

https://dspace.org/

Variable number of tandem repeat polymorphisms of man in Southern Africa.

Marques, Isabel Maria

1992

http://hdl.handle.net/10539/22190

Downloaded from a DSpace repository. https://dspace.org/

VARIABLE NUMBER OF TANDEM REPEAT POLYMORPHISMS OF MAN IN SOUTHERN AFRICA

ISABEL MARIA MARQUES

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

Johannesburg, 1992

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

COMMITTEE FOR RESEARCH ON HUMAN SUBJECTS (MEDICAL)
Ref: R14/49 (Registry)

CLEARANCE CERTIFICATE **PROTOCOL NUMBER** 17/8/92

PROJECT VARIABLE NUMBER OF TANDEM REPEAT POLYMORPHISMS OF MAN IN SOUTHERN AFRICA

INVESTIGATORS MRS IM MARQUES

DEPARTMENT HUMAN GENETICS, SAIMR

DATE CONSIDERED 28 AUGUST 1992

RECOMMENDATION OF THE COMMITTEE APPROVED

DATE 31 AUGUST 1992 CHAIRMAN .....
(Prof PE Cleaton-Jones)

* Guidelines for written "informed consent" attached where applicable.

==========

DECLARATION OF INVESTIGATOR(S)
To be completed in duplicate and **ONE COPY** returned to Miss Sharon Boshoff at Room 10-002, 10th Floor, Senate House, University.

I/we fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee.

DATE 7/9/92..... SIGNATURE .....

HECLEAR.WPS/files/smb

ABSTRACT

Two classes of hypervariable loci, VNTRs and $(CA)_n$ repeats, represent a rich source of highly polymorphic markers in the human genome. The main objective of this study was to assess their usefulness in elucidating the relationships between 17 southern African populations and to assess the feasibility of paternity testing using these hypervariable markers in the local context.

The allele frequency distribution of three VNTR loci (D2S44, D4S125 and APOB) and two $(CA)_n$ repeat loci (APOA2 and D21S168) were studied in 17 southern Africa populations representing four major racial groups, namely: Negroid, Khoisan, "Coloured" and Caucasoid.

The D2S44 locus was studied by the Southern blotting method, 928 chromosomes were analyzed and 28 different alleles scored, ranging in size from 8.0 to 20kb. The heterozygosity values ranged from 80.5% in the Caucasoids to 90.6% in a Negroid group.

The D4S125 locus was analyzed by the Southern blotting method. Twenty one alleles ranging in size from 2.6 to 5.0kb were scored on 866 chromosomes. Heterozygosity frequencies ranged from 66.5% to 91.9%. For this and the previous locus, deviations from H-W equilibrium were observed in the case of some population groups but this can possibly be attributed to small sample size and

errors in allele sizing.

The APOB locus was investigated by the PCR technique. A total of 1454 chromosomes were scored and 20 alleles ranging in size from 420 to 990bp were detected. The heterozygosity frequencies ranged between 69.2% in a Caucasoid group to 89.7% in a Negroid group. A deviation from the H-W equilibrium was observed for the San which was not unexpected.

Amplified DNA from the APOA2 locus from 1440 chromosomes detected 10 allelic fragments ranging in size from 129 to 147bp. Heterozygosity values ranged from 67.7% to 84.6%. No deviations from H-W equilibrium were observed.

At the D21S168 locus 1722 chromosomes were studied by PCR. Twelve alleles ranging in size from 100 to 122bp were scored. Heterozygosity values ranged from 74.5% to 85.7%. Deviation from H-W equilibrium was again observed for the San group as well as for the Jews and Lemba, this might be a reflection of small sample size.

Genetic distances between population groups were calculated using allele frequencies for each of the three VNTR loci and two $(CA)_n$ repeat loci separately and for the five hypervariable loci together. Genetic distances were subjected to cluster analysis and dendrograms constructed to show the affinities between the population groups. In the three VNTR loci and the

two $(CA)_n$ loci, the "Coloureds" were shown to be the most divergent group, but in the cluster obtained using the five hypervariable loci together, the populations of African origin cluster together, as expected, and away from the more divergent Caucasoids. The hypervariable loci may well have a place in the study of population affinities though the information generated will have to be treated with great care so as not to emphasize insignificant differences. Since the hypervariable loci have a higher rate of mutation, other methods of analysis may need to be devised for their accurate use in population genetics.

The hypervariable loci studied by PCR (APOB, APOA2, D21S168) were assessed for their use in solving paternity disputes. Average power of exclusion (PEX) values varied from 94% in the Caucasoid and "Coloured" groups to 96% in the Khoisan and 97% in the Negroid. Paternity indices (PI) and combined paternity indices (CPI) values were calculated for the different cases and these three loci had roughly the same power of discrimination as the 13 marker loci used together for paternity testing in our Department. It is suggested that two or three additional HVR probes would give far better resolution of problems than the existing method. DNA markers have an indisputable role in individual identification and, together with PCR, they constitute

an accurate, rapid and less expensive method of paternity testing.

DECLARATION

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



Isabel Maria Marques

day of September , 1992.

This work is dedicated to my parents for their encouragement throughout my academic career and for being such wonderful grandparents. To my husband for his constant support and to my beloved twin daughters for their patience and understanding when I could not devote more time to them.

ACKNOWLEDGEMENTS

I am in debt to all the people who so willingly donated blood for this study, as well as to the people in the Department of Human Genetics who were involved in the collection of blood samples. Without them this study would not have been possible.

My sincere gratitude to my Supervisors: Professor Trefor Jenkins, for the critical evaluation of my work and for giving me this opportunity and providing the necessary facilities in his Department for this project.
Dr Michele Ramsay who has unselfishly provided expert sound advice and given outstanding encouragement and enthusiasm throughout the period of these studies. It is indeed impossible to adequately express my gratitude for her unstinting patience, assistance and general influence.

My collegues in the Department of Human Genetics are thanked for their friendship, encouragement and helpful discussions during these past few years. A special word of thanks to Tony Lane, Mandy Spurdle, Himla Soodyall, Amanda Krause and Linda Moller.

I wish to thank Drs R. White and Y. Nakamura who so generously supplied me with the pYNH24 and pYNZ32 VNTR probes.

My great appreciation to the following people: Janet Robbie for the help in typing this manuscript, Mike Lanesman and Guy Hall for photographic prints, Etienne Swanepoel for assistance with some of the figures and Mrs Lilly Battaglia and the staff of the SAIMR library for their continued goodwill in obtaining publications for me.

I gratefully acknowledge the Medical Research Council for financial support, and the Director of the South African Institute for Medical Research for the facilities which made these studies possible.

TABLE OF CONTENTS PAGE

ABSTRACT.....ii

DECLARATION.....vi

DEDICATION.....vii

ACKNOWLEDGEMENTS.....viii

TABLE OF CONTENTS.....x

LIST OF TABLES.....xviii

LIST OF FIGURES.....xxv

LIST OF ABBREVIATIONS.....xxx

PUBLICATIONS.....xxxiv

CHAPTER ONE.....1

1 INTRODUCTION.....1

1.1 GENERAL CONSIDERATIONS.....1

1.2 SOUTHERN AFRICAN POPULATIONS AND THEIR AFFINITIES.....3

1.2.1 Nomenclature of the people of southern Africa.....7

1.2.2 The Khoisan People.....10

1.2.3 The southern African Negroids.....13

1.2.4 The southern African Caucasoids.....16

1.2.5 The southern African "Coloured" People.....19

1.3 GENETIC VARIATION AND THE POLYMORPHISM CONCEPT.....21

1.3.1 Protein Polymorphism.....25

1.3.2 DNA Polymorphism.....27

1.4 DNA POLYMORPHISMS.....29

1.4.1 Restriction Fragment Length Polymorphisms.....29

1.4.1.1 RFLP Aetiology.....31

1.4.1.1.1 Base Pair Substitutions.....32

1.4.1.1.2 DNA Rearrangements.....33

1.5 VARIABLE NUMBER OF TANDEM REPEATS (VNTRs).....34

1.5.1 Mechanisms of VNTR formation.....38

1.5.2 Functions of VNTRs.....42

1.5.3 Applications of VNTRs.....43

1.6 $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ REPEATS.....44

1.6.1 Mechanisms of formation of $(CA)_n$.....48

1.6.2 Function of $(CA)_n$.....49

1.6.3 Applications of $(CA)_n$.....52

1.6.4 Advantages of $(CA)_n$ typing.....52

1.7 APPLICATION OF POLYMORPHISMS.....53

1.7.1 Human Evolution.....53

1.7.2 Paternity testing.....55

1.7.2.1 Prior to DNA Technology.....55

1.7.2.2 After the advent of DNA technology.....58

1.7.2.3 Likelihood of Paternity.....62

1.7.3 Other Applications.....64
1.7.3.1 Mapping of the human genome.....64
1.7.3.2 Prenatal and Preclinical Diagnosis of diseases.....66
1.7.3.2.1 Direct diagnosis of mutations 68
1.7.3.2.2 Indirect diagnosis of disorders of unknown etiology.....70

1.8 AIMS OF THE STUDY.....71

CHAPTER TWO.....73

2 SUBJECTS, MATERIALS AND METHODS.....73

2.1 SUBJECTS.....73
2.1.1 The Caucasoid Population.....74
2.1.2 The southern African Negroid population.....74
2.1.3 The Khoisan population.....77
2.1.4 The Johannesburg "Coloured" Population.....77
2.1.5 Paternity Cases.....78

2.2 VENOUS BLOOD SAMPLE COLLECTION.....79

2.3 IMMORTALIZATION OF LYMPHOBLASTOID CELL LINES....80
2.3.1 Collection of samples and transport conditions.....80
2.3.2 Age of Sample.....80
2.3.3 Hypaque-Ficoll Density Gradient Separation of Blood.....81
2.3.4 Washing of the Mononuclear Cells.....82

2.3.5 Infection of Leucocytes with Epstein-Barr Virus.....82

2.3.5.1 Epstein-Barr Virus (EBV) Stock.....83

2.3.5.2 Cyclosporin A.....83

2.3.5.3 Transformation of B-cells with EBV.....84

2.3.6 Culturing and propagation of transformed cells.....85

2.3.7 Cyro-preservation of cells.....85

2.3.8 Thawing and reconstitution of lymphoblastoid cell lines.....86

2.4 DNA PROBES.....87

2.4.1 Plasmid transformation.....89

2.4.2 Preparation of plasmid DNA.....89

2.4.2.1 Amplification of plasmids.....89

2.4.2.2 Lysis by treatment with SDS.....90

2.4.2.3 Plasmid DNA purification by Caesium Chloride-Ethidium Bromide density gradient centrifugation.....91

2.4.3 Restriction analysis of recombinant plasmids.....92

2.5 ISOLATION OF HIGH MOLECULAR WEIGHT HUMAN GENOMIC DNA.....92

2.5.1 DNA extraction from white blood cells.....93

2.5.1.1 Method using phenol extraction.....93

2.5.2 DNA extraction from human lymphoblastoid cell lines.....95

2.5.3 DNA extraction from human tissue.....96

2.5.4 Analysis of human genomic DNA.....97

2.6 ANALYSIS OF HUMAN GENOMIC DNA BY SOUTHERN BLOT HYBRIDIZATION.....97

2.6.1 Restriction endonuclease digestion of genomic DNA.....98

2.6.2 Agarose gel electrophoresis.....100

2.6.3 Southern transfer of DNA fragments.....102

2.6.3.1 Transfer of DNA to nylon membranes.....104

2.6.4 Radiolabelling of DNA probes.....106

2.6.4.1 Oligolabelling of DNA fragments.....106

2.6.4.2 Removal of non-incorporated nucleotides.....108

2.6.5 Hybridization of labelled DNA to nylon southern transfers.....111

2.6.5.1 Prehybridization.....111

2.6.5.2 Hybridization.....112

2.6.5.3 Post-hybridization washing.....113

2.6.6 Autoradiography.....114

2.6.7 Stripping and re-use of membranes.....115

2.7 ANALYSIS OF HUMAN GENOMIC DNA BY POLYMERASE CHAIN REACTION.....116

2.7.1 Variable number of tandem repeat (VNTR) at the Apolipoprotein B locus.....117

2.7.1.1 Reaction mixture.....119

2.7.1.2 Primer selection.....119
2.7.1.3 Cycling parameters.....120
2.7.1.4 Gel electrophoresis.....120
2.7.1.5 Visualization of fragments.....120
2.7.2 $(dA\text{-}dC)_n.(dG.dT)_n$ repeats.....121
2.7.2.1 Reaction mixture.....122
2.7.2.2 Primer selection.....122
2.7.2.3 Cycling parameters.....123
2.7.2.4 Electrophoresis in a denaturing polyacrylamide DNA sequencing gel.....123
2.7.2.5 Autoradiography.....123

2.8 STATISTICAL METHODS.....124
2.8.1 Allele frequencies.....124
2.8.2 Heterozygosity.....124
2.8.3 Genetic distances.....125
2.8.4 Paternity index, combined paternity index, probability of paternity and average power of exclusion.....126

CHAPTER THREE.....129

3 RESULTS.....129

3.1 LYMPHOBLASTOID CELL LINES.....129
3.1.1 Time and temperature conditions before transformation.....129
3.1.2 Freezing of lymphocytes after separation...131

3.2 POLYMORPHISMS IDENTIFIED BY VNTR AND $(CA)_n$ REPEATS.....131

3.2.1 VNTRs and the polymorphisms they identify..134

3.2.1.1 VNTR at the D2S44 locus.....134

3.2.1.2 VNTR at the D4S125 locus.....153

3.2.1.3 VNTR at the APOB locus.....167

3.2.2 The $(CA)_n$ repeats and the polymorphisms they identify.....181

3.2.2.1 $(CA)_n$ repeat at the APOA2 locus.....181

3.2.2.2 $(CA)_n$ repeat at the D21S168 locus.....194

3.2.3 Populations affinities using different combinations of the hypervariable systems.....207

3.2.3.1 At the three VNTR loci.....207

3.2.3.2 At the two $(CA)_n$ repeat loci.....215

3.2.3.3 At the five hypervariable loci analysed together.....222

3.3 APPLICATION OF REPEAT LOCI TO PATERNITY DISPUTES.....229

CHAPTER FOUR.....237

4 DISCUSSION.....237

4.1 LYMPHOBLASTOID CELL LINES.....238

4.1.1 Ideal transportation conditions for transformation.....238

4.2 SOUTHERN BLOTTING COMPARED WITH PCR FOR THE DETECTION OF HYPERVARIABLE LOCI.....239

4.3 POPULATION STUDIES.....243

4.3.1 Basic assumptions for the study of population affinities using allele frequencies.....244

4.3.2 Allele frequencies among populations.....245

4.3.3 Heterozygosity.....249

4.3.4 Heterozygote defficiency.....252

4.3.5 Population affinities by cluster analysis.253

4.3.5.1 Three VNTR loci.....254

4.3.5.2 Two $(CA)_n$ repeat loci.....259

4.3.5.3 Five hypervariable loci.....260

4.4 PATERNITY TESTING.....262

4.5 HOW USEFUL ARE HYPERVARIABLE LOCI? 267

REFERENCES CITED.....269

APPENDIX A - SOURCE OF REAGENTS.....294

APPENDIX B - MEDIA AND SOLUTIONS.....296

LIST OF TABLES **PAGE**

TABLE 1.1 Population groups analysed in the present study.....8

TABLE 2.1 VNTR Probes.....93

TABLE 2.2 Polymorphic $(CA)_n$ detecting markers.....122

TABLE 3.1 Assessment of lymphocyte transformation on peripheral blood samples affected by parameters of age and temperature (T°C) during transportation, prior to transformation.....130

TABLE 3.2 Summary of the 5′ hypervariable loci examined in 17 southern African populations.....132

TABLE 3.3 Allele sizes detected at the different loci.....133

TABLE 3.4 Number and frequencies of binned alleles at the D2S44 locus in 17 southern African populations.....136

TABLE 3.5 Number and frequencies of the binned alleles at the D2S44 locus in the four major southern African populations.....139

TABLE 3.6 Observed and expected numbers of heterozygotes and homozygotes at the D2S44 locus in 17 southern African populations considered separately and pooled into larger groups.....143

TABLE 3.7 Heterozygosity rate at the D2S44 locus when 28 alleles were considered and when these were reduced to 13 bins. Seventeen populations were considered separately and pooled into larger groups.....144

TABLE 3.8 Genetic distances (x100) between 17 southern African populations using the pYNH24 hypervariable probe.....146

TABLE 3.9 Genetic distances (x100) between four major southern African populations using the pYNH24 hypervariable probe 147

TABLE 3.10 Number and frequencies of the binned alleles at the D4S125 locus in 17 southern African populations.....155

TABLE 3.11 Number and frequencies of the binned alleles at the D4S125 locus in the four major southern African populations.....158

TABLE 3.12 Differences between observed and expected numbers of heterozygotes at the D4S125 locus in 17 southern African populations considered separately and pooled into larger groups.....161

TABLE 3.13 Heterozygosity rate at the D4S125 locus when 21 alleles were considered and when these were reduced to 11 bins. Seventeen populations were considered separately and pooled into larger groups.....162

TABLE 3.14 Genetic distances (x100) between 17 southern African populations using the pYNZ32 hypervariable probe.....164

TABLE 3.15 Genetic distances (x100) between the four major southern African populations using the pYNZ32 hypervariable probe.....166

TABLE 3.16 Number and frequencies of the alleles at the APOB locus in 17 southern African populations.....169

TABLE 3.17 Number and allele frequencies at the APOB locus in the four main southern African populations.....172

TABLE 3.18 Observed and expected numbers of heterozygotes and homozygotes at the APOB locus in 17 southern African populations considered separately and pooled into larger groups.....175

TABLE 3.19 Heterozygosity rate at the APOB locus in 17 southern African populations considered separately and pooled into larger groups.....177

TABLE 3.20 Genetic distances (x100) between 17 southern African populations at the APOB locus.....178

TABLE 3.21 Genetic distances (x100) between the four major southern African populations at the APOB locus.....179

TABLE 3.22 Number and frequencies of the alleles at the APOA2 locus in 17 southern African populations.....184

TABLE 3.23 Number and frequencies of the alleles at the APOA2 locus in the four main southern African populations.....187

TABLE 3.24 Observed and expected numbers of heterozygotes and homozygotes at the APOA2 locus in 17 southern African populations considered separately and pooled into larger groups.....189

TABLE 3.25 Heterozygosity rate at the APOA2 locus in 17 southern African populations considered separately and pooled into larger groups.....191

TABLE 3.26 Genetic distances (x100) between 17 southern African populations at the APOA2 locus.....192

TABLE 3.27 Genetic distances (x100) between the four major southern African populations at the APOA2 locus.....193

