends forms the larger 30,7 kb fragment which can be seen in figure 3.2. Human DNA sequences present in the clones would be distinguishable from the Charon 4A sequences, since the human fragments replace the 7,8 and 6,6 kb Charon 4A fragments.

The sizes of the human EcoRI fragments within each clone were calculated using molecular weight standards (Lambda DNA digested with HindIII and EcoRI), and are listed in table 3.1. Not all the smaller fragments are visible in figure 3.1a but they can be seen in figure 3.2. A further 10 clones also contained human DNA inserts. Six of these clones did not detect any variation and the other 4 contained repetitive DNA sequences which gave results which could not be interpreted and therefore these recombinant phage have not been followed up.

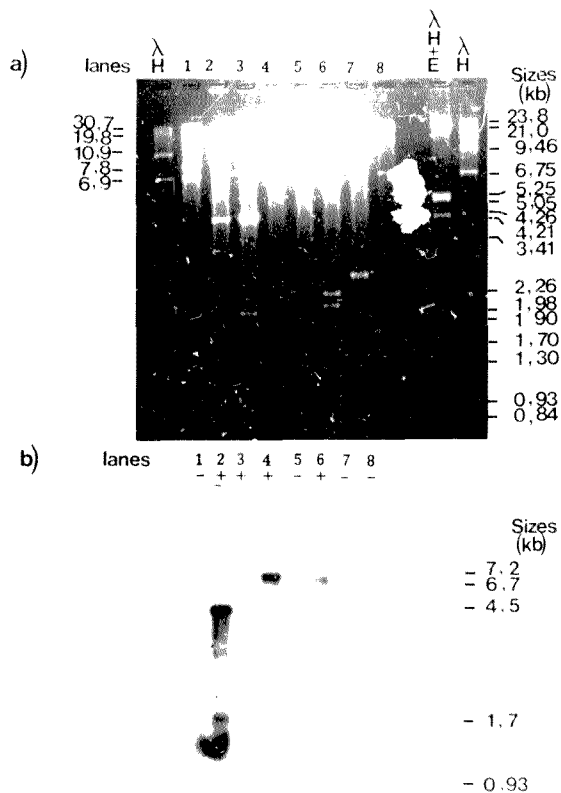


Figure 3.1. a) An ethidium bromide stained agarose gel containing bacteriophage clones, digested with *EcoRI*. Lanes 1-8 contain Lambda Charon 4A, probes VC-28, 61, 63, 64, 71, 75, and 85 respectively. Standard Lambda molecular weight markers, digested with *HindIII* (λ H) and *HindIII* + *EcoRI* (λ -H+E) are also shown. The sizes of the human inserts are given in Table 3.1.

b). An autoradiograph of the Southern blot of the gel in a) showing that some of the human inserts of clones VC28, VC61, VC63 and VC71 in lanes 2,3,4,6, have hybridized (+) to total human DNA. The dark mark in lane 2 is an artifact and the smearing is due to partial digestion of the DNA which is not visible in Figure 3.1a but is visible after hybridization.

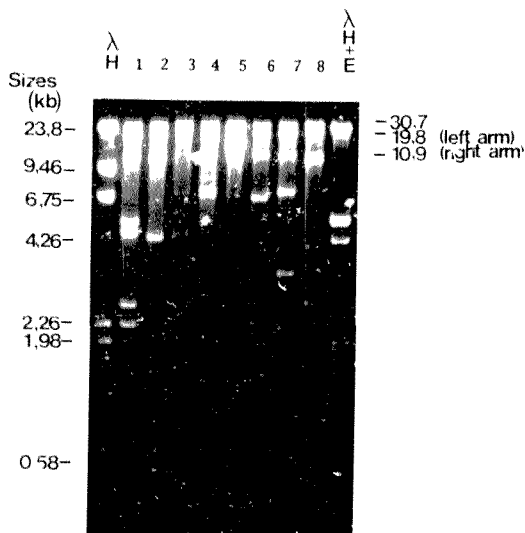


Figure 3.2. An ethidium Bromide stained gel containing lambda phage clones digested with EcoRI. Lanes 1-8 contain VC-14, 28, 61, 61, 64, 71, 85, and 75 respectively. Molecular weight markers are indicated. Note, in lane 5, the large human insert between the phage arms, and in lanes 1 and 8, the smaller human fragments of VC61 and VC85, are visible. Fragment sizes correspond to those given in Table 3.1.

TABLE 3.1. PROPERTIES OF THE EIGHT HUMAN DNA INSERTS IN BACTERIOPHAGE CHARON 4A

Probe name	EcoRI Fragments from insert (kb)	Total Size of insert (kb)	Fragments Sizes containing repetitive sequences (kb)
VC14	5,2; 4,9; 2,7 2,2; 1,0; 0,85 0,72	17,57	none
VC28	4,5; 1,7	6,2	both
VC61	4,50; 1,98; 1,0 0,93; <u>0,47</u> ; <u>0,40</u>	8,28	4,50 0,93
VC63	7,20; 6,10 1,30; 0,95	15,55	7,20 6,10
VC64	14,5; 1,25	15,75	none
VC71	6,7; 2,35 2,1; 0,84	11,99	6.7
VC75	10,0 ; 2,5 1,35; 0,47	14,32	none
VC85	6,7 ; 3,3 ; 2,15 1,7 ; 0,95; 0,84 <u>0,77</u>	16,41	none

The underlined fragments are visible in Figure 3.2 but not in figure 3.1a.

Further screening for repetitive human DNA sequences

A Southern blot, using a nylon membrane, was prepared from the agarose gel shown in figure 3.1a. The hybridization of this blot to radiolabelled total human genomic DNA indicated whether any of the human fragments contained repetitive DNA sequences. It can be seen from figure 3.1b that some of the human fragments have hybridized to total human DNA. The membrane in figure 3.1b has been washed at a moderate stringency (55°C in 0,1% SDS; 0,1X SSC) because even the repetitive DNA sequences would be removed at higher stringencies. When the clones were initially isolated, they were transferred to nitrocellulose and then screened under stringent conditions and therefore there was no detectable hybridization to total human DNA in the clones that were selected. Repetitive DNA sequences within these clones were only detected when nylon membranes were used and a high background, due to non-specific hybridization of the probe, was found. In order to get clear hybridization results those DNA fragments which contained repetitive sequences were identified and were excluded from the probe. The sizes of the human DNA fragments which contained repetitive sequences are indicated in table 3.1. Some of the human EcoRI fragments, which identified the RFLPs have been subcloned and are described in the following section.

3.1.2 Subcloning of certain Probes

The EcoRI fragments of clones VC63, VC61 and VC85 which detected specific RFLPs were identified, isolated and purified from a low melting temperature agarose gel. The vector chosen for subcloning was pUC19 and it was also digested with EcoRI for a simple "sticky-end" ligation reaction.

Successful subclones for VC61, VC63 and VC85 have been obtained, and are called pVC61, pVC63 and pVC85 respectively. Diagrams of the three subclones are given in figure 3.3. In figure 3.4a an EcoRI digest of DNA from each of the subclones is shown. In figures 3.4b, c, d, and e, the hybridization pattern of the original fragments are shown to correspond to those present in the new subclones. These subclones all successfully detected their respective RFLPs (see below).

In summary, then, eighty-five clones were isolated from a human genomic DNA library as potential single-copy human DNA probes. Restriction enzyme analysis indicated that 23 of these clones contained human DNA inserts. Of these, 14 still contained repetitive sequences within the clones and only 9 contained single-copy human DNA sequences. The overall percentage of single-copy sequences isolated in this study was therefore 0,6 %. A single-copy EcoRI fragment was isolated from 2 of the 14 clones, which contained repetitive DNA sequences, and detected variation in the human genome. Three of the clones which detect RFLPs have been subcloned into pUC19, a plasmid vector.

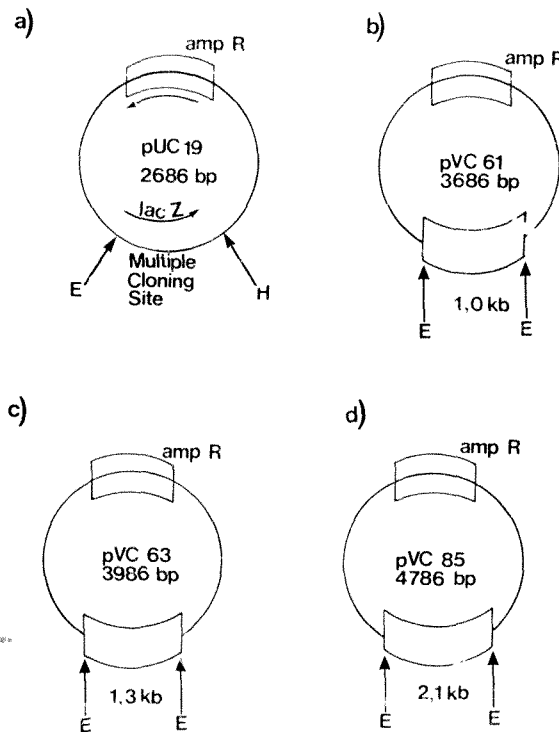


Figure 3.3. A schematic representation of pUC19 and the subclones are shown in (a). In b) pVC61, the subclone of VC61, containing a 1.0 kb *EcoRI* human inserts, in c.) pVC63, the subclone of VC63, containing a 1.3 kb *EcoRI* fragment and in d.) pVC85, the subclone of VC85 containing a 2.1 kb *EcoRI* fragment are illustrated. (*amp^r*- ampicillin resistance gene, *E*-*EcoRI* site, *H*-*HindIII* site, *lacZ*- β -galactosidase gene).

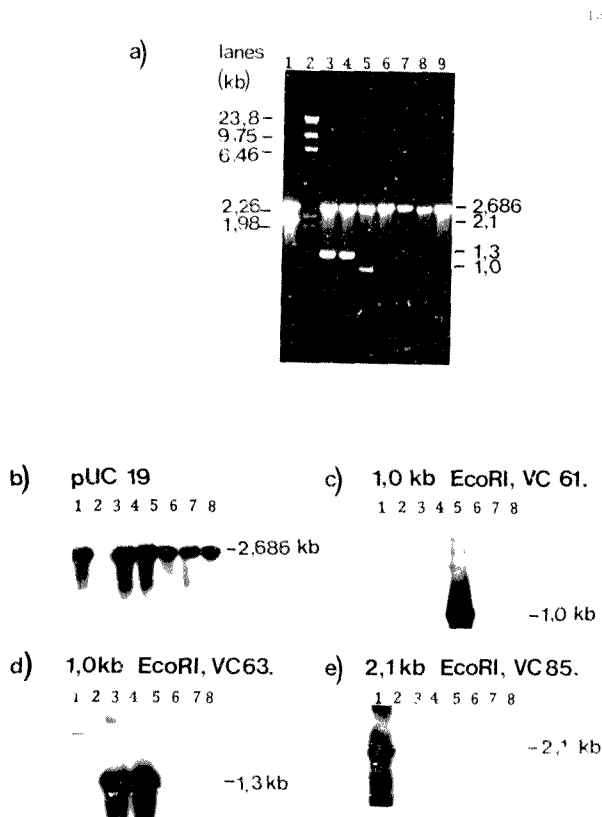


Figure 3.4. An ethidium bromide stained gel containing in lanes 1) pVC85, 2) Lambda HindIII marker, 3) pVC63(1), 4) pVC63(2), 5) pVC61 and pUC19 in lanes 6, 7, 8 and 9. In b) pUC19 has only hybridized to itself and not to the inserts. In c), d) and e) each fragment subcloned, has been hybridized to a Southern blot of the gel in a) and each has hybridized to their corresponding inserts. The dark smearing is due to partial digestion and excess DNA in the lanes.

3.2 RFLPs isolated in this study

3.2.1 Restriction Fragment Length Polymorphisms

Human DNA was isolated from random individuals. The average yield from 25 ml of whole blood was between 0,6 and 1,0 mg. Ten µg of human DNA was digested with each of the following fourteen restriction endonucleases; MspI, TaqI, BamHI, BglII, HinfIII, HindII, EcoRI, PvuII, PstI, SatI, StuI, AvaII, MboI and RsaI and Southern Blots were prepared.

The majority of RFLPs detected in this study were due to base-pair substitutions which resulted in the creation or loss of a recognition site for a restriction enzyme. These DNA changes were recognized by one restriction enzyme and generated simple two-allele polymorphisms. Family studies showed that all RFLPs discovered in this study were inherited in a co-dominant Mendelian fashion and these results are presented in the following paragraphs.

I. Basepair Substitution RFLPs

Probe VC14 detects two MspI Polymorphisms

This clone contains a 17,6 kb single-copy human DNA insert and it detects two MspI two-allele RFLPs (A and B) in the human genome. The allelic fragment sizes are (A1) 20,0 kb (A2) 11,6 kb and (B1) 14,0 kb (B2) 12,7 kb (figure 3.5a). In order to confirm this, the bacteriophage was digested

with MspI and two fragments, 11,0 and 8,5 kb were isolated. These fragments do not correspond directly to the size of the human fragments on a Southern blot. The clone was constructed using EcoRI linkers and hence digestion with any other enzyme would result in the flanking human fragments being attached to the left and right arms of the bacteriophage vector. The 11,0 kb fragment, when hybridized to a Southern blot of MspI digested human DNA, detected the 20,0 and 11,6 kb fragments (figure 3.5b). Similarly the 8,5 kb fragment hybridized to the 14,0 and 12,7 kb human fragments (figure 3.5c). There is some hybridization to the 11,6 kb MspI fragment, using the 8,5 kb MspI fragment which is due to the 8,5 kb MspI fragment not being isolated completely from the 11,0 kb MspI fragment.

Probe VC14 detects two polymorphic MspI sites which are outside of the region that was originally cloned and this is schematically represented in figure 3.6. Family studies using both MspI fragments confirm Mendelian inheritance of the specific alleles at each locus (figure 3.7a and 3.7b).

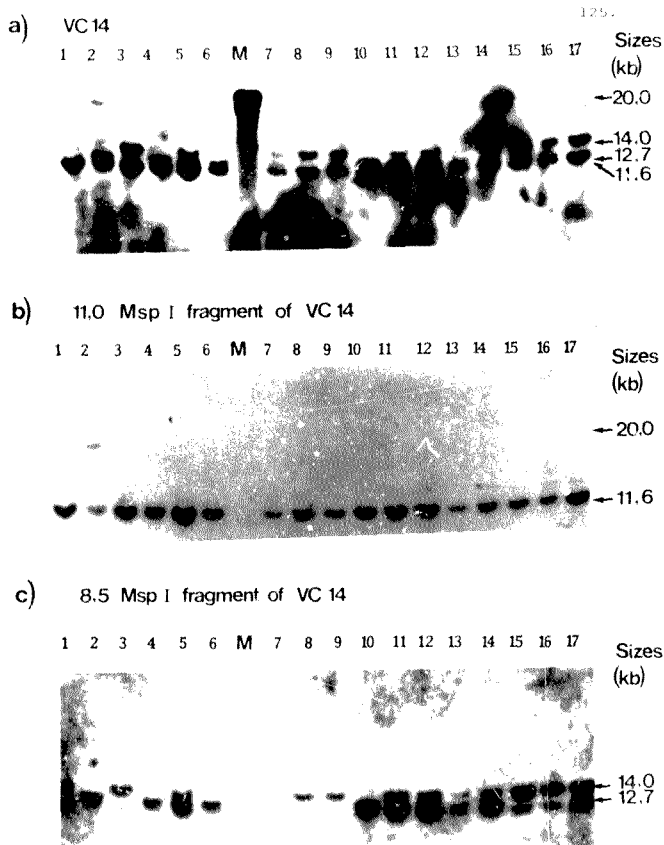


Figure 3.5. An autoradiograph of Southern blot of an MspI digest of DNA from random Bagroid individuals, hybridized to a) probe VC14, where the individual in lane 2 is the only heterozygote for the 20 and 11.6 kb fragments. b) The 11.0 MspI fragment of VC14, which detects the first RFLP at this locus. c) the 8.5 kb MspI fragment of VC14, which detects the second RFLP at this locus. There is slight hybridization to the 11.6 kb fragment possibly due to contamination of the probe with the 11.0 kb MspI fragment.

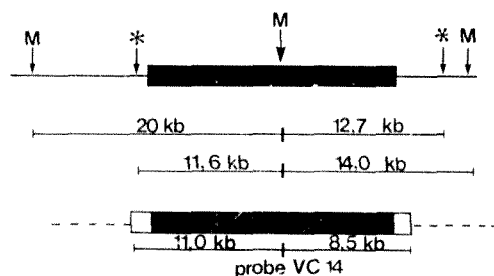
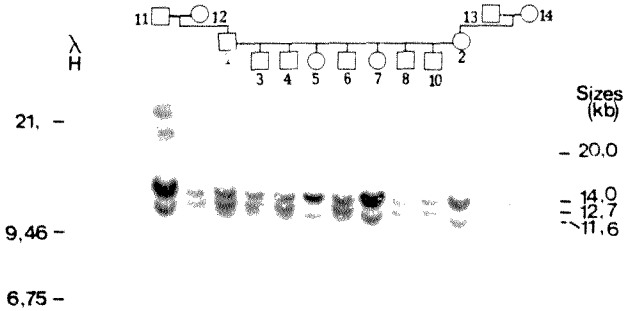


Figure 3.6. Schematic representation of the *Msp*I restriction pattern of probe VC14. The 11.0 kb *Msp*I fragment from the clone hybridizes to the 20.0 and 11.6 kb human genomic *Msp*I fragments. The 8.5 kb *Msp*I fragment hybridizes to the 12.7 and 14.0 kb human *Msp*I fragments. (*-polymorphic sites, M-*Msp*I).

a) VC 14



b) VC 14 11,0 Msp I fragment

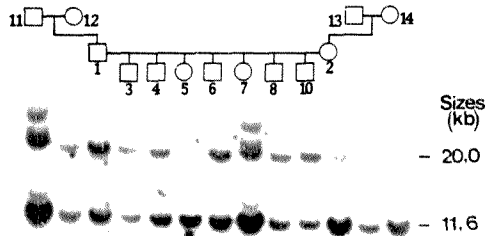


Figure 3.7. An autoradiograph of a Southern blot of an *MspI* digest on DNA from a three generation family (C.E.P.H. no. 1424), hybridized to a) probe VC14 which shows the segregation of the second *MspI* RFLP at this locus. The fragments of the first RFLP are only faintly visible. In b) the 11,0 kb *MspI* fragment showing the first RFLP at this locus. The extra bands in individuals 7, 11 and 14 which are visible due to partial digestion with the restriction enzyme.

Probe VC28 detects MspI and TaqI polymorphisms

This clone contains a 6,2 kb human DNA insert which contains moderately repetitive DNA sequences, since both the human EcoRI fragments hybridized to radiolabelled total human genomic DNA (figure 3.1b, page 116). This probe detected two simple two-allele RFLPs with the restriction enzymes TaqI and MspI. Stringent washing conditions (65°C in 0,05 X SSC 0,05% SDS for at least one hour) were however needed to remove the repetitive fragments.

The hybridization pattern of VC28 to an MspI digest of DNA from random Negroid individuals is shown in figure 3.8a and a number of constant fragments in addition to the polymorphic ones can be seen. In figure 3.8b the hybridization pattern of the 3,6 kb MspI fragment, which was isolated by digestion of VC28 with MspI, to the same blot is shown.

The two-allele MspI RFLP is apparent and the extra fragment in lane 6 in figure 3.8a, is no longer present in 3.8b and is therefore not part of the polymorphism. This extra fragment has not been seen in subsequent screening studies and could possibly have been due to either a partial digestion or it could represent a second variant fragment length detected with VC28. Family studies on this individual were unfortunately not possible.

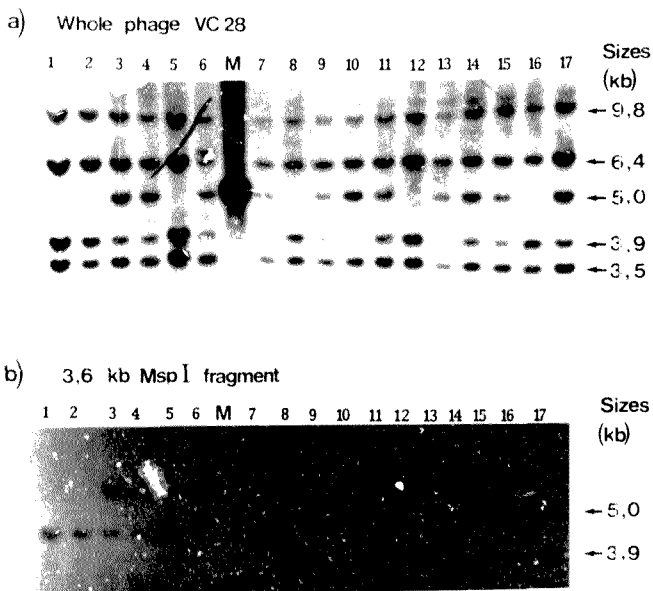


Figure 3.8. An autoradiograph of a Southern blot of DNA from random Negroid individuals digested with *Msp*I and hybridized to a) probe VC28 and b) the 3.6 kb *Msp*I fragment of VC28. In a) the two-allele RFLP, fragment sizes 5.0 and 3.9 kb are visible in addition to a number of constant fragments. The additional fragment present in lane 6 is not visible in b) and therefore is not part of this RFLP. The 5.0 kb *Hind*III/*Eco*RI fragment of the lambda molecular weight marker is visible (M). In b) the two-allele RFLP can clearly be seen. No labelled lambda DNA was added to the reaction and therefore no bands are visible in lane (M).

Figure 3.9 is an autoradiograph of a family study of the clone VC28 which was hybridized to a Southern blot of family DNA digested with MspI. The parents are homozygous for different alleles and the three children are all heterozygotes, thus confirming co-dominant Mendelian inheritance of this polymorphism.

The TaqI polymorphism can also be detected by the same 3,6 kb MspI fragment of probe VC28. In figure 3.10a an autoradiograph of the hybridization pattern of probe VC28 hybridized to a TaqI digest of DNA from random Caucasoid individuals is shown. A number of constant fragments are visible and an extra 3,6 kb fragment is present in some individuals. Hybridization of the 3,6 kb MspI fragment, to the same blot, revealed the polymorphic fragments. The 4,8 kb fragment not the 4,1 kb fragment, which is directly above the 3,6 kb fragment, is the allelic fragment of the RFLP (figure 3.10b).

Family studies have shown the Mendelian inheritance of this polymorphism, but to date no individual who is a homozygote for the smaller 3,6 kb fragment has been detected (figure 3.11). It is interesting to note that both the TaqI and MspI RFLPs, which have similar sized allelic fragments, have been detected within a 3,6 kb DNA sequence and further that both of these restriction enzymes contain a CpG dinucleotide sequence in their recognition site.

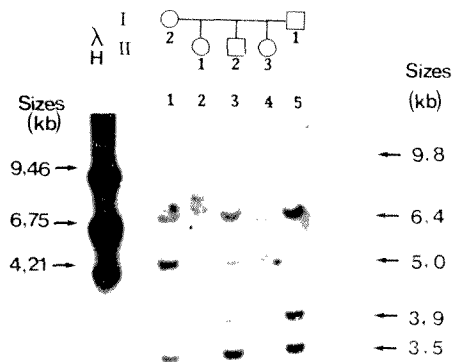
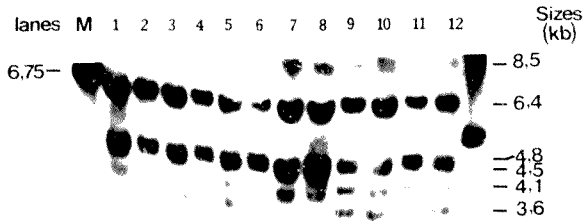


Figure 3.9. An autoradiograph of the bacteriophage clone VC28 hybridized to a Southern blot containing DNA, digested with *MSPI*, from a five member family. Each parent is homozygous for either the 5.0 or the 3.9 kb allele and the three children are clearly all heterozygotes. The faint extra fragments in lanes 2, 4 and in all lanes (around 4.0 kb) are due to partial digestion of the DNA. Fragment sizes of the lambda molecular weight marker are indicated.

a) VC 28



b) VC 28 3,6 kb Msp I fragment

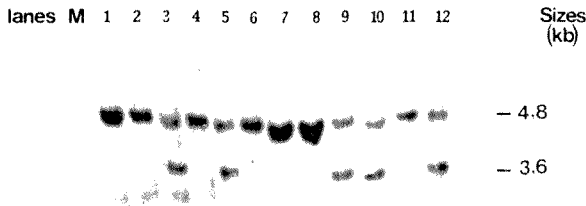


Figure 3.10. An autoradiograph of a Southern blot of random Caucasoid DNA digested with TaqI and hybridized to a) probe VC28 and b) the 3,6 kb MspI fragment of VC28. In a) the 6,75 kb fragment of the molecular weight marker is visible (M). In b) the two allelic fragments can be seen, the larger 4,8 kb and the smaller 3,6 kb fragment. The gel migrated unevenly and the bands are not straight.

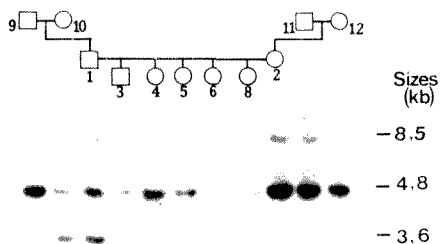


Figure 3.11. An autoradiograph of a Southern blot of *TaqI* digested DNA from a three generation family (C.E.P.H. no. 13294) and hybridized to the 3.6 kb *MspI* fragment of probe VC28. The 8.5 kb fragment is a constant fragment and faintly visible in all the lanes.

Probe VC61 detects a TaqI polymorphism

This clone contains a 8,28 kb human DNA insert. Two of the human EcoRI fragments within this clone hybridized to total human DNA (figure 3.1b, page 116). A high frequency two-allele TaqI polymorphism was detected and the variant fragments are 12,5 and 10,1 kb in length. There were no constant fragments (figure 3.12). The TaqI RFLP is visible even though the DNA has migrated unevenly in places. This RFLP was initially detected using the clone VC61 and by means of high stringency washing of the membrane.

In an attempt to get clearer results, the bacteriophage clone was digested with TaqI. Those TaqI fragments, which did not correspond to Charon 4A DNA TaqI fragments, were excised from a low melting temperature agarose gel and each was then in turn hybridized to TaqI human DNA blots. All of these fragments gave very strong background and the RFLP was not apparent. In order to get clear results, those human EcoRI fragments of the clone (1,98 and 1,0 kb), which did not hybridize to radiolabelled human DNA (figure 3.1b, page 116), were pooled and oligolabelled.

A study of a three generation family, showing co-dominant Mendelian inheritance of the alleles, can be seen in figure 3.13. The 1,0 kb human EcoRI fragment of VC61, which has been successfully subcloned into a plasmid, pUC19 (section 3.1.2) was used to produce this autoradiograph.

135.

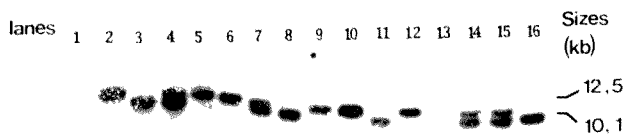


Figure 3.12. An autoradiograph of Southern blot of DNA from random Caucasoid individuals, digested with *TaqI* and hybridized to the 1.98 and 1.0 kb *EcoRI* fragments of probe VC61. The two-allele polymorphism can be seen. The DNA has migrated unevenly in the gel but this has not affected the results.