TABLE 3.28 Number and frequencies of the alleles at the D21S168 locus in 17 southern African populations.....196

TABLE 3.29 Number and frequencies of the alleles at the D21S168 locus in the four main southern African populations.....199

TABLE 3.30 Differencec between observed and expected numbers of heterozygotes at the D21S168 locus in 17 southern African populations considered separately and pooled into larger groups.....202

TABLE 3.31 Heterozygosity rate at the D21S168 locus in 17 southern African populations considered separately and pooled into larger groups.....203

TABLE 3.32 Genetic distances (x100) between 17 southern African populations at the D21S168 locus.....205

TABLE 3.33 Genetic distances (x100) between the four major southern African populations at the D21S168 locus.....206

TABLE 3.34 Genetic distances (x100) between 17 southern African populations at three hypervariable loci.....208

TABLE 3.35 Genetic distances (x100) for nine linguistic groups of southern African using allele frequencies at three VNTR loci.....209

TABLE 3.36 Genetic distance measurements (x100) for the four major southern African populations using allele frequencies at three VNTR loci.....210

TABLE 3.37 Genetic distances (x100) between 17 southern African populations at two $(CA)_n$ repeat loci.....216

TABLE 3.38 Genetic distances (x100) between nine linguistic groups of southern Africa using allele frequencies at two $(CA)_n$ repeat loci.....217

TABLE 3.39 Genetic distance measurements (x100) the four major southern African populations using allele frequencies at two $(CA)_n$ repeat loci.....218

TABLE 3.40 Genetic distances (x100) between 17 southern African populations at five hypervariable loci.....223

TABLE 3.41 Genetic distances (x100) nine linguistic groups of southern Africa using allele frequencies at five hypervariable loci.....224

TABLE 3.42 Genetic distance measurements (x100) for the four major southern African populations using allele frequencies at five hypervariable loci.....225

TABLE 3.43 Paternity indices (PI) and combined paternity indices (CPI) for DNA, HLA and serum protein polymorphisms in 23 paternity cases.....231

TABLE 3.44 Average power of exclusion (PEX) by the APOA2, D21S168 and APOB loci in the four major southern African population groups.....232

TABLE 3.45 Fragment sizes detected by southern blotting (S) or PCR (R) at 10 loci and PI values for case 1.....234

TABLE 3.46 Fragment sizes detected by PCR at nine loci and PI values for case 2.....235

LIST OF FIGURES

FIGURE 1.1 Map of Southern Africa showing the distribution of the ethnic groups comprising the populations investigated in the present study.....6

FIGURE 1.2 The Khoisan people.....11

FIGURE 1.3 Schematic representation of a VNTR locus.....36

FIGURE 2.1 Histopaque-1077 density gradient separation of blood.....81

FIGURE 2.2 An ethidium bromide stained agarose gel containing human DNA digested with the restriction enzyme *Pst*I and separated out by electrophoresis.....103

FIGURE 3.1 Hybridization pattern revealed by pYNH24 with *Pst*I. There is almost a continuous variation of fragment sizes but only some are indicated on the right hand side.....135

FIGURE 3.2 Frequency of binned alleles at the D2S44 locus in the four major southern African population groups.....141

FIGURE 3.3 Dendrogram based on binned frequencies at the D2S44 VNTR locus in 17 southern African populations.....149

FIGURE 3.4 Clustering implied by binned allele frequencies at the D2S44 VNTR locus in four major southern African populations.....150

FIGURE 3.5 Dendrogram based on binned frequencies at the D2S44 VNTR locus.....152

FIGURE 3.6 Dendrogram based on the frequencies of 28 alleles at the D2S44 VNTR locus...152

FIGURE 3.7 Hybridization pattern revealed by pYNZ32 with *Pst*I. There is almost a continuous variation of fragment sizes but only some are indicated on the right hand side.....154

FIGURE 3.8 Frequency of binned alleles at the D4S125 locus in the four major southern African population groups.....159

FIGURE 3.9 APOB VNTR amplification products of 15 random individuals after agarose gel electrophoresis and EtBr staining. Lanes 1 and 13 contain DNA size standards from 1kb ladder and lanes 7 and 20 contains DNA standards from ϕX174 digested with *Hae*III.....168

FIGURE 3.10 Frequencies of alleles at the APOB locus in the four major southern African population groups.....173

FIGURE 3.11 Autoradiogram of PCR products from 14 unrelated individuals primed with the Mfd3 marker at the APOA2 locus. Lanes 3, 11 and 12 contain DNA from two individuals, these were used as internal molecular size markers. There is a continuous variation of fragment sizes but only some sizes are indicated on the right hand side.....182

FIGURE 3.12 Frequencies of alleles at the APOA2 locus in the four major southern African population groups.....188

FIGURE 3.13 An autoradiogram showing PCR products at the D21S168 locus from twenty one unrelated individuals. Lane 1 contains a pBR322 molecular size marker. There is a continuous variation of fragment sizes but only some sizes in are on both sides of the autoradiograph.....195

FIGURE 3.14 Frequencies of alleles at the D21S168 locus in the four major southern African population groups.....200

FIGURE 3.15 Dendrogram based on allele frequencies at the three VNTR loci in 17 southern African populations.....211

FIGURE 3.16 Dendrogram based on allele frequencies at three VNTR loci in nine main linguistic groups of southern Africa....213

FIGURE 3.17 Dendrogram based on allele frequencies at the three VNTR loci in the four major population groups of southern Africa.....214

FIGURE 3.18 Dendrogram based on allele frequencies from two $(CA)_n$ repeat loci in 17 southern African population groups.....219

FIGURE 3.19 Dendrogram based on allele frequencies at the two $(CA)_n$ repeat loci for nine linguistic population groups.....220

FIGURE 3.20 Dendrogram based on allele frequencies at the two $(CA)_n$ repeat loci in the four major population groups of southern Africa.....221

FIGURE 3.21 Dendrogram based on allele frequencies at the five hypervariable loci considered together in 17 southern African populations.....226

FIGURE 3.22 Dendrogram based on allele frequencies at the five hypervariable loci considered together in nine main linguistic groups of southern Africa.....227

FIGURE 3.23 Dendrogram based on allele frequencies at the five hypervariable loci considered together in the four major population groups of southern Africa.....228

LIST OF ABBREVIATIONS

ACD	acid citrate dextrose
APOB	apoliprotein B
bp	base pair
BSA	bovine serum albumin
B95-8	marmoset cell line
°C	degrees celsius
CC	Caucasoids
CL	"Coloured"
Ci	Curie
cpm	counts per minute
CPI	combined paternity index
CyA	cyclosporin A
Da	Dama
dATP	deoxyadenine-5′-triphosphate
dCTP	deoxycytidine-5′-triphosphate
dGTP	deoxyguanine-5′-triphosphate
dNTP	deoxyribonucleotide triphosphate
dTTP	deoxythymidine-5′-triphosphate
DMSO	dymethyl sulphoxide
DNA	deoxyribonucleic acid
EBV	Epstein-Barr virus
EDTA	ethylene-diamine-tetra-acetic acid
EtBr	ethidium bromide
ETL	economic trait loci
FCS	fetal calf serum

g	grams
H	heterozygosity
He	Herero
HWE	Hardy-Weinberg equilibrium
Ind	South African Indians
kb	kilobase
KS	Khoisan
LB	luria Bertani medium
LCL	lymphoblastoid cell line
Lem	Lemba
M	molar
μg	microgram
μl	microlitre
mCi	millicurie
mg	milligram
min	minute
ml	millilitre
μM	micromolar
mM	millimolar
mRNA	mensager RNA
N	normal
Na	Nama
NG	Negroid
ng	nanogram
nm	nanomolar
OD_{600}	optical density of 600mm
p	short arm of a chromosome

PAGE	polyacrylamide gel electrophoresis
^{32}PdCTP	^{32}P deoxycytidine triphosphate
PBS	phosphate buffer saline
PCR	polymerase chain reaction
Pd	Pedi
PEX	average power of exclusion
PI	paternity index
PIC	polymorphism information content
q	long arm of a chromosome
RCP	relative chance of paternity
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RNAse	ribonuclease
rpm	revolutions per minute
RPMI	Roswell Park Memorial Institute medium
SDS	sodium dodecyl sulphate
SE	standard error
Sot	Sotho
SSC	standard saline citrate
ssDNA	salmon sperm DNA
SSPE	sodium sulphate phosphate EDTA
STR	short tandem repeats
Sw	Swazi
TBE	tris-boric acid EDTA
TCA	trichloroacetic acid
TE	tris-EDTA
Tris	tris-(hydroxymethyl) methylamine

Tso	Tsonga
Tsw	Tswana
UV	ultraviolet
Ve	Venda
VNTR	variable number of tandem repeats
Xh	Xhosa
Zu	Zulu
Wht	South African Europeans

PUBLICATIONS

Publications arising from this study:

MARQUES IM, RAMSAY M, JENKINS TJ. Study of three polymorphic variable number of tandem repeat (VNTR) loci in 17 southern African populations. (in preparation).

MARQUES IM, COLMAN M-A, STEVENS G, JENKINS TJ, RAMSAY M. Genetic variation at seven $(CA)_n$ repeat loci in southern Africa populations. (in preparation).

MARQUES IM, KRAUSE A, SPURDLE AB, JENKINS TJ. Paternity testing with deceased family members. (in preparation).

MARQUES IM, JENKINS TJ, LANE AB. Variation of three hypervariable loci in major southern African populations. (in preparation).

CHAPTER ONE

1 INTRODUCTION

1.1 GENERAL CONSIDERATIONS

There is abundant geographic variation in both morphology and gene frequency in most species. The extent of geographic variation results from a balance of forces tending to produce local genetic differentiation and forces tending to produce genetic homogeneity. Mutation, genetic drift due to finite population size and natural selection favouring adaptations to local environmental conditions will all lead to, the genetic differentiation of local populations. The movement of gametes, individuals and even entire populations, collectively called gene flow, will oppose population differentiation.

Gene flow is often regarded as a constraining force in evolution. Natural selection will tend to adapt a population to local environmental conditions but immigrants from other populations will introduce genes adapted to other conditions. In fact, gene flow between populations may prevent them from evolving into different species. Gene flow can, however, also be a creative force in evolution (Wright, 1931). The movement of individuals and even entire populations may spread superior genes and combinations of genes throughout a

species once the gene becomes common in the one location. The role that gene flow plays in a particular species depends both on the geographic distribution of that species and on the importance of other evolutionary forces.

Human populations are sometimes known as ethnic groups or "races". They are hard to define because human beings group themselves in a bewildering array of sets, some of them overlapping, all of them in a state of flux. But for much of its history, man, or perhaps not to offend the inflamable feminists, the human species, has been organized into tribes or groups of closely related individuals, where the language is often the unifying factor. Thus, language offers a rough classification of population groups.

The populations that can tell us the most about our evolutionary past and allow us a good understanding of human diversity are those that have been isolated. They are likely to be linguistically and culturally distinct and are often surrounded by geographic barriers. Isolated human populations contain much more informative genetic features than more recent urban ones, where admixture occurs more frequently.

In southern Africa a unique array of peoples are to be found, they represent three of the major races of

mankind and there is in addition a "mixed" or hybrid population. All these people live in a rich variety of ecological regions where unique topographical and climatological characteristics exist.

At present, however, the once stable and isolated human populations are being encroached upon by improvements in transportation and communication, population growth, famine and war. These conditions are causing the rapid merging of these unique populations with their neighbours. So, it is of the utmost urgency that all population geneticists worldwide, combine their manpower and economic resources to collect and immortalise genetic material from all these unique populations. It would be tragically ironic if, during the same period that the biological tools for understanding our species are developed, the opportunity for applying them in the understanding of our genetic heritage is squandered.

1.2 SOUTHERN AFRICAN POPULATIONS AND THEIR AFFINITIES

Africa has been called the "cradle of humanity" and this is supported by the finding of the fossilized remains of both major divisions of *Australopithecus*. *A. africanus* and *A.robustus* in southern Africa (Tobias, 1978).

Southern Africa has a wealth of anthropologically distinct populations living under diverse environmental conditions. It delights any scientist interested in

studying the effects of racial admixture on a variety of traits in a number of contrasting ecosystems. It allows also the study of the effects of migration, and the spread of cultural and social traditions, the promotion or inhibition of gene flow and gene exchange. Further, it enables the study of the population dynamics of biological evolution and linguistic developments which are governed by different informative mechanisms and unequal velocities of adaptation.

Before proceeding to a description of the main contemporary peoples of southern Africa I wish to describe, in very board terms, their historical and chronological movement into the region.

The San are the first known inhabitants of southern Africa, where they were widely distributed and pursued a hunter-gathering mode of subsistence. A division occured when a group acquired cattle and opted for the pastoral way of life, these become the Khoi (Hottentots) (Elphick, 1977). About two thousand years ago, successive migrations of pastoral Bantu-speaking Negroids began to migrate out of their hypothesized area of origin in West Africa (Guthrie, 1962; Van Warmelo, 1974; Huffman, 1982). One group migrated south-westerly via Zaire and Angola into the present day Namibia, and today the Herero and Ambo populations may be traced to these founders (Fig 1.1). A much larger group proceeded in an easterly direction into East Africa. Some of the

people in this group ventured into southern Africa and those who settled along coastal areas in South Africa are known as the Zulu and Xhosa chiefdoms of today.

Other Bantu-speakers who ventured inland are presently known as the Venda, Lemba and Sotho-Tswana ethnic groups. The locations of the various Negroid and Khoisan ethnic groups sampled in this investigation are shown in Fig 1.1. In the light of these historical events, common ancestry between all the Bantu-speaking Negroids of southern Africa is very plausible.

During the 15th century, the Portuguese circumnavigated the Cape of Good Hope at the southern tip of the African continent on their way to India discovering and opening the route to other explorers from European countries. In 1652, Jan van Riebeck established the first European settlement in the Cape of Good Hope. These early colonists originated in Holland and Germany and, joined by French Huguenot immigrants in 1685, became the progenitors of the present day Afrikaner community. Due to economic reasons at first, and later on, for political reasons, when in 1820 the English settlement disembarked in the Cape, the Afrikaners migrated in an eastern and then northerly direction. Initially, both African Negroids migrating southward and Caucasoids migrating in the opposite direction were unaware of each other but the two groups eventually met at Fish River (Fig 1.1). During the following years, the Khoisan

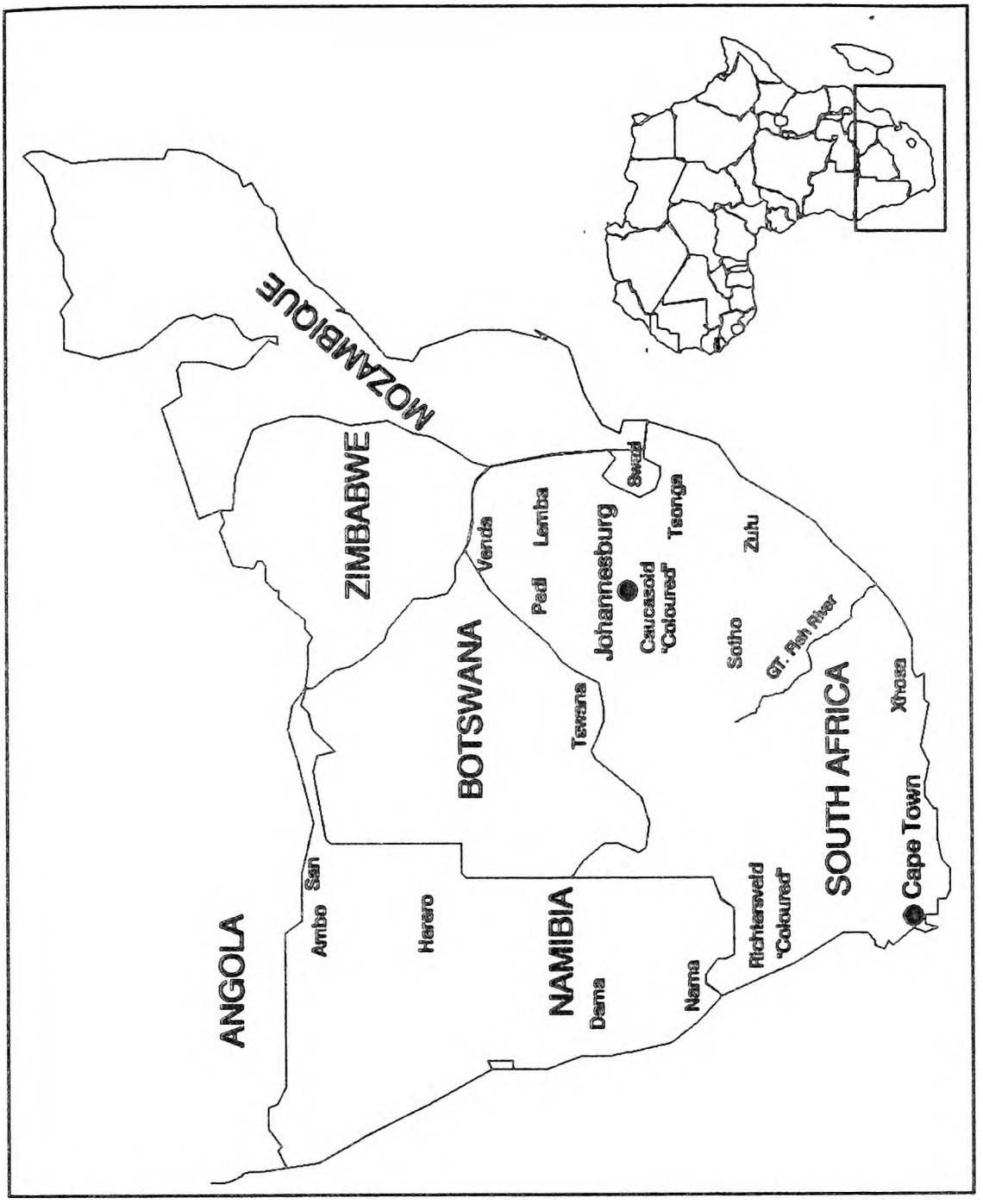


FIGURE 1.1 Map of southern Africa showing the distribution of the ethnic groups comprising the populations investigated in the present study.

people, the African Negroids and those of European stock were frequently in conflict in order to try to establish their supremacy. Following the European colonisation of the Cape, a hybrid population emerged from intermarriages between the Khoisan, Caucasoids and Negroids, forming the so-called "Coloured" or "Cape Coloured" people (Marais, 1939).

A representative sample of unrelated individuals of the Khoisan, Negroid, Caucasoid and "Coloured" peoples was chosen for the present study and these are specified in Table 1.1.

1.2.1 NOMENCLATURE OF THE PEOPLE OF SOUTHERN AFRICA

The existing nomenclature or the terminology referring to the inhabitants of southern Africa is still confusing and not entirely satisfactory to some researchers. The Black, Brown, Yellow and White people of South Africa are all South Africans but appropriate distinguishing names are required.

TABLE 1.1 Population groups analyzed in the present study

POPULATION GROUP	MAJOR DIVISION	ETHNIC GRP
NEGROID	Nguni	
		Zulu
		Swazi
		Xhosa
		Tsonga
	Sotho/Tswana	
		Sotho
		Tswana
		Pedi
	Lemba	
	Venda	
	Herero	
	Khoisan-speaking Negroid	Damara
KHOISAN		
	Khoikhoi	Nama
	San	
CAUCASOID	"European whites"	
	Asiatic Indians	
	Jews	Ashkenazi
"COLOURED"		

The indigenous Negroid peoples of South Africa and Namibia have been referred to as "Bantu". But the term "Bantu" is an unsatisfactory name to use, for various reasons, not only it is "disliked by those to whom it is applied" (Wilson and Thompson, 1969), but also, it has been shown to be biologically unjustified (Tobias, 1971). The term "Bantu" refers to linguistic affiliation and not physical type. The term "Negro" has always had an unacceptable political connotation associated with it and has been said in a UNESCO report to be derogatory

and unacceptable (Baird, 1970). The alternative term "Blacks" is generally preferred and is a politically acceptable term. However, for scientific and bio-medical purposes the inexactness of the term "Blacks" largely vitiates its usefulness.

In June of 1971 there was a conference on "The Peoples of Southern Africa" where discussions on various topics bearing upon the subject of the living peoples of the sub-continent took place. These topics included: social, cultural, physical, anthropological, linguistic, archaeological and genetic aspects (Jenkins and Tobias, 1977). It was recommended at the conference and is currently still accepted, that the description "South African Negro" could be used in the bio-medical and human biological literature, where a more generic term is required than the national or political unit (eg: Zulu, Xhosa, Sotho). "Bushman" and "Hottentot" would be referred to as San and Khoikhoi respectively, or as Khoisan when refered to both at the same time. The languages for the three populations would be referred to as Bantu, Bushman and Hottentot respectively. Their mode of subsistence could be any one of hunter/hunter-gatherer, herder/pastoralist, cultivator, peasant-villager or townsman. The four main population groups are going to be described in the next sections.

1.2.2 THE KHOISAN PEOPLE

The Khoisan people consist of two main groups: the pastoralist Khoi (the only surviving group is the Nama) formerly called "Hottentots" and the hunter-gatherer San or "Bushmen". The latter were people indigenous to the whole of southern and East Africa perhaps even extending as far north as the Mediterranean (Tobias, 1960; 1964). Today the San are virtually confined to the Kalahari basin, an area with only 127 to 380mm of rainfall annually therefore precluding agriculture. Their numbers today are very reduced but this does not seem to be due to biological extinction but rather due to assimilation by other groups.

The San share with the Khoi extremely unusual features that make them very different from other races. Some of these characteristic morphological features include the following: small stature (ranging from 1489mm to 1611mm); lumbar lordosis, due to postural peculiarities; fat deposits over the thighs (steatomeria) and over the buttocks (steatopygia); light yellowish-brown skin colour; tightly spiralled tufts of hair ("peppercorns"); slanting eyes with prominent epicanthia and other folds; very small ears with a high frequency of overrolled helices and commonly lobeless; broad flat noses and small flattish faces (Fig 1.2).

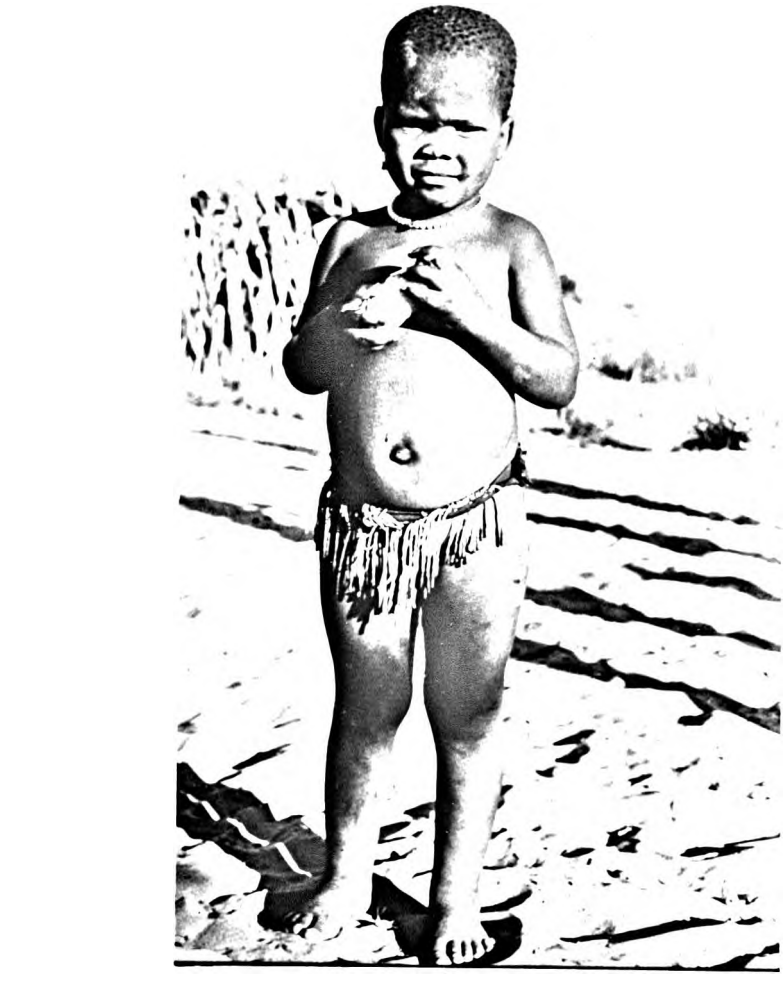




FIGURE 1.2 The Khoisan people

To judge by their modern close resemblances, the divergence between the Khoi and the San must have happened relatively late. The separation of the Khoisan into two main stocks must have begun only when one segment, the Khoi, accepted and started a herding lifestyle to exploit the bovine cattle of African provenance.

Genetic traits which were first identified in southern African Negroids were found to be commoner in the Khoisan and absent in Negroids from other parts of Africa. This suggest that Khoisan genes were introduced into southern African Negroids as a result of genetic admixture from the Khoisan.

A good example is that of the variant at the ABO red cell antigen locus, inappropriately called $A^{Bantu.}$ It is common in the San, rare in Negroids and absent in Caucasoids. The *Gm* allotype system has also been very informative, the typical Negroid haplotypes $Gm^{1,5,6,14}$ and $Gm^{1,5,13,14}$ are present in the San (Jenkins *et al*, 1970), but the San and the Khoi are unique among all indigenous sub-Saharan Africans in having the $Gm^{1,21}$ haplotype and the Khoi in having $Gm^{1,2,21}$ (Jenkins and Steinberg, 1966; Steinberg *et al*, 1975; Jenkins and Dunn, 1981). $Gm^{1,13}$ was found to be the commonest haplotype in the San (Jenkins *et al,* 1970) and the second most common in the Khoi (Steinberg *et al,* 1975) at a time when it had not been found in other African populations, but was shown to be

present among the Mongoloid people. It was therefore used to measure Khoisan admixture in the various Bantu-speaking populations giving results in keeping with cultural and linguistic characteristics (Jenkins *et al*, 1970).

The R^0 allele of the Rhesus system is present in high frequency in the Khoisan and Negroes but the *r* (or Rhesus negative) allele is virtually absent in the Khoisan and of low frequency in Negroids (Nurse *et al*, 1985).

Very little sickle cell haemoglobin is found in Negroids living south of the Kunene, Kavango or Limpopo rivers therefore one can assume that the Negroid people moved into southern Africa prior to the increase in frequency of the mutation in Central African response to selection by the malaria parasite. It is not surprising therefore that hamoglobin *S* is absent in the San. There is however a haemoglobin variant Hb D Bushman that is present in polymorphic frequencies in the San, but this variant does not appear to possess any selective advantage (Wade *et al*, 1967).

1.2.3 THE SOUTHERN AFRICAN NEGROIDS

The great majority of the present-day Negroid peoples of southern Africa descend from migrants from the North who brought with them a culture of settlement, agriculture, iron tools, linguistic elements and historical

tradition. It is therefore virtually impossible to place the southern African Negroid peoples in separate discrete categories. Despite all they have in common one can distinguish two stocks, the South-Eastern and the South-Western groups (Phillipson, 1977; Nurse, 1983). This thesis is mainly concerned with closely inter-related Nguni, Sotho-Tswana, Lemba, and Venda groups from the South-Eastern stock and the Herero and Damara groups from the South-Western one.

Anatomically the African Negroids have a brown to brownish black skin colour and black hair which is woolly or frizzy.

As discussed in the previous section, there is sufficient evidence to suggest that there has been genetic admixture between Khoisan and Negroid populations in southern Africa in prehistoric times. Using the $Gm^{1,13}$ haplotype as an index to measure gene flow from the Khoisan into Negroids, Jenkins *et al*,(1970) estimate that approximately 50% of the genes in the Cape Nguni were derived from the Khoisan, whilst the frequency of this haplotype was appreciably lower in other Negroid groups.

It has been known for a very long time that a proportion of Negroids were peculiarly resistant to infection by *Plasmodium vivax* malaria, but it was only with the demonstration by Miller *et al,* (1976) that there were good grounds for supposing that the Duffy blood group

antigens acted as receptors for *P. vivax* merozoites that the reason for this became clear. One of the sharpest distinctions between Negroids and the rest of humanity is the relative rarity of the Duffy blood group antigens Fy^a and Fy^b among the former and their commonness or universality in all other peoples. The San possess both Fy^a and Fy^b, the frequency of the latter tending to be slightly higher than that of the former, whereas in the southern African Negro populations the frequencies of both are so low that the contrast in their frequencies can hardly be significant. It is probable that all the Duffy positivity found in the southern African Negroes represents gene flow from the San (Nurse *et al*, 1985).

Sickle cell haemoglobin is virtually absent in Negroids south of the Limpopo River, with the exception of the Venda, but is found at high frequencies in western and central African Negroids. This suggests that the sickle cell trait probably occurred in Western and Central African people after the founders of the South-Eastern Bantu-speaking people had crossed the Limpopo into relatively malaria free areas. The Venda of the northern Transvaal, situated just south of the Limpopo River, are thought, from linguistic and cultural evidence (Jeffreys, 1967), to be the most recent immigrants into southern Africa, and some of their founders may have originated from the malaria belt region.

1.2.4 THE SOUTHERN AFRICAN CAUCASOIDS

The predominant origin of the Caucasoid people of the Cape was from Western Europe particulary the low-lying parts of North Western Europe. At the time of the founding in 1652 of the Dutch settlement at Table Bay, the men who formed the crews of the incoming ships were not only Dutch but also German, Irish and Scottish (Nurse *et al,* 1985). Immigration of Europeans to the Cape seems to have been a continuous process. In the seventeenth century there was a sudden influx of French Huguenots of great social diversity (Nurse *et al*, 1985). In the late seventeenth century slave trade from the west coast of Africa provided some slaves who stayed for a decade before being retransported to the East. This appears to have been long enough for the introduction of the characteristically West African gene for haemoglobin C (as well as that for HB S) into the Caucasoid gene pool, in which it today reaches appreciable but not polymorphic frequencies. Slaves were brought from Bengal, Ceylon, Madagascar, Mozambique and Malaysia. Descendents of these Malaysians together with admixture with other groups make up the "Cape Malay" community of today and some of their genes flowed into the forebearers of the present day Afrikaners.

In 1820 a significant influx of British settlers took place. This was followed during the latter half of the 19th century by more sporadic and less organized German,

Dutch and British immigrations as well as Ashkenazi Jews mainly from Lithuania. After the first World War there was an influx of Sephardic Jews and middle Eastern Christians: Lebanese, Syrians and Cypriots. From southern Europe there was a constant steady trickle of migration, reaching a peak in the years following the second World War (Nurse *et al*, 1985).

The Indians of South Africa do not comprise only Caucasoids, a good many of them might by virtue of their descent from both the Caucasoid and the "aboriginal" peoples of the sub-continent, be accounted hybrid. As with the Jews, the earliest immigrations of Indians happened effectively singly, and such descendants as they left were absorbed by either the Caucasoids or the hybrid population. It was not until the middle of the nineteenth century that Asiatic Indians arrived in any numbers and came to constitute a separate community (Nurse *et al*, 1985).

During the 19th century there was a large inland migration of Dutch (proto-Afrikaans) Caucasoid families from the Cape - The Great Trek - 1836. It would have been in these and subsequent generations that selection for a changed environment would have operated most strongly, and contributed most to the population structure of the southern African Caucasoid of today (Nurse *et al*, 1985).

With so varied a range of origin, the physical characters of the descendants of the seaborne migrants may be expected to cover a wide morphological spectrum most resembling an anthology of European types.

Botha (1972) in an analysis of English-speaking and Afrikaans-speaking populations at the Cape, based exclusively on the blood groups, concluded that the difference between them, and the extent of their differences from their parent populations in Europe, could be ascribed to different patterns of gene flow from non-Caucasoid peoples (as much as seven percent of African genes were found in these Caucasoid people). Investigations carried out in the laboratories of the Human Ecogenetics Research Unit in Johannesburg revealed occasional evidence of non-Caucasoid genetic contributions, mainly genes of probably Khoisan or Negro derivation such as A^{Bantu} in the ABO system the Gd^{A} and Gd^{A-} alleles of glucose-6-phosphate dehydrogenase system, PGM_2^2 at the second phosphoglucomutase locus, the Rhesus allele R^0, and haemoglobin C, were found in individual Afrikaners and a few English-speakers. This however did little to distort the characteristically North-Western European picture. Similarly, the Jewish population showed no significant differences from other Ashkenazi populations elsewhere (Lane *et al*, 1985). A small study of Indians in the Witwatersrand area produced gene frequencies comparable with the stated origins of the ancestors of the members of the samples

(Nurse *et al*, 1985).

With respect to heritable diseases, there are several conditions found at such high frequencies among the progeny of the seaborne migration and so rarely in the rest of the world that founder effect has certainly been responsible for their prominence. One of the most familiar examples is that of prophyria variegata, a dominantly inherited condition that affects 1 in 400 white South Africans and 1 in 250 Afrikaners, all of whom appear to be the direct descendants of one couple who emigrated from Holland in the 1680s (Dean, 1963). It has also been diagnosed in a number of "Coloured" patients, presumably descendants of the same couple.

After three centuries of evolution the population structure of the Afrikaners is still far from stable and there does not appear to be much prospect of its ever attaining uniformity.

1.2.5 THE SOUTHERN AFRICAN "COLOURED" PEOPLE

In South Africa the term "Coloured" is not referred to simply as someone of mixed ancestry, but someone into whose ancestry a Caucasoid element has entered. The other element may be Khoi, slave Negro, free tribal Negro or San, always ultimately a people originally African (Nurse *et al*, 1985).

Depending on the type of gene flow, three distinct categories of "Coloured" people can be distinguished:

The Cape "Coloureds", the Griqua and the Basters . The Cape "Coloureds" where despite the foregoing claim that the non-Caucasoid element is always ultimately African, it is not necessarily uniquely so a non-African, non-European strain like the Malay has entered into these people. The Griqua, enjoy less biological than social distinctness. The elements which have contributed to their genetic constitution are those which entered into the formation of the original Cape "Coloureds" though the Khoi part may well have been the most considerable. The Basters are hybrids between Khoi and Caucasoids. Unlike the Griqua, the Basters have cherished a marked pride of race. The children of a Baster who married a non-Baster hardly ever acquired Baster status (Nurse *et al*, 1985). The Griquas and the Basters have a political relationship whereas the Cape "Coloureds" and the Basters are biologically more closely related (Nurse *et al*, 1985).

The high concentrations of "Coloureds" in the Western Cape and in the Witwatersrand area and their proximity to medical centers has meant that a considerable amount of information on the health of these communities has become available.

Several inherited diseases in the hybrid community are undoubtedly due to founder effect. Lipoid proteinosis for example, is a rare autosomal recessive disease and of 200 cases reported worldwide 60 have been reported

from South Africa, twenty six of these from a small inbred hybrid community in little Namaqualand in the Northern Cape (Gordon and Botha, 1969). The remainder are found among the Caucasoid Afrikaners. The ancestry of all cases can be traced back to a couple who married at the Cape within a few years of the original settlement (Gordon *et al,* 1971; Heyl, 1970). Unlike the southern African Negroes, the hybrid peoples show a very low incidence of polydactyly and according to Stevenson *et al* (1966) this is about 0.98 per 1 000 births.

1.3 GENETIC VARIATION AND THE POLYMORPHISM CONCEPT

In the human body there are 50 000 - 100 000 different enzymes and proteins that are composed of amino acid sequences coded for by DNA sequences in the genome. Thirty percent of these gene-coding loci have been shown to have different alleles or variants among and within individuals.

Genetic variation may be partitioned into two categories depending on whether or not the variation produces an obvious phenotypic effect. Inherited variation in the human genome has been studied for two main reasons. Firstly, because of its relevance to inherited Mendelian 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 its role in evolution.

The degree of variation and the possible number of combinations of loci make it very unlikely that two individuals of a species, except in monozygotic twins, would be genetically identical. The interindividual diversity is usually generated by different combinations of the more common alleles over a range of polymorphic loci and a significant proportion is also derived from rare alleles.

The term "genetic polymorphism" has been used to describe the extent of genetic variation . In 1965, Ford defined polymorphisms as: "the occurrence together in the same habitat of two or more discontinuous forms or phases, of a species in such proportions that the rarest of them cannot be maintained merely by recurrent mutation". Since then many definitions of the term were put forward but perhaps one of the most precise ones is that of Bodmer and Cavalli-Sforza (1976) which defines genetic polymorphism as "the occurrence of two or more alleles for a particular locus when at least two alleles appear with frequencies of more than one percent". This implies that the proportion if identifiable heterozygotes of the locus is at least 2% of the population (ie. $2pq=2x0.01x0.99=0.0198$). Alleles present at a frequency of less than 1 percent are termed variants (Miesler, 1983).

Genetic polymorphism was first demonstrated to exist in man in 1900, when Karl Landsteiner, demonstrated agglutination of the red cells of a proportion of human subjects by employing the naturally occuring antibodies found in the serum of other individuals. But it was the perfecting of the technique of starch gel electrophoresis by Smithies, (1955) and Smithies and Walker, (1956) that was mainly responsible for what appeared to be the "polymorphism explosion" during the late 50's and 60's. During the late 70's polymorphisms at the DNA level were discovered setting up the beginning of a new and enormous "polymorphism explosion". Biologists have since then been trying to find the biological and evolutionary significance of genetic polymorphisms. The degree of allelic variation and widespread occurrence of polymorphisms has lead to great debates and controversies over evolutionary origins. The debate on the relative importance of natural selection and random genetic drift continues with the opposing views of the Selectionists and the Neutralists.