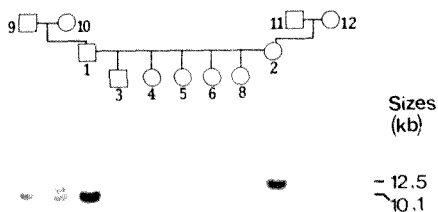


Figure 3.13. An autoradiograph of a Southern blot of a *TaqI* digest of a three generation family (C.E.P.H. no. 13294) hybridized to the 1.0 kb *EcoRI* fragment, isolated from the subclone pVC61. The parents are homozygous for different alleles and all the children are heterozygous.

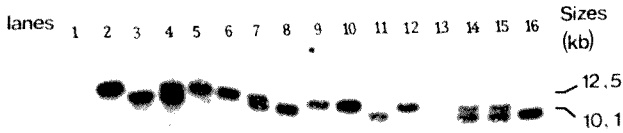


Figure 3.12. An autoradiograph of Southern blot of DNA from random Caucasoid individuals, digested with *TaqI* and hybridized to the 1.98 and 1.0 kb *EcoRI* fragments of probe VC61. The two-allele polymorphism can be seen. The DNA has migrated unevenly in the gel but this has not affected the results.

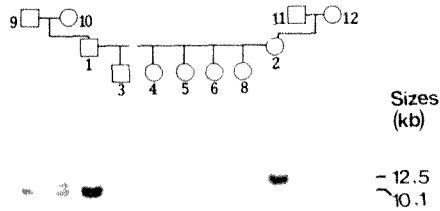


Figure 3.13. An autoradiograph of a Southern blot of a *TaqI* digest of a three generation family (C.E.P.H. no. 13294) hybridized to the 1.0 kb *EcoRI* fragment, isolated from the subclone pVC61. The parents are homozygous for different alleles and all the children are heterozygous.

Probe VC63 detects a TagI polymorphism

This clone contained a human DNA insert of 15,5 kb and two of the human EcoRI fragments (7,2 and 6,1 kb) hybridized to total human DNA (figure 3.1b, page 116). This probe detected a simple two-allele TagI RFLP. The two allelic fragments are 13,0 and 9,0 kb and a constant fragment of 10,5 kb is also found. In figure 3.14 an autoradiograph of a Southern blot using a nitrocellulose membrane, containing TagI digests of DNA from random Caucasoid individuals, hybridized to VC63 is shown.

These results were not repeatable when nylon membranes were used. The 1,3 and 0,95 kb human EcoRI fragments, which did not hybridize to human genomic DNA, were excised from a low melting temperature agarose gel, oligolabelled and used to obtain clear results. These two fragments did not detect the constant 10,5 kb fragment. In figure 3.15 an autoradiograph of a family study demonstrating the co-dominant Mendelian inheritance of the alleles of this RFLP is shown.

The constant fragment is no longer visible because the 1,3 kb human EcoRI fragment from the sub-clone of VC63, pVC63 was used. A schematic representation of the TagI restriction pattern which resulted in this polymorphism is illustrated in figure 3.16.

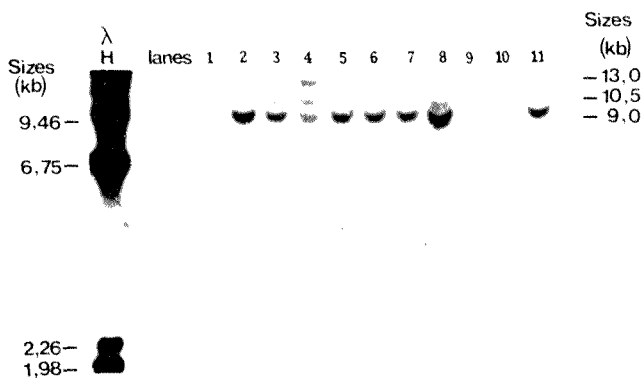


Figure 3.14. An autoradiograph of the probe VC63 hybridized to a Southern blot of DNA from random Caucasoid individuals, digested with *TagI*. A constant band of 10.5 kb, between the variant fragments, is visible. This Southern blot was made using nitrocellulose and it has been washed stringently. Some fragments of the lambda molecular weight marker are indicated.

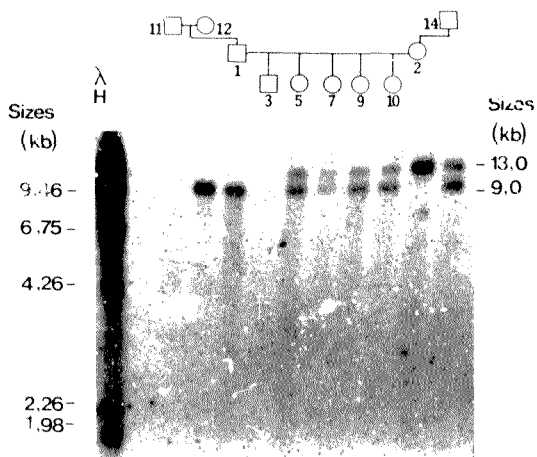


Figure 3.15. An autoradiograph of a Southern blot of *TaqI* digests of DNA from some individuals of a three generation family (C.E.P.H. no. 1416) which has been hybridized to the 1.3 kb *EcoRI* fragment of the subclone pVC63. Fragments from the lambda *HindIII* molecular weight marker are also shown. The constant band of 10.5 kb between the variant fragments is no longer visible.

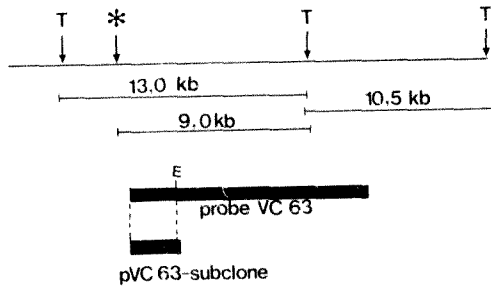


Figure 3.16. Schematic representation of the *TagI* restriction pattern of probe VC63. The position of the 1.3 kb *EcoRI* fragment which has been subcloned is shown (pVC63). It does not detect the 10.5 kb constant fragment because it is the outer lying fragment of the sequence originally cloned. (T-*TagI*, * -polymorphic *TagI* site).

Probe VC64 detects a TaqI polymorphism

This clone contains a 15,75 kb single-copy human DNA insert consisting of a 14,5 kb and 1,25 kb EcoRI fragment. The 14,5 kb fragment between the two phage arms (19,9 and 10,9 kb) can be seen in Figure 3.2 (page 117). This probe detects a simple two-allele TaqI RFLP. A TaqI site within the 15,5 kb fragment has either been created, thus resulting in two smaller fragments of 10,1 and 5,4 kb, or it has been lost between the 10,1 and 5,4 kb fragments, resulting in the larger 15,5 kb fragment. In figure 3.17, an autoradiograph of DNA from 6 random San individuals digested with TaqI and hybridized to VC64, can be seen.

A variant allele has also been detected, but unfortunately family studies were not possible to show the inheritance of this variant allele, since no relatives of this individual could be obtained. A further 59 San individuals have been screened with this probe and other 12,0 kb alleles have been found. This individual has probably gained a TaqI site which lies within the TaqI sites which gave rise to the 15,5 kb fragment. The TaqI restriction pattern of this probe is schematically illustrated in figure 3.18.

In figure 3.19, an autoradiograph of a family study showing the co-dominant Mendelian inheritance of the TaqI alleles is shown. The father, a heterozygote (result not shown), had three TaqI fragments and 4 of the 5 children have inherited the paternal allele represented by the 10,1 and 5,4 kb fragments since their mother (I-1) is a homozygote for the 15,5 kb allele.

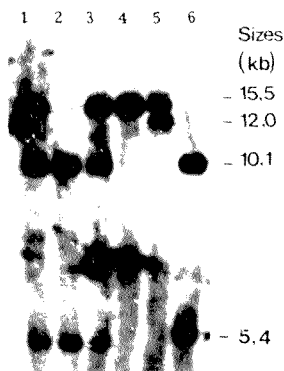


Figure 3.17. An autoradiograph of a Southern blot of DNA from random San individuals, digested with *TagI* and hybridized to probe VC64. One individual, lane 5, shows a variant pattern and has fragments of 12,0 kb and a 15,5 kb.

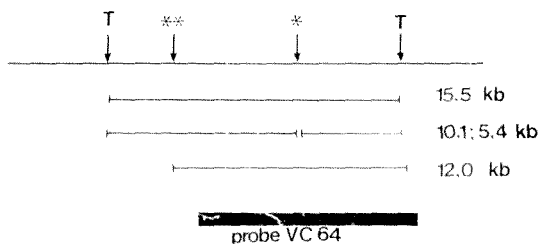


Figure 3.18. Schematic representation of the *TagI* polymorphism detected by VC64. The two smaller *TagI* fragments (10,1 and 5,4 kb) add up to the larger fragment (15,5 kb). A base pair change has probably resulted in the loss or gain of a *TagI* site. (* -polymorphic site, T-*TagI* ** - variant *TagI* site).

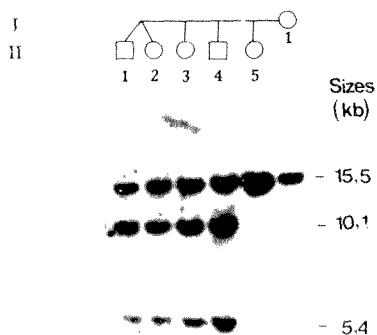


Figure 3.19. An autoradiograph of probe VC64 hybridized to a Southern blot of *TaqI* digests on DNA from a five child family. The father is a heterozygote having all three fragments (not shown) and the mother (I-1) is a homozygote for the 15.5 kb allele.

Probe VC75 detects a three allele AvaII polymorphism

This clone contains a 14,3 kb single copy human DNA insert. It was initially shown to detect a two-allele AvaII RFLP in all three Southern African populations, but when it was tested on a number of parents from the C.E.P.H. collaboration study, a third allele of 3,2 kb was found. This allele was present in a lower frequency than that of the 4,3 kb allele. Figure 3.20 shows the hybridization results of this probe to DNA from random Caucasoid individuals digested with AvaII. The three allelic fragments are 4,3; 3,5 and 3,2 kb, and a number of constant fragments can also be seen.

Figure 3.21a presents the results of a family study in which the co-dominant segregation pattern of 4,3 and 3,5 kb alleles can be seen. Both of the parents are heterozygous for the RFLP and one of the children (II-1) is a homozygote for the smaller 3,5 kb allele. Figure 3.21b shows the co-dominant segregation in three generations of the 3,5 and 3,2 kb alleles. This Southern blot was hybridized to the 1,35 EcoRI human fragment isolated from probe VC75. This fragment could be subcloned at a later stage since it detects only the RFLP and none of the constant fragments.

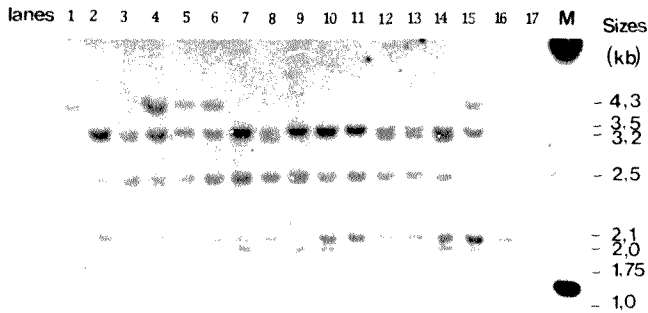


Figure 3.20. An autoradiograph of probe VC75 hybridized to a Southern Blot of *Ava*II digestions on DNA from random Caucasoid individuals. Three alleles represented by 4.3; 3.5 and 3.2 kb fragments (A1, A2 and A3 respectively) can be seen in addition to the constant fragments. The 5.24; 5.05 doublet and 1.32 kb fragments of the lambda *Hind*III/*Eco*RI molecular weight marker are present in lane M. In lane 1, a homozygote for A1, in lane 4 a heterozygote (A1A2) and in lane 8 a heterozygote (A2A3) can be seen.

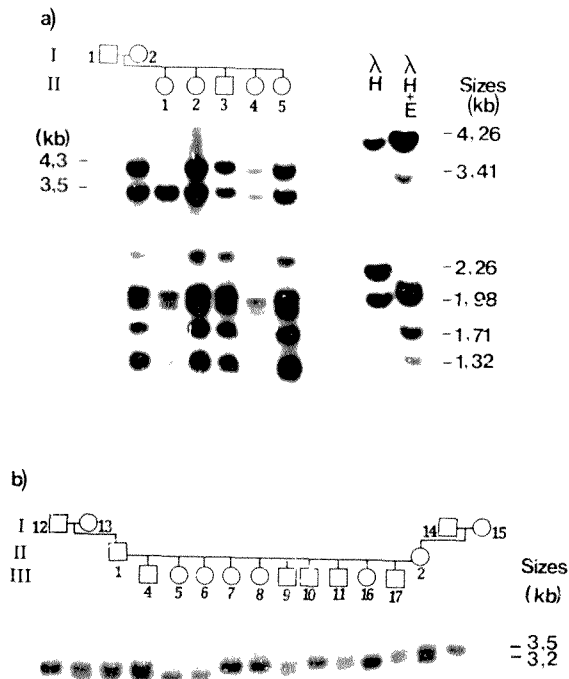


Figure 3.21. An autoradiograph of Southern blots of DNA digested with *Avall* from individuals of two families. In a) the segregation of the 4.3 and 3.5 kb alleles and the constant bands can be seen. In b) the segregation of the 3.5 and 3.2 kb alleles can be seen in family two (C.E.P.H. no 1311). There are no constant fragments present because the 1.35 kb *EcoRI* fragment from VC75 was used as a probe. In a) fragments sizes similar to the marker are in different positions on the gel. This is due to uneven running of the gel but the actual distances from the origin were the same.

II. DNA rearrangement-type RFLP detected by probe VC85

One of the more interesting findings in this study was the isolation of a probe which detected a DNA rearrangement-type RFLP. This type of RFLP is usually only detected if the rearrangement involves a region of DNA, probably greater than 300 bp, which is resolvable on a gel. A number of smaller rearrangements would be detected as base-pair substitution RFLPs.

Clone VC85 was an interesting probe discovered in this study. It contained a 16,41 single-copy human DNA insert. Initially nine enzymes were found to show variation with this probe and it was concluded that VC85 detected an insertion/deletion RFLP involving 500 bp of DNA because: Firstly, the variation in polymorphic fragment sizes was constantly 500 bp with the majority of restriction enzymes used. This is illustrated schematically in figure 3.22. Secondly, the same individuals were always heterozygous for the polymorphism with all the enzymes used.

In figure 3.23 the hybridization patterns of probe VC85 to DNA of random individuals digested with four restriction enzymes, StuI, PvuII, BglII and BamHI, and their fragment sizes are given. An insertion/deletion RFLP involving a constant amount of DNA behaves as di-allelic RFLP at a single locus.

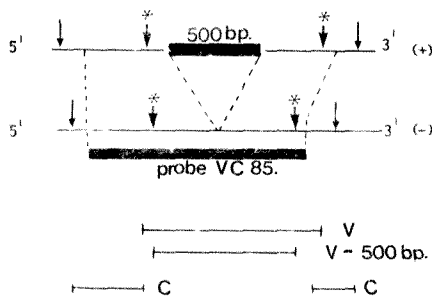


Figure 3.22. Schematic representation of the deletion/insertion-type RFLP detected by probe VC85. Any enzyme which has sites (*) 5' and 3' to the deletion/insertion region, will show a constant difference in length of 500 bp between the fragments (V), in addition to other constant fragments (C).

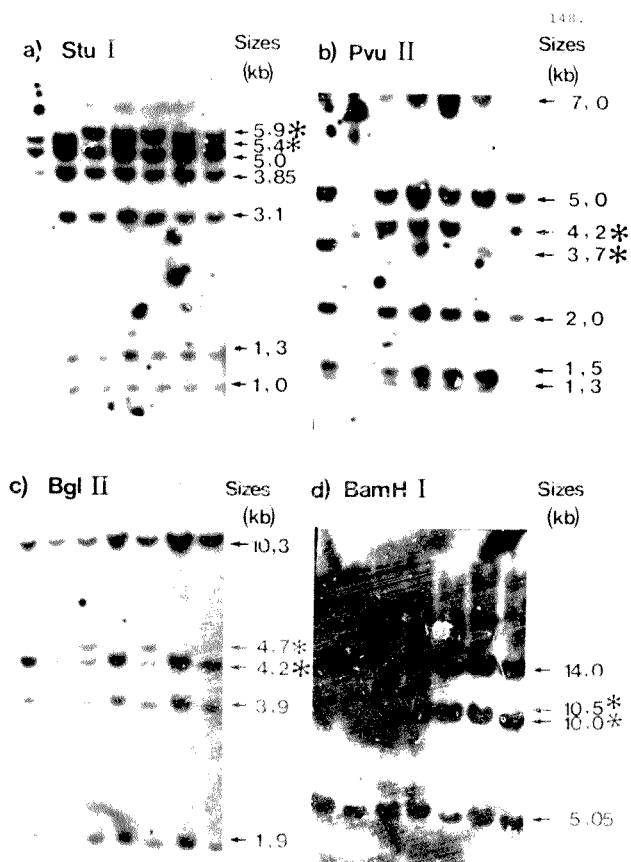


Figure 3.23. Autoradiographs of DNA from random individuals digested with four different restriction enzymes namely a) *Stu*I b) *Pvu*II c) *Bgl*II d) *Bam*H I, and hybridized to the probe VC85. The variant fragments (*) can be seen with each of these enzymes and the difference between the polymorphic fragment lengths was usually 500 bp. The fragment sizes for each enzyme are shown.

The enzymes MspI and TaqI both showed unusual restriction fragment length variation with this probe, but still corresponded to the insertion/deletion polymorphism. These results will be discussed in the section on characterization of probe VC85 (section 3.3).

The alleles generated by this insertion/deletion are inherited in a Mendelian fashion (figure 3.24).

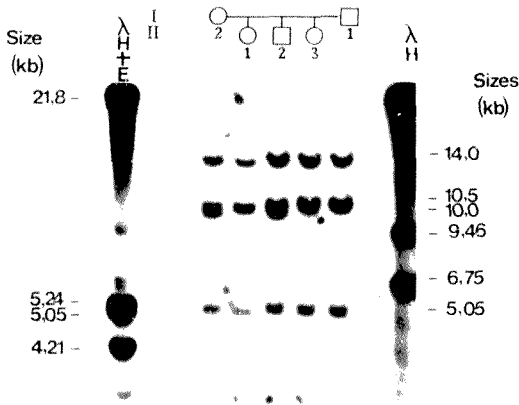


Figure 3.24. An autoradiograph of a Southern blot of DNA from a family, digested with *Bam*HI and hybridized to probe VCB1. The father (II-1) is a homozygote for the larger allele and the mother (II-2) a heterozygote. Constant fragments are also visible. The sizes of the fragments of two lambda molecular weight markers are shown.

3.2.2 Localization of the RFLPs detecting DNA sequence:

The tentative localization of all the RFLP detecting human DNA sequences isolated in this study was possible due to the generous donation of cell lines from a number of individuals. Lines XIID1dE, XII2D1f, XII2D1d were obtained from Dr. U. Francke, Yale University, U.S.A., lines 32-2-10 and 32-2-3 from Dr. T. Mohandas UCLA Medical Center, Torrance, C.A. U.S.A. and lines RBC13H-C, DMW5II-C, DMW13II-C and MDW55-C from Dr. A. E. Retief, University of Stellenbosch, S.A. All other lines were constructed by our Department (Dos Santos, 1986).

The somatic cell hybrids lines were maintained in the Cytogenetic Unit of the Department. Southern Blots were prepared from DNA extracted from these lines and the RFLP detecting sequences were hybridized to these blots. The hybridization pattern of presence or absence in a specific line could then be assessed according to the chromosomal content of the hybrid lines. In order to determine whether any cross-sequence homology of the human DNA probes exists, three important controls consisting of Chinese hamster, mouse and human DNA were also placed on these blots. In table 3.2 the results of the chromosomal analyses of the cell lines and the hybridization patterns of the different human RFLP-detecting sequences to the specific somatic cell lines are shown.

TABLE 3.2. CHROMOSOMAL CONTENT OF SOMATIC CELL HYBRID LINES AND CORRESPONDING HYBRIDIZATION PATTERN OF DNA PROBES.

Cell line	Chromosome content of each line																							Probe hybridization Results							
Cell lines code	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	VC14	VC28	VC61	VC63	VC64	VC75	VC85
10T1/2	3															p	*	*	*	*	*		p	*	*	*	*	*	*	*	*
10T1/3	2																*	*	*	*	*	*	p	*	*	*	*	*	*	*	*
10T1/4	3																*	*	*	*	*	*	p	*	*	*	*	*	*	*	*
10T1/5	4																								*	*	*	*	*	*	*
10T1/6	5																								*	*	*	*	*	*	*
10T1/7	6																								*	*	*	*	*	*	*
10T1/8	7																								*	*	*	*	*	*	*
10T1/9	8																								*	*	*	*	*	*	*
10T1/10	9																								*	*	*	*	*	*	*
10T1/11	10																								*	*	*	*	*	*	*
10T1/12	11																								*	*	*	*	*	*	*
10T1/13	12																								*	*	*	*	*	*	*
10T1/14	13																								*	*	*	*	*	*	*
10T1/15	14																								*	*	*	*	*	*	*
10T1/16	15																								*	*	*	*	*	*	*
10T1/17	16																								*	*	*	*	*	*	*
10T1/18	17																								*	*	*	*	*	*	*
10T1/19	18																								*	*	*	*	*	*	*
10T1/20	19																								*	*	*	*	*	*	*
10T1/21	20																								*	*	*	*	*	*	*
10T1/22	21																								*	*	*	*	*	*	*
10T1/23	22																								*	*	*	*	*	*	*
10T1/24	23																								*	*	*	*	*	*	*
10T1/25	24																								*	*	*	*	*	*	*

(+) indicates the chromosome was present in at least 15% of the cells analysed. A (-) indicates that the chromosome was not visible cytogenetically. (*) - cell line contained a micrometaphase (12p or 4p). (B) - line contained an isochromosome for 6p. Blank cells where lines were not tested. ? - presence of chromosome not definite. c - chromosome Xq22-qter. In section on products, the hybridization pattern corresponds to that given in section of localization of each probe. (Refer to the text) - (p, X) is translocation products. In section on hybridization the (+) and (-) indicate presence or absence of probe to that specific line.

Most of the RFLPs reported in this study were localized using one Southern blot, from which it was not possible to distinguish between chromosomes 7 and 20. Two other Southern blots using new lines were prepared and these helped in the assignment of some of the probes. Tentative assignments without any discrepancies were obtained for the probes, VC14, VC61, VC64 and VC85.

Probe VC14

This probe hybridized to all lines containing chromosome 6, (lines 2, 8 and 9) and to the human male control (figure 3.25). This probe was hybridized to other lines (20 and 21) which contained an isochromosome of the short arm of chromosome 6 and the probe was present in these two lines thus indicating that the probe is on chromosome 6p.

Probe VC61

Probe VC61 has been tentatively localized to chromosome 2. This probe can be seen to hybridize to lines 2, 4 and to the male control in figure 3.26. A few other lines were also analysed and there was hybridization to line 19 (which contains all the chromosomes except the Y chromosome) and line 22 (table 3.2). The only chromosome consistently present in all these lines was chromosome 2.

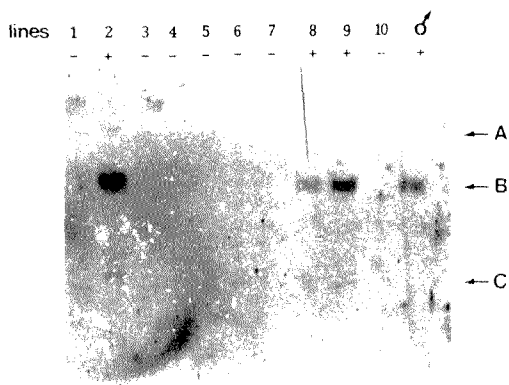


Figure 3.25. An autoradiograph of probe VC14 hybridized to a Southern blot of somatic cell hybrid DNA digested with *Hind*III. The probe has hybridized (+) to lines 2, 8, 9 and to a human male control. The arrows, A, B, and C indicate the human fragments. Fragments A and C are much fainter and are not clearly seen in all the positive lines. There was no hybridization to RAG or Chinese hamster control DNA (not shown).

lines

λ	1	2	3	4	5	6	7	8	9	10	ϕ	M	C
H	-	+	-	+	-	-	-	-	-	-	-	+	-

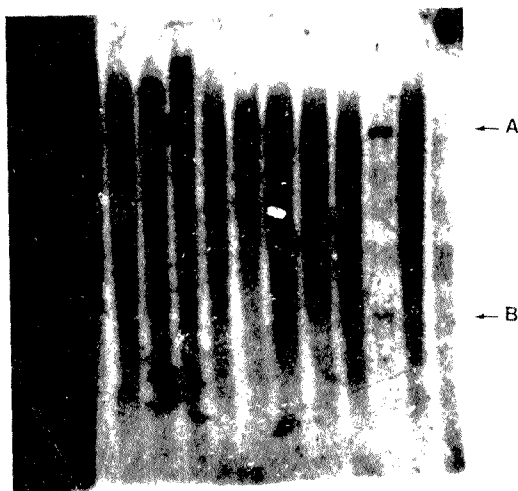


Figure 3.26. An autoradiograph of Southern blot of somatic cell hybrid DNA digested with *Hind*III and hybridized to probe VC61. There is definite hybridization (+) to lines 2, 4 and to the human male control, indicated by arrows A and B. The strong background is due to repetitive DNA sequences in the probe. The dark shadow, on the left, was due to another X-ray plate being placed over the blot during exposure. There is possibly some sequence homology to mouse (M) but not to Chinese hamster (C) DNA.

Probe VC64

This probe has been tentatively localized to chromosome 1. The probe has hybridized to lines 2, 9, 11, 12, 13, 14 and 19. There is cross sequence homology to mouse DNA and possibly to Chinese hamster DNA, but this cannot be clearly seen in figure 3.27. This probe was later shown to be linked to probe VC85 which is also found on chromosome 1.

Probe VC85

This probe has been assigned to chromosome 1. In figure 3.28 it can be seen that VC85 has hybridized to somatic cell hybrid lines 2 and 9 and to the human control, who was a heterozygote for the RFLP. Further hybridization results indicated that the probe was also present on a chromosome common to lines 11, 12, 13 and 14 (table 3.2).

Other fragments not corresponding to the human control fragments have hybridized to the probe. The two fragments present in lines 1, 2, 3, 8, 9, and 10 are thought to be Chinese hamster fragments, because all these lines were formed by fusion with Chinese hamster cells, even though the Chinese hamster control line (Wq3h) does not clearly show these fragments. This could be due to partial degradation of the Chinese hamster DNA on the blot. The extra fragments in lines 4, 5, 6 and 7, which are all mouse hybrid lines, correspond to fragments present in the RAG line (control mouse DNA).

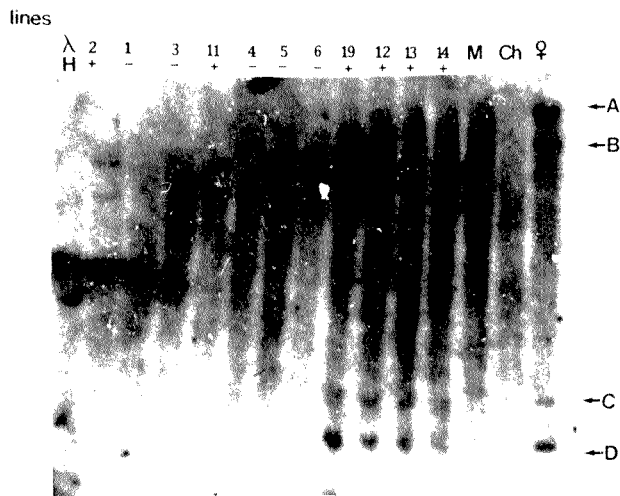


Figure 3.27. An autoradiograph of Southern blot of DNA of somatic cell hybrid lines digested with *Hind*III and hybridized to probe VC64. The probe is present in those lines which show all of the human fragments (A, B, C, and D,) namely lines 2, 12, 13, 14 and 19 and faintly in 11. Cross hybridization to mouse (M) DNA can be seen in lines 4, 5, 6, 12, 13 and 14. Hybridization to Chinese hamster DNA (Ch) is not clear. Some fragments of the lambda molecular weight marker are faintly visible.

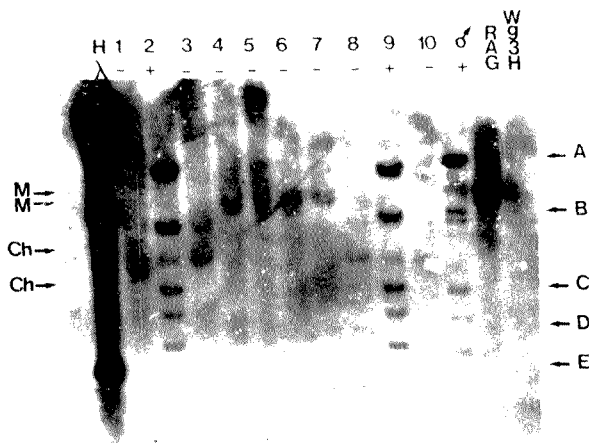


Figure 3.28. An autoradiograph of a Southern blot of somatic cell hybrid DNA digested with *Hind*III and hybridized to probe VC85. The human fragments of the control male, A, B, C, D, and E, are also present in lines 2 and 9. The extra fragment in the human control is due to the deletion/insertion RFLP where the control was a heterozygote at this locus. The extra bands in lines 1, 2, 3, 8, 9, and 10, which do not correspond to the human fragments, represent cross-sequence homology to Chinese hamster DNA and are marked Ch. The control Chinese hamster DNA (WgH) does not show these fragments clearly. Two further bands which also do not correspond to the human control are present in lines 4, 5, 6, 7 and in the RAG line (M), and indicate cross-sequence homology to mouse DNA. These lines were all constructed with mouse cells.

This human DNA probe therefore shows cross-species homology to both mouse and Chinese hamster. This suggests that it could be a conserved mammalian sequence. As was mentioned above probe VC64 and VC85 have been shown to be linked. Both were independently assigned to chromosome 1 and also show some cross-species homology to mouse and Chinese hamster DNA.

The following probes have also been tentatively assigned. A number of discordances which will be discussed however do exist.

Probe VC28

This probe has been tentatively localized to chromosome 8 since it has hybridized to lines 2, 4, 5, 8 and 10 and to the human male control (figure 3.29). One discordancy which was present is that lines 8 and 10 have either chromosome 8 or 4 present respectively, but not both. This indicates that if chromosome 8 is present in line 10, it is at a lower frequency than could be detected cytogenetically. The probe has further been shown to be present in lines, 16 and 18 which both have chromosome 4 and 8 (table 3.2).

The probe has however not hybridized to lines 11, 12 and 17, all of which contain chromosome 4. Since this probe has moderately repetitive sequences, one cannot exclude the possibility that it could be duplicated on chromosome 4 and 8.

This human DNA probe therefore shows cross-species homology to both mouse and Chinese hamster. This suggests that it could be a conserved mammalian sequence. As was mentioned above probe VC64 and VC85 have been shown to be linked. Both were independently assigned to chromosome 1 and also show some cross-species homology to mouse and Chinese hamster DNA.

The following probes have also been tentatively assigned. A number of discordances which will be discussed however do exist.

Probe VC28

This probe has been tentatively localized to chromosome 8 since it has hybridized to lines 2, 4, 5, 8 and 10 and to the human male control (figure 3.29). One discordancy which was present is that lines 8 and 10 have either chromosome 8 or 4 present respectively, but not both. This indicates that if chromosome 8 is present in line 10, it is at a lower frequency than could be detected cytogenetically. The probe has further been shown to be present in lines, 16 and 18 which both have chromosome 4 and 8 (table 3.2).

The probe has however not hybridized to lines 11, 12 and 17, all of which contain chromosome 4. Since this probe has moderately repetitive sequences, one cannot exclude the possibility that it could be duplicated on chromosome 4 and 8.

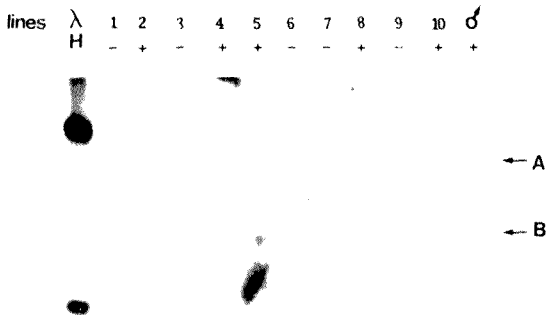


Figure 3.29. An autoradiograph of probe VC28 hybridized to Southern blot of a HindIII digest of somatic cell hybrid DNA. The probe is present (+) in lines 2, 4, 5, 8, and the human male control, indicated by the arrows A and B. The dark spots which are visible on the photograph are artifacts. There was no hybridization to RAG (mouse) or Chinese hamster control DNA (not shown).

It has also become apparent that lines 11 and 12 need more analysis because too many discrepancies have arisen with these lines as is illustrated by the assignment of probe VC63 to chromosome 4 (see below).

Probe VC63

This probe has been tentatively localized to chromosome 4 when the human EcoRI fragments were hybridized to the somatic cell hybrid panel blot. The first blot, containing lines 1 to 10 gave an indication localization of chromosome 4 since the probe hybridized to lines 2, 4, 5, 6, and 10. The assignment to 4 was not definite because the probe has also hybridized to line 6 and it only has a micromarker, which has been identified as either 4p or 12p. Further hybridization of this probe to a number of other lines has shown that it is present in lines 16, 17, 18 and 19 (table 3.2). This localization to chromosome 4 was not conclusive since the probe has not hybridized to lines 11 and 12, which both contain chromosome 4. These two lines are presently being re-analyzed but results are not yet available. No other chromosomes correspond to its hybridization pattern. Chromosome 12 and 22 have both been excluded because there are 5 and 6 discordant lines respectively.

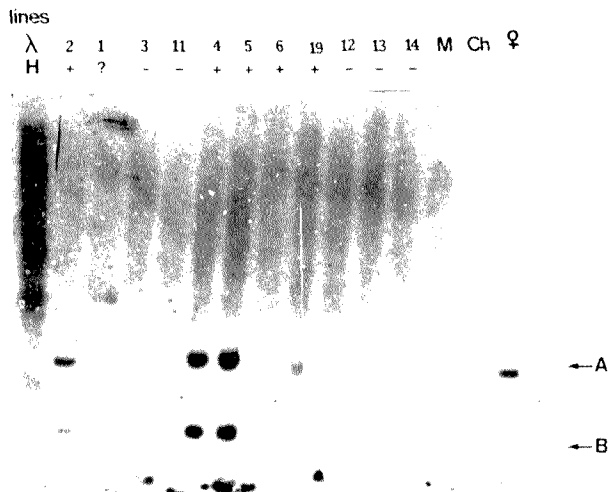


Figure 3.30. An autoradiograph of Southern blot of DNA from somatic cell lines digested with *Hind*III and hybridized to the 1.3 and 0.95 kb *Eco*RI fragments, of probe VC63. The probe is present in lines 2, 4, 5, 6, 19 and in the human female control, indicated by arrows A and B. There is no hybridization to mouse (M) or Chinese hamster controls (Ch).

Probe VC75

This probe has been tentatively localized to chromosome 7. The hybridization pattern of the first Southern blot containing lines 1 to 10, could not distinguish between chromosomes 7 and 20 (table 3.2). The probe hybridized to lines 6, 7, 8 and 10. The probe was then hybridized to other somatic cell hybrid lines which could distinguish between chromosome 7 and 20 (figure 3.31). These results indicated that the probe was present in lines 6, 11, 12, 14, 19, 22, 23 and 24 and therefore was more likely to be on chromosome 7.

A problem existed because there were three discordances in the hybridization pattern. Firstly, lines 11 and 12 cytogenetically show the presence of chromosome 7, but isozyme studies using the enzyme β -glucuronidase do not confirm the presence of chromosome 7 in these two lines. Secondly, line 14, does not appear cytogenetically or enzymatically to have chromosome 7. The cytogenetic data has been taken to be more accurate in this case because chromosome 7 was definitely present in lines 11 and 12 when they were re-evaluated. It has been concluded that this probe is more likely to be on chromosome 7 since it is present in 2 lines which definitely have chromosome 7 in a very high percentage and no chromosome 20 present.

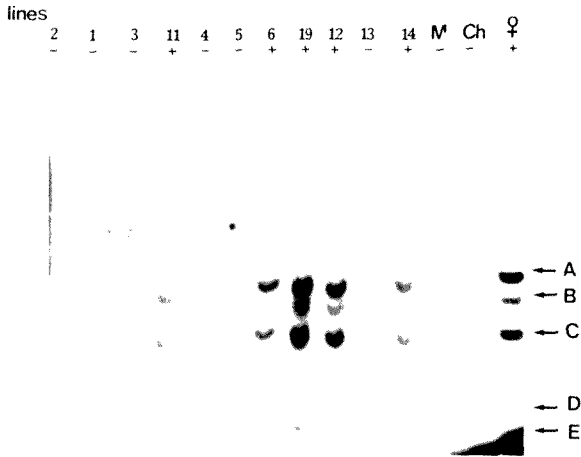


Figure 3.31. An autoradiograph of Southern blot of DNA from somatic cell hybrid lines hybridized to probe VC75. The probe is present in lines 6, 11, 12, 13, 14, 19 and human control as indicated by arrows A-E. There is no hybridization to the mouse (M) or Chinese hamster (Ch) DNA.

3.2.3 Summary

In this study, nine RFLPs have been detected using 7 cloned human DNA sequences. The provisional localizations of all the above RFLP-detecting human DNA sequences have been found to be on some of the largest eight chromosomes. These chromosomes represent 20,22% of human DNA that would have been present in the genomic library therefore these results are to be expected. Two probes have been found to map to chromosome 1. A summary of the RFLPs isolated, the sizes of their alleles and their provisional chromosomal localizations are summarized in Table 3.3 below.

TABLE 3.3. SUMMARY OF RESTRICTION FRAGMENT LENGTH POLYMORPHISMS DETECTED AND THEIR CHROMOSOMAL LOCALIZATIONS.

Restriction Enzymes	Probe Name	Number of alleles	Allele Sizes (kb)			Chromosome location	No. of Lines tested
			(A1)	(A2)	(A3)		
TaqI	VC61	2	12,5	10,1	--	2	24
	VC63	2	13,0	9,0	--	74p	24
	VC64	2	15,5	10,1	--	1	14
			5,4				
	VC28	2	4,8	3,6	--	8	24
HspI	VC14A	2	20,0	11,6	--	76p	20
	VC14B	2	14,0	12,7	--	76p	20
	VC28	2	5,0	3,9	--	8	24
	VC75	3	4,3	3,5	3,2	7	24
BamHI*	VC85	2	10,5	10,0	--	1	14

* - Many other enzymes can be used to show this deletion / insertion RFLP.

3.3 Characterization of unusual RFLP, detected by VC85

The unusual RFLP VC85 has been restriction mapped using some of the enzymes which showed the variation. Table 3.4 lists the human fragment sizes which were detected after hybridization to the probe VC85. In order to restriction map the probe, it was digested with several enzymes. A control of Charon 4A DNA was also digested with the same enzymes. This allowed for the identification of the bacteriophage fragments generated from the left and right arms as well as those which were still joined to the human insert could also be identified.

Figure 3.32 shows an ethidium bromide stained agarose gel of clone VC85 and Charon 4A digested with some restriction enzymes and table 3.5 lists these fragment sizes. The DNA in the gel in figure 3.32 was transferred to a nylon membrane, by the method of Southern, and was been hybridized to each of the 7 human EcoRI fragments from VC85. These EcoRI fragments have also been hybridized to a Southern blot containing the DNA of one individual digested with fourteen different restriction enzymes. These results helped in the ordering of the human fragments along the sequence of DNA.

TABLE 3.4. HUMAN FRAGMENT SIZES (kb) DETECTED BY PRIME VCBS AFTER SOUTHERN BLOT ANALYSIS

RESTRICTION ENZYMES							
BamHI	BglII	HindIII	EcoRI	PvuII	StuI	SacI	TaqI
14,0	10,3	10,0	18,0	7,0	<u>5,2</u>	9,5	17,0* 11,0
<u>10,5</u>	<u>6,7</u>	<u>5,3</u>	3,4	5,0	<u>5,6</u>	8,5	2,4 5,8
<u>10,0</u>	<u>6,2</u>	<u>4,8</u>	<u>2,6</u>	<u>4,2</u>	5,0	<u>6,05</u>	2,6
5,2	3,9	3,0	<u>2,1</u>	<u>3,7</u>	3,85	5,5	<u>1,5</u>
	1,0	1,9	1,85	2,0	3,1	2,8	<u>1,6</u>
			1,3	1,5	1,3	<u>2,65</u>	1,2
			1,2	1,3	1,0		
			1,1				

Underlined fragments show variation when hybridized to probe VCBS.

* - No variation detected due to lack of fragment separation.

** - Fragment length difference is not 500 bp, see text.

TABLE 3.5. DNA FRAGMENT SIZES FOUND IN PHAGE CLONE VCBS

MspI	EcoRI	BamHI	HindIII	BglII	B+Bg	E+H	E+Bg
R8,0	J50,7	L19,0	L19,5	L19,0	14,0	19,0	5,8
5,5	L19,94	8,0	7,0	C4,8	6,0	5,5	3,3
2,7	R11,04	J7,0	J5,7	R4,0	5,1	5,0	1,8
L1,3	3,3	C5,56	4,2	3,9	5,0	2,1	1,3
	2,1	C0,5	3,0	3,6	4,5	1,7	1,2
+SF	1,85	R8,0	1,65	p3,47	4,0	1,5	1,1
	0,85	C3,93		C2,16	2,0	1,4	1,0
	0,75	C1,35		J1,7	1,1	1,2	0,9
	0,45	C0,5		C1,31	1,0	1,1	
				0,36	0,20	0,5	
				0,06		0,6	
						0,4	

C - phage sequences, L - left phage arm + human sequences.
J - phage arms, joined at their cos sites, B - human sequences + right phage arm, p - partial digests.
Those with no code are human fragments, SF - other smaller fragments.

The restriction map of VC85 is schematically illustrated in figure 3.33. The slight fragment size differences between the internal fragments of the phage VC85 and human genomic fragments, are due to differences in the migration of the DNA fragments and the amount of DNA present in the gel. Molecular weight markers in the genomic blots were placed in the outer lanes and independently hybridized to Lambda DNA since the Charon 4A derivative does not detect all of the wildtype Lambda DNA fragments.

Some of the fragments detected after digestion of phage VC85 would contain portions of the phage arms as indicated in table 3.5. The 2.1 kb human EcoRI fragment from the clone has been isolated and subcloned and it detects the RFLP with all the enzymes used. The human fragment which was cloned into VC85 represents the allele which does not have the 500 bp region. This could be deduced from the EcoRI and the MspI restriction patterns in the clone.

After restriction mapping of VC85 the unusual MspI and TaqI restriction patterns could be explained as being due to the presence of restriction sites for both these enzymes within the deleted/inserted region. Those individuals who were heterozygotes for the insertion/deletion were also heterozygotes for the MspI and TaqI RFLPs. The differences in the fragment sizes generated by these enzymes did not correspond to 500 bp.

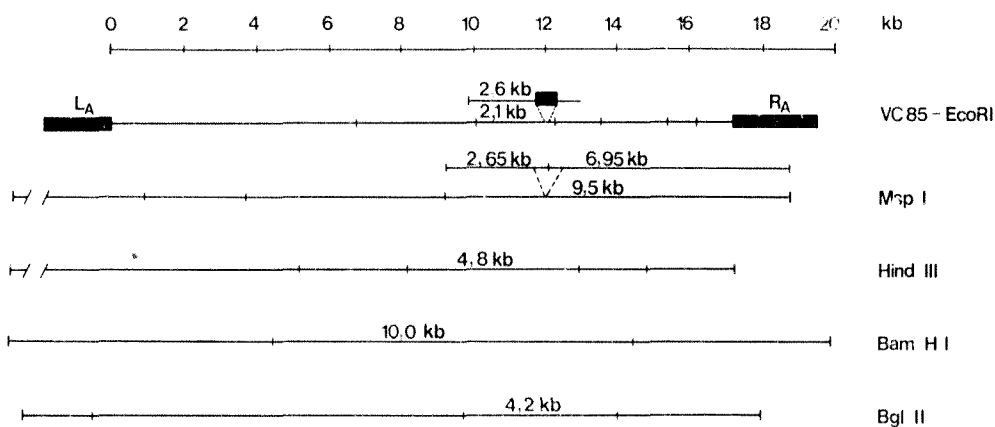


Figure 3.33. A partial restriction map of VC85. The human fragments sizes correspond to those of the smaller allele, and are given in Table 3.4. The order of each human fragment is given with respect to the EcoRI fragments of the clone and the position of the deletion/insertion region. R_A and L_A represent the right and left arms of the phage. The position of the *Msp*I site within the inserted/ deleted region is indicated.

The unusual pattern generated using MspI

An autoradiograph of VC85 hybridized to a Southern blot containing DNA from random Caucasoid individuals digested with MspI is shown in figure 3.34. Those individuals, previously shown to be homozygous for the absence of the deleted/inserted region, have a 9,5 kb MspI fragment in addition to constant fragments. Individuals who are homozygous for the presence of the deleted/inserted region have two fragments of 6,9 and 2,65 kb. Individuals previously shown to be heterozygotes have all 3 MspI fragments. A schematic representation of the MspI restriction pattern is illustrated in figure 3.35. The presence of an MspI site within the 500 bp insertion/deletion region would account for the two fragments in individuals who have the presence of the inserted/deleted region. Those individuals who lack this region are homozygous for absence of this region, therefore lack the MspI sites and therefore have a larger 9,5 kb fragment.

The unusual pattern generated by TaqI

In figure 3.36, the variation in fragment sizes after TaqI digestion and hybridization to VC85 can be seen. The difference in fragment length is only 100 bp. Those individuals who are homozygous for absence of the deleted/inserted region have the larger 1,5 kb TaqI fragment while individuals homozygous for the presence of this region have a smaller 1,4 kb fragment. Constant fragments are also visible.



Figure 3.34. An autoradiograph of a Southern blot of DNA from random Caucasoid individuals, digested with *MspI*, and hybridized to probe VC85. The varying fragments are 9.5, 6.9 and 2.65 kb, and the constant fragments (C) are 8.5, 5.5 and 2.85 kb in length. Individuals in lanes 1, 2 and 7 are homozygous for absence of the deletion/insertion region, in lanes 5 and 6, they are heterozygous at this locus, and in lanes 3 and 4 are homozygous for the presence of this region. In lane 2 the fainter larger fragments are due to partial digestion with the restriction enzyme.

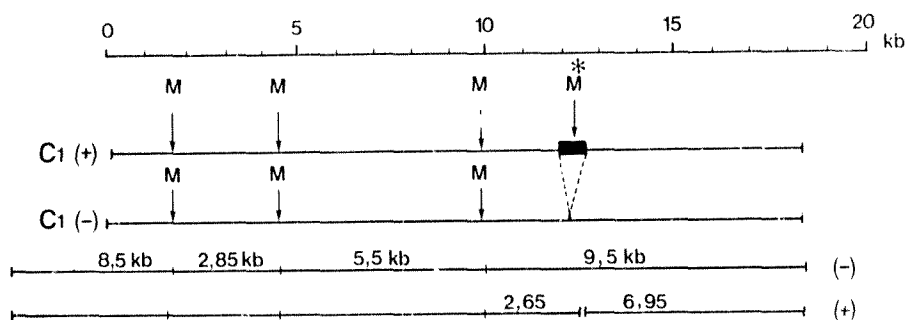


Figure 3.35. Restriction map showing *MspI* sites in the human genome detected by probe VC85. ■ represents the 500 bp deletion/insertion region containing an *MspI* restriction site. C1(+) indicates that the region is present on the chromosome while C1(-) indicates its absence. The sizes of the human fragments which are detected on a Southern blot are shown.

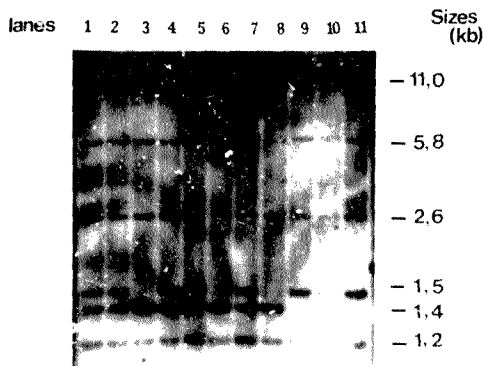
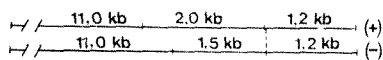
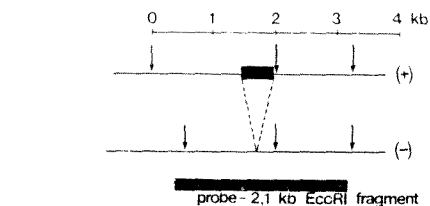


Figure 3.36. An autoradiograph of a Southern blot of DNA from random Caucasoid individuals, digested with *TagI* and hybridized to probe VC85. Variation, corresponding to the deletion/insertion RFLP, is between fragments of 1,5 and 1,4 kb. Individuals with the 1,5 kb fragment are homozygous for absence of the region and those individuals who are homozygous for the presence of this region have a 1,4 kb fragment.

The 2,1 kb EcoRI fragment, isolated from the bacteriophage clone, was hybridized to a blot containing DNA from random individuals digested with TagI. This fragment hybridized to the 11,0; 1,5; 1,4 and 1,2 kb TagI fragments. A schematic representation of the TagI pattern is shown in figure 3.37. A 600 bp fragment which could form part of the polymorphism has not been detected and either has not transferred correctly or there is another TagI site within the region. This explanation allows the TagI restriction result to fit into the deletion/insertion RFLP pattern.

a) Expected.



b) Hypothesis from observations.

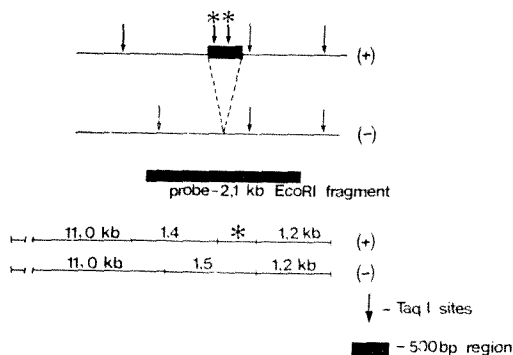


Figure 3.37. Schematic representation of the TaqI restriction pattern of VC85 around the deletion/insertion region. The expected pattern is shown in a). In b) the postulated 600 bp fragment indicated by a (*) was not visible using a high percentage gel and a corresponding explanation would be, that more than one TaqI site is present within the deleted/inserted region and therefore the fragments which will be generated are smaller than 600 bp and were not detected on a Southern blot.

3.4 Population Studies

When an RFLP had been found in the Caucasoid population the frequencies of the alleles were determined in the Caucasoid and two other southern African Populations, namely the Southern African Bantu-speaking Negroids and San. The C.E.P.H. study was also used as a source of data for the frequencies of the RFLPs in a large number of Caucasoid individuals. The C.E.P.H. families consisted of large Utah (Mormon), French and Venezuelan families. It had been agreed, at the C.E.P.H. meeting in Paris, September 1986, that the population data on Caucasoid individuals could be obtained by pooling data on the random individuals from the French and Utah families. This would probably be representative of a Western European population. The data from the Southern African Caucasoids could probably be pooled with this group but has been kept separate in the present study.

Population data was obtained by determining the genotypes of all random individuals with all the RFLP-detecting probes identified by this study. The following calculations were performed:

- i. Allele frequencies were calculated according to the accepted gene counting methods:

$$p = \frac{\text{number of chromosomes of a particular allele size}}{\text{number of chromosomes tested}}$$

$$q = 1 - p$$

where p represents the allele frequency of allele A1 and q the allele frequency of allele A2.

- ii. The polymorphic information content (PIC) giving an indication of the number of informative matings was calculated using the following equation.

$$PIC = 1 - p^2 - q^2 - 2p^2q^2$$

When there are more than two alleles the PIC value is calculated according to the expanded basic binomial equation from above. For two-allele systems the maximum PIC value is 0.38. As the number of alleles increase, the PIC value increases towards 1.

- iii. The standard errors (SE) for the above allele frequencies were calculated according to the following formula:

$$SE = \sqrt{(p(1-p)/N)}$$

where N = the number of chromosomes studied

The results of the population studies are listed in table 3.6. The frequencies of both alleles are given since the rarer allele in one population was sometimes the more common allele in another population.

- iv. The allele frequencies between the populations were compared using two by two contingency tables and the computer programme Epistat. The results and p-values are given in table 3.7. The allele frequencies in the populations have been tested and found to be in Hardy-Weinburg equilibrium.
- v. The population data has also been analyzed by computing the genetic distances between the different population groups based on the allele frequencies generated at each polymorphic site. The results of this analysis are listed in table 3.8 and a dendrogram is illustrated in figure 3.38.

In summary the population results show that nearly all the RFLPs isolated, are polymorphic in all three populations studied and the allele frequencies in the different populations show a number of significant differences between population groups.

TABLE 3.6. ALLELE FREQUENCIES OF DNA SEQUENCE POLYMORPHISMS IN VARIOUS POPULATIONS

Probe name	Restriction Enzyme	Allele Size (kb)	Allele frequencies in the different population groups studied										CEPH ³ Caucasoids	
			South African Caucasoids	PIC ¹	(N) ²	South African Negroids	PIC	(N)	San		PIC	(N)		(N)
VC14	HspI	(A1) 20,0	0,125 ± 0,06	0,20	(32)	0,013 ± 0,02	0,02	(40)	---	---	---	(40)	0,096 ± 0,03	0,16 (104)
		(A2) 11,6	0,875 ± 0,06			0,987 ± 0,02			1,00				0,904 ± 0,03	
	HspI	(B1) 14,0	0,875 ± 0,06	0,20	(32)	0,592 ± 0,06	0,36	(76)	0,742 ± 0,05	0,30	(62)		0,848 ± 0,06	0,23 (66)
		(B2) 12,0	0,125 ± 0,06			0,408 ± 0,06			0,258 ± 0,05				0,152 ± 0,06	
VC28	HspI	(A1) 5,0	0,500 ± 0,08	0,38	(36)	0,550 ± 0,07	0,37	(40)	0,066 ± 0,03	0,12	(76)		0,433 ± 0,06	0,37 (120)
		(A2) 3,9	0,500 ± 0,08			0,450 ± 0,07			0,934 ± 0,03				0,567 ± 0,06	
	TaqI	(B1) 4,8	0,805 ± 0,07	0,26	(36)	1,00	---	(40)	1,00	---	(40)		0,922 ± 0,01	0,13 (116)
		(B2) 3,6	0,195 ± 0,07			---			---				0,078 ± 0,04	
VC61	TaqI	(A1) 12,5	0,472 ± 0,08	0,37	(36)	0,510 ± 0,05	0,38	(102)	0,595 ± 0,06	0,37	(74)		0,528 ± 0,06	0,37 (116)
		(A2) 10,1	0,528 ± 0,08			0,490 ± 0,05			0,405 ± 0,06				0,472 ± 0,06	
VC63	TaqI	(A1) 13,0	0,238 ± 0,07	0,29	(42)	0,321 ± 0,05	0,34	(76)	0,692 ± 0,05	0,34	(70)		0,147 ± 0,05	0,22 (116)
		(A2) 9,0	0,762 ± 0,07			0,679 ± 0,05			0,308 ± 0,05				0,853 ± 0,05	
VC64	TaqI	(A1) 15,5	0,778 ± 0,07	0,29	(36)	0,590 ± 0,05	0,37	(100)	0,475 ± 0,05	0,37	(120)		0,819 ± 0,04	0,25 (116)
		(A2) 10,1; 5,4	0,222 ± 0,07			0,410 ± 0,05			0,525 ± 0,05				0,181 ± 0,04	
VC75	AvaII	(A1) 4,9	0,211 ± 0,07	0,28	(38)	0,553 ± 0,05	0,37	(94)	0,717 ± 0,05	0,33	(92)		0,180 ± 0,02	0,43 (116)
		(A2) 3,5	0,789 ± 0,07			0,447 ± 0,05			0,283 ± 0,05				0,690 ± 0,02	
		(A3) 3,2	---			---			---				0,150 ± 0,02	
VC85	BamHI	(A1) 10,5	0,522 ± 0,07	0,38	(46)	0,202 ± 0,04	0,26	(84)	0,171 ± 0,04	0,24	(76)		0,400 ± 0,06	0,36 (120)
		(A2) 10,0	0,478 ± 0,07			0,798 ± 0,04			0,829 ± 0,04				0,600 ± 0,06	

1. PIC - polymorphism information content.

2. N = the number of chromosomes studied.

3. CEPH - Caucasoids are combined data of Utah and French parents.

4. - excluding one variant.