The "Selectionists" maintain that allelic distributions are primarily the consequence of the action of natural selection over the course of many previous generations, on the flux of new mutant alleles which are being continuously generated at a low but steady rate. Mutants that occur in particular genotypic combinations are selectively advantageous and consequently generated

at a low but steady rate. Mutants that occur in particular genotypic combinations are selectively advantageous and consequently tend to reach higher frequencies and give rise to polymorphisms. Thus polymorphisms for the selectionists are always a consequence of differential selection (Fisher, 1930).

The "Neutralists" provide quite a different explanation for the matter; they suggest that the alleles which determine the enzyme and protein variants commonly observed in natural populations, particularly those that make up the polymorphisms which are selectively neutral or near neutral and their frequencies and distribution, are simply the consequence of random genetic drift (Kimura, 1968; Kimura and Ohta, 1971; Crow, 1973; Kimura, 1983).

Harris (1980) presents a critical evaluation of both these views and concludes that it is no longer possible to accept in their purest forms either the selectionist or the neutralist positions. To account for the extraordinary degree of individual diversity within species, which has been revealed by modern biochemical techniques and to relate this variation to differences between species and to evolution seems to require elements derived from both positions.

1.3.1 PROTEIN POLYMORPHISM

When techniques such as starch gel electrophoresis were developed, it became possible to identify polymorphisms in structural proteins and a host of different enzymes. A polymorphism in human haptoglobin was one of the first to be recognised by starch gel electrophoresis, when serum from different individuals was shown to have distinct characteristic patterns (Smithies, 1955; Connell *et al,* 1962; Smithies *et al,* 1962).

In the 1960's it was not known how common the protein polymorphism was, so, two groups set out to determine the incidence of this phenomenon. Harris and Hopkinson, studied variation in man and Hubby and Lewontin studied variation in *Drosophila pseudoobscura*. 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 polymorphisms were ubiquitous in naturally occurring populations (Harris and Hopkinson, 1972; Lewontin and Hubby, 1966).

By 1977, 104 different human enzyme loci had been screened for electrophoretic variation and 32% of these loci showed variation in at least 1 of the major population groups (Caucasoids, Negroids, Asiatic Indians and Orientals). Some of the electrophoretic polymorphisms were limited specifically to one population group (Hopkinson *et al,* 1976). These authors

obtained an estimate of the average heterozygosity per locus for common alleles that result in electrophoretic differences of 6.3%. This is obviously still an underestimate, as theoretically, one would expect approximately one third of the possible protein variants with single aminoacid substitutions to show a change in electrophoretic mobility (Harris, 1980). Neel, (1984) considered the impact of a number of 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. From enzyme studies, the heterozygosity was estimated to be 17-18%. These studies used 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 protein 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).

1.3.2 DNA POLYMORPHISMS

Polymorphisms at the DNA level can occur in intervening sequences, in the flanking regions of structural genes or in the coding regions of the genes themselves. In some cases it has been possible to infer the nature of the DNA changes which give rise to the polymorphic differences at the protein level. Only mutational differences in the DNA which are reflected after transcription and translation in protein differences will of course, be detected by this approach. DNA mutations that are not reflected in differences at the protein level can be divided into three main classes:
(i) Single base change mutations which alter a particular codon to a different one but which nevertheless codes for the same aminoacid.
(ii) Mutations which alter the regions of the gene which are transcribed into the 5′ or 3′ untranslated regions of the mRNA but are not translated into proteins.
(iii) Mutations in intervening sequences which are transcribed into RNA but are subsequently removed during RNA processing, so that they do not occur in mRNA.

DNA polymorphism was first used in 1974 as a tool for genetic analysis. It was used to map the temperature sensitive mutations of adenoviruses (Grodzicker *et al*, 1974). Kan and Dozy, (1978a) detected with the restriction enzyme *Hpa*I, the first human genomic RFLP at a site 3′ to the ß-globin gene.

Studies of DNA sequence variation enabled further estimation of human heterozygosity. It was initially estimated from data on the β-globin gene cluster region, that one variable site per 100 nucleotides was found to exist (Jeffreys, 1979). This estimate was amended to 1 in 200 nucleotides using population genetics theory (Ewens *et al,* 1981) leading to the suggestion of one heterozygous nucleotide pair per 1000 nucleotide pairs. Further data on the β-globin cluster region changed the estimate to 1 in 500bp (Kazazian *et al,* 1983). Similar results were subsequently obtained by Murray *et al,* (1984) for the human growth hormone gene cluster, and by Matteson *et al,* (1985) for the α_1-antitrypsin gene locus.

Cooper and Schmidtke, (1984) calculated human heterozygosity based on the formula developed by Nei, (1975). The average heterozygosity rate for the autosomes was calculated to be 0.00257. Using additional data Cooper *et al,* (1985) recalculated their estimate of heterozygosity for the autosomes to be 0.0039. This is in good agreement with the estimate of Hofker *et al,* (1986) of 0.0034. The latter authors estimate X chromosome heterozygosity to be 0.0009. Although a lower level of heterozygosity was found on the X-chromosome (Murray *et al,* 1982; Bakker *et al,* 1983; Aldridge *et al,* 1984) it was not significantly different from the rest of the genome (Cooper *et al,* 1985). "Ohno's Law" (Ohno, 1967) has predicted the conservation of the X-chromosome from identical localization of gene sequences in

different species. One explanation of this conservation may be that the X-chromosome contains a higher proportion of coding and/or regulatory sequences which are not at liberty to accumulate polymorphic variants to the same degree as the autosomes. To correlate heterozygosity at the DNA level with that observed at the protein level, one has to look at the distribution of RFLPs throughout the genome. A high degree of polymorphism is expected from the noncoding regions, when compared to the coding regions, because in the former nucleotide changes do not affect aminoacid sequences. Subsequently Kazazian *et al,* (1983) pointed out that of the 12 RFLPs identified in the β-globin complex, seven were in flanking DNA, three in intervening sequences, one in a pseudogene sequence and only one in the coding sequence of the β-globin gene.

1.4 DNA POLYMORPHISMS

1.4.1 RESTRICTION FRAGMENT LENGTH POLYMORPHISMS

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 by that enzyme. In terms of the classical defenition of genetic polymorphisms, the rarer form ("allele") must have attained a frequency of at least one percent. These genotypic changes can be recognised by the altered mobility of restriction

fragments after agarose gel electrophoresis.

RFLPs are inherited as codominant Mendelian traits and can occur in any region of the genome thus providing an inexhaustible source of genetic markers. These DNA fragments can be detected by hybridization using the method of Southern (1975), or by the polymerase chain reaction (PCR) (Saiki *et al*, 1988).

DNA polymorphisms have revolutionized human genetic analysis and have found general use in antenatal diagnosis (Weatherall, 1991), mapping of human linkage group (Botstein *et al*, 1980; Murray *et al*, 1982; Gusella *et al*, 1983) and analysis of the role of mitotic nondisjunction and recombination in inherited cancer (Cavanee *et al*, 1983; Fearson *et al*, 1984; Koufos *et al*, 1984; Orkin *et al*, 1984).

RFLPs were found in several human DNA systems, for example, in the region of the human globin genes, heterozygosity for a *Pst*I restriction site was found within the second intervening sequence of the δ globin gene (Lawn *et al*, 1978). Soon after this discovery, Kan and Dozy, (1978a) applied the technique of restriction endonuclease mapping to the human globin genes in a small population survey in the U.S.A. An RFLP, due to variation in some Negroid individuals, at a site 3′ to the β-globin gene was detected with the restriction enzyme *Hpa*I (Kan and Dozy, 1978a). In Caucasoids, the β-globin gene was found on a 7.6kb fragment but in

American Negroids and Asians variation was found, giving 7.6, 7.0 or 13.0kb fragments (Kan and Dozy, 1978a). The 13,0kb fragment was found to be associated with the sickle cell gene in some Negroid populations. This polymorphism was exploited to make possible the first prenatal diagnosis of a foetus at risk for sickle cell anaemia (Kan and Dozy, 1978b).

Since then an abundance of RFLPs associated with the β- and γ-globin genes have been examined (Jeffreys, 1979; Kazazian *et al,* 1983), and more recently the polymerase chain reaction (PCR) and the amplification refractory mutation system (ARMS) have been applied with great success for the detection of individual β-thalassaemia mutations in heterozygous parents and "at risk" fetuses (Old *et al*, 1990).

1.4.1.1 RFLP AETIOLOGY

RFLPs may be caused by two major types of mutations. The first is caused by one or more base pair (bp) changes leading to the formation or loss of a recognition site and the second is due to a DNA structural rearrangement of a block of nucleotides, such as a deletion, insertion, translocation, inversion or a duplication, including those caused by a variable number of tandem repeat sequences (VNTR).

1.4.1.1.1 BASE PAIR SUBSTITUTIONS

The most common form of RFLP is caused by mutations which lead to base pair changes in the genome, thus altering the restriction enzyme recognition sequences (Barker *et al*, 1984). The mutations are either transitions, where a purine is mutated to a purine or a pyrimidine to a pyrimidine, or transversions when the change is from a purine to a pyrimidine or a pyrimidine to a purine. The restriction fragment length so generated is increased or decreased depending on whether a recognition site was created or lost by the mutation.

Polymorphisms based on single base substitutions which are responsible for the addition or loss of a restriction site are limited in principle to two alleles and a maximum polymorphism information content of 0.375 (Botstein *et al*, 1980). By combining several base pair substitution sites into haplotypes, the information content of a locus can be improved. For a given number of polymorphic restriction sites, there is a maximum of 2^n possible combinations of sites, with an expected frequency for each of these haplotypes equal to the product of the probability for every individual site. The polymorphic restriction sites in the vicinity of the β-globin gene cluster provides a good example (Antonarakis *et al*, 1985).

1.4.1.1.2 DNA REARRANGEMENTS

DNA arrangements alter the sequence of the bases along a region of the DNA, thereby altering the positions of the recognition sites for a number of restriction enzymes. If an enzyme has a recognition site within the inserted or deleted region, the generated DNA fragments will behave as base pair substitution-type RFLPs where a site has been created within a larger fragment generating two smaller fragments or vice-versa.

Insertion-deletion polymorphisms are apparently rarer than base pair substitution RFLPs, but their PIC values are generally higher (Skolnick and White, 1982). The first arbitrary RFLP to be identified, by Wyman and White in 1980, was a polymorphism arising from a DNA rearrangement. Hybridization of a single copy 16kb genomic probe to an *EcoR*I restriction digest of DNA from 17 individuals revealed at least 8 different fragments in the size range of 14-29kb. Because of difficult resolution of bands, an accurate account of the number of alleles was not possible but 75% of the individuals tested were heterozygotes and the PIC of the locus considered to be 0.849 (Skolnick and White, 1982).

Certain regions of the genome may be more involved in rearrangements than others and have been termed "hypervariable". Polymorphisms at these loci are due to differences in the number of tandemly repeated sequences within a restriction fragment. These polymorphisms will

be discussed in detail in the sections 1.5 and 1.6.

1.5 VARIABLE NUMBER OF TANDEM REPEATS (VNTRs)

Approximately 30% of the human genome consists of a number of DNA segments characterized by a variable number of short tandemly repeated sequences (STR) (Britten and Kohne, 1968; Craig *et al,* 1988). Some of these repeated sequences are organised in tandem arrays, others are interspersed among less frequently represented sequences including structural genes. Loci with repeated motifs of a few base pairs (eg: 2 bp repeats) are often referred to as microsatellites (Litt and Luty, 1989), or simple sequence length polymorphisms (Tautz, 1989), while those with longer motifs (±15 bp) are referred to as minisatellites (Jeffreys *et al,* 1985a) or variable number of tandem repeats (VNTRs) (Nakamura *et al,* 1987). A simple VNTR locus can be separated into two distinct regions, the core and the flanking region. The core region contains at least one short sequence of basepairs called the "repeat". An almost invariant core sequence, GGGCAGGAXG, seems to be common to nearly all of the VNTR sequences (Jeffreys *et al,* 1985a). Nakamura *et al,* (1987) identified GNNGTGGG as a sequence common to 10 VNTR markers and Wong *et al,* (1987) established the core sequences of several highly polymorphic loci to be GGAGGTGGGCAGG AR(A or G)G, which also includes GNNGTGGG. Ondek *et al,* (1988) reported that the consensus sequence of many enhancer sequences,

$GNTGTGG^{AAA}_{TTT}$ is similar to GNNGTGGG suggesting a possible functional significance to these sequences.

The copy number variation of these core sequences reveals genetic variation several orders of magnitude greater than the classical serological and biochemical genetic markers.

Surrounding the core of tandem repeats, is a region of DNA with a sequence of base pairs different from the repeat. Cutting the DNA molecule into pieces by a restriction enzyme that does not cut the repeat produces a restriction fragment with a core of repeats and a sequence of base pairs on each side of the core called the "flanking region". Variation in the base pair sequence outside the tandem repeats can alter the restriction enzyme sites, yielding variation in the size of the flanking region and thus in the sizes of the fragments. In general, studies of RFLPs from loci that are not VNTR loci detect only a few alleles per locus (Nei, 1978) whereas for some VNTR loci this number might vary from a few to more than 70 alleles (Wong *et al*, 1986). Differences in the actual size of two restriction fragments could arise, as already mentioned, from differences in the sizes of the flanking regions or from the number of tandem repeats or even both (Fig 1.3).

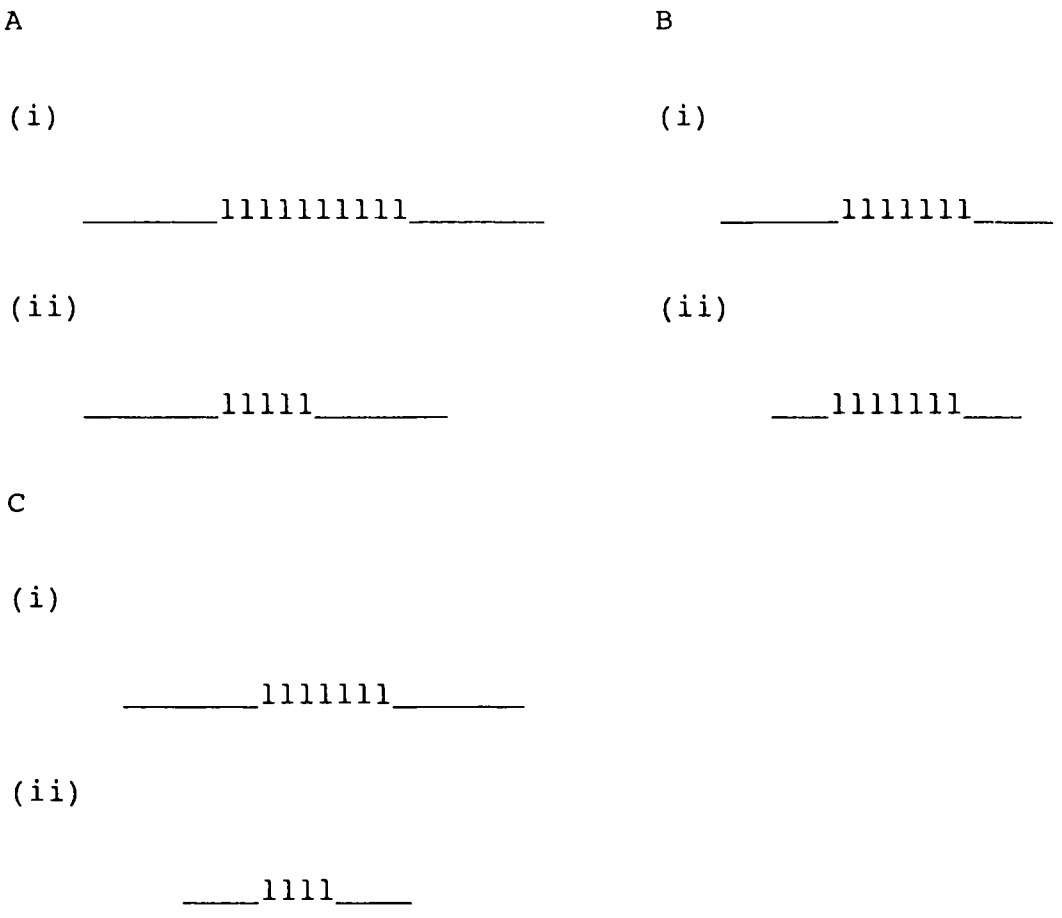


Fig. 1.3 Schematic representation of a VNTR locus. (A(i)-C(ii)) are restriction fragments. The vertical bars represent repeats and the horizontal lines are the flanking regions. Restriction fragments A (i) and (ii) differ from each other in the number of repeats and fragments B(i) and (ii) have the same number of repeats but differ in the flanking regions and fragments C(i) and (ii) differ from each other in the number of repeats and flanking regions.

A number of the major families of human interspersed repeated DNA sequences have been identified on the bases of their genomic organization, length, sequence of the repeat unit and copy number. For example hypervariable regions were described flanking several structural loci - the insulin gene (Bell *et al,* 1982, Ullrich *et al,* 1982), the Harvey-Ras oncogene (Capon *et al,* 1983), the α-related globin genes (Higgs *et al,* 1981; Proudfoot *et al,* 1982, Goodbourn *et al*, 1983), the zeta-globin gene and pseudogene (Goodbourn *et al*, 1983, Proudfoot *et al,* 1982) and the myoglobin gene (Jeffreys *et al*, 1985a). The partial homology that was noted among the repeats of these loci led investigators (Goodbourn et al, 1983, Proudfoot *et al,* 1982) to postulate that something inherent in the sequence may allow tandem expansion at the independent loci and that these loci may belong to a group of related sequences dispersed throughout the genome.

It was soon recognised that the human genome contains a large number of these polymorphic segments, numbering possibly in the thousands, and they have come to be known as "minisatellites" (Jeffreys et al, 1985a). Chromosomal localizations of many VNTR markers showed the loci to be scattered on many chromosomes and that there seems to be a bias toward location near telomeric regions (Donis-Keller *et al,* 1987; Kondoleon *et al,* 1987; Nakamura *et al,* 1988a; Royle *et al*, 1988). This bias however does not seem to be present in some other

species, such as mouse (Julier *et al*, 1990) and cattle (George *et al*, 1991).