TABLE 3.7. STATISTICAL ANALYSIS OF GENOTYPE DATA BETWEEN POPULATIONS

Probe name	Caucasoids vs Negroids	Caucasoids vs San	Negroids vs San	S.A. Cauc. vs C.E.P.M. Caucasoids
	χ^2 (p)	χ^2 (p)	χ^2 (p)	χ^2 (p)
VC14A	2,24 0,32 ---	---	---	3,09 0,24
VC14B	11,60 * 6,50 0,04	2,86 0,24	5,76 0,06	
VC20M	0,10 0,95 1,60 **	23,20 **	0,71 0,70	
VC28T	---	---	---	3,21 0,07
VC61	0,24 0,29 0,04 0,98	1,99 0,37	1,76 0,41	
VC63	14,50 ** 56,00 **	21,89 **	2,75 0,25	
VC64	13,02 ** 13,02 **	3,06 0,22	1,05 0,59	
VC75	34,80 ** 57,40 **	6,24 0,04	4,81 0,44	
VC85	16,60 ** 18,80 **	1,90 0,39	5,34 0,07	

** p-value < 0,01 : result is significant

p > 0,05 : no significant difference

0,01 < p < 0,05 need to increase sample size before conclusion is made.

 χ^2 : Chi squared value.

TABLE 3.8. GENETIC DISTANCES BETWEEN POPULATION

POPULATIONS	S.A. Negroids	San	C.E.P.H. Caucasoid
S.A. Caucasoids	77	157	23
S.A. Negroids	---	99	71
San	---	---	149

Cluster Analysis

S.A. vs C.E.P.H. Caucasoids 23,36

S.A. Negroids vs San 59,27

S.A. Caucasoids vs S.A. Negroids 70,54

A dendrogram has been constructed using the above values and can be seen in figure 3.38.

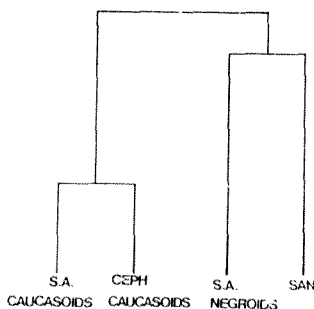


Figure 3.38. Dendrogram indicating the genetic distance between the four population groups studied, based their allele frequencies of the RFLPs, detected using seven arbitrary DNA probes. The allele frequencies and genetic distances are listed in table 3.6 and 3.8 respectively.

3.5 Estimates of genetic heterozygosity

The degree of polymorphism in the human genome can be estimated using the information generated from restriction endonuclease data. A formula was developed by Ewens *et al.*, (1981) based on population genetics theory to estimate genetic variation at the DNA level. The one assumption which is made is that if a restriction site is lost it is not due to changes in more than one nucleotide. The equation is extended to include restriction endonucleases that have four and six base pair recognition sites.

For this study the enzyme *AvaII* recognizes 5 base pairs and thus has also being included in the equation. In the equation k_4 , k_5 and k_6 represent the number of cleavage sites which show polymorphism in a population when restriction enzymes recognizing four, five and six base pair sequences are used. The total number of cleavage sites for all restriction enzymes used are indicated by m_4 , m_5 and m_6 . From table 3.9. the minimum number of sites identified by each probe/enzyme combination (number of fragments + 1), the number of polymorphic sites and the number of individuals screened are listed.

The following equation was used to calculate genetic heterozygosity:

$$\frac{k_4 + k_5 + k_6}{8m_4 + 10m_5 + 12m_6}$$

TABLE 3.2 ESTIMATION OF GENETIC HETEROGENEITY AT THE Loci IDENTIFIED BY THE PROBES ISOLATED IN THIS STUDY (See text for explanation).

Enzyme	Probes						No. of nucleotides recognized
	VC14	VC28	VC61	VC63	VC64	VC75	VC85
AvaII	---	5 (7)	9 (7)	---	6 (7)	8 (7)*	6 (7)
BamHI	---	2 (7)	3 (7)	3 (7)	3 (7)	3 (7)	3 (7)
BglII	4 (7)	6 (7)	4 (7)	5 (7)	4 (7)	6 (7)	5 (7)
EcoRI	6 (7)	3 (7)	7 (7)	3 (7)	2 (7)	4 (7)	5 (7)
HindIII	7 (7)	8 (7)	7 (7)	---	4 (7)	4 (7)	9 (7)
HindIII	6 (7)	4 (7)	4 (7)	4 (7)	4 (7)	---	6 (7)
HhaI	---	9 (7)	6 (7)	3 (7)	7 (7)	---	8 (7)
HpaI	3 (7)*	6 (7)*	4 (7)	2 (7)	5 (7)	6 (7)	5 (7)
PstI	4 (7)	4 (7)	6 (7)	---	6 (7)	6 (7)	5 (7)
PvuII	---	6 (7)	4 (7)	3 (7)	---	6 (7)	7 (7)
RsaI	---	7 (7)	---	---	7 (7)	---	6 (7)
SstI	5 (7)	4 (7)	3 (7)	---	---	2 (7)	3 (7)
StuI	2 (7)	7 (7)	3 (7)	2 (7)	4 (7)	3 (7)	6 (7)
TaqI	6 (7)	5 (7)*	4 (7)*	3 (7)*	5 (7)*	4 (7)	6 (7)

Total no. of sites

ma	9	28	14	0	24	11	25	2m+12n
m5	0	5	0	0	6	9	7	2m5+36
mo	35	49	50	20	27	51	51	2m6+283
k4	2	2	1	1	1	0		2k4+8
k5	0	0	0	0	0	1	1	2k5+1
k6	0	0	0	0	0	0		2k6+0

Site difference frequency

$\frac{k4+k5+k6}{8m4+10m5+12m6}$	$\frac{2}{492}$	$\frac{2}{862}$	$\frac{1}{694}$	$\frac{1}{312}$	$\frac{1}{648}$	$\frac{1}{790}$	$\frac{1}{847}$	
	0.004	0.0023	0.0014	0.0032	0.0015	0.0013	0.0012	

Average heterozygosity

$$\frac{2k4 + 2k5 + 2k6}{8m4 + 10m5 + 12m6} = \frac{9}{4716} = 0.002$$

Numbers in brackets refer to number of individuals in the initial screening study
(*) indicates the presence of an RFLP.

TABLE 3.9 ESTIMATION OF GENETIC HETEROGENEITY AT THE LDC1 IDENTIFIED BY THE PROBES ISOLATED IN THIS STUDY (See text for explanation).

Enzyme	VC14	VC2B	VC61	Probes VC63	VC64	VC75	VC85	No. of nucleotides recognized
Avall	---	5 (7)	9 (7)	---	6 (7)	8 (7)*	6 (7)	5
BamHI	---	2 (7)	3 (7)	3 (7)	3 (7)	3 (7)	3 (7)	6
BglII	4 (7)	6 (7)	4 (7)	5 (7)	4 (7)	6 (7)	5 (7)	6
EcoRI	6 (7)	3 (7)	7 (7)	3 (7)	2 (7)	4 (7)	5 (7)	6
HincII	7 (7)	8 (7)	7 (7)	---	4 (7)	4 (7)	9 (7)	6
HindIII	6 (7)	4 (7)	4 (7)	4 (7)	4 (7)	---	6 (7)	6
MboI	---	9 (7)	6 (7)	3 (7)	7 (7)	---	8 (7)	4
MspI	3 (7)*	6 (7)*	5 (7)	2 (7)	5 (7)	6 (7)	5 (7)	4
PstI	4 (7)	4 (7)	6 (7)	---	6 (7)	6 (7)	5 (7)	6
PvuII	---	6 (7)	4 (7)	3 (7)	---	6 (7)	7 (7)	4
RsaI	---	7 (7)	---	---	7 (7)	---	6 (7)	4
SatI	5 (7)	4 (7)	3 (7)	---	---	2 (7)	3 (7)	6
SmaI	2 (7)	7 (7)	3 (7)	2 (7)	4 (7)	3 (7)	6 (7)	6
TaqI	6 (7)	5 (7)*	4 (7)*	3 (7)*	5 (7)*	4 (7)	6 (7)	4
Total no. of sites								
m4	9	28	14	9	24	11	25	Dm4=120
m5	0	5	9	0	6	9	7	Dm5=36
m6	35	49	50	20	27	51	51	Dm6=283
Site difference frequency								
$\frac{k4+k5+k6}{8m4+10m5+12m6}$	$\frac{2}{492}$	$\frac{2}{862}$	$\frac{1}{694}$	$\frac{1}{372}$	$\frac{1}{648}$	$\frac{1}{790}$	$\frac{1}{847}$	
	0.004	0.0023	0.0014	0.0032	0.0015	0.0013	0.0012	
Average heterozygosity								
$\frac{Dm4+Dm5+Dm6}{8m4+10m5+12m6} \times \frac{2}{4716}$	= 0.002							

Numbers in brackets refer to number of individuals in the initial screening study.
(*) indicates the presence of an RFLP.

For the probe VC85 which detected a deletion/insertion type RFLP it has been included as only one polymorphic site. The overall estimate of genetic heterozygosity was calculated to be 0.002. This estimate was based on the average number of individuals screened which was usually seven. Further this estimate has not included the data on 6 other sequences which did not detect any variation because the probes gave high background and results were not always easily interpretable. If these were included they would decrease this estimate even further.

3.6 Linkage study-preliminary results

The C.E.P.H. collaborative study was started in July 1986, when the first DNA samples arrived from Paris. The samples consisted mostly of the parents of the large families and some of their children. The DNA of the parents was typed with 7 of the RFLP-detecting probes (VC14, VC28, VC61, VC63, VC64, VC75, and VC85). Family studies were done in informative families, where at least one parent was heterozygous and where there were at least four children. Eight such families were available in our laboratory. The results for 6 of the RFLPs were placed into a computer database, which was submitted for the construction of the first database to the organizers in January 1987. These results have been summarized in table 3.10.

The only linkage studies which were possible were between the probes that had been isolated in this Department. This was because no data on the protein polymorphisms or any DNA markers on the families were available.

TABLE 3.10. SUMMARY OF RESULTS OF THE C.E.P.H. STUDY

Probe	Enzyme	No. of parents Typed	No. of informative families	No. of families typed
VC14A	HspI	62	12	3
VC14B	MspI	43	12	3
VC28M	MspI	70	29	7
VC28T	TaqI	68	9	1
VC61	TaqI	68	24	2
VC63	TaqI	68	24	3
VC64	TaqI	68	23	6
VC75	AvaII	67	28	6
VC85	StuI*	72	31	25

* - This probe can be typed with many enzymes.
C.E.P.H. - Centre for the Study of Human Polymorphism.

Linkage analysis was performed, using the programme LINKAGE between VC64 and VC85 since both these probes had been assigned to chromosome 1. These probes were found to be linked with a LOD score of 6,2 at θ of 0,0. No recombinants were observed. This result was obtained on 6 informative families consisting of 60 individuals.

3.7 Summary

This chapter has reported all the results obtained in this study. Seven of the probes found in the Caucasoid population were also polymorphic in the Bantu-speaking Negroid and San populations. Five of the RFLPs found in this study have rarer allele frequencies greater than 0,2 in all the population groups. The highest PIC value (0,43) was obtained for VC75 in the Caucasoid population due to the discovery of a third allele. The probe VC85 has been partially restriction mapped because it represents an unusual type of DNA sequence variation. This probe, VC85, was also found to be linked to a second arbitrary probe, VC64. All of the probes in this study have been provisionally assigned to chromosomes. Six of the RFLPs have already been placed into the C.E.P.H. programme and it is hoped that some will be found to be linked to other genetic markers.

CHAPTER FOUR

DISCUSSION

4 CHAPTER FOUR - DISCUSSION

The main aim of this study was to isolate and characterize DNA sequences detecting RFLPs in the populations of Southern Africa.

4.1 The Polymorphism Concept

Polymorphism is not a new concept and has been shown to characterize many soluble proteins and to be useful as genetic markers (Harris and Hopkinson, 1972). The term "polymorphism" was given to variation present in a population, in which the frequency of the rarer allele was at least one percent (Bodmer and Cavalli-Sforza, 1976). Human and other natural populations were studied for the presence of enzyme variants in the 1960s and it was concluded that polymorphism was a general phenomenon although a few proteins were only polymorphic in specific populations (Harris, 1966; Lewontin and Hubby, 1966; Harris and Hopkinson, 1972). It appears that some protein polymorphisms which occur in some populations may give a clue as to the selective processes producing the polymorphisms. For example the sickle-cell variant of haemoglobin appears to have been selected for in hyperendemic malarial regions (Allison, 1954).

Classical genetic markers, which are usually due to altered protein products have already been used as markers on particular chromosomes. However, due to the conservation of coding regions, the frequency of the variant alleles at such

loci is usually low. Only 4 out of 29 classical blood group and protein markers were highly informative with a PIC value greater than 0,5 and a further 15 were reasonably informative with a PIC value between 0,25 and 0,50 (Botstein et al., 1980). These classical genetic markers were detected because a DNA mutation had led to a change in an amino-acid which had altered the size or the overall charge of the polypeptide chain.

Based on data from protein studies any individual was likely to be heterozygous at 6% of his loci (Harris and Hopkinson, 1972). This estimate was later revised to 12 to 13% when improved techniques for detection of protein variation were also considered (Neel, 1984).

The analysis of variation, at the DNA level, using restriction endonucleases, has allowed for a new level of study of genetic polymorphism. Restriction enzymes and their site-specific cleavage of DNA sequences have become important tools for the determination of DNA sequence variation within and around the coding regions of genes. The first estimates of human heterozygosity were made by Jeffreys (1979) and he concluded that 1 in every 100 base pairs varied between normal individuals. His data was re-analyzed by Ewens et al., (1981) and the estimate was amended to 1 heterozygote base pair per 200 nucleotides. A more recent estimate of genetic heterozygosity, based on the variation detected by random DNA sequences, of presently unknown function, is that one in every 300 bp could vary between normal individuals (Cooper et al., 1985).

Estimates of heterozygosity from this study, using the formula proposed by Ewens et al., (1981) is that one in every 500 base pairs varies between random individuals. These sequences are of unknown function and therefore the effects of selection, if any, are unknown. These results are slightly lower than those mentioned above. This could be due to the fact that initially only 7 individuals were screened using 14 different enzymes and therefore any variation present at a frequency of less than 0.15 might not have been detected. In addition, if the fragment length variation was below the limit of resolution or if small restriction fragments were not detected, these would account for a lower estimate of heterozygosity. The rearrangement detected by VC85 has been accounted for as a single mutation. MspI and TaqI sites appear to be present within this region but these sites are not polymorphic. Furthermore, this method of estimation does not account for the number of individuals screened. Larger numbers of individuals using fewer enzymes had been screened by Jeffreys (1979) and also by Cooper et al., (1985). Genomic DNA sequence variation was observed during the characterization of β -like globin genes and the first human RFLP was discovered, in American Blacks, 3' to the human β -globin gene, using the restriction enzyme HpaI (Kan and Dozy, 1978a). This HpaI RFLP has only been found in Negroid populations and appears to have increased in frequency due to selection for the $Hb\beta^S$ gene in hyperendemic malarial regions of Africa. Since 1978, many more RFLPs have been found using cloned genes or single-copy DNA sequences as probes (Cooper and Schmidtke, 1986).

Are RFLPs randomly distributed throughout the genome?

From the data on proteins and DNA sequences it became apparent that RFLPs were not distributed evenly throughout the genome but they appeared to be clustered in the non-coding regions of the genome. Evidence has been provided by studies around the human β -globin-like cluster, where out of 12 RFLPs detected only one was within a coding sequence (Kazazian *et al.*, 1983). Furthermore, in the 16,5 kb region containing the $\delta\beta$ -globin genes, 0,22% of the nucleotides were polymorphic sites within exons compared with 0,58% in other regions (Poncz *et al.*, 1983). It was expected that, based on estimates of heterozygosity, at least 6 RFLPs should have been detected within exons of the β -globin-like gene cluster.

The finding that some forms of thalassemia were due to mutations within non-coding regions of the genome further emphasized the non-randomness of distribution of RFLPs (Kazazian *et al.*, 1983). A study in *Drosophila* showed that only one out of 43 polymorphic sites was found to be within a sequence coding for an aminoacid, in 11 cloned alcohol dehydrogenase genes. All the others were distributed in the non-coding regions around the gene (Kreitman, 1983). The non-randomness of distribution of RFLPs reduced the disproportion that was observed between heterozygosity at the DNA and protein level. This uneven distribution of sequence variation demonstrated that although RFLPs are generated by random mutation throughout the genome, selection is acting on those mutations in the genome which cause unfavourable alterations in coding sequences. There

is a need to unravel the complex interaction between the patterns of mutation and patterns of selection and it has been suggested that the answers may lie in the understanding of the nature of mutation at the DNA level (Neel, 1984).

Some polymorphisms are undoubtedly due to random genetic drift (Kimura, 1969a, 1969b; Crow, 1973). The variation present in the population represents a balance between the generations of new alleles by mutation and the elimination of others by chance or random genetic drift. One special case of random genetic drift is known as the "founder effect". A founder generation may have by chance a different pool of genes from that of the parent population, and therefore some of the alleles carried by the founders may not be exactly representative of the original population. If one or more of the founders was by chance heterozygous for an allele, which was rare in the original population, this allele may become relatively common in the new population. The polymorphic frequencies of the deleterious gene for familial hypercholesterolaemia (FH), which is unusually common in the Afrikaaner population in South Africa, has been attributed to the founder effect. The carrier rate for this gene has been estimated to be about 1 in 100. If one or two of the original founders possessed the gene for FH, its presence can be attributed to the founder effect or random genetic drift which occurred due to the small sample size of the original population (Jenkins et al., 1980).

4.2 The Isolation of Single-Copy Regions of the Genome

The human genome is organized, like other eukaryotic genomes, into highly and mid-repetitive sequences interspersed with single-copy regions (Schmid and Deininger, 1975). The single-copy regions contain genes, which are transcribed into proteins, and other sequences which currently are of unknown function. These single-copy sequences were recognized as potential genetic markers when such a region, of unknown function, was isolated from a genomic library and used as a probe to detect variation between normal individuals (Wyman and White, 1980). A highly polymorphic locus was identified, where 75% of individuals screened were found to be heterozygotes. The majority of the RFLPs isolated by 1984 indicated that such highly polymorphic loci were not found to be characteristic of the human genome (Skolnick et al., 1984). This will be discussed further below.

In 1980, Botstein et al., proposed the construction of a human genetic linkage map using polymorphic DNA markers. These markers would consist of cloned DNA sequences, which may or may not code for a gene, but which detected variation in the genome. Each clone would define a genetic marker at a unique locus on a chromosome and could be used for the study of linkage to other DNA or protein markers. It was proposed that 150 evenly-spaced highly polymorphic markers would be needed to cover the human genome in order to find linkage with any unknown genetic disease locus. This number was revised to 1 500 because many of the markers would not meet the optimum requirements set up by Botstein et al.,

(1980) including PIC values greater than 0,5 and distances between the markers of 0,20 cM and distributed randomly across the genome (Lange and Boehnke, 1982). Many more RFLPs would therefore need to be isolated before a "saturated" map could be obtained (White et al., 1985).

Strategies used to isolate single-copy regions to detect RFLPs

The present study has isolated single-copy human DNA clones from a human HaeIII/AluI genomic library, which contained inserts of between 7 and 20 kb in bacteriophage Charon 4A. The library was screened for unique sequences by means of hybridization to radiolabelled total human DNA and those clones which did not appear to hybridize to the human DNA were isolated. Two percent of the clones which were screened contained potential single-copy human sequences. However restriction enzyme analysis of the DNA, which was isolated from these clones, indicated that 73% of these clones did not contain any human DNA but only the parental Charon 4A EcoRI restriction pattern was present.

It seems that during the library construction the two internal EcoRI fragments (7,8 and 6,9 kb) of the phage, which form the "stuffer region", were not completely separated from the 30,7 kb fragment, which represents the left and right phage arms joined at their cohesive ends. The Charon 4A phage genome was either reconstructed during ligation or the phage vector DNA had not been completely digested prior to separation of the phage arms (Maniatis et

al., 1978). An alternative explanation would be that over time those Charon 4A phage lacking human inserts have become amplified in the library.

One could probably have avoided isolation of these Charon 4A phage by selective screening for mutant β -galactosidase bacteriophage. Those phage which contained human inserts would have lost the β -galactosidase gene function and clear plaques instead of blue would have been detected when using IPTG and Xgal in the plating medium (Blattner *et al.*, 1977).

The DNA of the potential single-copy human clones was digested with *EcoRI* which identified those containing human inserts and also the number of internal *EcoRI* sites present in the sequence because the library had been constructed into the Charon 4A *EcoRI* sites. Furthermore, if the clone represents a single-copy region, all but two of these human *EcoRI* fragments in the phage clone, should correspond to the hybridization pattern of this clone to a Southern blot of human *EcoRI* digests. The additional two fragments represent the outer fragments of the insert and would be larger on this genomic blot, because the human DNA library was prepared from *HaeIII*/*AluI* digests. The true genomic *EcoRI* site is 5' and 3' to the cloned sequence (Lawn *et al.*, 1978; Feder *et al.*, 1985).

Southern blot analysis was used to determine whether any of the internal human fragments contained repetitive sequences. Despite the stringent screening procedures, some of these clones were still found to contain repetitive sequences. The final percentage of single-copy sequences

isolated was about 0,6% . This result is less than that reported by Botstein et al., (1980) (1 to 3%) and Kao et al., (1982) (1%) but greater than that reported by Feder et al., (1985) (0,15%). This difference could be due to the different restriction enzymes used in the construction of the genomic libraries.

The higher percentages of single-copy clones were obtained from libraries where the human DNA was partially digested with EcoRI while the lower percentages were obtained from libraries constructed using HaeIII/AluI. The repetitive family known as Alu is known to contain a recognition site for the restriction enzyme AluI and a number of the clones would therefore contain part of this sequence in the flanking regions. Alu sequences were estimated by Crampton et al., (1981) to be present in 94% of the clones isolated from the Maniatis human genomic library. This could account for the lower percentage of single-copy probes isolated in the present study. The human insert sizes ranged from about 7 to 17 kb and therefore from estimates of the distribution of repetitive sequences it is not unexpected that a low percentage of single-copy sequences were found (Schmid and Deininger, 1975).

Newer strategies to isolate RFLPs

During the initial screening procedures only those clones which were free from repetitive sequences were used to search for RFLPs. New strategies using probes containing repetitive sequences have now, however, been developed. The repetitive DNA is blocked out by a pre-reassociation

reaction with sheared human genomic DNA prior to hybridization (Litt and White, 1985; Sealey et al., 1985; Buroker et al., 1986). The efficiency of using this new technique, to screen for polymorphisms in single-copy regions of the genome was comparable to that of previous techniques (Barke et al., 1985; Schumm et al., 1985).

When cosmid clones, which contained up to 45 kb of human DNA, were used as probes with the pre-reassociation technique a much larger region was screened for RFLPs and more polymorphic sites were found. The only problem with this approach was that the alleles of the RFLPs at the different loci were often found to be in linkage disequilibrium and therefore were inherited together as a single locus. These loci would not be as informative as other loci which showed linkage equilibrium (Litt and White, 1985; Buroker et al., 1986). Such loci need to be haplotyped and can become important for anthropological studies since they are usually inherited together as a unit and certain haplotypes are found to be specific for particular populations.

Hypervariable regions in the human genome were discovered, and these single-copy regions varied in size according to the number of times that a core sequence was tandemly repeated. The first arbitrary single-copy DNA sequence, pAW10¹, was found to detect a RFLP which had more than 8 different alleles (Wyman and White, 1980). The polymorphism has been found to be due to two separate polymorphic regions and each is composed of oligonucleotide repeats, arranged in tandem arrays. The polymorphism is the result of variation

in the copy number of the repeat unit. The site of minor polymorphism is due to a 60 mer repeat unit (Wyman et al., 1986). Other hypervariable regions have been found 5' to the human insulin gene (Bell et al., 1982), at several positions around the human α -globin gene cluster (Higgs et al., 1986), within the immunoglobulin heavy chain genes (Mignone et al., 1983), and the HaRas gene (Goldfarb et al., 1982).

Sequencing studies have revealed that all these regions varied due to short tandem repeats and that a number of the core sequences found at different loci have similar sequences (Jarman et al., 1987). Such regions result in highly polymorphic loci in the genome and these repeat families have now been termed variable number of tandem repeats or VNTRs (Nakamura et al., 1987). The core sequence of the tandem repeat that was found within the intron of the myoglobin gene, was used to screen human genomic DNA. It was found to simultaneously detect many loci each of which showed specific polymorphic patterns and resulted in a "DNA fingerprint" for every individual (Jeffreys et al., 1985a). This led to the discovery that these core sequences were probably homologous to many other similar regions throughout the genome. The large number of alleles present at these loci are thought to arise from the propensity of the repeat sequences to engage in unequal exchange during crossing-over at meiosis or these sequences could undergo "slippage" during DNA replication (Jeffreys et al., 1985a; Jarman et al., 1987). It has further been suggested that VNTRs might encode hotspots for recombinational activity (Jeffreys et al., 1985a).

Synthetic oligonucleotide probes which correspond to the internal core sequences of the δ pervariable regions have been constructed and used to screen genomic libraries under low stringencies to isolate other such regions in the genome. A number of highly variable loci have been found using these sequences and they are, as expected, mostly due to tandem repeats (Nakamura *et al.*, 1987). The identification of these loci using synthetic oligonucleotide sequences is an order of magnitude more efficient compared with random clones (White *et al.*, 1986).

It has also become apparent that these types of sequences were not stable in *recA+* hosts and were therefore unlikely to be present in the libraries that were screened for single-copy sequences because they would have been lost during library amplification. This has only become apparent after two groups spent a great deal of time trying to clone the region which was responsible for the variation detected by pAW101 (Wyman *et al.*, 1985) and the hypervariable locus 3' to the α -globin (Jarman *et al.*, 1987).