1.5.1 MECHANISMS OF VNTR FORMATION

The mechanism(s) by which the number of repeats varies, leading to a wide spectrum of new alleles, is still speculative. There are suggestions that they are presumed to arise by mitotic or meiotic unequal exchange between misaligned repeats (Jeffreys *et al*, 1985a) or by DNA slippage at replication forks leading to the gain or loss of repeat units (Smith, 1976; Jarman *et al*, 1986; Jeffreys *et al*, 1986; Nakamura *et al*, 1987; Wong *et al*, 1987).

Jeffreys *et al*, (1985a) showed that some minisatellites have substancial repeat divergence, suggesting that these regions are not actively undergoing sequence homogenization by unequal exchange (Dover, 1982). Some highly polymorphic minisatellites also show high repeat copy number together with substantial sequence homogeneity of repeats. In addition it was also noticed that base substitutions in the repeat units of hypervariable minisatellites tend to be present in more than one repeat which indicates that these minisatellites are actively and repeatedly engaging in unequal exchange, resulting in the diffusion of novel base substitutions across more than one repeat unit (Dover, 1982).

If the "minisatellite" core sequences are indeed recombination hot spots (Jeffreys *et al*, 1985a) then a high frequency of crossing over could generate the variability by a high rate of unequal crossing over at the repeated sequences. This model allows flexibility for the generation of new alleles, both larger and smaller. It also provides an explanation for the apparent homogenization of base substitutions to other repeat units within the tandem array. A prediction of this model is that flanking markers surrounding the VNTR would be recombinant following an event that generates a new allele. Wolff *et al*, (1988) tested this prediction that flanking markers would be recombinant and concluded that no apparent exchange of flanking DNA sequences took place between homologous chromosomes in the generation of the new allele. Therefore they excluded simple models of crossing over that require flanking markers to be recombinant. Their data did not allow the exclusion of models that predicted two independent crossing over events within the short interval from the VNTR to the flanking base differences or gene conversion events. The latter allows single base exchanges to be involved into both the original and recombinant flanking repeat units. Wolff's observations were also consistent with a variety of pathways for the generation of new alleles such as unequal sister chromatid exchange, polymerase slippage and loop out deletion. If allelic variation at VNTR loci were due to

events between sister chromatids or within one chromatid then the homogeneity found within the series of tandem repeats between two alleles remains unexplained. Some variation in base pair sequence of the tandem repeat units is often observed and these same base substitutions frequently appear in a regular pattern in other units within the tandem array (Gill *et al,* 1985; Bell *et al,* 1982; Goodbourn *et al*, 1983; Proudfoot *et al,* 1982). This mechanism of single chromatid or sister chromatid exchange could lead to such homogenization of repeat units but only within the particular chromosome or allele (Wolff *et al*, 1988).

According to Deka *et al*, (1991) there may be more than one mechanism of production of new VNTR alleles and they might occur at different rates. Size variations involving a large change in allele size may be a relatively infrequent event whereas size variations of smaller magnitude could be more frequent. It can be postulated that the former type of change could be triggered by unequal recombination and/or sister chromatid exchange while the smaller shifts of allele sizes can be caused by some form of replication slippage or mutations within core sequences. While Wolff *et al,* (1989) demonstrated that unequal recombination is not a general mechanism of production of new VNTR-alleles, mathematically simulated "mutational" changes produced by unequal sister chromatid exchange could be modelled by the Infinite Allele Model (IAM) (Kimura and Crow,

1964). This model provides a test of the past effects of natural selection (Ewens, 1972; Clark, 1987). In a finite population with a constant rate of mutation to an infinite series of novel alleles as for example in the case of VNTRs, there will be a steady-state configuration of allele frequencies representing a balance between the mutational gain of new alleles and the loss of alleles by drift. A sample from a population may deviate from a true allele distribution by having an excessive frequency of the most common allele, and many very rare alleles, or by having an excessively even distribution. Small size shifts on the other hand may be due to replication slippage and can be forward-backward in nature and such changes can be reasonably explained by the Stepwise Mutation Model (Kimura and Ohta, 1978). This model has a significant property that distinguishes it from the conventional allele model: as the population size increases indefinitely while the product of the effective population size and the mutation rate is kept at a fixed value, the mean number of different alleles contained in the population rapidly reaches a plateau which is not much larger than the effective number of alleles (reciprocal of homozygosity).

1.5.2 FUNCTIONS OF VNTRs

It has been proposed that some of the sequences are involved in the regulation of expression of single copy sequences (Davidson and Britten, 1979) or that these DNAs may be of the "parasitic DNA" type (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Another explanation, as Jeffreys *et al*, (1985a) suggested, is that the VNTR sequences might encode hot spots for recombinational activity and thus account for their mutability through a very high rate of unequal exchange. They concluded that, after noting that the core sequence of the tandem repeats contain strong homology in length and G richness to *chi*, the cross-over hotspot initiator sequence GCTGGTGG of *Escherichia coli*. This sequence is believed to initiate generalized recombination by serving as the recognition sequence for the binding of the rec BC gene product, endonuclease V (Smith *et al*, 1981; Smith, 1983). It remains to be seen whether the core sequence is the functional homologue of *chi*. Steinmetz *et al*, (1986) also described a recombination hot spot in the mouse genome that contains tandem repeats with homology to the "minisatellite" core sequence of Jeffreys. Spandidos and Holmes, (1987) showed that the VNTR sequence at the 3′ flanking region of the *H-ras* gene has a function in transcriptional control. Ondek *et al*, (1988) reported that the consensus sequence of many enhancer sequences ($GNTGTGG^{AAA}_{TTT}$) is

similar to GNNGTGGG therefore showing a possible significance for the repeats and VNTR sequences present in a flanking sequence present in a flanking region or in an intron of certain genes (e.g. insulin *H-ras* and retinoblastoma), where they might have have a role in control of gene expression.

A fuller assessment of the numbers and classes of core sequences has to be done by sequence analysis of a large spectrum of minisatellites to pinpoint and make less speculative the function of the repeat sequences in the human genome.

1.5.3 APPLICATIONS OF VNTRs

Given that most RFLPs are diallelic and would be uninformative in most pedigrees, the prior odds of detecting linkage between a disease locus and a random marker in a given pedigree is even lower. This problem would be easily circunvented by using more highly polymorphic markers such as tandem repeats.

VNTRs, due to their great variability, have been instrumental in the genetic mapping of several disease causing genes, genetic linkage studies (Nakamura *et al*, 1988a; Risch, 1990), as tools for individual identification, forensic DNA fingerprinting (Jeffreys *et al*, 1985b; Giusti *et al*, 1986; Kanter *et al*, 1986; Coleman *et al*, 1990; Budowle *et al*, 1991a), in paternity diagnosis (Jeffreys *et al*, 1985c); Baird *et al*, 1986;

Jeffreys, 1987; Allen *et al*, 1989), bone marrow transplantation (Thein *et al*, 1986), in genetic counselling (Nakamura *et al*, 1987), population genetics (Lynch, 1988; Balaz *et al*, 1989; Devlin *et al*, 1990; Deka *et al*, 1991) and the addressing of a variety of biological issues including imprinting, loss of heterozygosity in malignancies (Nakamura *et al*, 1988b) and diagnosis of twin zygosity (Akane *et al*, 1991).

In animal genetics, more particularly in Bovidea, highly polymorphic markers such as VNTRs can similarly be used for individual identification and paternity diagnosis relying less on information of biochemical polymorphisms and blood groups systems and for mapping of the so-called economic trait loci (ETL) or genes involved in the determination of production traits (Georges *et al*, 1991).

1.6 $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ REPEATS

A subclass of eukaryote tandemly repeated DNA contains very short simple sequence repeats such as $(dC\text{-}dA)_n.(dG\text{-}dT)_n$, which can be referred to as $(CA)_n$ repeats, $(GT)_n$ repeats, short tandem repeats (STRs) or microsatellites (Miesfield *et al*, 1981; Sun *et al*, 1984; Gross and Garrard, 1986; Weber and May, 1989; Edwards *et al*, 1992). The number of copies of $(AC)_n$ repetitive sequences in different eukaryotic species varies from 100 copies in yeast to 200 000 copies in salmon (Hamada and Kakunaga, 1982). Mammals such as humans have an

estimated 50 000 - 100 000 interspersed $(CA)_n$ blocks with the range of (n) being 15 - 30 (Hamada and Kakunaga, 1982; Hamada *et al*, 1984a; Jeang and Hayward, 1983; Braaten *et al*, 1988). Uniform spacing of the $(CA)_n$ blocks throughout the genome would place them every 30 - 60kb (Weber and May, 1989; Stallings *et al*, 1991). These numbers of $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences in the genome have been estimated by hybridization of poly (dC-dA).poly (dG-dT) to large insert human genomic DNA libraries (Miesfeld *et al*, 1981; Hamada *et al*, 1982; Litt and Luty, 1989) or by hybridization of the same polymer to dot blots of total human DNA (Hamada and Kakunaga, 1982; Hamada *et al*, 1982; Sun *et al*, 1984). Estimates of the copy number ranged from 32 000 to 35 000 for colony hybridization and from 50 000 to 130 000 for the dot blot hybridization. Because all these estimates were determined by hybridization, they apply only to sequences with about 12 or more repeats (Weber, 1990). Knowledge of the approximate total number of $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences within the human genome, the length distribution of the repeat sequences and the dependence of informativeness upon repeat length permitted Weber, (1990) to calculate roughly the numbers of potential $(CA)_n$ markers in the genome as a function of informativeness. He calculated that there would be 35 000 $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences per haploid genome. Seven percent of the PCR primers tested yielded 3 alleles or more in single individuals, rendering them

useless polymorphisms and therefore they were excluded. From these and other observations Weber, (1990) calculated that about 12 000 $(dC\text{-}dA)_n.(dG\text{-}dT)$ sequences could be found in the human genome with PIC values of > 0.70, occurring approximately every 0.3 to 0.5 cM.

Several examples are known for which the number of repeats within a block of simple sequence tandemly repeated DNA varies among individuals within a species. The variation in the number of repeats appears to be a universal feature of eukaryote DNA regardless of the length of the repeat unit. Examples of these include, primate alphoid satellite DNA (Tyler-Smith and Brown, 1987; Waye *et al*, 1987), tandem repeat units, varying in length from 3bp to 243bp, within the malaria parasite antigen genes (Kemp *et al*, 1987; Weber, 1988) and minisatellites which have a repeat unit in the range of about 9 - 60bp (Jeffreys *et al*, 1985a); Nakamura *et al*, 1987). In addition to $(CA)_n$ dinucleotide repeats, other abundant simple sequence tandem repeats such as $(TC)_n$ and $(A)_n$ exist dispersed in mammalian genomes (Crabtree and Kant, 1982; Cheng *et al*, 1982; Htun *et al*, 1984) and seem to be pause or arrest sites for DNA replication and amplification (Baran *et al*, 1991).

Although GT repetitive sequences are widespread in eukaryotic genomes, it is not known whether the positions of $(CA)_n$ repeats have been extensively conserved along the length of the chromosomes.

Pardue *et al*, (1987) found by *in situ* hybridization that polytene chromosomes of distantly related Drosophila species had similar distributions of $(CA)_n$ repetitive sequences. Braaten *et al*, (1988) reported that similar restriction fragments containing rRNA genes from mouse, rat and human contained $(CA)_n$ repeat sequences and they concluded that the position of the $(CA)_n$ site has been conserved. The sequence of the human and rat insulin-like growth factor were studied by Rotwein *et al*, (1986) and Shimatsu and Rotwein, (1987) respectively and it showed a conservation of (CA) repeat positions. Stallings *et al*, (1991) observed that conservation of $(CA)_n$ positions can only be definitively established by DNA sequence analysis because homologous genes in different species may contain (CA) repetitive sequences in close proximity but distinctly different positions. An example of (CA) repetitive sequences that occur within different locations of the same gene in different species is the urokinase plasminogen activator gene and haptoglobin genes. Stallings *et al*, (1991) concluded that there does not appear to be extensive conservation of $(CA)_n$ positions in evolutionarily distant species and that the few observed examples of conservation of $(CA)_n$ positions indicated that some $(CA)_n$ sites are of very ancient origin.

With respect to informativeness it has been shown to increase as the average number of repeats at that locus increases. It is very important to notice that

informativeness for sequences with about 10 or fewer repeats is very low and, attempts to develop polymorphisms for such sequences are probably not worthwhile. For sequences with about 11 - 15 repeats, the informativeness of the resulting polymorphisms is variable. However for sequences with 16 or more repeats, moderately to highly informative polymorphisms always occur (Weber, 1990). Imperfections within the repeat sequence tend to reduce the informativeness of the resulting polymorphism relative to what could be expected for the total number of repeats.

1.6.1 MECHANISM OF FORMATION OF $(CA)_n$

If selection does not play a major role in the determination of $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequence length (and the wide variation in lengths of the blocks of repeats indicate that it probably does not), then heterozygosity at the $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphism loci should be determined primarily by mutation rate (Jeffreys *et al*, 1988b; Kimura, 1989). Knowledge of the molecular mechanism of mutation is necessary for the construction of hypotheses concerning the mutation rate of repeat sequences. Several different mutation mechanisms for tandem simple sequence repeats have been proposed but almost all invoke the process of strand slippage during replication or recombination (Streisinger and Owen, 1985; Levinson and Gutman, 1987a; Wolff *et al*, 1989 and Coggins and O'Prey, 1989).

The rate of strand slippage increases with increasing lengths of blocks of repeats for $(A)_n$ sequences in bacteriophage T4 (Streisinger and Owen, 1985), for $(CA)_n$ sequences in M13 (Levinson and Gutman, 1987b) and for $(CTCTT)_n$ sequences in *Escherichia coli* (Murphy *et al*, 1989).

Mutation of human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences *via* strand slippage is therefore consistent with the observed dependence of marker informativeness (or heterozygosity) on numbers of repeats. Apparently the rate of strand slippage for $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences jumps, as the number of repeats exceeds about 10 (Weber, 1990). The lower informativeness of repeat sequences containing imperfections may be due to the fact that strand slippage of these sequences produces structures with non-complementary bases which are energetically less favourable than analogous structures derived from perfect repeat sequences (Weber, 1990).

1.6.2 FUNCTION OF $(CA)_n$

The fact that $(CA)_n$ repetitive sequences occur in genomes of all eukaryotes thus far tested suggests that they possess some function. This assumption was first put forward when it was shown that the $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences could assume a Z-DNA conformation in negatively supercoiled plasmids (Nordheim and Rich, 1983; Hamada *et al*, 1984a), however, there is no evidence for this occurring in eukaryotic cells (Tautz

and Renz, 1984; Gross *et al*, 1985). The predominance of the CA/GT and low proportion of CG repeat types added to the fact that the former type is often conserved across a wide range of species suggested that some mechanism possibly due to a functional role, is involved in the maintenance of at least some microsatellites (Moore *et al*, 1991).

The absence of GT repetitive sequences in bacteria (Gross and Garrard, 1986) could be an indication that such repeats play a role in packaging and condensing DNA into eukaryotic chromosomes. GT repetitive sequences are much more frequent in euchromatin than in heterochromatin, so these repeats could be the determining factor that distinguishes constitutive heterochromatin from euchromatin or facultative heterochromatin. GT repetitive sequences may provide a DNA conformation that can be condensed and decondensed during different parts of the cell cycle (Stallings *et al*, 1991).

It has also been proposed that these repeated sequences serve as hot spots for recombination (Pardue *et al*, 1987; Slightom *et al*, 1980). A large amount of unequal crossing over may take place within GT repetitive sequences since there is a great amount of variation in the length of these sequences at the same locus in humans (Weber and May, 1989). Although such observations do not prove the hyper-recombinability of these repeats,

they are consistent with such a hypothesis. Also consistent with this hypothesis is the observation that the genome of SV40 contains a short $(GT)_n$ repeat and that this virus frequently integrates into the rat genome near a GT repeat. The fact that B-satellite regions of Drosophila polytene chromosomes, which do not undergo recombination, are deficient in GT repeats further suggests that GT repetitive sequences play a role in recombination. However, observations made by Stallings *et al*, (1991) show no correlation of increased recombination with increased numbers of $(GT)_n$ repeats. In general, the mouse genome undergoes significantly less recombination than the human genome, yet mice contains a greater number of $(GT)_n$ repeats. They went further and suggested that GT repetitive sequences could be a necessary but not sufficient condition for a recombinational hotspot to occur.

Finally there has been suggestions that these $(CA)_n$ blocks participate in gene regulation (Hamada *et al*, 1984b). The $(GT)_n$ repeat could enhance gene activity from a distance independent of orientation, but more effective transcriptional enhancement resulted from the GT repeat being closer to the promoter sequence. A recent study has shown a down-regulatory role for the DNA microsatellite (GT) type upstream of the rat prolactin gene (Naylor and Clark, 1990).

In conclusion one can say that, any possible functional significance of the $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ repeats remains speculative and open to discussion.

1.6.3 APPLICATIONS OF $(CA)_n$

Because of the very large number of $(CA)_n$ blocks in the human genome, the frequency with which they occur, the high degree of polymorphism displayed and their random distribution across the genome (Luty *et al*, 1990) make them potentially very useful as DNA markers in gene mapping studies and linkage studies, not only in humans but also in other mammals (Weber and May, 1989; Lewis *et al*, 1990; Petersen *et al*, 1990 and Moore *et al*, 1991), and in DNA typing (Edwards *et al*, 1991). The literature is rapidly growing on repeats within several genes which are being used for diagnostic work (Taylor *et al*, 1989; Lalloz *et al*, 1991; Petersen *et al*, 1991).

In addition, GT repetitive sequences are being utilized as landmarks in physical mapping studies that involve fingerprinting cosmid clones (Stallings *et al*, 1990).

1.6.4 ADVANTAGES OF $(CA)_n$ TYPING

(CA) typing is done using the polymerase chain reaction which has several advantages over standard blotting and hybridization techniques. It requires less DNA is faster and less expensive. Two or more microsatellites may be analysed simultaneously (Weber and May, 1989; George *et al*, 1991) opening new possibilities for the genetic

analysis of large numbers of samples.

1.7 APPLICATIONS OF POLYMORPHISMS

The advent of DNA cloning, Southern blot detection of single copy genes, the polymerase chain reaction and sequencing of nucleotides have revolutionized human genetics over the last 15 years. It has provided the opportunity for the construction of genetic linkage maps, the mapping of several diseases to specific chromosomal regions and prenatal and preclinical diagnosis of a number of single gene disorders. Other applications of RFLPs include studies on the population genetics of linked haplotypes of RFLPs detecting loci which have lead to the discovery of meiotic recombination hotspots in gene clusters (Chakravarti *et al*, 1984) and which have illuminated our understanding of the emergence of *Homo sapiens* and the diversification of human races (Wainscoat *et al*, 1986).

1.7.1 HUMAN EVOLUTION

While the practical advantages of repeat (VNTR) polymorphisms over the classical isozyme and RFLP loci is widely documented through the fast growing literature, relatively little is known about the distribution of alleles at VNTR loci in ethnically well-defined human populations. Despite their demonstrated usefulness in paternity testing, forensic medicine and genetic linkage studies, the population genetic

characteristics of VNTR polymorphisms are less precisely defined compared to the classical genetic markers. The most noteworthy studies in this regard are those of Baird *et al,* (1986) which have shown extensive variation in three ethnic groups (American blacks, Caucasoids and Hispanics) using the two highly polymorphic DNA sequences at the D14S1 locus and the one flanking the HRAS-1 region. Flint *et al,* (1989) showed that Melanesians have high heterozygosity compared with Polynesians at a number of VNTR loci, implying therefore that the latter have passed through several population bottlenecks. Balaz *et al*, (1989) showed significant differences between American blacks, Caucasoids and Hispanics when hypervariable probes at the D2S44, D14S1 and D15S13 loci were used, and small differences for the D17S79 and DXYS14 loci. Deka *et al*, (1991) investigated the allelic distribution of six well characterized VNTR loci (D1S57, RB1, D1S77, D1S61, α-globin 5'HVR, D1S76) in three ethnically defined population groups, the Kachari of Northeast India, Kalam and Gainj speakers of Papua New Guinea and Dogrib Indians of Canada. He suggested that more than one mechanism for the production of new VNTR alleles is needed and indicated that increased heterozygosity at VNTR loci in comparison to protein and blood group loci may lead to a more accurate estimate of genetic distances. Edwards *et al,* (1992) did the analysis of a multilocus genotype survey of 97-380 chromosomes in U.S. Blacks, White, Mexican-

American and Asian populations at five STR loci located on chromosomes 1, 4, 11 and X. They found unimodal, bimodal and complex distributions among the populations as a whole.

Although these studies indicate that the polymorphisms revealed by VNTR loci can be used to address several interesting population genetic issues, the unavailability of data on other VNTR loci among genetically diverse populations limits the scope of characterizing the population genetic features of the human race using these types of polymorphisms. This will be expanded in the discussion.

In an attempt to contribute to this field, the present study analyses the allelic distribution of three VNTR loci and two $(CA)_n$ repeat loci among 17 southern African population groups.

1.7.2 PATERNITY TESTING

Once information using VNTR loci was obtained for the different populations, another way of using the data was to apply it to assessing its usefulness in paternity testing.

1.7.2.1 PRIOR TO DNA TECHNOLOY

A need to confirm family relationships or to resolve disputed parentage has always existed. Methods for resolving the problems this need has created can be

traced back to biblical days. The first published case was one of disputed maternity which was resolved by King Solomon (Scriptures I Kings 3: 16-27). When his proposed approach was to cut the child, whose lineage was in question, in half so each claimant could share a part, only one claim was withdrawn leaving King Solomon with no doubt in deciding who was the true mother.

The Chinese, according to folklore from the 12th and 13th centuries employed interesting and unique blood tests to ascertain genealogic disputes. If a deceased person was involved in a dispute, blood would be dropped from the claimed relative on to the skeleton of the deceased and the relationship would be considered if the blood was absorbed into the skeleton. If two living individuals were involved, blood from each was dripped into a bucket of water, if the drops of blood flowed together and mixed, the relationship was "confirmed". Of course, people soon realised that these methods were not at all accurate and the search continued for a test that would give them an ultimate and infallible answer.

Progress in developing a reliable means of determining "blood" relationships has been very slow. In 1900, the delineation of ABO blood group types by Landsteiner and his students had dramatic importance not only for the would-be transfusionists but also for contemporary geneticists.

Further investigation by Von Dungel and Hirszfeld in 1910 gave proof that these traits were in fact inherited and therefore Mendel's concepts could be applied to humans. The next significant event occurred in 1927 when the MN system was discovered by Landsteiner and Livine, (1928). This was followed by the discovery of the Rh-Hr system in 1939-1940 by Levine and Stetson, (1939) and,independently,by Landsteiner and Wiener, (1940). The use of these systems allowed exclusion from paternity among 55% of falsely accused men (Shaw, 1983). In 1950, this percentage rose to 70% with the discovery of the Kell, Duffy and Kidd systems. The most celebrated court case on blood grouping and paternity was that of Charlie Chaplin in California in 1946. Although the tests indisputably established that he could not be the father the court ordered him to pay child support! At that time, blood group evidence was not considered to be conclusive in California courts. The jury was therefore allowed to weigh the blood group test results along with other evidence presented by the plaintif, Ms: Berry, in the determination of liability (Shaw, 1983).

It was not until the late' 60's, early' 70's, that the Human Leucocyte Antigens (HLA) system was recognised, and used to bring more credibility and legal acceptance to identity testing. It was calculated that the HLA system alone could distinguish wrongly accused men in approximately 93% of the cases (Walker *et al,* 1987). Until the development of methods of DNA typing,the use

of blood groups, red cell enzyme, serum proteins and HLA types has provided a practical way of excluding at least 95% of putative fathers. This, however, leaves judges, attorneys, parents and children with as much as a five-percent margin of doubt, a margin that often clouds issues and bogs down court cases.

There are approximately 62 immunologic and biochemical systems that have potential application in paternity testing. If all were to be used in each case of questionable parentage, it would be theoretically possible to establish a non paternity in over 99.9% of cases. Because of the time and expense involved, this is not feasible and is therefore not recommended. Thus, the quest for the ultimate paternity test has continued. Ideally such a test would make use of easily distinguishable determinants and would include close to 100% of true fathers and exclude close to 100% of wrongly accused men.

1.7.2.2 AFTER THE ADVENT OF DNA TECHNOLOGY

The application of modern DNA typing techniques to paternity testing can (accepting that the use of Bayes' theorem is valid) provide probabilities of paternity greater than 99.9%. Evidence of this nature is so strong that it cannot be ignored and will streamline and facilitate court proceedings.

RFLPs are inherited as co-dominant Mendelian traits and can be detected by Southern blotting methods as well as by PCR amplification techniques. The ability to detect RFLPs in DNA derived from biological evidence is making a major impact on the field of individual identification in paternity testing and forensics (Gill *et al*, 1985; Thein *et al*, 1986; Baird *et al*, 1986; Nakamura *et al*, 1987; Jeffreys, 1987; Cahill *et al*, 1990; Coleman *et al*, 1990; Budowle *et al*, 1991a; Helminen *et al*, 1991).

Polymorphic systems suitable for use in cases of disputed parentage should meet a number of requirements:

(i) Gene frequencies in the relevant populations have to be known when calculating the "paternity index" and "probability of paternity".

(ii) The true genotypes of the tested individuals have to be distinguishable.

(iii) The frequency of mutations must be known.

By using a battery of DNA probes and restriction endonucleases, sufficient data can be obtained to produce a composite profile which is unique to an individual, except in the case of monozygotic twins. It should be kept in mind that this potential may be limited by the number of loci analysed and by the quantity and quality of the DNA available for typing.

Of particular value and power are the loci at which VNTRs occurs (Wyman and White, 1980; Jeffreys *et al*, 1985c; Nakamura *et al*, 1987). These hypervariable DNA

regions provide the best avenue to exclude an individual who has been falsely accused in a paternity suit or has been falsely accused of commiting a crime (Jeffreys *et al*, 1985b; Baird *et al*, 1986, Allen *et al*, 1989; Yokoi *et al*, 1990). When the DNA evidence does not exclude the accused the next step is to calculate the probability of observing the genotype in an individual drawn at random from the population. Estimates of the frequencies of the alleles of the loci investigated is therefore important if not essential. This may however be rather difficult to achieve for VNTR loci whose alleles can differ from one another by the length of a single repeat of the core element, and the resolution of such closely spaced bands is close to the resolving capabilities of the gel electrophoresis system used. Thus the true number of alleles at a locus may remain unknown. This limitation will be reflected in a frequency data base and thus the level of resolution of the gel system will affect the conclusions regarding inclusion of alleged fathers. It is therefore important to realise the resolving capabilities of the electrophoretic systems used and to be aware of the fact that the fragment sizes cannot be determined without correction for "measurement error". Various approaches have been developed (Baird *et al*, 1986; Budowle *et al*, 1991a; Devlin *et al*, 1991) and we opted for the fixed-bin analysis where alleles of similar size are grouped into frequency "bins". (Reeders *et al*, 1985; Wainscoat *et al*, 1986; Budowle *et al*,

1991a). The effect of binning is to underestimate the true number of alleles and hence reduce the discriminatory power of the system, this will be discussed in the next section.

DNA technology is now entering a new phase of development as interest is focused on the polymerase chain reaction (PCR). Paternity testing is among the more recent PCR applications, with various methods used to analyse the amplified DNA (Gyllensten and Erlich, 1988; Saiki *et al*, 1989; Jeffreys *et al*, 1986; Horn *et al*, 1989; Helmuth *et al*, 1990).

One of the most promising approaches to paternity appears to be PCR amplification of VNTR loci and direct electrophoretic characterization of the products. This approach does not require Southern blotting and hybridization with specific ^{32}P-labelled on non-radioactive labelled probes. A novel aspect is postmortem paternity testing using PCR amplification (Howard *et al*, 1991). As with any PCR amplification, its use in paternity testing should be the subject of careful evaluation and control. The extreme sensitivity of PCR is responsible for both its power and its weakness as a technique. PCR may be difficult to optimize and standardize. Both false positive and false negative results are possible because of contamination and copying errors, thus special procedures to address these issues should be implemented in every laboratory

which uses PCR (White *et al*, 1989; Kwok and Higuchi, 1989).

1.7.2.3 LIKELIHOOD OF PATERNITY

Paternity testing always resulted in one of two possible outcomes:

- either the alleged father could not have fathered the child and so the laboratory does not need to proceed with the evaluation of the case, exclusion is in most cases considered definitive or,
- if the alleged father cannot be excluded, this does not prove paternity since there is a chance that other men exist who have copies of all of the paternally derived genes. The genetics odds ratio or paternity index (PI), favouring the putative father over that of a random man of the same race is calculated in such cases (Kaiser and Seker, 1985). If the group of observed paternal genetic markers in the child is collectively rare in the population, only a limited number of men in the general population could father the child and therefore the likelihood that the alleged father is the biological father is high. On the other hand if the parentally derived alleles are very common in the general population, the likelihood of paternity is low and more systems have to be used to provide an acceptable PI.

The power of exclusion (PEX) of a given probe can be estimated both theoretically and empirically. However, the large number of alleles at the hypervariable region loci, make theoretical estimates difficult. In such cases, an empirical method can be used to estimate the PEX value. It involves the random selection of paternity cases for which probe data are available and the substitution of the alleged father of case 1 for the alleged father of case 2. This moving of alleged fathers up one case number is repeated for all cases, the last one moves to the first case. Thus non-fathers are present in each trio. The frequency of exclusions of known non-fathers is thus estimated.

One further consideration in the analysis of paternity is the mutation rate at the locus under examination. Critical analysis of the mutation rates for a large number of VNTR loci is still lacking and the only recommendation is that only loci with low mutation rates should be used for paternity testing. The possibility of occurrence of recombinant alleles is important for the interpretation of paternity test results since they may lead to false exclusion of the alleged fathers or produce a maternal exclusion. Therefore it is essential that cases of parental exclusion be confirmed by more than one locus (Balazs *et al*, 1989).

1.7.3 OTHER APPLICATIONS

RFLPs, VNTRs and other "repeat" polymorphisms have, for the first time, provided an almost unlimited source of very useful genetic markers in man. The construction of detailed linkage maps of human chromosomes has become a real possibility and the gigantic task of mapping the human genome is well underway. It is hoped that by the end of year 2005, the complete human genome sequence will be accomplished allowing futher studies on the function of genomic sequences, causes of genetic diseases, multifactorial as well as monogenic, human variation and many other topics relevant to the well being of Humanity!

1.7.3.1 MAPPING THE HUMAN GENOME

The goal of the Human Genome Project (HGP) is nothing less than a complete understanding of the genetic basis of *Homo sapiens*, including the genetic basis of disease in every population group and in every geographic region. It combines skills from the fields of biological and technological research, thus promoting futher interactions between scientific disciplines. The combination of these skills should stimulate many advances in both pure and applied fields of research.

The human genetic blue print - the human genome - contains an estimated 100 000 genes. Only 2 percent of these genes have so far been pinpointed to specific

chromosomal locations and only about 600 genetic diseases are understood at the molecular level (Human Gene Mapping 11, 1991). The HGP aims to locate the position of all these genes and to read the genetic information encoded in them, including the aberrant information in genes responsible for diseases.

The impact this will have in the developed world includes the detailed understanding of many diseases such as cancer, cardiovascular disease, Alzheimer's disease, arthritis and autoimmune disorders and the interactions between environmental and genetic factors.

Complete maps and extensive sequencing data will also help us to understand and control many diseases that plague the developing world, in addition to those diseases prevalent in developed nations. The most efficient ways of improving the quality of life in developing countries is more likely to be by improving the general standards of living. This can be done through nutrition, housing and sanitation but the fight against malaria, Chagas disease, Schistosomiasis and many other infectious diseases will not be completely successful without an understanding of the organisms that cause them.

In addition to work on the human genome mapping project, sequencing of the genomes of "model organisms" will also be undertaken. The idea is to provide a base of comparative data against which human genome information

can more effectively be interpreted.

Although some critics argue that the HGP will damage biomedical research (Rechsteiner, 1991), it can be argued that it will have a great positive effect since it will provide new research opportunities through its discoveries.

1.7.3.2 PRENATAL AND PRECLINICAL DIAGNOSIS OF DISEASES

Although over 4300 Mendelian traits associated with the pathology of human inherited diseases have been catalogued (Mckusick, 1988), only a small fraction have been fully characterized and can be diagnosed either antenatally or preclinically.

It is now over a decade since Kan and Dozy, (1978a) described their novel procedure for the antenatal diagnosis of sickle cell disease by gene analysis using DNA derived from amniotic fluid cells. Their suggestion (1978b) that "this method could be useful in the antenatal diagnosis of other genetic disorders" has proved to be an enormous understatement. This was followed, a few years later by the identification and direct analysis of the single base-pair changes within the beta-globin gene causing sickle cell disease (Wilson *et al*, 1982; Chang and Kan, 1982; Orkin *et al*, 1982) and β-thalassaemia (Chang and Kan, 1980). These techniques have now been applied to a wide range of different human inherited disorders. The direct analysis of the mutant

gene which is very reliable, informative and hence the most desirable means of disease detection (detection of the mutation itself) and the indirect or gene "tracking" analysis which relies on the use of RFLPs either within, or closely linked to disease loci, have become almost routine. Cooper and Schmidtke, (1991) have compiled reports listing human inherited diseases which can be analysed and diagnosed at the DNA level. It serves both the genetic counsellor and the clinical geneticist as a quick reference source to current diagnostic possibilities in a very fast moving field, although such a list is out of date before it appears in print!

Amniocentesis which is usually performed between the twelth and the eighteenth week of gestation, has been the most widely used method of obtaining material for prenatal diagnosis by DNA analysis or other biochemical methods. Within the last five years, great strides have been made in chorionic villus sampling for earlier diagnosis (9-11 weeks) which permits the option of late first trimester termination if the results indicate an affected foetus. Fetal blood sampling, from the 15th week of gestation is also now more widely used. Chordocentesis offers the advantage that rapid karyotyping (Shah *et al*, 1990) as well as DNA analysis may be performed much more rapidly than with amniocentesis. Other laboratory tests can be performed which would not be possible with an amniocentesis, as for example in the case of the rare but severe

oligohydramnios (Ulm *et al*, 1990).

For prenatal diagnosis by linkage analysis, family studies must be carried out in order to determine which of the parental chromosomes carry the mutant alleles. Parents must be heterozygous at the marker locus for there to be enough information for the test to be carried out and, although this is a limitation to the procedure, the increasing number of highly polymorphic RFLPs becoming available are improving the possibility of diagnosis. A second limitation is that the precision of the risk estimate is limited by the frequency of recombination between the marker and disease locus. Linkage must be very close in order to have sufficiently definitive risk estimates.

1.7.3.2.1 DIRECT DIAGNOSIS OF MUTATIONS

Gross insertions, deletions or rearrangements where more than several hundred nucleotides are involved are easily discernible because they cause differences in the size of restriction fragments generated by endonuclease digestion of genomic DNA or cause total absence of the band if the gene of interest is entirely deleted.

Point mutations can be detected if they interrupt a known restriction enzyme site or create a new one, thereby altering the size of the restriction fragments observed. The sickle cell mutation is probably the best known example of a single nucleotide alteration removing

a recognition site for a restriction endonuclease (Chang and Kan, 1982; Orkin *et al*, 1982). A single base pair change can also be detected by using chemically synthesized specific oligonucleotide probes.

The direct detection of known mutations is also possible by using the polymerase chain reaction (PCR) technique followed by sequencing of the DNA thus amplified. This approach was used by Cutting *et al*, (1990) when studying the nucleotide sequences of exons 9, 10, 11, 12, 20 21 and 22 of the nucleotide-binding folds (NBF) of the cystic fibrosis transmembrane conductance regulator.

Another way of direct detection of a known mutation is by a modification of the PCR technique called the amplification refractory mutation system (ARMS) which is based on specific priming of the PCR (Newton *et al*, 1989a and b; Old *et al*, 1990). To diagnose a specific mutation one requires two oligonucleotide primers identical in sequence except for the terminal 3′ nucleotides. One primer is complementary to the normal DNA sequence and the other to the changed nucleotide in the mutant DNA. For a primer to act as a template for the DNA polymerase enzyme the terminal 3′ nucleotide has to be perfectly matched to the target DNA sequence. Under carefully controlled conditions a primer with its terminal 3′ nucleotide mismatched, will not anneal properly and no amplification occurs.

1.7.3.2.2 INDIRECT DIAGNOSIS OF DISORDERS OF UNKNOWN AETIOLOGY

Anonymous DNA fragments and candidate gene probes have been used for linkage analysis and, when close linkage is found, they have been applied successfully for prenatal diagnosis of single gene disorders.

This approach is also known as the method of "positional cloning", where polymorphic markers and large families with affected members are used to link a DNA fragment to a disease locus. Once linkage has been found, various techniques may be used to identify the critical portion of the DNA in order to identify the gene and its product and eventually, the genetic mutation(s) responsible for the disease phenotype. Although this idea is extremely attractive, each disease locus presents a new set of problems in order to identify the gene and it can take years before a gene is found. One example is that of the Huntington's disease locus. It is now over eight years since the locus was assigned by RFLP linkage analysis (Gusella *et al,* 1983) but, as yet, the gene that is responsible for the condition has not been found, despite the intense work that has been carried out by several groups. As in the case of the Huntington disease locus, there are many other disorders where assignments by linkage or other method have been achieved and it will only be a matter of time before the disease causing gene(s) are cloned.