In this study these newer strategies to search for genomic variation have not been employed. The study was initiated in January, 1985 and therefore all the potential DNA probes had already been isolated prior to reports of these newer strategies. Furthermore, some of these probes had already been found to detect polymorphisms and had been characterized in different populations.

In this study some of the probes which contained repetitive human sequences were used to screen for RFLPs, since it was found that the non-specific hybridization of the repetitive DNA sequences could be removed with stringent post-hybridization washing conditions. Three RFLPs were detected with such probes. These clones, however, gave high non-specific background when nylon (instead of nitrocellulose) membranes were used. Subsequently the non-repetitive human EcoRI fragments within these clones were identified and used to detect the described RFLPs.

A number of RFLPs which were discovered in this study would therefore have been missed if these clones had been ignored due to initial selection criteria i.e. any clones which showed positive hybridization to total human DNA were excluded from the study. These single-copy RFLP-detecting sequences are therefore adjacent to repetitive sequences in the genome. In the isolation of single-copy regions from chromosome-specific libraries, which were constructed using somatic cell hybrids, the clones were isolated on the basis of hybridization to total human DNA and non-hybridization to rodent DNA. The positive clones were digested with a restriction enzyme and single-copy human fragments, within the clones, were identified. All RFLPs detected by this method were therefore adjacent to repetitive sequences (Gusella et al., 1980; Davies et al., 1981; Aldridge et al., 1984). A similar strategy has been used by Krumlauf et al., (1982) for screening flow-sorted chromosome-specific libraries.

These results indicate that single-copy regions which are adjacent or flanked by repetitive sequences may show a higher incidence of polymorphism than other single-copy regions in the genome (Calabretta *et al.*, 1982). This may be due to low selective pressures in such regions and therefore mutation rates are higher in nearby and adjacent single-copy regions. However, until the function of such repetitive regions in the genome are known the reason for this mechanism will remain unclear.

Some of the human *Eco*RI fragments which detected RFLPs in the present study have been subcloned into a plasmid vector in order to facilitate the use of the probes and to remove the repetitive sequences. By subcloning the RFLP-detecting fragments the probe is more specific for the region containing the RFLP and constant fragments are no longer revealed after the Southern blotting procedure. Plasmid vectors are also more stable and we feel that the plasmid DNA extraction technique is more convenient to perform than isolation of DNA from bacteriophage.

Larger genomic probes have been reported to be more efficient than smaller probes in the screening for variation in the human genome since they cover a larger region of the genome during one hybridization experiment (Feder *et al.*, 1985). Daiger *et al.*, (1984) has demonstrated that the cDNA clone for argininosuccinate synthetase (AS) also detects AS-like sequences on eleven different chromosomes in the genome, including the X and the Y chromosomes. Thus during one hybridization experiment at least eleven different sites across the genome are being evaluated for autosomal and

sex-linked RFLPs. Pseudogenes are also expected to reveal more polymorphism than the functional gene because of relaxed selective pressures on such regions.

4.3 Strategies Used to Screen for RFLPs

The strategy used to screen for RFLPs was based on a number of reports in the literature. A relatively small panel of individuals (6-7) was screened thereby increasing the likelihood of detecting RFLPs with high allele frequencies. Skolnick and White (1982) have shown that when such a small panel of individuals is screened, the majority of useful polymorphisms detected would be unlikely to have rarer alleles at frequencies of less than 0.15. Our studies confirmed this. Only 7 individuals were screened and all the polymorphisms found in the Caucasoid population had rarer alleles present at frequencies greater than 0.125. The enzymes used to screen for RFLPs in this study were chosen because they had been shown in previous studies to be useful (Skolnick and White, 1982; Wjisman, 1984; Feder et al., 1985).

Those enzymes which contain the CpG dinucleotide sequences in their recognition sites have been reported to detect RFLPs more frequently because the cytosine residue in this dinucleotide sequence is often methylated. Studies in E. coli demonstrated that the spontaneous deamination of methylated cytosine residues results in a thymine residue which would not be repaired (Coulondre, 1978; Duncan and Miller, 1980). Many of the first RFLPs reported in humans

were found with MspI and TaqI, which both have CpG dinucleotide sequences in their recognition sequence: (Skolnick and White, 1982; Barker et al., 1984).

Another explanation for the relatively large number of RFLPs revealed by MspI and TaqI could be that a larger number of studies included these enzymes (Cooper and Schmidtke, 1984; Skolnick et al., 1984). The newer strategies which aimed at the isolation of highly informative RFLPs, using cosmids or synthetic oligonucleotides as probes, used MspI and TaqI for the initial screening studies (Litt and White, 1985; Nakamura et al., 1987). Seven out of nine RFLPs discovered in this study were detected using TaqI and MspI which agrees closely with previous reports.

Other restriction enzymes were evaluated by Feder et al., (1985) in order to assess their importance in screening for variation and used to test the model proposed by Wjisman (1984). This model predicted the efficiency of restriction enzymes for detecting variant sites, on the basis of their recognition and potential site frequencies which were computed from the genomic dinucleotide distribution (Wjisman, 1984; Feder et al., 1985).

At the Human Gene Mapping Conference, 8, Knowlton et al., (1985b) reported that after using a much larger number of probes to screen for RFLPs, the detection rate using both MspI and TaqI was still much higher than for other enzymes. The enzyme RsaI, which has a four base pair recognition site has also been shown to detect a higher number of RFLPs (Feder et al., 1985; Knowlton et al., 1985b). This enzyme

were found with MspI and TaqI, which both have CpG dinucleotide sequences in their recognition sequences (Skolnick and White, 1982; Barker et al., 1984).

Another explanation for the relatively large number of RFLPs revealed by MspI and TaqI could be that a larger number of studies included these enzymes (Cooper and Schmidtke, 1984; Skolnick et al., 1984). The newer strategies which aimed at the isolation of highly informative RFLPs, using cosmids or synthetic oligonucleotides as probes, used MspI and TaqI for the initial screening studies (Litt and White, 1985; Nakamura et al., 1987). Seven out of nine RFLPs discovered in this study were detected using TaqI and MspI which closely with previous reports.

Other restriction enzymes were evaluated by Feder et al., (1985) in order to assess their importance in screening for variation and used to test the model proposed by Wjisman (1984). This model predicted the efficiency of restriction enzymes for detecting variant sites, on the basis of their recognition and potential site frequencies which were computed from the genomic dinucleotide distribution (Wjisman, 1984; Feder et al., 1985).

At the Human Gene Mapping Conference, 8, Knowlton et al., (1985b) reported that after using a much larger number of probes to screen for RFLPs, the detection rate using both MspI and TaqI was still much higher than for other enzymes. The enzyme RsaI, which has a four base pair recognition site has also been shown to detect a higher number of RFLPs (Feder et al., 1985; Knowlton et al., 1985b). This enzyme

was used in the present study but no RFLPs were detected. This could have been due to the smaller fragments generated which are not always detected after Southern blot analysis.

In 1980, Botstein et al., calculated that when using a restriction enzyme containing a 4 base pair recognition site the probability that RFLP would be detected was 12,35% with a lower limit of 0,34. If an enzyme recognizing a 6 base pair site was used it would be 17,7% with a lower limit of 0,5%. When using larger genomic probes, those enzymes which have 4 base pair recognition sequences should be more useful in revealing polymorphism due to single base pair changes in the DNA since more sites in the genome would be screened at any one time. This strategy would not be useful unless very efficient transfer and detection methods were available for the smaller fragments (Skolnick and White, 1982; Feder et al., 1985).

4.4 Types of RFLPs isolated

In this study seven cloned DNA sequences revealed nine RFLPs, eight of which were due to basepair substitution and one was due to a DNA-rearrangement involving 500 bp of DNA. Four of the RFLPs were detected using the restriction enzyme TaqI, three with MspI and one with AvaII. Two probes each detected two RFLPs and one probe detected a three-allele RFLP.

The majority of RFLPs reported at Human Gene Mapping ,7, (1983) were base pair substitution type RFLPs and only 4 out of 156 were due to DNA rearrangements (Skolnick et al., 1984). However, those DNA rearrangements which involved fewer than 200 bp of DNA would not be detected by the methods used unless a restriction enzyme site was altered within the rearranged DNA sequence. Such RFLPs would probably have been grouped in the base pair substitution category (Skolnick and White, 1982; Skolnick et al., 1984). Since the introduction of newer strategies a number of results from such techniques were reported at the Human Gene Mapping 8 (1985). In addition to these reported, about 1500 clones have been screened by Schumm et al., (1985) and 500 new RFLPs have been identified. Of these a number had PIC values greater than 0,7 and were shown to be due to DNA rearrangements or multiple restriction sites detected by one probe (Braman et al., 1985).

In this study 50% of the DNA sequences, used to screen for variation, detected RFLPs which is not significantly higher than the 30% reported by Barker et al., (1985). A large number of base pair sites was screened using 14 enzymes but no further polymorphic sites using these probes have been detected. This was probably due to the small number of individuals used for screening purposes, where it was anticipated that a number of polymorphic variants having frequencies of less than 0,15 would have been missed (Skolnick and White, 1982).

Some regions of the human genome have been labelled "cold" since they fail to show any significant polymorphism. The X-linked Factor VIII gene has been studied intensively for polymorphism, because it is the locus involved in the commoner human genetic disease Haemophilia A, but only one RFLP was detected after 37 restriction enzymes were used (Gitschier et al., 1984). Two highly polymorphic X-linked sequences have however now been linked to haemophilia A and are useful in family studies (Antonarakis et al., 1983b).

Autosomal "cold" regions have been described and one example is the gene for thyroglobulin, on chromosome where only one RFLP having a rare allele frequency of 0.022 was found after screening 90 individuals with the thyroglobulin gene probe and using 15 different restriction enzymes (Baas et al., 1984).

The DNA sequences isolated in this study have not been extensively characterized since the sequences are initially only useful as markers and are presently of unknown function. The possibility that any of these sequences represents a coding sequence is low since only about 1% of the genome appears to have a coding function. The rest of the genome has been described as "junk" DNA (Orgel and Crick, 1980; Doolittle and Sapienza, 1980).

In the present study four TagI RFLPs have been found and three of these, VC28, VC61 and VC63 are adjacent to repetitive sequences because flanking fragments within the clone have hybridized to human genomic DNA. The probe VC28 detected two RFLPs, both with enzymes which have a Cpg

dinucleotide sequence in their recognition sequence. These sites are close to each other because they are both detected by the 3.6 kb MspI fragment isolated from the clone. This result indicated that this sequence could be a CpG rich region but one that is not conserved since two RFLPs have been discovered in it.

MspI has also been reported to detect a large number of RFLPs but care must be taken not to misinterpret partial digestion which has been found to occur with enzyme. DNA containing meCCGG and meCmeCGG sequences has been reported to be resistant to digestion using MspI. These are rare sites in the human genome and should not contribute to partial digestion with MspI (Busslinger *et al.*, 1983). In the present study three RFLPs were detected using MspI. The first high frequency MspI RFLP was detected with probe VC28 and the two other MspI RFLPs were both detected using probe VC14 in a region of DNA spanning approximately 35 kb.

One three-allele RFLP was detected in this study using probe VC75. Initially, only two alleles were detected but during screening of the C.E.P.H. families a third allele was found. This third allele may have been missed in the southern African Caucasoid individuals because of incomplete separation of the two smaller alleles.

One interesting RFLP found in this study was due to a DNA rearrangement involving 500 bp of DNA. This type of RFLP is not as common as base pair substitution RFLPs but the majority of DNA rearrangements involving fewer than 200 bp would be unlikely to be detected unless an enzyme site is

situated within the deleted/inserted region. The detection of smaller DNA rearrangements depends on the screening technique and the resolution of the fragments generated by restriction digests (Botstein *et al.*, 1980 ; Skolnick and White, 1982).

The DNA rearrangement was possible to resolve by gel electrophoresis when restriction enzymes which have sites 5' and 3' to this region were used. There was always a constant difference of 500 bp in the fragment lengths of the alleles found in heterozygotes. Those enzymes which did not have sites close to the deletion did not permit the resolution of the two fragments in heterozygotes because the fragments were too large to be separated by conventional electrophoresis. One such enzyme was *Sst*I where the size of the fragment which contained the deletion/insertion region was 17,0 kb. Similarly, those enzymes which have sites within the deleted/inserted region would be unlikely to be detected because the small size of the resulting fragments. This could be the reason why the RFLP was not detected using the restriction enzymes *Ava*II, *Mbo*I, *Bsa*I and *Alu*I.

Two restriction enzymes, *Msp*I and *Taq*I, recognized sites within the deleted /inserted region. This was concluded from the partial restriction mapping of this region and the unusual length variation detected by Southern blot analysis. These results are interesting because *Taq*I and *Msp*I, which both have CpG nucleotides in their recognition sequences, have internal sites in the deletion/insertion region indicating that it could be a CpG rich sequence.

A 2,1 kb EcoRI fragment from this clone detects the polymorphism with all the enzymes. The size of this EcoRI fragment (2,1 kb) and the MspI restriction pattern of the cloned sequence indicate that the DNA sequence which was isolated from the library represents the smaller allele. Furthermore, when the 2,1 kb EcoRI fragment was hybridized to the MspI human digests, the 6,5 kb MspI fragment did not hybridize to this sequence. Therefore, it can be inferred that the deletion extends into the adjacent 1,3 kb EcoRI fragment of the clone, or that another MspI site exists within the deleted /inserted region. This site would then be between the sites giving rise to the two visible MspI fragments. A similar explanation can be presented for the unusual TaqI restriction pattern observed, that is, more than one TaqI site exists within this region.

The RFLP detected by VC35 is an interesting polymorphism but will only be useful as a single two-allele locus. Its usefulness lies, however, in the fact that the majority of restriction enzymes can be used to reveal the polymorphism and, therefore, no specific enzyme digests for a linkage study would need to be prepared for this probe. The presence of such a DNA rearrangement in normal individuals confirms the view that the genome contains "junk" DNA. Such a region may be selected against because in the more "ancient" populations, namely the San and Negroids, this region is more often absent than present.

Two DNA rearrangements involving 500 bp have been reported for the oncogenes c-fms and c-abl (Xu et al., 1985; Xu and Galibert, 1986). The RFLPs detected are within the intron

3' to the codon of the phosphate receptor for the tyrosine residue in the c-abl gene and in the intron 5' to the phosphate receptor of the tyrosine residue of the c-fms gene. A similar deletion has also been reported for the c-erbB gene in chickens (Raines et al., 1985). Although the RFLP at the c-fms locus follows a Mendelian pattern of inheritance there appears to be a selective pressure in favour of the heterozygotes in the offspring of a heterozygote and a homozygote for the smaller allele (ab x bb). The RFLP detected by VC85 does not appear to be selected for in a similar manner.

4.5 Localization of RFLP detecting probes

For effective gene mapping, arbitrary DNA sequences detecting polymorphisms need to be localized to specific regions of the genome. The localization of the DNA sequences isolated in this study was achieved using a panel of somatic cell hybrids. In situ hybridization has to date been of limited use in our department and will not be discussed here. The idea of using a panel of hybrid cell lines was originally devised to correlate the presence or absence of enzymatically detected human markers with the chromosome content of each somatic cell hybrid (Ruddle, 1973; Burgerhout, 1978).

As molecular techniques advanced, DNA from different somatic cell hybrid lines has been used to form a clone panel for mapping of DNA sequences. The pattern given by hybridizing a probe to a human DNA control was compared with the pattern

of hybridization of this probe to the DNA from a panel of somatic cell hybrids, which had been digested with the same restriction enzyme. Those clones which have an identical pattern of hybridization to the human sequence are scored as positive for the sequence. Controls of mouse and Chinese hamster DNA are very important because cross-species homology of the human DNA sequence with the rodent lines could exist. This must not be misinterpreted as positive hybridization to the specific hybrid line (Kamarck et al., 1984).

The ideal panel would contain a sufficient number of lines so that the presence or absence of the probe to a specific chromosome would be clearly defined. However this ideal has not been realized since many lines appear to retain certain groups of chromosomes (reviewed by Ruddie and Creagan, 1975). The somatic cell hybrid lines which were used in this study had been characterized cytogenetically for the presence of human chromosomes and these results were confirmed by isozyme studies.

The human DNA sequences which were to be localized, were initially hybridized to a single Southern blot containing 10 different somatic cell hybrid lines which almost formed a complete panel. Unfortunately, one could not distinguish between chromosomes 7 and 20 using this panel, but provisional assignments for probes VC85 and VC64 to chromosome 1, VC61 to chromosome 2 and VC14 to chromosome 6 have been made. Additional studies showed that VC14 was present in two hybrid lines which contained an isochromosome for chromosome 6p and thus a regional assignment can be

made. Furthermore, the assignment of both VC64 and VC85 to the same chromosome was made prior to the linkage study in which they were found to be closely linked. The human DNA sequence of VC85 has also cross-hybridized to Chinese hamster and mouse DNA, thus indicating that it could be a conserved region and even a potential coding sequence.

The localization of VC28, VC63 and VC75 was not definitive using the first panel and other lines were then analyzed. Probe VC75 was assigned to either chromosome 7 or 20 using the initial panel. One hybrid line which showed definite hybridization to VC75, did not appear cytogenetically to contain a chromosome 7. The probe VC75 has been tested against other hybrid lines and was found to be present in at least two other lines which contained chromosome 7 and lacked chromosome 20. It has therefore been concluded that this probe is on chromosome 7, even though a single discrepancy exists. In order to confirm this result DNA probes which are known to be on chromosome 7 can be hybridized to the lines in question and could help to solve the discrepancy.

The assignment of probe VC63 to chromosome 4 has been made on the basis of the initial result using the first 10 lines. In order to confirm this result, the micromarker present in one line needs to be identified as chromosome 4p and not 12p. When confirmed this could contribute to a regional assignment for this probe. Two lines are discordant with this assignment and they could either have micro-deletions on chromosome 4, which were not visible cytogenetically, or else an insufficient number of cells

were analyzed to get a true indication of the chromosomal content of the lines. If these two lines are excluded from the hybridization results, the assignment would be conclusive for chromosome 4 in 22 hybrid lines.

The assignment of VC23 to chromosome 8 has been tentatively made despite the one discrepancy in which the probe has hybridized to a hybrid line which has not been cytogenetically assessed as containing chromosome 8. This could be due to the presence of chromosome 8 in an undetectable frequency in the lines analyzed and the increased sensitivity of DNA hybridization studies compared with the cytogenetic analysis. From studies using further lines the probe appears to be on chromosome 8.

The above assignments of the DNA sequences, which were isolated in this study, must be regarded as provisional. Further work will be needed to clarify the human chromosomal content of the somatic cell hybrid lines constructed in the department. Most of these probes have been placed into the C.E.P.H. linkage study and their assignment could be confirmed independently of *in situ* hybridisation or somatic cell hybrid studies, if they are found to be linked to other assigned classical or DNA markers (Skolnick *et al.*, 1984).

4.6 RFLPs used for Population Studies

It has been shown that rarely can individual loci be sufficiently informative for conclusions about the ethnic origins of individuals to be made. The accumulation of

information from many gene loci should allow for the reconstruction of the evolutionary history of population groups. In 1964, 5 loci were used to construct a tree of descent, of human ethnic groups, consisting of 15 population groups (Cavalli-Sforza and Edwards, 1964). Such results gave information about the existing relationships among the population groups as well as the evolutionary history with respect to the origin of the populations. Initially, the accumulation of data for such studies was slow since the number of available classical markers was limited. The discovery of RFLPs will ensure that these sorts of studies will proceed more rapidly.

The reconstruction of the early history of the peoples of southern Africa is possible using gene markers, when these data are considered together with archaeological, ethnological and linguistic findings. In southern Africa there are two indigenous population groups, namely the Khoisan and the Negroids. The Caucasoid and Mongoloid individuals represent two recent immigrant populations. Fossilized remains have been found of both major divisions of Australopithecus, A. africanus and A. robustus, in southern Africa (Tobias, 1978). The more recent rhodesiensis man has been proposed to be a stage intermediate to both the Negro and Khoisan peoples. From skeletal evidence it has been concluded that if the Khoisan and Negroids descended from a single common stock the divergence was already present at least 20 000 to 30 000 years ago (Nurse et al., 1985).

The peoples of southern African have been grouped on the basis of socio-cultural, physical, anthropological, linguistic, archaeological and genetic criteria into three entities, namely, the San, Khoikhoi and the Negro. Most of the San speak Bushman languages and are, or were until recently, hunter-gatherers, the Khoikhoi speak a language termed Hottentot and are herders, and the Negroids speak Bantu languages and are cultivators, peasant and townsmen (Jenkins and Nurse, 1977).

Morphological findings and genetic evidence indicate that the Khoi and San shared a relatively recent common progenitor and may be grouped into one known as the Khoisan (Singer and Weiner, 1963; Jenkins, 1987). Analysis at the genetic level using serological markers has demonstrated that some alleles are restricted to the Khoisan people and others are characteristic of the Negroids. Further traits are widely spread and support the argument for the common origin of these peoples. The origin of the Negroids is not defined due to discrepancies in the fossil record (discussed by Nurse et al., 1985). The oldest Negroid remains appear to have been found in the Border Cave, South Africa and these suggest that the origin of the Negroid group of peoples could have been as long as 110 000 years ago (de Villiers and Fatti, 1982). The migration of the Bantu-speaking peoples into the subcontinent appears from linguistic and archaeological evidence to have occurred rapidly and the different tribal groups are closely related. These intertribal relationships can be tested at

the DNA level using gene markers and the relationships of these peoples to the San and Khoikhoi peoples can also be assessed in this way.

The frequencies of the classical protein and blood group markers have already been studied intensively in the indigenous populations of southern Africa (Jenkins, 1972; Jenkins and Dunn, 1981; Nurse *et al.*, 1985). The results have indicated that the Bantu-speaking populations are closely related to one another and are likely to have had a common ancestor, whereas the San are more distantly related to the Bantu-speaking populations. The initial studies between these populations have been limited by the availability of protein markers. However a number of polymorphic traits specific to these peoples have been found (Jenkins and Dunn, 1981) and studies using DNA markers could be used to confirm and amplify these findings. The advent of recombinant DNA technology has removed this paucity of genetic markers because more than 1000 RFLPs have already been reported (see proceedings of the Human Gene Mapping conference held in 1985; Willard *et al.*, 1985).

RFLPs specific for the Y chromosome (Page *et al.*, 1982; Bishop *et al.*, 1984, Casanova *et al.*, 1985) will be important to determine the source of the male contribution to a particular population and will be useful to study gene flow from one population to another. Mitochondrial DNA represents the female counterpart to Y chromosomal RFLPs. The rate of gene substitutions are thought to be ten times

higher in mitochondrial DNA than in nuclear DNA and special statistical methods will be needed to interpret the genetic distance analyses between population groups (Jenkins, 1987).

In this study three distinct groups of people have been studied, namely the San, Negroid and the Caucasoids using the polymorphisms isolated. Data on random Caucasoid individuals was obtained after the screening of the French and Utah parents from the C.E.P.H. study. The allele frequencies obtained for the southern African Caucasoids were not significantly different from the C.E.P.H. sample when the results were compared using two by two contingency tables. This was expected because all the polymorphisms reported have been found in all three population groups which suggests that they could represent very old mutations and hence the frequencies obtained for Caucasoids of Western European origin should not differ.

Even though the RFLPs in this study were identified by screening Caucasoid individuals 7 of the 9 RFLPs were found to be present in the other two population groups. The frequency of the minor allele was greater than 0,2 for 6 of the RFLPs detected in the Caucasoid population, and the PIC values ranged from 0,26 to 0,43. In the Negroid population, 7 of the RFLPs detected had the rarer allele at frequencies greater than 0,2 with PIC values ranging between 0,26 to 0,37. Finally, in the San population, 5 of the RFLPs have rarer allele frequencies greater than 0,2 with PIC values ranging between 0,24 to 0,37. The PIC values fall into the range described as "reasonably informative" (Botstein et al., 1980).

Only 1 polymorphism, that was detected by probe VC28 with the restriction enzyme TaqI, was found to be monomorphic in the Negroid and San populations. The rarer allele was present in low frequency ($0,195 \pm 0,07$) in the Caucasoid population. From preliminary haplotype studies, it appears to be a recent mutation in the Caucasoid population which has arisen after the divergence of the racial groups and occurred on the chromosome which contains the larger 4,8 kb MspI allele also detected with the same probe.

The only probe which showed no significant differences in allele frequencies among all three populations was VC61. Both alleles were equally common with frequencies around 0,50 in all populations. This probe has the maximum PIC value (0,37) for a two-allele system in all three of the populations studied. Either it represents a neutral polymorphism, or both alleles are under different selective pressures and are being maintained in the population groups.

One of the two MspI polymorphisms detected using VC14 has allele frequencies in the Caucasoid population which are not significantly different from that of the Negroid population. The rarer allele frequency is $0,13 \pm 0,02$ in the Negroids but this allele has not yet been found in the San population. This polymorphism has either arisen in both these population groups independently after the divergence of the racial groups and represents parallel evolution or, in the San population, it has been selected against or driven to fixation due to random genetic drift. The alleles of the second MspI RFLP show frequencies which are significant.

different between the Caucasoids and Negroids ($X^2 = 11,6$; $p < 0,01$; $df=2$) and yet there is no significant difference in the frequencies between the Caucasoids and San. The allele frequency difference between the San and Negroid obtained at this locus are not conclusive for the number of individuals studied ($p=0,04$) and therefore there is a need to increase the sample size.

Three probes, VC64, VC75 and VC35, detected RFLPs for which the allele frequencies were similar in the Negroid and San populations but which were significantly different from the allele frequencies obtained in the Caucasoid population. The minor allele of VC64 in the Negroids ($0,410 \pm 0,05$) is more common than that in the Caucasoids ($0,222 \pm 0,07$) while in the San it is the more common allele ($0,525 \pm 0,03$). The PIC values for this probe are 0,37 in the San and Negroid compared with 0,29 in the Caucasoid population.

The allele frequencies obtained with probe VC85, which detects a DNA rearrangement, show differences between populations similar to those found for probe VC64; the PIC value in the Caucasoid population is 0,37 and in the San, it is 0,24. The larger allele i.e. the one having the deleted/inserted region present, was more common in the southern African Caucasoids, but in the Negroids and San populations it was the smaller allele, i.e. the one for the absence of this region, which was the more common allele.

The results for VC85 and VC64 can be combined into a haplotype because linkage studies revealed that the two loci were tightly linked. Haplotype studies indicated that the

common haplotype in the Caucasoids was -/+, in the Negroids it was -, + and in the San -/- and +/- were both found. The first (-) or (+) indicates the absence or presence of the TaqI site of VC64 and the second (+) or (-) indicates the presence or absence of the 500 bp region. Of the ten possible genotypes obtained using the two probes, 8 have been found in the Caucasoid population, 6 in the Negroid population and only 5 in the San population.

Although of these two loci are linked the alleles do not appear to be in linkage disequilibrium. The observed frequencies of each of the haplotypes is not significantly different from the expected, which was calculated from the product of the allele frequencies at each locus. All four haplotypes have been found in all populations studied. If one considers that the San and Negroid populations are the older populations it appears that this 500 bp region could be selected against.

In the Caucasoid population three alleles have been found for probe VC75 whereas only two have been found in the Negroid and San populations. The allele frequencies are significantly different between the Caucasoid population and the Negroid populations and between the San population and the Caucasoid population ($\chi^2 = 34.8$ and 57.4 respectively; $p < 0.01$ for 5 degrees of freedom). In the Caucasoids, the larger allele is one of the two rarer alleles while in the San and Negroids it is the more common allele. The presence of the third allele detected in Caucasoid individuals raised the PIC from 0.28 to 0.43. This third allele has probably arisen in the Caucasoid population after the divergence of

the racial groups and is a population specific or "private" polymorphism. The PIC values are close to the maximum value for a two-allele locus in the Negroid populations (0,37), decreasing to 0,33 in the San. This could be a useful locus on chromosome 7.