The mapping and isolation of genes responsible for many genetic disorders mark the beginning of new and very challenging chapters dealing with a better understanding, management and prognosis of diseases. The study of single locus diseases are paving the way to understanding and unravelling polygenic traits.

1.8 AIMS OF THE PROJECT

The variety of southern African populations provides an unique opportunity to study polymorphisms in different populations. With the ongoing interest in evolution of man in Africa, polymorphisms at the hypervariable regions of the genome are expected to shed more light in this fascinating field.

The use of highly polymorphic loci, including HVR and $(CA)_n$ regions, have not yet been used widely in the study of genetic variation in human populations. They present a more complex set of markers where alleles may not necessarily have arisen only once and where the mutation rate is thought to be increased. The main objective of this study was to assess their usefulness in the further elucidation of the relationships between southern African populations. More specifically the aims can be listed as follows:

(i) To study allele distributions of three HVR (VNTR) and two $(CA)_n$ loci in 17 different southern African populations representing four major racial groups namely: Negroid, Khoisan, "Coloured" and Caucasoid.

(ii) To use this information to study population affinities based on inter- and intra-population variation with a view to reconstructing the ethnohistory of southern Africa.

(iii) To assess the feasibility of parentage testing using these hypervariable probes in the local context.

CHAPTER TWO

2 SUBJECTS, MATERIALS AND METHODS

This chapter discusses the materials and the techniques used in this study which were modified throughout in order to give clearer, quicker and more consistent results. Only the improved methods are given. A list of the sources of reagents are provided in Appendix A and the components of buffers, media and other solutions are given in appendix B.

Human genomic DNA was isolated from buffy coat samples, muscle or brain tissue. The latter two were from deceased subjects where paternity testing was needed. The DNA was analysed in 3 ways: (i) by restriction endonuclease digestion, Southern gel transfer and hybridization to specific VNTR probes. (ii) By the polymerase chain reaction (PCR) method, polyacrylamide gel electrophoresis (PAGE) and autoradiography. (iii) By PCR, agarose gel electrophoresis and direct visualization under Ultra Violet (UV) light.

2.1 SUBJECTS

The subjects of this study were southern African Caucasoids, who are recent immigrants, the 2 main indigenous races, the Negroids and the Khoisan, and the hybrid population group, the so-called "Coloureds".

Fourty six paternity cases involving biological mother, child or children and putative fathers were also investigated. Blood samples were collected from individuals who gave informed consent and in the case of children under the age of 16 the consent was obtained from the parents.

2.1.1 THE CAUCASOID POPULATION

The Caucasoid population in this study consisted of the Afrikaans-speaking, the Jewish (Jew), the Indian (Ind) and the English-speaking groups. The second and third groups are thought to be more biologically homogeneous compared with the English and the Afrikaans-speaking (Wht) since they tend not to marry outside their culture. The English-speaking people tend to mix more often and have Afrikaner, British and other European ancestors. All the groups share a biological affinity and are therefore grouped as the Caucasoid population. Blood samples were obtained from unrelated individuals living in the metropolitan city of Johannesburg.

2.1.2 THE SOUTHERN AFRICAN NEGROID POPULATION

In Southern Africa one finds many different tribal groups of Bantu-speaking Negroids. Due to urbanization many of these groups are no longer separated as distinct entities, and intertribal marriages frequently occur. In this study, samples were obtained from individuals

belonging to different tribal groups. They include the following Nguni-speaking people: The Zulus (Zu) who are descendants of a number of independent Nguni chiefdoms and live in an area which encompasses the Natal and Kwazulu. The Xhosa (Xh) people who are found in the Eastern Cape. The Swazi (Sw) people who live in the independent country of Swaziland and the Tsonga (Tso) people who established themselves in the escarpment of the northern Drakensberg. All these groups are similar in terms of their language and customs.

The Sotho-Tswana group of Bantu-speaking Negroids consist of three groups, namely: the Northern Sotho or Pedi (Pd) who live in the northern Transvaal, the southern Sotho (Sot) who live in the southern and eastern Transvaal and the Tswana (Tsw) who are from a wide area extending beyond the northern Cape province and the Transvaal to the Republic of Botswana. Although quite dispersed these groups share customs and their language is similar.

The Venda (Ve) Bantu speaking group, inhabit an isolated geographical area on the northern Transvaal, and have remained a distinct group with a language of their own that separates them from the other South-eastern groups (Wilson and Thompson, 1969).

The Lemba (Lem) are an itinerant Negroid tribe of metalworkers, potters and merchants who allegedly possess markedly semitic features and who exhibit a number of morphological and cultural traits that distinguish them sharply from their neighbours and suggest an East-African and Arab related origin (Van Warmelo, 1974). They are Bantu speakers living mainly among the Venda and the Karanga in the nothern Transvaal (Nurse *et al*, 1985).

The Herero (He) people are Western Bantu speaking Negroids living in Hereroland which share common borders with Bushmanland in the North, Botswana in the East and South-East and the rest with Namibia. Because of inhospitable land, a considerable number of Herero people live outside the homeland although they keep a close contact with "home".

The Damara (Da) people, are scattered around the Northern frindge of the Kalahari desert, have typical Negroid physical features but linguistically they are close to the Nama of whom they were slaves. Because of this they are sometimes referred to "Black Bushmen" (Nurse *et al*, 1985).

Blood samples were collected from Negro individuals belonging to the different chiefdoms. Most samples were obtained from urban dwellers in and around Johannesburg

and the ethnic affinities claimed by the individuals were accepted. Other samples like the Herero, Lemba, Venda and Damara were obtained during field trips to their area of origin, organised by the Department.

2.1.3 THE KHOISAN POPULATION

This fine example of hunter-gatherer peoples have occupied South and South-eastern Africa for many centuries. The incoming of the Negroid and Caucasoid groups to the region led to the confinement and isolation of the San and the Nama (Na) to the in-hospitable Kalahari Desert, making them distinct groups where some mutant genes and polymorphic loci have attained high frequencies due to random genetic drift, a decreased population size and geographic isolation. Blood samples from random San and Nama individuals were obtained during Departmental field studies in Namibia.

2.1.4 THE JOHANNESBURG "COLOURED" POPULATION

In South Africa the term "Coloured" is not simply someone of mixed ancestry but someone into whose ancestry a Caucasoid element has entered. The other elements may be Khoi, slave Negro, free tribal Negro or San, always ultimately a people originally African (Nurse *et al*, 1985).

The "Coloured" population in this study is made up of healthy blood donors living in the metropolitan area of Johannesburg.

2.1.5 PATERNITY CASES

Fourty six parentage dispute cases consisting of biological mother, child or children and putative father were investigated. These were taken at random from the cases that were assessed, on a routine basis, for paternity in our Department. Two unnusual cases of paternity dispute were solved using only DNA technology only, because one of the parties in each case was deceased.

CASE 1

A 29 year old Caucasoid male was killed in a road accident and the body was taken to the State mortuary for postmortem examination. At the time of the incident, the man was separated from his wife who was 12 weeks pregnant. She claimed that the deceased was the father of the child but his parents disputed this. The deceased's large estate was one of the central issues in this dispute.

Our laboratory was approached by the parents of the deceased 10 days after the accident to ascertain whether tissue from the deceased could be obtained in order to

carry out paternity tests that would exclude him from being the father of the unborn child. The body had been kept in a refregirator and muscle tissue from the left thigh was obtained and stored at -70°C until the child was born. When the male child was seven months old, blood was collected from him and his mother and DNA extracted from these samples and the preserved muscle.

CASE 2

A female Caucasoid foetus was delivered prematurely and died three days later. The mother's husband suspected his wife of infidelity and he was not prepared to pay the medical expenses surrounding the birth of the deceased infant if she was not his child. Legal proceedings were instituted and the body exhumed 14 days after burial. Muscle and brain tissue were removed for DNA extraction. Blood for DNA extraction was also collected from the mother of the child, the husband and the man accused of being the child's biological father.

2.2 VENOUS BLOOD SAMPLE COLLECTION

Venous blood samples were collected from randomly chosen consenting individuals into acid citrate dextrose (ACD) vacutainers. Usually 36-48mls of blood was obtained per adult individual, but in the case of small children appropriately small volumes of blood were collected. White blood cell were separated from 12mls of blood on

a hypaque-ficoll gradient and immortalized with Epstein-Barr virus (EBV). In this way a continuous source of DNA was made readly available. Buffy coats were stored at -20°C or -70°C for further processing if required.

2.3 IMMORTALIZATION OF LYMPHOBLASTOID CELL LINES

The success of lymphocyte transformation is dependent on a number of parameters namely: (i) age of sample at the time of transformation (ii) temperature during transport and (iii) freezing conditions for lymphocytes prior to transformation at a later stage.

2.3.1 COLLECTION OF SAMPLES AND TRANSPORT CONDITIONS

Venous blood samples (6-18mls) were collected into ACD vacutainers. Each sample was thoroughly mixed with the anticoagulant, and kept in a cooler bag containing frozen ice packs. Care was taken so that the blood samples did not actually touch the frozen ice packs. On arrival at the camp or the laboratory they were transferred to a fridge (4°C) until ready to be processed. In the case of a field trip the samples were transported to the laboratory in cooler bags.

2.3.2 AGE OF SAMPLE

The age of the sample is defined as the period from which the sample was obtained from the donor, (day 0), to the time the sample was processed in the laboratory.

Tests were done to determine the maximum age of a sample which would still allow efficient transformation. Samples were kept for 0, 24, 48, 72 and 96 hours before being proccessed.

2.3.3 HYPAQUE-FICOLL DENSITY GRADIENT SEPARATION OF BLOOD

Using a Pasteur pipette, 6mls of whole blood was layered on top of 3mls of Histopaque-1077 pre-warmed to 37°C and centrifuged at 1 400 rpm for 20 minutes. During centrifugation, erythrocytes and granulocytes are aggregated by ficoll and rapidly sedimented, whereas, lymphocytes and other mononuclear cells remain at the plasma Histopaque-1077 interface. (Fig 2.1).

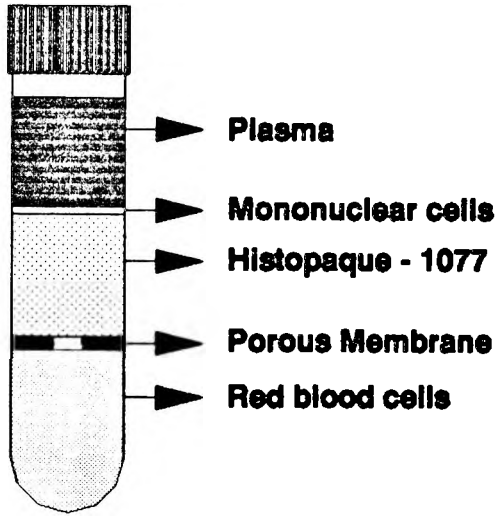


FIG 2.1 Histopaque-1077 density gradient separation of blood (adapted from Leucosep TM)

2.3.4 WASHING OF THE MONONUCLEAR CELLS

Most of the plasma was discarded and the mononuclear cell layer was transferred to a sterile tissue culture tube containing 10mls of Hanks balanced salt solution. The suspension was centrifuged for 10 minutes at 1 100 rpm and this washing step was repeated twice. Erythrocyte contamination was usually negligible and most extraneous platelets were removed by low speed centrifugation during the washing step. However, if a very bloody pellet was obtained after the first wash, it would be resuspended in 5mls of cold lysing solution (appendix B), and left on ice for 5 minutes. This would be followed by centrifugation and 2 extra washes in Hanks. Except for very few cases this lysing procedure was enough to produce clean white pellets.

2.3.5 INFECTION OF LEUCOCYTES WITH EPSTEIN-BARR VIRUS

Once the lymphocytes were separated from whole blood, they were resuspended in "Transformation medium" (RPMI-1640 containing 20% heat inactivated fetal calf serum (FCS), 1% L-glutamine, 100 μg/ml streptomycin, 100 IE/ml Penicillin and 2mg/ml cyclosporin A) and infected with Epstein-Barr virus.

2.3.5.1 EPSTEIN-BARR VIRUS (EBV) STOCK

EBV is a human herpes-virus which is endemic in all human populations. It has a double stranded DNA genome of about 172 000 base pairs which is linear in the virus particle, but circularizes inside the infected cells (Baer *et al*, 1984). The commercially available marmoset cell-line, B95-8, which harbours this virus was originally established by infecting marmoset B-lymphocyte with EBV from a human patient with infectious mononucleosis. This B95-8 line was grown in RPMI-1640 medium supplemented with 10% heat inactivated FCS. After 4 - 6 days in culture the supernatant was collected and centrifuged at 1 600 rpm for 10 minutes, following which the supernatant was filtered through a 0.22 μm sterivex-GS filter (Millipore) and the filtrate stored at 4°C until required. Using the method described below, virus containing supernatant was tested for its capacity to immortalize lymphoid cells. It was shown that the EBV kept its transforming ability for 4 - 6 months at 4°C.

2.3.5.2 CYCLOSPORIN A

Cyclosporin A (Cy A) is a selective immunosuppressive agent of great importance in human organ transplantations. The underlying mechanism is the specific inhibition of DNA polymerase II in T

patients treated with Cy A developed B cell lymphomas indicating that Cy A markedly facilitates the outgrowth of B lymphoblastoid cells in vivo. It was therefore suggested that Cy A could also inhibit B cell regression after EBV transformation *in vitro* (Bird *et al*, 1981). This is the case if Cy A is added within the first 24 hours after infection with EBV. Addition of Cy A 72 hours after initiation of cultures has no effect because it acts only during T cell activation and at that time T cells are already activated and functionally matured (Palacios, 1981; Tosato *et al*, 1982). Two mg/ml of cyclosporin A was added to the transformation medium and this was used to feed the cell cultures over a period of 2-3 weeks until clumps could be isolated by collection.

2.3.5.3 TRANSFORMATION OF B CELLS WITH EBV

Leucocytes at a concentration of $2x10^9$ cells/ml were resuspended in 1.5mls of transformation medium. This mixture was transferred to a $25cm^2$ sterile tissue culture flask (NUNC) and infected with 1.5ml EBV containing supernantant followed by an overnight incubation of 37°C in a 5% CO_2 atmosphere. Three mls of transformation medium were added on each of the 2 subsequent days.

2.3.6 CULTURING AND PROPAGATION OF TRANSFORMED CELLS

After the first three days of infection, the cultures were fed with 1 - 1.5mls of transformation medium every second day until clumps, which is the criteria indicating successful transformation, attained a size that could be isolated by collection. At this stage the transformation medium was replaced by RPMI-1640 containing 10% heat inactivated FCS and antibiotics and the cultures were subcultivated.

2.3.7 CRYO-PRESERVATION OF CELLS

Once the cells reached the stage of subcultivation they were ready for storage. Cells were collected in a tissue culture tube or spun down for 10 minutes at 1 100 rpm. The pellet was resuspended at a concentration of $5\text{-}10x10^6$ cell/ml in ice-cold freezing mixture (90% heat inactivated FCS + 10% DMSO). The suspension was transferred to freezing ampoules (NUNC) and placed in crushed ice for 30 minutes, followed by 30 minutes in dry ice and were then stored in liquid Nitrogen until needed for chromosome analysis or DNA extraction. For each sample, freezing was done on 10 different occasions.

If conditions in the laboratory were not ideal for transformation immediately after receiving the samples, for example due to a heavy workload, samples were

manipulated until the end of stage 2.3.4, frozen down as above and left in liquid nitrogen until ready to be transformed.

2.3.8 THAWING AND RECONSTITUTION OF LYMPHOBLASTOID CELL LINES

When the cells kept in liquid nitrogen were needed for transformation or expantion, the ampoules were removed from -196°C and immediately transferred into a 37°C water bath where they were thawed rapidly. These cells were added to 10mls of cold Hanks balanced salt solution in order to dilute the hamful effects of DMSO. The culture was centrifuged for 8 minutes at 1 100 rpm and the DMSO containing medium discarded. If the cells were in the lymphoblastoid stage (LCL) the pellet was resuspended in 5mls of RPMI-1640 supplemented with 10% heat inactivated FCS and transferred to a 25cm^2 tissue culture flask where they were incubated for 24 hours at 37°C in a 5% CO_2 atmosphere. The cells were usually ready for subcultivation 24 to 48 hours later. If cells were to be transformed, the pellet after being washed was subjected to the procedure described on section 2.3.5.3.

2.4 DNA PROBES

Two probes which detect VNTRs and were analysed by the method of Southern were used in this study.

pYNH24 PROBE

Probe pYNH24 (D2S44 locus) is a cloned human DNA sequence that has 30bp repeats, and it has been assigned to chromosome 2 by multi-point linkage analysis (Lathrop *et al,* 1987; Nakamura *et al,* 1987). This locus, is polymorphic in the flanking region when genomic DNA is cut with the restriction enzyme *Pst*I (Devlin *et al,* 1991). In each individual it generates either two polymorphic fragment sizes (heterozygote) or one (homozygote). The distribution of allele frequency estimates for this locus is usually multimodal (Devlin *et al,* 1991).

Odelberg *et al,* (1989) proposed that due to the rich polymorphic nature of locus D2S44 it would have a high power of exclusion and therefore would be one of the most powerful loci for parentage testing. This opinion however was not completly shared by Van Eede *et al,* (1990), who used *Taq*I to detect the VNTR. This author was more cautions and concluded that the DNA polymorphism obtained with this probe and enzyme should not be used in cases of disputed parentage if the paternity index has to be calculated, because exact

allele frequencies cannot be obtained

pYNZ32 PROBE

Probe pYNZ32 (D4S125 locus) has been assigned to chromosome 4p by linkage analysis (Nakamura *et al*, 1988c). This probe was reported to reveal a six allele VNTR identified by several enzymes including *Taq*I and *Pst*I, and of having a heterozygosity of 58% with *Taq*I in 94 unrelated Caucasians (Nakamura *et al*, 1988c). YNZ32 has been found to be closely linked to the locus responsible for Huntington disease (HD) located on the short arm of chromosome 4 (MacDonald *et al*, 1989). Although other VNTR markers have been mapped closer to the HD gene, pYNZ32 exhibits a higher degree of polymorphism, and may be one of the few loci unambiguously known to be centromeric to the disease gene, thus YNZ32 can play an important role in clinical determination of risk for this disease (Richards *et al*, 1991).

Both probes pYNH24 and pYNZ32 were kindly supplied by Drs R. White and Y. Nakamura from the Howard Hughes Medical Institute, Salt Lake City.Since each probe recognises a single locus, the DNA of an individual has either 2 polymorphic fragment sizes (heterozygotes) or one (homozygotes).