The MspI RFLP detected using probe VC28 has similar allele frequencies in the Caucaoid and Negroid populations but is barely polymorphic in the San population. This polymorphism is a reasonably informative marker for the Caucasoids and Negroids having PIC values of 0,38 and 0,37 respectively, but in the San population the PIC value is only 0,12 so it will not be very useful in this population. This result in the San population could be due to random genetic drift where the relative isolation of this small population has allowed for one of the alleles to increase in frequency. A similar result has been found for probe VC63 where the rarer allele in the Caucasoids ($0,238 \pm 0,07$) and Negroids ($0,321 \pm 0,05$) is the more common allele in the San population ($0,692 \pm 0,05$). These results are significantly different between all three populations.

The RFLPs isolated in this study are presently of unknown function and, therefore, no conclusions about any selective pressures for their presence in a population can be made. If these are, in fact, neutral markers they could become important for studying the history of populations. Selective markers may well indicate the ecological history of the populations (Cavalli-Sforza et al., 1986). When

using a small array of genetic markers it is not possible to come to firm conclusions about the relationships between populations.

A genetic distance analysis was carried out using the specific allele frequencies at the 9 polymorphic sites in four population groups namely the southern African Caucasoids, C.E.P.H. Caucasoids, the southern African Bantu-speaking Negroids and the San. The results indicated that the southern African Caucasoids are closely related to the C.E.P.H. Caucasoids. The small distance which was found between the two Caucasoid populations can be attributed to the third allele detected by VC75 in the C.E.P.H. Caucasoids, which has not yet been found in the S.A. Caucasoids. The Negroid and San populations are more closely related and share a more recent common ancestor than with the Caucasoids. These results were expected because although the San are an isolated population they share common ancestry with the Negroids of southern Africa (Nurse et al., 1985).

Previous studies on these populations using DNA markers have been very limited. However, the allele frequencies of the highly polymorphic locus, pAW101, have been determined in the San and Negroid populations (Bowcock, 1984). It appears that such highly polymorphic loci may not be as useful for population studies as polymorphic loci with fewer alleles. It was not possible to compare the frequencies of the different alleles between the populations studied because of the large number of alleles found at the locus. It is also

not known whether such loci are functionally important nor how frequently new alleles are generated by non-homologous crossing-over at such a region (Bowcock, 1984).

Furthermore, polymorphisms which are present in more than one population do not always indicate a single origin of the polymorphism prior to the divergence of the human racial groups but could indicate that the mutation has occurred more than once in the population groups and parallel evolution has occurred. Further, evidence exists that certain regions can accommodate mutations more readily than others and hence the specific polymorphism has arisen more than once in different population groups. This is best exemplified by haplotype studies around the Hb β S gene which indicate that the mutation has occurred at least three times in Africa (Pagnier *et al.*, 1984). Polymorphisms are useful to mark specific individuals or communities but may not be shed light on the question of the origin of the particular population.

4.7 Contribution to mapping of the genome

The proposed mapping of the human genome would not be possible unless collaboration between many different centres existed. The Centre for the Study of Human Polymorphism (C.E.P.H.) was established in Paris under the direction of Professor J. Dausett. This centre aims to co-ordinate all available data relevant to the mapping of the human genome (Marx, 1985). In any large scale linkage study one of the most important requirements is a source of suitable family

material and such material has been made available through the efforts of Dr. J. Dausett, in Paris and Dr. R. L. White, in Salt Lake City, Utah. Large multigeneration families have been obtained and their lymphocytes immortalized by Epstein Barr Virus (EBV) transformation.

DNA from a set of forty large families has been sent out to be typed by collaborators with all their available probes detecting RFLPs. These data have been pooled together into a database which can be analyzed using the computer programme LINKAGE (Lathrop and Lalouel, 1984a, 1984b). This programme improves on previous linkage analyses, such as the sequential tests of Morton (1955a) and LIPED (Ott, 1974), since multiple point linkage is possible using this programme. The preliminary results of the C.E.P.H. study were recently made available when the first database was compiled. This database contains information on 171 probe-enzyme systems and 16 "classical" markers. The database is far from complete because a number of collaborators, including our laboratory, do not yet have DNA samples from all the individuals.

Six probes from this study have already being placed into the C.E.P.H. database. All available DNA samples were typed with VC61, VC63 and VC64, the three probes which detect TagI RFLPs, VC28 which detects an MspI RFLP, and VC75 which detects the three-allele AvaII RFLP. The probe VC35, which detects the DNA rearrangement, was typed using the restriction enzyme StuI. The other three RFLPs described have been used to type the parents but family studies have not yet been completed. These studies will be completed

when more DNA becomes available but these results are beyond the scope of this thesis. The most important contribution is that these six probes have been placed into a database and that they will be considered in the construction of the human genetic linkage map.

Two probes, VC64 and VC85, which were both localized to chromosome 1 using a somatic cell hybrid panel were used to type the available C.E.P.H. families. Linkage studies were performed between these two probes and they were found to be closely linked having a LOD score of 6.2 at a recombination fraction of 0.00. The prior probability of synteny of any two sequences is 1/18.5 as calculated from the chromosomal lengths of the autosomes (discussed by Renwick, 1971).

These two probes do not contain part of the same DNA sequence since they do not cross-hybridize with each other. Clone VC64 has one internal EcoRI site while VC85 has six. Clone VC64 detected a low frequency TaqI RFLP in the Caucasoid population while VC85 detects a DNA rearrangement which is of higher frequency in the Caucasoid population. The human DNA fragments on Southern Blots differ for all enzymes used when hybridized to the two probes. This result could reflect the existence of a problem in the human DNA library from which these two sequences were isolated. However, if one considers that chromosome 1 contains 4.32% of the autosomal genome, and that it is expected to be proportionately represented in the genomic libraries, that were constructed, then this explanation is less likely.

The initial stages in the construction of the human linkage map will proceed more rapidly than the final stages, which will deal with those areas which will be more difficult to map. Certain regions of the genome will become saturated with markers while other regions which have fewer polymorphic markers will need further attention. These "bare" regions could then be mapped using walking techniques or any of those regionally assigned DNA sequences, which presently do not detect any polymorphisms, but could be used to search for RFLPs using a wider range of enzymes. These sequences have already had human gene mapping workshop numbers assigned to them at the Human Gene Mapping Conferences in 1981, 1983 and 1985. The isolation of large contiguous DNA molecules of up to 200 kL combined with pulse-field gel electrophoresis will facilitate the mapping of such regions (Shaw, 1986). Chromosome-specific libraries will also be important for the final stages of mapping. Already at Human Gene Mapping, 8, preliminary maps for all the chromosomes have been reported (De La Chapelle, 1985).

Those chromosomes which contain the loci for specific disease genes have been intensively studied. The X chromosome has 68 polymorphic DNA sequences localized to it and therefore almost any disease locus on the X chromosome should be able to be identified using these markers (Gusella, 1986). Chromosome 4 has more than thirty polymorphic markers due to the interest in Huntington's Disease which has been localized to chromosome 4p (Gusella et al., 1986). Twenty nine polymorphic DNA sequences have been isolated for chromosome 21, 13 for Chromosome 13 and 21 have been found on chromosome 11. In strategies to search

for such disease gene loci a large number of RFLPs specific for a particular chromosome are usually generated (Willard et al., 1985).

Highly polymorphic hypervariable regions in the genome such as those found near the insulin locus and 3' to the α -globin locus have already shown their value in linkage studies. Adult polycystic kidney disease has been linked to the hypervariable region 3' to the α -globin gene cluster on chromosome 16 (Reeders et al., 1985). One form of bipolar affective disorder have been found to be linked to the HVR's on chromosome 11p, found at the insulin and Ha-ras gene loci (Egeland et al., 1987). The search for and isolation of many more such loci will be of immense value in other linkage studies (Nakamura et al., 1987).

The construction of the human gene map is proceeding well but it will require continuing major collaborative efforts between many laboratories in order to complete the map. It is now probably realistic to expect the final linkage map of the human genome to be completed by 1990. Only one of the probes isolated in this study comes close to the originally required PIC value of 0.5 to make it a sufficiently useful marker. However if one considers the haplotype data between VC64 and VC85 this linkage group forms a more informative marker with a PIC value of 0.53.

Linkage studies using pooled probes, each detecting alleles of different sizes have been suggested by Feder et al., (1985). Four RFLP-detecting sequences which only detected those fragments actually involved in the different RFLPs

were pooled in the hybridization study and each RFLP could be independently scored. It would be possible to use in such a linkage experiment the four probes which detected TagI polymorphisms in this study. The TagI digested DNA would however need to be adequately separated by electrophoresis and all probes must be equally labelled, so that the fragments detected are of equal intensity. This "pooled" probe would efficiently reduce the time spent typing the C.E.P.H. families and also be of value in any linkage study since results for all four loci could be obtained after one hybridization experiment.

4.8 Future studies

The results presented in this study will be enhanced by linkage studies in the C.E.P.H. programme. These probes will either be mapped by linkage to already known markers or else they could become markers in regions which presently have very few. All the probes which have been placed into the C.E.P.H. programme will also be made available without any restrictions to interested collaborators.

At the present moment all the DNA sequences are of interest because of their potential as genetic markers in single-copy regions of the human genome. The probe VC85 was found to detect a DNA rearrangement of 500 bp found in normal individuals. The cross-species homology, which was an incidental finding, made during the somatic cell hybrid studies suggests that it may be a coding region conserved between the species.

A study has been initiated by Cavalli-Sforza et al., (1986) to examine genetic variation in the human species using DNA markers. This type of study is very time consuming, since sufficient material on each population group must be accumulated. In addition, anthropological studies are becoming more difficult because isolated, traditional population groups are decreasing in size due to intermarriage and migration. It is important, therefore, that such populations be ascertained and their lymphocytes immortalized. Small family units consisting of two parents and two children are being collected so that haplotypes can be established and included in the study (Cavalli-Sforza et al., 1986). Results on 47 DNA markers from 30 genes or chromosomal regions have now been investigated in five populations groups. The overall results favour a primary split between Eurasia and Africa but the conclusion is not statistically significant (Bowcock et al., 1987). A much larger number of markers still needs to be studied to reach significance.

Southern Africa provides us with a unique group of different populations. In this study the different divisions within the San and Negroid populations have not been considered but future studies could contribute to our understanding of the relationships within these different population groups. Anthropological studies cannot be completed in a short period of time but the accumulation of data for many protein and DNA markers will facilitate this type of research. The

DNA polymorphisms which have been used in this study could be extended to type all other population groups for comparative purposes.

After the human linkage map has been completed, RFLPs will be used as the starting point for major linkage studies in the search for the loci of many inherited disorders for which the biochemical defect is unknown. In addition prenatal diagnosis in these kinds of diseases will be possible using RFLPs.

Conclusions

In this study nine RFLPs have been discovered using seven single-copy DNA sequences. Each of the polymorphisms has been confirmed to be inherited in a co-dominant Mendelian fashion by means of family studies. The RFLP-detecting sequences have all been provisionally assigned to specific chromosomes. These RFLPs have been studied in some of the different population groups inhabiting Southern Africa. Linkage analysis using these polymorphisms will be possible in the near future when the C.E.P.H. study is completed. A contribution to human genetics has been made by this study in that seven arbitrary loci, which together detect nine polymorphisms in the human genome, have been discovered. It is hoped that one or more of them will be useful as a genetic markers on a specific chromosome in the "completed" human linkage map.

APPENDICES

APPENDIX A

REAGENT	MANUFACTURER
Acetic acid (glacial)	univAR
Adenosine 5' triphosphate	Sigma
Agarose	Seakem
Antibiotics-Ampicillin	Sigma
-Chloramphenicol	Sigma
-Tetracycline	Sigma
[α - ^{32} P] dCTP	Amersham
Bacto-Agar	Difco
Bacto Tryptone	Difco
Bacto Yeast Extract	Sigma
Bromophenol Blue	Sigma / Merck
Cesium Chloride	Sigma
Chloroform	univAR
Core buffer	Bethesda Research Laboratories
2'-deoxycytidine 5'-triphosphate	Sigma
D-L dithiothreitol (DTT)	Sigma
Deoxyribonuclease (DNase I)	Amersham
Deoxyribonucleic acid Polymerase I (DNA pol I)	Amersham / Boehringer Mannheim

Dodecyl hydrogen sodium sulphate (SDS)	Merck
Ethanol	Merck
Ethidium Bromide	Boehringer Mannheim
Ethylene diamine tetra acetic acid (EDTA)	NDH Chemicals
Ficoll	Sigma
Formamide	Fluka AG
Hepes	Sigma
8-Hydroxyquinoline	Sigma
Hydrochloric acid (HCl)	univAR
Isopropyl alcohol	SMM chemicals
IPTG	Sigma
Klenow Fragment	Amersham
Lambda deoxyribonucleic acid (DNA)	Boehringer Mannheim, Worthington
Ligase (T4)	Amersham
Lysozyme	Sigma
β -2-mercaptoethanol	Merck
Magnesium acetate	Merck
Magnesium Chloride (MgCl)	Anal R
Orange G	Sigma

Proteinase K	Boehringer Mannheim
Phenol	univAR
Poly A	Sigma
Polyvinylpyrrolidone	PVP
Potassium Chloride (KCl)	univAR
Pyrophosphate	Sigma

Restriction endonucleases

i. <u>AvaII</u>	Amersham, Boehringer Mannheim
ii. <u>BamHI</u>	Anglian, Amersham, Boehringer Mannheim
iii. <u>BglII</u>	Anglian, Amersham
iv. <u>EcoRI</u>	Amersham, Boehringer Mannheim, Anglian
v. <u>HincII</u>	Amersham
vi. <u>HindII</u>	Boehringer Mannheim
vii. <u>HindIII</u>	Amersham, Boehringer Mannheim
viii. <u>MboI</u>	Bethesda Research Laboratories
ix. <u>MspI</u>	Boehringer Mannheim, Bethesda Research Laboratories
x. <u>PstI</u>	Amersham
xi. <u>PvuII</u>	Amersham
xii. <u>RsaI</u>	Boehringer Mannheim

xiii. <u>SacI</u>	Boehringer Mannheim, Amersham
xiv. <u>SstI</u>	Bethesda Research Laboratories
xv. <u>StuI</u>	Boehringer Mannheim
xvi. <u>TaqI</u>	Boehringer Mannheim
xvii. <u>TthI</u>	Amersham
xix. <u>XbaI</u>	Boehringer Mannheim

Spermidine trihydrochloride	Sigma
Salmon sperm Deoxyribonucleic acid	Sigma
Sephadex G-50	Pharmacia Fine Laboratories
Sodium Acetate	Merck, Sigma
Sodium Chloride	univAR
Sodium Hydroxide	univAR
Sodium Lauryl Sulphate	univAP
Sucrose	PAL chemicals

T4 ligase	Amersham
Trichloroacetic acid	univAR
Tri-Sodium citrate	univAR
Thymidine	Sigma
Tris-base	
Tris-HCl	Boehringer Mannheim
Xgal	Boehringer Mannheim

APPENDIX B

MEDIA

antibiotics

Ampicillin- 10 mg/ml stock. Store at -20 °C

Chloramphenicol- 34 mg/ml of absolute ethanol. Use fresh.

Tetracycline- 10 mg/ml stock. Store at -20 °C in a dark container

Denaturing Solution

	NaCl (M)	NaOH (M)
i. Biotinylated (PALL)	1,5	0,5
ii. Zetaprobe (Biorad)	1,0	0,4
iii. Zetabind (AMF Cuno)	0,6	0,2
iv. Nitrocellulose	1,0	0,5

Denhardt's Solution

0,2 % Ficoll

0,2 % Polyvinylpyrrolidone

0,2 % BSA

Filter through a disposable filter.

Store at -20 °C

Formamide

Deionize using 5 % amberlite.
Stir for 30 minutes.
Filter through Whatman filter paper.
Repeat.
Store at -20 °C.

Ficoll Dye

50 % sucrose
50 mM EDTA pH 7,0
0,1 % Bromophenol Blue
10 % Ficoll
Autoclave.

Hybridization Solution (2X stock)

0,1 M Hepes pH 7,0
20 µg/ml Poly A
6X SSC
100 µg/ml denatured, sonicated sperm DNA
10X Denhardt's Solution
10 % Dextran Sulphate
Store at -20 °C.
Dilute prior to hybridization using 50 % deionised
formamide.
Filter if not clear.

Luria broth

20 g Tryptone
 10 g yeast extract
 20 g NaCl
 Make up to 2 l.
 Autoclave.

Luria Agar

Add 10-12 g of agar /l Luria broth.
 Autoclave.

Lysozyme

Prepare 10 mg/ml fresh solution using 0,25 M Tris-HCl. Adjust pH to 8,0.

Neutralizing Solution

	Na Acetate (M)	NaCl (M)	Tris-HCl	pH
a. Biodyne	3	-	-	5,5
b. Zetabind	-	1,5	0,5	7,5
c. Zetaprobe	-	3,0	0,5	7,4
d. Nitrocellulose	-	3,0	0,5	7,0

Nick Translation buffer

50 mM Tris HCl pH 7,5
5 mM Magnesium chloride
10 mM β -2-mercaptoethanol
0,2 mM dATP
0,2 mM dGTP }- when using hot dCTP
0,2 mM dTTP
Store at -20 °C.

NZCYM

10 g NZ amine
5 g NaCl
5 g Yeast
1 g Casaminoacids /l
pH 7,5

Preparation of organic solvents

- i. **Phenol:** Redistillation of organic phenol at 160°C removed contaminants which cause breakdown or cross-linking of RNA and DNA. The phenol was stored at 4°C in a dark container. A required volume of phenol was melted down and extracted several times with equal volumes of 0,1 M Tris-HCl, pH 8,0 until the pH of the aqueous phase was greater than 7,6. An antioxidant, hydroxyquinoline was added to a final concentration of 0,1 %.

Hydroxyquinoline is a partial inhibitor of RNase activity and a weak chelator of metal ions. The saturated phenol was stored at 4°C in a dark container.

- ii. Chloroform : isoamyl alcohol: A 24:1 ratio of chloroform to isoamyl alcohol was prepared. Isoamyl alcohol helps in the separation of the organic and the aqueous phases.

Oligolabelling buffer

Solution A: 1,25 M Tris-HCl
125 mM Magnesium Chloride
250 mM β -mercaptoethanol
0,5 mM dNTP's-dCTP
Store at -20°C

Solution B: 2 M Hepes, pH 6,6.
Store at -20 °C.

Solution C: p(dN)6 random hexadeoxynucleotides, dissolved at 90 OD units per ml in 1 mM Tris-HCl, 1 mM EDTA, pH 7,5. Stored at -20°C.

OLB: Mix solutions A,B and C in a ratio of 100:250:150.
Store at -20 °C.

Proteinase K

10 mg/ml stock solution.

Incubate at 37 °C for 1 hour

Store at -20 °C.

Prehybridization Buffer (2X)

0.1 M hepes pH 7,0

6X SSC

Poly A (20 µg/ml)

5X Denhardt's solution

100 µg/ml denatured sonicated salmon sperm DNA

Store at -20 °C.

Dilute prior to prehybridization using equal volumes of
deionised formamide.

Salmon sperm DNA

Dissolve to a concentration of 10 mg/ml.

Sonicate for 15 second bursts in every minute.

Repeat five times.

Boil for 10 minutes.

Store at -20 °C.

SSC

0,15 M NaCl

0,015 Na Citrate

STE

10 mM Tris HCl pH 8.0
100 mM NaCl
1 mM EDTA
Adjust pH to 8.0.

TE Buffer

10 mM Tris- HCl
1 mM EDTA
Adjust pH to 8.0.

TEA buffer

40 mM Tris Base
2 mM EDTA
20 mM Glacial Acetic acid
Adjust the pH to 8.1.
Autoclave.

General Notes

1. Use distilled water to make up solutions and store solutions at room temperature unless otherwise stated.
2. All glassware is autoclaved prior to use.

APPENDIX C

MOLECULAR WEIGHT MARKERS

Sizes of fragments in kilobases (kb)

Lambda HindIIILambda HindIII + EcoRI

23,70	21,80
9,46	5,24
6,75	5,05
4,26	4,21
2,26	3,41
1,98	1,98
0,58	1,90
	1,57
	1,32
	0,93
	0,84
	0,58

REFERENCES

ALDRIDGE, J. KUNKEL, L. BRUNS, G. TANTRAVAH, U. LALANDE, M. BREWSTER, T. ET AL (1984). A strategy to reveal high-frequency RFLPs along the X chromosome. *Am. J. Hum. Genet.*, 36:546-564.

ALLISON, A.C. (1954). Protection afforded by the sickle cell trait against malarial infection. *Br. Med. J.*, 1: 290-294.

AMOS, D.B. KOSTYU, D.D. (1980). HLA- A central immunological agency of man. *Ad. Hum. Genet.*, 10:137-208.

ANDERSON, S.A.T. BANKIER, B.G. BARRELL, M.H.L. DE BRUIJN, A.R. COULSON, J. DROUIN, I.C. ET AL. (1981). Sequence and organisation of the human mitochondrial genome. *Nature*, 290:457-465.

ANTONARAKIS, S.E. KAZAZIAN, H.H. JR, ORKIN, S.H. (1985a). DNA polymorphism and the molecular pathology of the human globin gene clusters. *Hum. Genet.*, 69:1-14.

ANTONARAKIS, S.E. WARBER, P.G. KUMAR, S.D. PATEL, A.S. KAZAZIAN, H.H. MELLIS, M.A. ET AL. (1985b). Detection of molecular defects and carriers by DNA analysis. *New Engl. J. Med.*, 313:842-848.

BAAS, F. BIKKER, H. VAN OMMEN, G.J. DE VIJLDER, J.J.M. (1984). Unusual scarcity of restriction site polymorphism in the human thyroglobulin gene. A linkage study suggesting autosomal dominance of a defective thyroglobulin allele. *Hum. Genet.*, 67:301-305.

BAKKER, E. WIECKER, P. BEVERSTOCK, G.C. PEARSON, P.L. (1983). Recombinant DNA techniques for mapping the human X chromosome. *Clin. Genet.*, 23:225.

BALAZS, I. PURRELO, M. RUBENSTEIN, P. ALHADEFF, R. SINISCALCO, M. (1982). Highly polymorphic DNA site D14S1 maps to the region of the Burkitt lymphoma translocation and is closely linked to the heavy chain gamma-1 immunoglobulin locus. *Proc. Natl. Acad. Sci. (USA)*, 79:7395-7399.

BARKER, D. HELMAS, C. AKOTS, G. DONIS-KELLER, H. (1985). The analysis of probes containing repetitive DNA using probe-prehybridization. *Cytogenet. Cell Genet.*, 40:576.

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

- BARSKI, G. SOREIL, S. CORNEFERT, F. (1961). Hybrid type cells in combined cultures of two different mammalian cell strains. *J. Natl. Clin. Inst.* 26:1269-1291.
- BARTON, P. MALCOLM, S. MURPHY, C. FERGUSON-SMITH, M.A. (1982). Localization of the human α -globin gene cluster to the short arm of chromosome 16 (16p12-16pter) by hybridisation in situ. *J. Mol. Biol.*, 156:269-278.
- BASTIE-SIGEAC, F. LUCOTTE, G. (1983). Optimal use of restriction enzymes in the analysis of human DNA polymorphism. *Hum. Genet.*, 63:162-165.
- BEAUCHAMP, R.S. MITCHELL, A.R. BUCKLAND, R.A. BOSTOCK, C.J. (1979). Specific arrangements of human satellite III DNA sequences in human chromosomes. *Chromosoma*, 71:153-166.
- BELL, G.I. PICTET, R. RUTTER, W.J. (1980). Analysis of the regions flanking the human insulin gene and sequences of an Alu family member. *Nucleic Acid Res.*, 8:4091-4109.
- BELL, G.I. KARAM, J.H. RUTTER, W.J. (1981). Polymorphic DNA region adjacent to the 5' end of the human insulin gene. *Nucleic Acid Res.*, 8: 4091-4109.
- BELL, G.I. SELBY, M.J. RUTTER, W.J. (1982). The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature*, 295:31-35.
- BELL, G.I. HORITA, S. KARAN, J.H. (1984). A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes*, 33:176-183.
- BELL, J. HALDANE, J.B.S. (1937). The linkage between the genes for colour-blindness and haemophilia in man. *Ann. Eugen. Lond.*, 14:10-31.
- BENADE, L.E. NIERMAN, W.C. (1985). Gene banking at the ATCC: a resource for geneticists. *Cytogenet. Cell Genet.*, 40: 580A.
- BENDER, W. SPIERER, P. HOGNESS, D.S. (1983). Chromosomal walking and jumping to isolate DNA from the ace and rosy and the bithorax complex in *Drosophila melanogaster*. *J. Mol. Biol.*, 168:17-33.
- BINTON, W.D. DAVIS, R.W. (1977). Screening lambda gt recombinant clones by hybridisation to single plaques in situ. *Science*, 196:180-182.

- BENZER, S. (1961). On the topography of genetic fine structure. *Proc. Natl. Acad. Sci. (USA)*, 47:403-416.
- BIRD, A.P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acid Res.* 8:1499-1504.
- BISHOP, C. WEISSENBAACH, J. CASANOVA, M. BERNHEIM, A. FELLOUS, M. (1984). DNA sequences and analysis of the human Y chromosome. In *Progress in Cytogenetics*, Yunis, ed. Academic Press.
- BISHOP, D.T. WILLIAMSON, J.A. SKOLNICK, M.H. (1983). A model for restriction fragment length distributions. *Am. J. Hum. Genet.*, 35:795-815.
- BISHOP, J.O. (1974). The gene numbers game. *Cell*, 2:81-86.
- BLATTNER, F.R. BLECHL, A.E. DENNISON-THOMPSON, K. FABER, H.E. RICHARDS, J.E. SLIGHTOM, J.L. ET (1978). Cloning human fetal γ -globin and mouse α -type globin DNA: Preparation and screening of shotgun collections. *Science*, 202:1279-1284.
- BLATTNER, F.R. WILLIAMS, B.G. BLECHL, A.E. DENNISTON-THOMPSON, K. FABER, H.E. FURLONG, L. GRUNW (1977). Charon phages : Safer derivatives of bacteriophage lambda for DNA cloning. *Science*, 196:161-169.
- BOBROW, M. CROSS, J. (1974). Differential staining of mouse and human interspecific cell hybrids. *Nature*, 251:77-79.
- BODMER, W.F. CAVALLI-SFORZA, L. (1976). Genetics, Evolution and Man. San Francisco, W.H. Freeman and Company.
- BODMER, W.F. THOMPSON, G. (1977). Population genetics and Evolution of the HLA system. In "HLA and Disease" J. Dausett and Svegaard, A (eds.), 280-295, Munksgaard, Copenhagen.
- BOTSTEIN, D. WHITE, R.L. SKOLNICK, M. DAVIES, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, 32:314-331.
- BOWCOCK, A. (1984). Identification of restriction fragment length polymorphisms in southern African human populations. PhD Thesis. Univ. of Witwatersrand.
- BOWCOCK, A.M. BUCCI, C. HERBERT, J.M. KIDD, J.R. KIDD, K.K. FRIEDLANDER, J.S. CAVALLI-SFORZA, L. (1987). Study of 47 DNA markers in five population from four continents. *Gene Geography* 1:47-64.

BOYD, W.C. (1950). Genetics and the Races of Man. Little Brown. Boston, MA.

BRAMAN, J. BARKER, D. SCHUMM, J. KNOWLTON, R. DONIS-KELLER, H. (1985). Characterization of very highly polymorphic RFLP probes. *Cytogenet. Cell Genet.*, 40: 589A.

BRITTEN, R.J. KOHNE, D.E. (1968). Repeated sequences in DNA. *Science*, 161:529-540.

BURGERHOUT, W.G. (1978). A standard procedure for gene assignment by use of somatic cell hybrids. *Cytogenet. Cell Genet.*, 22:689-693.

BUROKER, N.E. MAGENIS, R.E. WELIKY, K. LITT, M. (1986). Four restriction fragment length polymorphisms revealed by probes from a single cosmid map to human chromosome 12q. *Hum. Genet.*, 72:86-94.

BUSSLINGER, M. DEBOER, E. WRIGHT, S. GROSVELD, F.G. FLAVELL, R.A. (1983). The sequence GGCMCGG is resistant to MspI cleavage. *Nucleic Acid Res.*, 11:3559-3569.

CALABRETTA, B. ROBERSON, D.L. BARRERA-SALDANA, H.A. LEMBROU, T.P. SAUNDERS, G.F. (1982). Genome instability in a region of human DNA enriched in Alu repeat sequences. *Nature*, 296:219-225.

CAMI, B. KOURILSKY, P. (1978). Screening of cloned recombinant DNA in bacteria by in situ colony hybridisation. *Nucleic Acid Res.*, 5:2381-2390.

CANN, R.L. STONEKING, M. WILSON, A.C. (1987). Mitochondrial DNA and human evolution. *Nature*, 325:31-36.

CASANOVA, M. LEROY, P. BOUCEKKINE, C. WEISSENBAACH, J. BISHOP, C. FELLOUS, M. PURRELLO, M. FIORI (1985). A human Y-linked DNA polymorphism and its potential for estimating evolutionary distances. *Science*, 230:1403-1406.

CARLE, G.F. FRANK, M. OLSON, M.V. (1986). Electrophoretic separations of large DNA molecules by periodic inversion of the electric field. *Science*, 232:65-67.

CARRANO, A.V. GRAY, J.W. LANGLOIS, R.G. BURKHART-SCHULTZ, K.J. VAN DILLA, M.A. (1979). Measurement and purification of human chromosomes by flow cytometry and sorting. *Proc. Natl. Acad. Sci. (USA)*, 76:1382-1384.

CARSON, P.E. FLANAGAN, C.I. ICKES, C.E. ALVING, A.S. (1956). Enzymatic deficiency in primaquine sensitive erythrocytes. *Science*, 124:484-485.

CASPERSSON, T. LOMAKKA, G. ZECH, L. (1971). The 24 fluorescence patterns of the human metaphase chromosomes - distinguishing characters and variability. *Hereditas*, 67:89-102.

CASPERSSON, T. ZECH, L. JOHANSSON, C. MODEST, E.J. (1970). Identification of human chromosomes by DNA-binding fluorescent agents. *Chromosoma*, 30:215-227.

CAVALLI-SFORZA, L.L. EDWARDS, A.W.F. (1964). Analysis of human evolution. *Genetics today. Proceedings of the XI International Congress of Genetics*. 2:923, The Hague, Netherlands, 1963.

CAVALLI-SFORZA, L.L. KIDD, J.R. KIDD, K.K. BUCCI, C. BOWCOCK, A.M. HEWLETT, B.S. FRIEDLANDER, J. (1986). DNA markers and genetic variation in the human species. *Cold Spring Harbor Symp. Quant. Biol.*, LI: 411-417.

CHAKRAVARTI, A. PHILLIPS, J.A. MELLITS, K.H. BUETOW, K.H. SEEBURG, P.H. (1984). Patterns of polymorphism and linkage disequilibrium suggest independent origins of the human growth hormone gene cluster. *Proc. Natl. Acad. Sci. (USA)*, 81:6085-6089.

CHANDLER, M.E. YUNIS, J.J. (1978). A high resolution in situ hybridisation technique for the direct visualization of labelled G-banded early metaphase and prophase chromosomes. *Cytogenet. Cell Genet.*, 22:352-356.

COLLINS, F.S. WEISMANN, S.M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: a circularisation method. *Proc. Natl. Acad. Sci. (USA)*, 81:6812-6816.

COLLINS, J. HOHN, B. (1980). Cosmids, a type of plasmid gene cloning vector, that is packagable in vitro into bacteriophage lambda heads. *Proc. Natl. Acad. Sci. (USA)* 75:4242-4246.

CONNELLY, P.M. RIVAS, M. L. (1980). Linkage analysis in man. *Ad. Hum. Genet.*, 10:209-266.

CONNELL, G.E. DIXON, G.H. SMITHIES, O. (1962). Subdivision of the three common haaptoglobin types based on hidden differences. *Nature*, 193:505-506.

COOKE, H.J. (1979). Cloning of human satellite III DNA: different components are on different chromosomes. *Nucleic Acid Res.*, 6:3177-3197.

COOPER, D.N. (1983). Eukaryotic DNA methylation. *Hum. Genet.*, 64: 319-333.

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

COOPER, D.N. SCHMIDTKE, J. (1986). Diagnosis of genetic disease using recombinant DNA. *Hum. Genet.*, 73:1-11.

COOPER, D.N. SMITH, B.A. COOKE, H.J. NIEMANN, S. SCHMIDTKE, J. (1985). An estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.*, 69: 201-205.

COULONDRE, C. MILLER, J.H. FARABAUGH, P.J. GILBERT, W. (1978). Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature*, 274:775-780.

COX, D. WOO, S.L. MANSFIELD, T. (1985). DNA restriction fragments associated with α -antitrypsin indicate a single origin for the deficiency allele PIZ. *Nature*, 316:79-81.

CRAMPTON, J.M. DAVIES, K.E. KNAPP, T.F. (1981). The occurrence of families of repetitive sequences in a library of cloned cDNA from human lymphocytes. *Nucleic Acid Res.*, 9:3821-3834.

CREAGAN, R.P. RUDDLE, F.H. (1975). The clone panel: a systematic approach to gene mapping using interspecific somatic cell hybrids. *Cytogenet. Cell Genet.*, 14:282-286.

CROW, J.F. (1973). The dilemma of nearly neutral mutations: How important are they for evolution and human welfare? *J. Hered.*, 63:306-316.

DAHL, H.H. FLAVELL, R.A. GROSVELD, F.G. (1981). The use of genomic libraries for the isolation and study of eukaryotic genes. In *Genetic Engineering 2*. Williamson, R (eds.) Academic Press. London.

DAIGER, S.P. HOFFMAN, N.S. WILDIN, R.S. SU, T.S. (1984). Multiple independent restriction site polymorphisms in human DNA detected with a probe to argininosuccinate synthetase (AS). *Am. J. Hum. Genet.*, 36:736-749.

DAVIES, K.E. PEARSON, P.L. HARPER, P. MURRAY, J.M. O'BRIEN, T. SARFARAZI, M. WILLIAMSON, R. (1983). Linkage analysis of two cloned DNA sequences flanking the Duchenne muscular dystrophy locus on the short arm of the human X chromosome. *Nucleic Acid Res.* 11:2303-2312.

DAVIES, K.E. YOUNG, B.D. ELLES, R.G. HILL, M.E. WILLIAMSON, R. (1981). Cloning of a representative genomic library of the human X chromosome after sorting by flow cytometry. *Nature*, 293:374-381.

DE LA CHAPELLE (1985). The 1985 human gene map and human gene mapping in 1985. *Cytogenet. Cell Genet.*, 40:1-7.

DE VILLIERS, H. FATTI, L.P. (1982). The antiquity of the Negro. *S.A. J. Science* 78: 321-332.

DEISSEROTH, A. NIENHUIS, A. LAWRENCE, J. GILES, R. TURNER, P. RUDDLE, F. (1978). Chromosomal localisation of human β -globin gene on human chromosome 11 in somatic cell hybrids. *Proc. Natl. Acad. Sci. (USA)*, 75:1456-1460.

DEISSEROTH, A. NIENHUIS, A. TURNER, P. VELEZ, R. ANDERSON, W.F. RUDDLE, F. ET AL. (1977). Localisation of the human α -globin structural gene to chromosome 16 in somatic cell hybrids by molecular hybridisation assay. *Cell*, 12:205-218.

DEMARTINVILLE, B. WYMAN, A. WHITE, R.L. FRANCKE, U. (1982). Assignment of the first random restriction fragment length polymorphism (RFLP) Locus (D14S1) to a region of human chromosome 14. *Am. J. Hum. Genet.*, 34:216-226.

DENHARDT, D.T. (1966). A membrane-filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Comm.* 23:641-646.

DETERA-WADLEIGH, S.D. BERRETTINI, W.H. GOLDIN, L.R. BOORMA, D. ANDERSON, S. GERSHOW, E.S. (1987). Close linkage of the c-Harvey-ras-1 and the insulin gene to affective disorder is ruled out in 3 North American Pedigrees. *Nature*, 325:806-808.

DONAHUE, R.P. BIAS, W.B. RENWICK, J.H. MCKUSICK, V.A. (1968). Probable assignment of the Duffy blood group locus to chromosome 1 in man. *Proc. Natl. Acad. Sci. (USA)*, 61:949-955.

DONLON, T.A. LITT, M. NEWCOM, R. MAGENIS, R.E. (1983). Localization of the restriction fragment length polymorphism D14S1 (pAW-101) to chromosome 14q32.1-32.2 by in situ hybridisation. *Am. J. Hum. Genet.* 35:1097-1106.

DONNIS-KELLER, H. BARKER, D. KNOWLTON, R.G. SCHUMM, J.W. BRAMAN, J.C. GREEN, P. (1986). Highly polymorphic RFLP probes as diagnostic tools. Cold Spring Harbor Symp. Quant. Biol., LI:317-324.

DOOLITTLE, W.F. SAPIENZA, C. (1980). Selfish gene, the phenotype paradigm and genome evolution. Nature, 284:601-603.

DOS SANTOS, M.D.R.N. (1986). Gene assignment: interspecific somatic cell hybridization. MSc Thesis. Univ. Witwatersrand.

DRAYNA, D. DAVIES, K.E. HARTLEY, D.A. WILLIAMSON, R. WHITE, R.L. (1984). Genetic mapping of the human X chromosome using restriction fragment length polymorphisms. Proc. Natl. Acad. Sci. (USA), 81:2836-2839.

DUNCAN, B.K. MILLER, J.H. (1980). Mutagenic deamination of cytosine residues in DNA. Nature, 287:560-561.

EGELAND, J.A. GERHARD, D.S. PAULS, D.L. SUSSEX, J.N. KIDD, K.K. ALLEN, C.R. HOSTETTER, A.M. HOU (1987). Bipolar affective disorders linked to DNA markers on chromosome 11. Nature, 325:783-787.

EHRlich, M. GAMA-SOSA, M. HUANG, L.H. MIDGETT, R.M. KUO, K.C. MCCUNE, R.A. GEHRKE, C. (1982). Amount and distribution of 5'-methylcytosine in human DNA from different types of tissues or cells. Nucleic Acids Res., 10:2709-2721.

EIBURG, H. MOHR, J. NIELSEN, S. SIMONSEN, N. (1983). Genetics and linkage relationships of the C3 polymorphism. Discovery of the C3-Se linkage and assignment of LES-C3-DM-Se-PEPD-Lu syteny to chromosome 19. Clin. Genet., 24:159-170.

EIBURG, H. MOHR, J. SCHMIEGLOW, K. NIELSEN, L.S. WILLIAMSON, R. (1985). Linkage relationship of paraoxonase (PON) with others: indication of PON-cystic fibrosis syteny. Clin. Genet., 28: 265-271.

EICHWALD, E.J. SILMSER, C.R. (1955). Untitled communication. Transplant Bull., 2:148-149.

ELSTON, R.C. LANGE, K. (1975). The prior probability of autosomal linkage. Ann. Hum. Genet., 38:341-349.

EPHRUSSI, B. WEISS, M.C. (1965). Interspecific hybridization of somatic cells. Proc. Natl. Acad. Sci. (USA), 53:1040-1042.

ESTIVILL, X. FARRALL, M. SCAMBLER, P.J. BELL, G.M. HAWLEY, K.M.F. LENCH, N.J. BATES, G. ET AL. (1987). A candidate for the cystic fibrosis locus isolated by selection for methylation-free islands. *Nature*, 326:840-845.

EVANS, H.J. BUCKLAND, R.A. PARDUE, M. (1974). Location of the genes coding for 18S and 28S ribosomal RNA in the human genome. *Chromosoma*, 48:405-426.

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

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

FEINBERG, A.P. VOGELSTEIN, B. (1983). High specific labelling of DNA restriction endonuclease fragments. *Analyt. Biochem.*, 132:6-13.

FERGUSON-SMITH M.A. AITKEN, D.A. (1982). The contribution of chromosome aberrations to the precision of human gene mapping. *Cytogetet. Cell Genet.*, 32:24-42.

FESTENSTEIN, H. AWAD, J. HITMAN, G.A. CUTBUSH, S. GROVES, A.V. CASSELL, P. OLLIER, W. SACHS, J. (1986). New HLA DNA polymorphisms associated with autoimmune diseases. *Nature*, 322:64-67.

FEUERSTEIN, B.G. PATTABIRAMAN, N. MARTON, L.J. (1986). Spermine-DNA interactions: A theoretical study. *Proc. Natl. Acad. Sci. (USA)*, 83:5948-5952.

FISHER, R.A. (1930). The genetical theory of natural selection. Claredon Press, Oxford.

FORD, C.E. HAMERTON, J.L. (1956). The chromosomes of man. *Nature*, 178:1020-1023.

FORD, E.B. (1940). Polymorphism and Taxonomy. The New Systematics. Claredon Press, Oxford.

FOX, A.L. (1932). The relationship between chemical constitution and taste. *Proc. Natl. Acad. Sci. (USA)*, 18:115-120.

- FRITSCH, E.F. SHEN, C.K.J. LAWN, R.M. MANIATIS, T. (1980). The organisation of repetitive sequences in mammalian globin gene clusters. Cold Spring Harbor Symp. Quant. Biol., 45:761-774.
- FROMMER, M. PROSSER, J. TKACHUK, D. REISER, A.H. VINCENT, P.C. (1982). Simple repeated sequences in human satellite DNA. Nucleic Acid Res., 10:547-563.
- FUCHS, R. BLAKESLEY, R. (1983). Guide to the use of type II restriction endonucleases. Methods in enzymology. Vol 100. Recombinant DNA. Part 13. Wu, R.
- GERHARD, D.S. KAWASAKI, E.S. BANCROFT, F.C. SZARO, P. (1981). Localization of a unique gene by direct hybridisation in situ. Proc. Natl. Acad. Sci. (USA), 78:3758-3759.
- GERSHON, D. SACHS, L. (1963). Properties of a somatic hybrid between mouse with different genotypes. Nature, 198: 912-913.
- GILES, E. AMBROSE, S.H. (1986). Are we all out of Africa?. Nature, 322:21-22.
- GILL, P. JEFFREYS, A.J. WERRETT, D.J. (1985). Forensic application of DNA 'fingerprints'. Nature, 318:577-579.
- GITSCHIER, J. DRAYNA, D. TUDDENHAM, E.G.D. WHITE, R.L. LAWN, R.M. (1984). Genetic mapping and diagnosis of haemophilia A achieved through a BclI polymorphism in the factor VIII gene. Nature, 314:738-740.
- GOLDFARB, M. SHIMIZU, K. PERUCHO, M. WIGLER, M. (1982). Isolation and preliminary characterisation of a human transforming gene from T24 bladder carcinoma cells. Nature, 296:404-409.
- GOODBOURN, S.E.Y. HIGGS, D.R. WAINSCOAT, J.S. CLEGG, J.B. WEATHERALL, D.J. (1983). Highly variable regions of DNA flank the human α globin genes. Nucleic Acid Res., 9:4213-4224.
- GOSDEN, J.R. MITCHELL, R.A. BUCKLAND, R.A. CLAYTON, R.P. EVANS, H. (1975). The location of four human satellite DNAs on human chromosomes. Exp. Cell Res., 92:148-158.
- GREENE, A.E. (1978). Report from the human genetic mutant cell repository to the fourth international workshop on the human gene mapping (Winnipeg, 1977). Cytogenet. Cell Genet., 22:684-685.

GREENWOOD, J.J.D. (1986). Occam and mankind's genetic bottleneck. *Nature*, 324:21-22.

GRODZICKER, T. WILLIAMS, J. SHARP, P. SIMBROOK, J. (1974). Physical mapping of temperature-sensitive mutations of adenoviruses. *Cold Spring Harbor Symp. Quant. Biol.*, 39:439-446.

GRUNSTEIN, M. HOGNESS, D.S. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. (USA)*, 72:3961-3965.

GUSELLA, J.F. KEYS, C. VARSANYI-BREINER, A. KAO, F. JONES, C. PUCK, T.T. HOUSMAN, D. (1980). Isolation and localization of DNA segments from specific human chromosomes. *Proc. Natl. Acad. Sci. (USA)*, 77:2822-2833.

GUSELLA, J.F. WEXLER, N.S. CONNEALLY, P.M. NAYLOR, S.L. ANDERSON, M.A. TANZI, R.E. WATKINS, P.C. (1983). A polymorphic marker genetically linked to Huntington's disease. *Nature*, 306:234-238.

GUSELLA, J.F. (1986). Recombinant DNA techniques in the diagnosis of inherited disorders. *J. Clin. Invest.*, 77:1723-1726.

GUSELLA, J.F. TANZI, R.F. BADER, P.I. PHELAN, M.C. SEVENSON, R. HAYDEN, M. ET AL (1985). Deletion of Huntington's disease linked G8(D4S10) locus in Wolf-Hirschhorn syndrome. *Nature*, 318:75-78.

HALDANE, J.B. (1962). Conditions for stable polymorphisms at an autosomal locus. *Nature*, 193:1108.

HARPER, M.E. BARRERA-SALDANA, H.A. SAUNDERS, G.F. (1982). Chromosomal localization of the human placental lactogen-growth hormone gene cluster to 17q22-24. *Am. J. Hum. Genet.*, 34:227-234.

HARPER, M.E. SAUNDERS, G.F. (1981a). Localisation of single copy DNA sequences on G-banded human chromosomes by in situ hybridisation. *Chromosoma*, 83:431-439.

HARPER, M.E. ULLRICH, A. SAUNDERS, G.F. (1981b). Localization of the human α -uligin gene to the distal end of the short arm of chromosome 11. *Proc. Natl. Acad. Sci. (USA)*, 78:4458-4460.

HARRIS, H. (1980). The principles of human biochemical genetics. Elsevier/North Holland. New York 3rd ed.

HA RIS, H. (1966). Enzyme polymorphisms in man. *Proc. Roy. Soc. Biol.*, 164:298.

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

HARRIS, H. HOPKINS D.A. ROBSON, E.B. (1962). Two dimensional electrophoresis of pseudo-cholinesterase components in normal human serum. *Nature*, 196:1296-1298.

HARRIS, H. KAIMUS, H. (1949). The measurement of taste sensitivity to phenylthiourea (P.T.C.). *Ann. Eugen. Lond.*, 15:24-31.

HARRIS, H. WATKINS, J.F. (1965). Hybrid cells derived from mouse and man: artificial heterokaryons of mammalian cells from different species. *Nature*, 205:640-646.

HATLEN, L. ATTARDI, G. (1971). Proportion of the HeLa cell genome complementary to transfer RNA and 5S RNA. *J. Mol. Biol.*, 56:535-553.

HELENTJARIS, T. GESTELAND, R. (1983). Evaluation of random cDNA clones as probes for restriction fragment polymorphisms. *J. Mol. Appl. Genet.*, 2:237-247.

HIGGS, D.R. GOODBOURN, S.E.Y. WAINSCOAT, J.S. CLEGG, J.B. WEATHERALL, D.J. (1981). High variable region of DNA flank the human α -globin genes. *Nucleic Acid Res.*, 17:4213-4224.

HIGGS, D.R. WAINSCOAT, J.S. FLINT, J. HILL, A.V.S. THEIN, S.L. NICOLLS, R.D. TEAL, H. AYYUB, H. (1986). Analysis of the human α -globin gene cluster reveals a highly informative genetic locus. *Proc. Natl. Acad. Sci. (USA)*, 83:5165-5169.

HIGUCHI, R. PADDOCK, G.V. WALL, R. SALSER, W. (1976). A general method for cloning eukaryotic structural gene sequences. *Proc. Natl. Acad. Sci. (USA)*, 73:3146-3150.

HILL, A.V.S. WAINSCOAT, J.S. (1986). The evolution of the α - and β -globin gene clusters in human populations. *Hum. Genet.*, 73:16-23.

HOCKWALD, R.S. ARNOLD, J. CLAYMAN, J.E. ALVING, S.A. (1952). Toxicity of primaquine in Negroes. *JAMA*, 149:1568-1570.

HODGKINSON, S. SHERRINGTON, R. GUILING, H. MARCHBANKS, R. REEDERS, S. MALLET, J. MCINNIS, M. PE (1987). Molecular Genetic evidence for heterogeneity in manic depression. *Nature*, 325:805-806.

HOHN, B. COLLINS, J. (1980). A small cosmid for efficient cloning of large DNA fragments. *Gene*, 11:291-298.

HOHN, B. MURRAY, K. (1977). Packaging recombinant DNA molecules into bacteriophage particles in vitro. *Proc. Natl. Acad. Sci. (USA)*, 74:3259-3263.

HOPKINSON, D.A. EDWARDS, V.H. HARRIS, H. (1976). The distribution of subunit numbers and subunit sizes of enzymes: a study of the products of 100 gene loci. *Ann. Hum. Genet. London* 39:393-411.

HOUCK, C.M. RINEHART, F.P. SCHMID, C.W. (1979). A ubiquitous family of repeated DNA sequences in the human genome. *J. Mol. Biol.*, 132:289-306.

HSU, T.C. (1979). Human and mammalian cytogenetics. An historical perspective. Springer-Verlag, New York.

HUBBY, J.L. LEWONTIN, R.C. (1966). A molecular approach to the study of genetic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics*, 54:577-594.

HUMAN GENE MAPPING 6 (1981). Sixth International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet.*, 34:Vols 1-4.

HUMAN GENE MAPPING 7 (1983). Seventh International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet.*, 32:Vols 1-4. (1984).

HUMAN GENE MAPPING 8 (1985). Eighth International Workshop on Human Gene Mapping. *Cytogenet. Cell Genet.*, 40:Vols 1-4.

HUTCHISON III, C. NEWBOLD, J. POTTER, S. EDGEELL, M. (1974). Maternal inheritance of mammalian mitochondrial DNA. *Nature*, 251:536-537.

JACOBS, P.A. STRONG, J.A. (1959). A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature*, 183:302.

JARMAN, A.P. NICHOLLS, R.D. WEATHERALL, D.J. CLEGG, J.B. HIGGS, D.R. (1987). Molecular characterisation of the hypervariable region downstream of the human α -globin gene cluster. *EMBO. J.*, 5:1857-1863.

JEANTEUR, P. ATTARDI, G. (1969). Relationship between Hela cell ribosomal RNA and its precursors studied by high resolution RNA-DNA hybridisation. *J. Mol. Biol.*, 45:305-324.

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

JEFFREYS, A.J. (1981). Recent studies of gene evolution using recombinant DNA. In *Genetic Engineering* 2. 1-45. Williamson, R. (ed) Academic Press. London.

JEFFREYS, A.J. WILSON, V. THEIN, S.W. (1985a). Hypervariable 'minisatellite' regions in the human DNA. *Nature*, 314:67-73.

JEFFREYS, A.J. BROOKFIELD, F.J.Y. SEMEONOFF, R. (1985b). Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, 317:818-819.

JELINEK, W.R. TOOMEY, T.P. LEINWAND, L. DUNCAN, C.H. BIRD, P.A. CHOUDARY, F.V. ET AL. (1980). Ubiquitous, interspersed repeated sequences in mammalian genomes. *Proc. Natl. Acad. Sci. (USA)*, 77: 1398-1408.

JENKINS, T. (1972). Genetic polymorphisms of man in southern Africa. MD thesis, Univ. of London.

JENKINS, T. (1987). The prehistory of the san and the khoikhoi as recorded in their blood. Contemporary studies on Khoisan, in honour of Oswin Kohler on the occasion of his 75th birthday. Rainer Vossen and Klaus Kenthmann. (eds.) Helmut Buske Verlag. Hamburg.

JENKINS, T. DUNN, D.S. (1981). Haematological genetics in the tropics. *Clinics in haematology* 10:1029-1050.

JENKINS, T. NICHOLLS, E. GORDON, E. MENDELSON, D. SEFTEL, H.C. ANDREW, M.J.A. (1986). Familial hypercholesterolemia - a common genetic disorder in the Afrikaans population. *S. A. Med. J.*, 57:943-947.

JENKINS, T. TOBIAS, P.V. (1977). Nomenclature of population groups in Southern Africa. *African Studies*, 36,49-55.

JONES, K.W. PROSSER, J. (1971). The chromosomal location of human satellite DNA III. *Chromosoma*, 42:445-451.

KALMUS, H. (1965). Diagnosis and genetics of defective colour vision. Pergamon Press. Oxford.

KAMARCK, M.E. BARKER, P.E. MILLER, R.L. RUDDLE, F.H. (1984). Somatic cell mapping panels. *Exp. Cell Res.*, 152:1-14.

KAN, Y.W. DOZY, A.M. (1978a). Polymorphism of DNA sequence adjacent to the human β -globin structural gene: relationship to sickle mutation. *Proc. Natl. Acad. Sci. (USA)*, 75:5631-5635.

KAN, Y.W. DOZY, A.M. (1978b). Antenatal diagnosis of sickle-cell anaemia by DNA analysis of amniotic-fluid cells. *Lancet*, ii:910-911.

KAN, Y.W. DOZY, A.M. (1980). Evolution of the human hemoglobin S and C genes in world populations. *Science*, 209:388-391.

KAO, F. HARTZ, J. LAW, M.L. DAVIDSON, J.N. (1982). Isolation and chromosomal localisation of unique DNA sequences from a human genomic library. *Proc. Natl. Acad. Sci. (USA)*, 79:865-869.

KAZAZIAN, H.H. JR. ORKIN, S.H. BOEHM, C.D. GOFF, S.C. WONG, C. DOWLING, C.E. NEWBURGER, P.E. KN (1986). Characterization of a spontaneous mutation to a β -thalassemia allele. *Am. J. Hum. Genet.*, 38:860.

KAZAZIAN, H.H. JR. CHAKRAVARTI, A. ORKIN, S.H. ANTONARAKIS, S.E. (1983). Evolution of genes and proteins. Nei, M. Koehn, R. K. Sunderland, Mass. Sinauer, p.137-146.

KELLY, R.B. COZZARELLI, N.R. DEUTSCHER, M.P. LEHMAN, I.R. KORNBERG, A. (1970). Enzymatic synthesis of deoxyribonucleic acid. *J. Biol. Chem.*, 245:189-193.

KELLY, T.J. SMITH, H.O. (1970). A restriction enzyme from *Hemophilus influenzae*. *J. Mol. Biol.*, 51:393-409.

KESSLER, C. NEUMAIER, P.S. WOLF, W. (1985). Recognition sequences of restriction endonucleases and methylases-a review. *Gene*, 13:1-102.

KHANDJIAN, E.W. (1987). Optimized hybridization of DNA blotted and fixed to nitrocellulose and nylon membranes. *Biotechnology*, 5:165-167.

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 isocalleles. *Genet. Res. Camb.*, 11:247.

KIMURA, M. (1983). The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KIMURA, M. OHTA, T. (1971). Protein polymorphism as a phase of molecular evolution. *Nature*, 229:467-469.

KIYAMA, R. MATSUI, H. OISHI, M. (1986). A repetitive DNA family (SauJA family) in human chromosomes: extra chromosomal DNA and DNA polymorphism. *Proc. Natl. Acad. Sci. (USA)*, 83:4665-4669.

KNOWLTON, R.G. COPEN-HAGUENAUER, O. VAN CONG, N. FREZAL, J. BROWN, V.A. BARKER, D. BRAMAN, D. T. (1985a). A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature*, 318:380-382.

KNOWLTON, R. SCHUMM, J. BRAMAN, J. BARKER, D. DONIS-KELLER, H. (1985b). Relative efficiencies of restriction endonucleases in detecting RFLPs. *Cytogenet. Cell Genet.*, 40:670A.

KNOWLTON, R. BROWN, V.A. BRAMAN, J.C. BARKER, D. SCHUMM, J.W. MURRAY, C. TAKVORIAN, T. RITZ, J. (1986). Use of highly polymorphic DNA for genotypic analysis following bone marrow transplantation. *Blood* 68:378-385.

KREITMAN, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature*, 304: 412-417.

KRUMLAUF, R. JEANPIERRE, M. YOUNG, B.D. (1982). Construction and characterisation of genomic libraries from specific human chromosomes. *Proc. Natl. Acad. Sci. (USA)*, 79:2971-2975.

KUNKEL AND CO-AUTHORS (1986). Analysis of deletions in DNA of patients with Becker and Duchenne muscular dystrophy. *Nature*, 322:73-77.

KUNKEL, L.M. MONACO, A.P. MIDDLESWORTH, W. OCHS, H.D. LATT, S.A. (1985). Specific cloning of DNA fragments from the DNA from a patient with an X-chromosome deletion. *Proc. Natl. Acad. Sci. (USA)*, 82:4778-4782.

KUNKEL, L.M. SMITH, K.D. BOYER, S.H. (1976). Human Y-chromosome-specific reiterated DNA. *Science*, 191:190.

KUNKEL, L.M. TANTRAVAHU, U. LATT, S.M. (1982). Chromosomal localization of the human X of DNA segments cloned from flow sorted chromosomes. *Nucleic Acid Res.* 10:1557-1578.

LANDSTEINER, K. (1901). Ueber agglutinationserscheinungen normalen menschlichen blutes. *Wein. klin. Wschr.* 14: 1132.

LANGE, K. BOEHNKE, M. (1982). How many genes will it take to span the human genome?. *Am. J. Hum. Genet.*, 34:842-845.

LATHROP, G.M. LALOUEL, J.M. (1984a). Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.*, 36:460-465.

LATHROP, G.M. LALOUEL, J.M. JULIER, C. OTT, J. (1984b). Strategies for multilocus linkage analysis in humans. *Natl. Acad. Sci. (USA)*, 81:3443-3446.

LAWLER, S.D. SANDLER, M. (1954). Data on linkage analysis in man: Elliptocytosis and bloodgroups. IV families, 5, 6, 7. *Ann. Eugen. Lond.*, 18:328-334.

LAWN, R.M. FRITSCH, E.F. PARKER, R.C. BLAKE, G. MANIATIS, T. (1978). The isolation and characterisation of linked δ - and β -globin genes from a cloned library of human DNA. *Cell*, 15:1157-1174.

LFACH, D.R.F. STAHL, F. (1983). Variability in lambda phage carrying a perfect palindrome in the absence of recombination nucleases. *Nature*, 305:448-451.

LEARY, J.J. BRIGATI, D.J. WARD, D.C. (1983). Rapid and sensitive colorimetric method for visualizing biotinlabeled DNA probes hybridized to DNA or RNA immobilized on nitrocellulose: bio-blot. *Proc. Natl. Acad. Sci. (USA)*, 80:4045-4049.

LEBO, R.V. CARRANO, A.V. BURKHART-SCHULTZ, K. DOZY, A.M. YU, L. KAN, Y.W. (1979). Assignment of human β -, γ -, and δ -globin genes to the short arm of chromosome 11 by chromosome sorting and DNA restriction analysis. *Proc. Natl. Acad. Sci. (USA)*, 76:5804-5808.

LERMAN, M.I. THAYER, R.E. SINCEP, M. (1983). KpnI family of long interspersed repeated DNA sequences in primates: polymorphism of family members and evidence for transcription. *Proc. Natl. Acad. Sci. (USA)*, 80:3966-2970

LEWONTIN, R.C. HUBBY, J.L. (1966). A molecular approach to the study of the study of genetic heterozygosity in natural population. II. Amount of variation and degree of heterozygosity in natural population. *Genetics*, 54:595-609.

LITT, M. WHITE, R.L. (1985). A highly polymorphic locus in human DNA revealed by cosmid-derived probes. *Proc. Natl. Acad. Sci. (USA)*, 82:6206-6210.

LITTLE, P. (1986). Finding the defective gene. *Nature*, 321:558-559.

LITTLEFIELD, G.W. (1964). Three degrees of guanylic acid-inosinic acid pyrophosphorylase deficiency in mouse fibroblasts. *Nature*, 203:1142-1144.

MANIATIS, T. HARDISON, R.C. LACY, E. LAUER J. O'CONNELL, C. QUON, D. (1978). The isolation of structural genes from libraries of eukaryotic DNA. *Cell*, 15:687-701.

MANIATIS, T. KEE, S.G. EFSTRATIADIS, A. KAFATOS, F.C. (1976). Amplification and characterisation of a β -globin gene synthesized in vitro. *Cell*, 8:163-182.

MANIATIS, T. FRITSCH, E.F. SAMBROOK, J. (1982). Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory.

MANUELIDIS, L. (1978). Chromosomal localization of complex and simple repeated human DNAs. *Chromosoma*, 66:23-32.

MARX, J.L. (1985). Putting the human genome on the map. *Science*, 229:150-151.

MCKUSICK, V.A. (1979). The anatomy of the human genome. *J. Hered.*, 71:370-391.

MCKUSICK, V.A. RUDDLE, F.H. (1977). The status of the gene map of the human chromosome. *Science*, 196: 390-403.

MENDELSON, M.L. MAYALL, B.H. BOGART, E. MCGEE II, D.H. PERRY, B.H. (1973). DNA content and DNA-based centromeric index of the 24 human chromosomes. *Science*, 179:1126-1129.

MIESLER, M.H. (1983). DNA polymorphisms. *Nature*, 302:108.

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

- MILLER, O.J. ALLFREDDICE, P.W. MILLER, D.A. BREG, W.R. MIGEON, B.R. (1971). Human thymidine kinase gene locus: assignment to chromosome 17 in a hybrid of man and mouse cells. *Science*, 173:244-245.
- MILLER, R.L. KIDD, K.K. FFRANCKE, U. PARTRIDGE, C. RUDDLE, F.H. (1985). The Yale human gene mapping library. *Cytogenet. Cell Genet.*, 40:699A.
- MILLER, R.L. PARTRIDGE, C.W. KIDD, K.K. RUDDLE, F.H. (1984). The Yale human gene mapping library. *Cytogenet. Cell Genet.*, 32:394-397.
- MITCHELL, A.R. BEAUCHAMP, R.S. BOSTOCK, C.J. (1979). A study of sequence homologies in four satellite DNAs of man. *J. Mol. Biol.*, 135:127-149.
- MOHR, J. (1951). A search for linkage between Lutheran blood group and other hereditary characters. *Acta Path. Microbiol. Scand.* 28:207-210.
- MONACO, A.P. NEVE, R.L. COLLETTI-FEENER, C. BERTELSON, KURNIT, D.M. KUNKEL, L.M. (1986). Isolation of cDNA for portions of the Duchenne muscular dystrophy gene. *Nature*, 323:646-650.
- MORTON, N.E. (1955a). Sequential tests for the detection of linkage. *Am. J. Hum. Genet.*, 7:277-318.
- MORTON, N.E. (1955b). The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am. J. Hum. Genet.*, 8:80-96.
- MURRAY, J.C. DEMOPULOS, C.M. LAWN, R.M. MOTULSKY, A.G. (1983). Molecular genetics of human serum albumin: restriction enzyme fragment length polymorphisms and analbuminemia. *Proc. Natl. Acad. Sci. (USA)*, 80: 5951-5955.
- MURRAY, J.M. DAVIES, K.E. HARPER, P.S. MEREDITH, L. MUELLER, C.R. WILLIAMSON, R. (1982). Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular Dystrophy. *Nature*, 300:69-71.
- MUSICH, P.R. DYKES, R.J. (1986). A long interspersed (LINE) DNA exhibiting polymorphic patterns in human genomes. *Proc. Natl. Acad. Sci. (USA)*, 83:4854-4858.
- NAKAMURA, Y. LEPPERT, M. O'CONNELL, P. WOLFE, R. HOLM, T. CULVER, M. MARTIN, C. FUJIMOTO, E. HO (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science*, 235:1616-1620.

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. (1975). Molecular population genetics and evolution. North Holland Publishing company, Amsterdam..

NURSE, G.T. WEINER, J.S. JENKINS, T. (1985). The peoples of Southern Africa and their affinities. Claredon Press, Oxford.

O'BRIEN, S.J. (1973). On estimating functional gene number in eukaryotes. *Nature (New Bio.)*, 242:52-54.

O'FARRELL, P.H. (1975). High resolution 2D electrophoresis of proteins. *J. Biol. Chem.*, 250:4007-4021.

OHNO, S. (1967). Sex chromosomes and sex-linked genes. Springer, Berlin.

OHATA, T. KIMURA, M. (1971). Functional organization of genetic material as a product of molecular evolution. *Nature*, 233:118-119.

ORGEL, L.E. CRICK, F.H. (1980). Selfish DNA: The ultimate parasite. *Nature*, 284:604-607.

OTT, J. (1974). Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.*, 26:568-597.

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

PAGNIER, J. MEARS, G. DUNDA-B. LKHODJA, O. SCHAEFER-REGO, K.E. BELDFORD, C. NAGEL, R.L. LABIE, D. (1984). Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *Proc. Natl. Acad. Sci. (USA)*, 81:1721-1725.

PARDUE, M.L. GALL, J.G. (1969a). Molecular hybridisation of radioactive DNA to the DNA of cytological preparations. *Proc. Natl. Acad. Sci. (USA)*, 64:600-604.

PARDUE, M.L. GALL, J.G. (1969b). Formation and detection of RNA-DNA hybrid molecules in cytological preparations. *Proc. Natl. Acad. Sci. (USA)*, 65:378-383.

PARDUE, M.L. GALL, J.G. (1970). Chromosomal localisation of mouse satellite DNA. *Science*, 160:1356-1358.

PEARSON, P.L. BAKKER, E. FLAVELL, R.A. (1981). Considerations in designing an efficient strategy for localizing unique sequence DNA fragments to human chromosomes. *Cytogenet. Cell Genet.*, 32:308A.

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

PORTER, I.H. SCHULTZE, J. MCKUSICK, V.A. (1962). Genetical linkage between the loci for glucose-6-phosphate dehydrogenase deficiency and colourblindness in American Negroes. *Ann. Hum. Genet.*, 26:107-122.

POTTER, S.S. NEWBOLD, J.E. HUTCHINSON III, C.A. EDGELL, M.H. (1975). Specific cleavage analysis of mammalian mitochondrial DNA. *Proc. Natl. Acad. Sci. (USA)*, 72:4496-4500.

PROUDFOOT, N.J. GIL, A. MANIATIS, T. (1982). The structure of the zeta-globin gene and closely linked, nearly identical pseudogene. *Cell*, 31:553-563.

RACE, R.R. SANGER, R. (1975). Blood groups in man. 6th edition. Blackwell, Oxford.

RADLOFF, R. BAUER, W. VINOGRAD, J. (1967). A dye-buoyant density method for the detection and isolation of closed circular duplex DNA. The closed circular DNA in HeLa cells. *Proc. Natl. Acad. Sci. (USA)*, 57:1514.

RAINES, M.A. LEWIS, W.G. CRITLEDEN, L.B. KUNG, H.J. (1985). C-erbB activation in avian leukosis virus induced erythroblastosis clustered integration sites and arrangement of the provirus in the c-erbB alleles. *Proc. Natl. Acad. Sci. (USA)*, 82:2287-2291.

RAPLEY, S. ROBSON, E.B. HARPIS, H. SMITH, S.M. (1968). Data on the incidence, segregation and linkage relationships of the adenylate kinase (AK) polymorphism. *Ann. Hum. Genet.*, 31:237-242.

REED, K.C. MANN, D.A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.*, 13:7207-7221.

REEDERS, S.T. BREUNING, M.H. CORNEY, G. JEREMIAH, S.J. MEERA KHAN, P. DAVIES, K.E. HOPKINSON, D. (1986). Two genetic markers closely linked to adult polycystic kidney disease on chromosome 16. *B. Med. J.*, 292:851-853.

- REEDERS, S.T. BREUNING, M.H. DAVIES, K.E. NICHOLLS, R.D. JARMAN, A.P. HIGGS, D.R. WEATHERALL, D. (1985). A highly polymorphic DNA marker linked to adult polycystic kidney disease. *Nature*, 317:542-544.
- RENWICK, J.H. (1969). Progress in mapping human autosomes. *Brit. Med. Bull.*, 25:65-73.
- RENWICK, J.H. (1971). The mapping of human chromosomes. *Ann. Rev. Genet.*, 4: 81-120.
- RENWICK, J.H. LAWLER, S.D. (1955). Linkage between ABO and nail patella loci. *Ann. Hum. Genet.*, Lond 19:312-331.
- RIGBY, P.W.J. DIECKMANN, M. RHODES, C. BERG, P. (1977). Labeling deoxyribonucleotides to high specific activity in vitro by nick translation with DNA Polymerase I. *J. Mol. Biol.*, 113:237-251.
- RODRIGUEZ, R.L. TAIT, R. SHINE, F. BOLIVAR, H. HEYNEKER, M. BOYER, H.W. (1977). Characterization of tetracycline and ampicillin resistant plasmid cloning vehicles. in. *Molecular cloning of recombinant DNA*. Scott, W.A. Werner, R. ed.
- RUDDLE, F.H. (1973). Linkage analysis in man by somatic cell genetics. *Nature*, 242:165-169.
- RUDDLE, F.H. (1981). A new era in mammalian gene mapping : somatic cell genetics and recombinant DNA methodologies. *Nature*, 294:115-120.
- RUDDLE, F.H. (1984). The William Allan Memorial address: Reverse Genetics and beyond. *Am. J. Hum. Genet.*, 36:944-953.
- RUDDLE, F.H. CREAGAN, R.P. (1975). Parasexual approaches to the genetics of man. *Ann. Rev. Genet.*, 9:407-486.
- RUDDLE, F.H. RICCIUTI, F. MCMORRIS, F.A. TISCHFIELD, J. CREAGAN, R. DARLINGTON, G. CHEN, T. (1972). Somatic cell genetic assignment of peptidase C and the Rh linkage group to chromosome A1 in man. *Science*, 176:1429-1431.
- SAUNDERS, G.F. HSU, T.C. GETZ, M.J. LEE SIMES, E. ARRIGHI, F.E. (1972). Locations of a human satellite DNA in human chromosomes. *Nature, New Biology* 236:244-246.
- SCHAFER, B. WHITE, R. (1983). Three random loci in the human genome with base pair-change polymorphisms. *Cytogenet. Cell Genet.*, 32:314A.

SCHMECKPEPER, B.J. SMITH, K.D. DORMAN, B.P. RUDDLE, F.H. TALBOT, JR., C.C. (1979). Partial purification and characterization of DNA from the human X chromosome. *Proc. Natl. Acad. Sci. (USA)*, 76:6525-6528.

SCHMID, C.W. DEININGER, P.L. (1975). Sequence organization of the human genome. *Cell*, 6: 345-358.

SCHMID, C.W. JELINEK, W.R. (1982). The Alu family of dispersed repetitive sequences. *Science*, 216:1065-1070.

SCHMIDTKE, J. KRAWCZAK, M. COOPER, D.N. (1985). Human gene cloning: the storm before the lull? *Nature*, 322:119.

SCHUMM, J. KNOWLTON, K. BRAMAN, J. BARKER, D. VOVIS, G. AKOTS, G. BROWN, V. GRAVIUS, T. HELMAS, (1985). Detection of more than 500 single-copy RFLPs by random screening. *Cytogenet. Cell Genet.*, 40:739A.

SEALEY, P.G. WHITAKER, P.A. SOUTHERN, E.M. (1985). Removal of repeated sequences from hybridisation probes. *Nucleic Acids Res.*, 13:1905-1922.

SHAFIT-ZAGARDO, B. MIAIO, J.J. BROWN, F.L. (1982). KpnI families of long, interspersed repetitive DNAs in human and other primate genomes. *Nucleic Acids Res.*, 10:3175-3192.

SHARP, P.A. SUGDEN, B. SAMBROOK, J. (1973). Detection of two restriction endonucleases activities in haemophilus parainfluenzae using analytical agarose. *Biochemistry*, 12:3055.

SHAW, D.J. (1986). A new strategy for mapping the human genome. *J. Med. Genet.*, 23:421-424.

SHINE, J. SEEBURG, P.H. MARTIAL, J.A. BAXTER, J.D. GOODMAN, H. (1977). Construction and analysis of recombinant DNA for human chorionic somatomammotrophin. *Nature*, 270:494-499.

SHOWS, T.B. SAKAGUCHI, A.Y. NAYLOR, S.L. (1982). Mapping the human genome, cloned genes, DNA polymorphisms, and inherited disease. *Adv. Hum. Genet.*, 12:341-451.

SINGER, M.F. (1982). SINES and LINES: Highly repeated short and long interspersed sequences in mammalian genomes. *Cell*, 28:433-434.

SINGER, R. WEINER, J.S. (1963). Biological aspects of some indigenous African populations. *South-western Journal of Anthropology*, 19:168-176.

SKOLNICK, M. WHIFF, R. (1982). Strategies for detecting and characterising restriction fragment length polymorphisms (RFLPs). *Cytogenet. Cell Genet.*, 32:58-67.

SKOLNICK, M. WILLARD, H.F. MENLOVE, L.A. (1984). Report of the committee on human gene mapping by recombinant DNA techniques. *Cytogenet. Cell. Genet.*, 32:210-273.

SMITH, C.L. CANTOR, C.R. (1986). Pulsed-field gel electrophoresis of large DNA molecules. *Nature*, 319:701-702.

SMITH, H. WILCOX, K.W. (1970). A restriction enzyme from *Haemophilus influenzae*. *J. Mol. Biol.*, 51:379-391.

SMITHIES, O. (1955). Zone electrophoresis in starch gels: group variation in the serum proteins of normal human adults. *Biochem. J.*, 61:629-641.

SMITHIES, O. CONNELL, G.E. DIXON, G.H. (1962). Inheritance of haptoglobin subtypes. *Am. J. Hum. Genet.*, 14:14-21.

SMITHIES, O. HILLER, O. (1959). The genetic control of transferrin in humans. *Biochem. J.*, 72:121-126.

SNOUSE, P.E. CHAKRABORTY, R. (1986). The use of restriction fragment length polymorphisms in paternity analysis. *Am. J. Hum. Genet.*, 38:918-939.

SORGE, J. GILBERT, T. WEST, C. WESTWOOD, B. BEUTLER, E. (1985). Heterogeneity in type I Gaucher disease demonstrated by restriction mapping of the gene. *Proc. Natl. Acad. Sci. (USA)*, 82:5442-5445.

SOUTHERN, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.*, 98:303-317.

SOUTHERN, E.M. (1979). Gel electrophoresis of restriction fragments. *Methods in enzymology* 68:152-176.

STEFFENSEN, D. DUFFEY, P. PRENSKY, W. (1974). Localisation of 5S ribosomal RNA genes on human chromosome 1. *Nature*, 252:741-743.

STRAYER, D. HEINTZ, N. ROEDER, R. GILLESPIE, D. (1983). Three organizations of human DNA. *Proc. Natl. Acad. Sci. (USA)*, 80:4773-4774.

SUMNER, A.T. EVANS, H.J. BUCKLAND, R.A. (1971). New technique for distinguishing between human chromosomes. *Nature, New Biology* 232:31-32.

SVENSSON, H. (1962). Isoelectric fractionation, analysis, and characterization of ampholytes in natural pH gradients. III. Description of a paratus for electrolysis in columns stabilizing by. *Arch. Biochem. Supp* 1:132-138.

SWARTZ, D.C. CANTOR, C.R. (1984). Separation of yeast chromosomes-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 37:67-75.

SYKES, B.G. (1983). Quick genomic extraction method. *Lancet*, II:787-788.

THOMAS, M. CAMERON, J.R. DAVIS, R.W. (1974). Viable molecular hybrids of bacteriophage lambda and eukaryotic DNA. *Proc. Natl. Acad. Sci. (USA)*, 71:4579-4583.

THOMPSON, G. (1977). The effect of a selected loci on a linked neutral loci. *Genetics*, 85:753-758.

TILGHMAN, S.M. TIEMEIER, D.C. SEIDMAN, J.G. PETERLIN, B.M. SULLIVAN, M. MAIZEL, J.V. LEDER. P. (1978). Intervening sequence of DNA identified in the structural portion of a mouse β -globin gene. *Proc. Natl. Acad. Sci. (USA)*, 75:725-729.

TJIO, J.H. LEVAN, A. (1956). The chromosome numbers of man. *Hereditas*, 42:1-6.

TOBIAS, P.V. (1978). Early Hominoids of Africa. Jolly, C.J. London Duckworth.

TONEGAWA, S. MAXAM, A.M. TIZARD, R. BERARD, O. GILBERT, W. (1978). Sequence of a mouse germline gene for variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. (USA)*, 75:1485-1489.

TSUI, L.C. BUCHWALD, M. BARKER, D. BRAMAN, J.W. KNOWLTON, J. SCHUMM, J. EIBERG, H. MOHR, J. ET (1985). Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science*, 230:1054-1057.

VAN DER PLOEG, L.H.T. FLAVELL, R.A. (1980). DNA methylation in the human $\gamma\delta$ -globin locus in erythroid and non-erythroid tissues. *Cell*, 19:947-958.

VIEIRA, J. MESSING, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with universal primers. *Gene*, 19:259-268.

VOGELSTEIN, B. FEARSON, E.R. HAMILTON, S.R. FEINBURG, A.P. (1985). Use of restriction fragment length polymorphisms to determine the clonal origin of human tumours. *Science*, 227:642-644.

WAHL, G.M. STERN, M. STARK, G.R. (1979). Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulphate. *Proc. Natl. Acad. Sci. (USA)*, 76:3683-3687.

WAINSCOT, J.S. HILL, A.V.S. BOYCE, A.L. FLINT, J. HERNANDEZ, M. THEIN, S.L. OLD, J.M. LYNCH, J. (1986). Evolutionary relationships of human populations from an analysis of nuclear polymorphisms. *Nature*, 319:491-493.

WAINWRIGHT, B.J. SCAMBLER, P.J. SCHMIDTKE, J. WATSON, E.A. LAW, H. FARRALL, M. COOKE, H.J. EIRE (1985). Localization of cystic fibrosis to human chromosome 7 cen-q22. *Nature*, 318:384-385.

WEISS, M.C. GREEN, H. (1967). Human-mouse hybrid cell lines containing partial complements of human chromosomes and functioning human genes. *Proc. Natl. Acad. Sci. (USA)*, 58:1104-1111.

WELLAUER, P.K. DAW'D, I.B. (1979). Isolation and sequence organisation of human ribosomal DNA. *J. Mol. Biol.*, 128:289-303.

WHITE, R. LEPPERT, M. BISHOP, T. BARKER, D. BERKOWITZ, J. BROWN, C. CALLAHAN, P. HOLM, T. JEROM (1985a). Construction of linkage maps with DNA markers for human chromosomes. *Nature*, 313:101-105.

WHITE, R. WOODWARD, S. LEPPERT, M. O'CONNEL, P. HOFF, M. HERBST, J. LALOUEL, J. DEAN, M. VAN DE (1985b). A closely linked genetic marker for cystic fibrosis. *Nature*, 318:382-384.

WHITE, R. LEPPERT, M. O'CONNELL, P. NAKAMURA, Y. JULIER, C. WOODWARD, S. SILVA, A. WOLFF, R. LA (1986). Construction of the human genetic linkage map: 1. Progress and perspectives. *Cold Spring Harbor Symp. Quant. Biol.*, 51:29-38.

WHITE, R.L. BARKER, D. HOLM, T. BERKOWITZ, J. LEPPERT, M. CAVENEE, W. LEACH, P. DRAYNA, D. (1983). Approaches to linkage analysis in the human. *Banbury Report*, 14:235-249.

WIEACKER, P. DAVIES, K.E. COOKE, H.J. PEARSON, P.L. WILLIAMSON, R. BHATTACHARYA, S. ET AL. (1984). Towards a complete linkage map of the human X chromosome: Regional assignment of 16 cloned single-copy DNA sequences employing a panel of somatic cell hybrids. *Am. J. Hum. Genet.*, 36:265-276.

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

WILLARD, H.F. (1984). Chromosome-specific organisation of the human aliphoid repeated DNA. *Cytogenet. Cell Genet.*, 37:608-609.

WILLARD, H.F. SKOLNICK, M.H. PEARSON, P.L. MANDEL, (1985). Report of the committee on human gene mapping by recombinant techniques. *Cytogenet. Cell Genet.*, 40:360-489.

WILSON, M.C. MELLI, M. (1977). Determination of the number of histone genes in human DNA. *J. Mol. Biol.*, 119:511-535.

WRIGHT, S. (1966). Polyallelic random drift in relation to evolution. *Proc. Natl. Acad. Sci. (USA)*, 55:1074.

WYMAN, A.R. MULHOLLAND, J. BOTSTEIN, D. (1986). Oligonucleotide repeats involved in the highly polymorphic locus D14S1. *Am. J. Hum. Genet.*, 39S:672A.

WYMAN, A.R. WHITE, R. (1980). A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. (USA)*, 77:6754-5758.

WYMAN, A.R. WOLFE, L. BOTSTEIN, D. (1984). Isolation of the highly polymorphic locus D14S1 and other lethal/unstable human sequences in a new cloning host. *Am. J. Hum. Genet.*, 36:159S.

WYMAN, A.R. WOLFE, L. BOTSTEIN, D. (1985). Propagation of some human DNA sequences in bacteriophage Lambda vectors requires mutant *Escherichia coli* hosts. *Proc. Natl. Acad. Sci. (USA)*, 82:2880-2884.

XU, D.Q. GALIBERT, F. (1986). Restriction fragment length polymorphism caused by a deletion within the human c-abl gene (ABL). *Proc. Natl. Acad. Sci. (USA)*, 83:3447-3450.

XU, D.Q. GUILHOT, S. GALIBERT, F. (1985). Restriction fragment length polymorphism of the human c-fms gene. *Proc. Natl. Acad. Sci. (USA)*, 82:2862-2865.

YOUNG, B.D. FERGUSON, M.A. SILLAR, R. BOYD, E. (1981).
High-resolution analysis of human peripheral lymphocyte
chromosomes by flow cytometry. Proc. Natl. Acad. Sci.
(USA), 78:7727-7731.

YUNIS, J.J. SAWYER, J.R. BALL, D.W. (1978).
Characterisation of banding patterns of metaphase-prophase
G-banded chromosomes and their use in gene mapping.
Cytogenet. Cell Genet., 22:679-683.

YOUNG, B.D. FERGUSON, M.A. SILLAR, R. BOYD, E. (1981).
High-resolution analysis of human peripheral lymphocyte
chromosomes by flow cytometry. Proc. Natl. Acad. Sci.
(USA), 78:7727-7731.

YUNIS, J.J. SAWYER, J.R. BALL, D.W. (1978).
Characterisation of banding patterns of metaphase-prophase
G-banded chromosomes and their use in gene mapping.
Cytogenet. Cell Genet., 22:679-683.

Author Conn Vera

Name of thesis The isolation and localization of arbitrary restriction fragment length polymorphisms in Southern African populations. 1987

PUBLISHER:

University of the Witwatersrand, Johannesburg

©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.