2.4.1 PLASMID TRANSFORMATION

Plasmid transformation is the process by which bacterial host cells take up naked DNA molecules, that are added to the surrounding medium. In this study, the recombinant plasmid DNA was transformed into *E. coli* HB101 when it was received as DNA. The $CaCl_2$ method for making *E. coli* cells competent was used (Mandel and Higa, 1970).

2.4.2 PREPARATION OF PLASMID DNA

The preparation of plasmid DNA includes amplification in the rich medium, luria broth (LB) lysis by sodium dodecyl sulphate (SDS) (Maniatis *et al*, 1982) and plasmid DNA purification by CsCl/EtBr density gradient centrifugation.

2.4.2.1 AMPLIFICATION OF PLASMIDS

The stock culture was streaked out on selection medium and a single colony with the correct antibiotic resistance markers was chosen to inoculate 10ml of LB medium containing the appropriate antibiotic. The culture was incubated overnight at 37°C with vigorous shaking to allow aereation and 0.1ml of this overnight culture was then incubated into a 25ml LB containing the antibiotic and incubated at 37°C for 2-3 hours or until an OD_{600} of 0.6 was reached (late log phase). The 25ml

culture was inoculated into 500ml LB pre-warmed to 37°C and containing the appropriate antibiotic in a 2 litre flask. Incubation proceeded for about four hours until a OD_{600} of approximately 0.4 was obtained.

2.4.2.2 LYSIS BY TREATMENT WITH SDS

The bacterial cells were harvested by centrifugation at 2 750 rpm for 20 minutes at 4°C and the supernatant discarded. In order to obtain cell lysis, the bacterial pellet was resuspended in 6mls of solution I containing lysozyme by sucking the suspension up and down the suspension or by vortexing it. It is essential to ensure that the bacterial pellet is completely dispersed in the above solution. It was then transferred to 200ml flasks and left on ice for 20 minutes. Twelve mls of 2N NaOH, 1% SDS were added, mix thoroughly but gently and left for a further 10 minutes on ice, at this stage the mixture becomes very viscous because the cell membranes are being lysed. In order to isolate cell debris, 7.5ml 3M sodium acetate pH 4.6 was added and left on ice for 20 minutes. The solution was then centrifuged at 15 000 rpm (Beckman JA20 rotor) for 15 minutes at 4°C. The smaller plasmid DNA molecules remain in the supernatant which was recovered and further treated with 50μl of 1mg/ml RNase for 20 minutes at 37°C in order to digest the RNA in the clear cell lysate.

Deproteinization was achieved by extracting the supernatant twice with phenol:chloroform:isoamylalcohol (25:24:1) and once with chloroform:isoamylalcohol (24:1). The DNA was precipitated by adding 2 volumes of ice-cold 95% ethanol. It was left at -70°C for 15 minutes (or overnight at -20°C) and was then recovered by centrifugation. The pellet was washed with 70% ethanol and dried. The DNA was dissolved in an appropriate volume (4ml) of TE (pH8.0) (Maniatis *et al*, 1982).

2.4.2.3 PLASMID DNA PURIFICATION BY CAESIUM CHLORIDE-ETHIDIUM BROMIDE DENSITY GRADIENT CENTRIFUGATION

CsCl/EtBr density gradient centrifugation is a very efficient method for separating supercoiled plasmid DNA from linear bacterial DNA, protein and RNA. The gradient was performed as follows: Exactly 1g of solid CsCl and 80μl of 10mg/ml EtBr were added per ml of DNA solution. This was transferred to a 5ml centrifuge tube, sealed and spun in a fixed angle rotor (Beckman Ti65) at 45 000 rpm for 36 hours at 20°C. After centrifugation two bands were apparent, a very thin upper band of chromosomal DNA and a thick lower band of plasmid DNA. The latter was removed by inserting a syringe into the side of the tube (Maniatis *et al*, 1982).

The ethidium bromide was removed from the plasmid sample by repeated extraction with isoamylalcohol. Once there was no trace of EtBr, the CsCl could be removed by dyalisis against TE (pH8.0). The amount of DNA was estimated on an agarose gel and the integrity of the plasmid was checked by appropriate restriction enzyme digestion and agarose gel electrophoresis with know DNA size markers.

2.4.3 RESTRICTION ANALYSIS OF RECOMBINANT PLASMIDS

In order to check the integrity of the recombinant plasmids, they were appropriately digested and the DNA fragments were electrophoresed in agarose.

Table 2.1 indicates the size of the insert, the vectors into which they have been cloned and the restriction endonucleases with which the fragment were excised. An aliquot of the digestion mixture was analysed and the fragments were electrophoresed on a 1% low melting point agarose gel until the bands were separated.

2.5 ISOLATION OF HIGH MOLECULAR WEIGHT HUMAN GENOMIC DNA

The extraction of high molecular weight genomic DNA from peripheral blood leukocytes, from EBV transformed LCLs and from tissue is discussed in the following sections.

TABLE 2.1 VNTR Probes

Probe	Vector	Size of Vector(Kb)	Size of Insert(kb)	Enzymes
pYNH24	pUC18	2.686	2.0	*Acc*I or *Pst*I/*Xba*I
pYNZ32	pUC18	2.686	3.3	*Hinc*II or *Pst*I/*Xba*I

2.5.1 DNA EXTRACTION FROM WHITE BLOOD CELLS

2.5.1.1 METHOD USING PHENOL EXTRACTION

The blood in the ACD vacutainers was centrifuged at 2 750 rpm for 10 minutes. The plasma was removed. The buffy coat or leukocyte layer was removed with a Pasteur pipette into a plastic ampoule and stored at -20°C or -70°C until required for the extraction of DNA.

The method of Sykes, (1983), was used with slight modifications, for DNA extraction from either whole blood or buffy coats. When whole blood is used, as in the case of paternity determination where only 2 tubes per subject is available, it is essential to remove the plasma and freeze the remaining cells before beginning the extraction procedure.

Thawed whole blood or buffy coats were decanted into a sterile 50ml plastic centrifuge tube (NUNC) and an equal

volume of cold 0.9% NaCl, 0.2% Triton X-100 was added. Triton X-100 is a non-ionic detergent which initiates cell lysis by the removal of lipid molecules from the cell membrane, thereby disrupting it. The mixture was centrifuged at 2 750 rpm for 15 minutes at 4°C and the supernatant was very carefully poured off, leaving a soft red-brown pellet. It was resuspended by shaking in 30ml of the NaCl/Triton-X mixture. Centrifugation at this stage usually resulted in a soft pale pink/white pellet after removal of the supernatant. If the pellet is still very red at this stage, the NaCl/Triton X-100 wash was repeated.

Further lysis was achieved by the addition of a lysing buffer (7M urea, 0.3M NaCl, 10mM EDTA, 10mM Tris HCl, pH7.5), drop by drop. The pellet was dispersed into this solution using a large diameter glass rod. Small volumes of the buffer was added to a total volume of 10mls. Finally lysis was completed by adding 2ml of 10% SDS and incubating at 37°C for 10 minutes.

Deproteinization was achieved with two organic solvents as it has been found to be more efficient than treatment with only one. Phenol and chloroform denature and precipitate proteins efficiently while leaving the nucleic acids in aqueous solution. Some RNA molecules, especially the mRNA, will be removed by the phenol treatment, but most will be retained with the DNA in the

aqueous layer (Brown, 1986). This however did not prove to be a problem when subsequently using the DNA.

Protein was extracted from the lysed white blood cells (12ml volume) with 10ml of phenol and 5ml of chloroform:isoamylalcohol. The samples were thoroughly mixed and centrifuged for 15 minutes at 2 750 rpm. The aqueous phase, which was separated from the organic phase, was collected and the phenol:chloroform: isoamylalcohol step was repeated. The aqueous phase was then extracted once with 5ml of chloroform: isoamylalcohol and after centrifugation it was removed into a clean Nunc tube. The DNA was precipitated using 2 ½ volumes of ethanol, rinsed in 70% ethanol, dried and dissolved in 0.5-1ml of TE buffer pH 8.0 for at least 24 hours on a rotation wheel. The DNA solutions were stored at 4°C for immediate use or at -70°C for more permanent storage.

2.5.2 DNA EXTRACTION FROM HUMAN LYMPHOBLASTOID CELL LINES

Lymphoblastoid cell lines derived from the method described in 2.3, consist of non-adherent fast growing cells. The culture medium was removed by centrifugation, at 2 750 rpm for 10 minutes, and the pellet resuspended and washed twice with PBS in order to remove any traces of the foetal calf serum that was part of the culture

medium.

Cell lysis was achieved by resuspending the pellet in the residual PBS from the last wash and adding, 5mls of lysis buffer (same as in 2.5.1) and 25μl proteinase K (10mg/ml). This was incubated at 45°C for 10 minutes, then at 68°C for 30 minutes and finally at 37°C for 90 minutes.

Deproteinazition was achieved by the two organic solvents phenol and chloroform. The samples were mixed vigorously and thoroughly with 10mls of phenol and 5mls of chloroform:isoamylalcohol and centrifuged for 15 minutes at 2 750 rpm. The aqueous phase was collected and the phenol:chloroform:isoamylalcohol step was repeated. The aqueous phase was then extracted twice with 5-10mls of chloroform:isoamylalcohol and the DNA precipitated and collected as above.

2.5.3 DNA EXTRACTION FROM HUMAN TISSUE

DNA was extracted from muscle or brain tissue by the method described in 2.5.1.1. which was slightly adapted. The tissue was frozen at -20°C at least overnight or at -70°C until required. The weight was noted before the homogenization step in lysing buffer, in order to allow the calculation of the DNA yeild at the end. Futher lysis was achieved by adding 10% SDS and incubating at 37°C for 10 minutes. The deproteinization step was

described in 2.5.1.1. The yield obtained from tissue was lower than the one obtained from buffy coats.

2.5.4 ANALYSIS OF HUMAN GENOMIC DNA

Five μl of each DNA sample was mixed with 5μl of Ficoll Dye. They were loaded, together with a serial dilution of a DNA of known concentration, onto an agarose gel containing ethidium bromide and electrophoresed in TBE buffer at 60 volts for 60 minutes. The gel was then put on a UV-transilluminator (Chromato-vue Transilluminator model TC312A) to check that the DNA was high molecular weight DNA and to get a rough estimate of the concentration.

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

2.6 ANALYSIS OF HUMAN GENOMIC DNA BY SOUTHERN BLOT HYBRIDIZATION

The most common means of qualitatively assessing the presence of a particular polynucleotide sequence in a

mixture of nucleic acids is by the molecular hybridization of the latter to membrane bound DNA (Cannon *et al*, 1985).

This is achieved by restriction endonuclease digestion of high molecular weight human genomic DNA, agarose gel electrophoresis of the digested fragments, their transfer into nylon membranes by the method of Southern blotting, hybridization of the membrane with specific radioactively labelled DNA probes and visualization of results by autoradiography.

2.6.1 RESTRICTION ENDONUCLEASE DIGESTION OF GENOMIC DNA

Restriction endonuclease are DNases that recognize specific oligonucleotide sequences, make double-standarded 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). There are three classes of restriction endonuclease (Type I, II, and III) derived from restriction modification systems in prokaryotes (Old and Primrose, 1980), but the one used in this study fell into the Type II category. Restriction enzymes are commercially available and their optimum assay conditions are given by the manufactures.

The restriction enzymes are kept in a storage buffer at -20°C.

Only one restriction enzyme, *Pst*I, was used for the digestion of genomic DNA in this study, since it identifies polymorphisms with high percentage of heterozygotes with both probes. *Pst*I was isolated from the bacteria *Providencia Stuartii* and the recognition sequence is CTGCA/G.

DNA restriction digests were carried out in 1.5ml plastic Eppendorf tubes at 37°C. Ten to fifteen μg of human genomic DNA were digested in a 50μl final volume. A 10 times concentrated enzyme buffer, as specified by the supplier, was added to the DNA sample. The last ingredient added was the restriction enzyme at 2 to 5 units per μg of human genomic DNA. (One unit of enzyme activity is defined as the amount of enzyme required to digest to completion 1μg of bacteriophage Lambda DNA in 1 hour at the specific temperature). In addition, 1μl of spermidine trihydrochloride was added to the reaction mixture to a final concentration of 4mM. Spermidine trihydrochloride is a polyamine and causes a B to Z conformation change in the DNA (Feuerstein *et al*, 1986).

The incubation proceeded from 4 to 16 hours, after which an aliquot of the reaction mixture (usually one tenth of the total volume) mixed with 5μl of loading dye (Ficoll)

was electrophoresed a short distance on a trial gel to determine whether the restriction reaction was complete. An even smear with little or no very high molecular weight DNA was considered to be completly digested (Fig 2.2). When the digestion was incomplete, more enzyme was added and the digestion mix incubated at 37°C for another 2 to 4 hours, this was usually successful.

Reactions were stopped by the addition of one-tenth volume of a ten times concentraction of bromophenol blue loading dye (0.1% bromophenol blue, 0.1M EDTA, 50% Glycerol) and heating the reaction mixture to 65°C for 5 minutes and then immediate transfering to ice. The chelation of Mg^{2+} by EDTA in the loading dye is an effective means of inhibiting the action of the *Pst*I enzyme. Incubation of the reaction mixture at 65°C just prior to load the gel, ensures distinct reproducible DNA fragment patterns. This is achieved by the dissociation of bound proteins and reduction in DNA-DNA associations such as "sticky ends" caused by certain enzymes (Fuchs and Blakesley, 1983).

2.6.2 AGAROSE GEL ELECTROPHORESIS

Agarose gel electrophoresis is a standard method used for separating DNA fragments according to their molecular weight. Agarose creates a matrix that acts as a filter in slowing down larger fragments and allowing

smaller fragments to migrate further. The migration of the fragments is inversely proportional to the logarithm of the molecular weight. Fragments ranging from 1 to 30kb can be effectively separated in agarose gels, whereas fragments smaller than 1kb are more effectively separated by polyacrylamide electrophoresis (Southern, 1979).

The agarose used was Seakem HGT and the concentration of the agarose in the gel varied, depending on the sizes of restriction fragments to be separated, the smaller the sizes, the greater the concentration of agarose used. The agarose was dissolved while stirring on a magnetic hot plate, in 1xTBE buffer (Appendix B). In order to visualise the DNA fragments after electrophoresis, ethidium bromide was added to a final concentration of 0.3μg/ml. The gel was poured onto a perspex plate 19.5cm long by 18.0cm wide, to a depth of 5-7mm. A set of loading wells, each 5mm wide and 1mm thick were 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. The buffer of choice was TBE since it is well-known to have a good buffering capacity especially during extended electrophoresis (Maniątis *et al*, 1982).

Samples of 5 to 10μg of genomic DNA and 20 ng of bacteriophage DNA marker and 10μg of sonicated salmon sperm were mixed with a running dye (ficoll dye) and

loaded into slots and electrophoresed at 40 volts (constant voltage) for 18-20 hours. The buffer was stirred 3-4 times during the running period. The marker was used to allow the estimation of the human genomic fragment size.

DNA in the gel was visualised over ultraviolet light on a transilluminator (chromato-vue Transilluminator model TC312A) ethidium bromide molecules intercalate into the DNA and fluoresce under ultraviolet light. Molecular weight markers, run simultaneously, were used to determine the size of the hybridizing genomic DNA fragments. These include bacteriophage Lambda C1857 digested with *Hind*III or a double digest with *Hind*III and *EcoR*I.

Gels were photographed under ultraviolet illumination on polaroid type 667 (positive) film using a yellow filter (Fig 2.2).

2.6.3 SOUTHERN TRANSFER OF DNA FRAGMENTS

In 1975 Southern described a technique enabling the capillary transfer of DNA restriction fragments from an agarose gel to an appropriate membrane. The Southern blotting technique is fundamental to the analysis of genome organisation and expression. The most used method of transfer involves DNA binding to a nylon membrane by passing a salt solution through the denatured DNA into

the membrane. The denatured DNA is retained by the membrane and the binding procedure is completed by drying the membrane.

A modification of the method of Southern, (1975) was used to transfer DNA fragments to the nylon membrane Hybond N (Amersham).

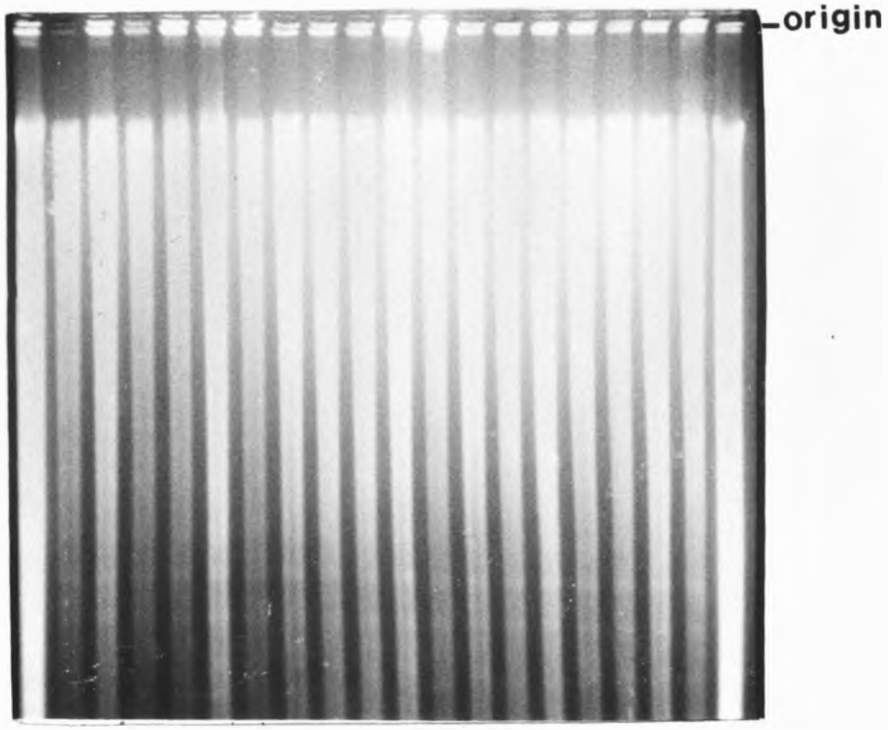


FIG 2.2 An ethidium bromide stained agarose gel containing human DNA, digested with the restriction enzyme *Pst*I and separated out by electrophoresis. The origin where the samples were loaded, is indicated.

2.6.3.1 TRANSFER OF DNA TO NYLON MEMBRANES

The nylon membrane Hybond N is of high durability and can be used repeatedly for hybridization to different probes.

Following electrophoresis the gel was photographed under UV illumination and rapidly removed from the transilluminator to avoid excessive nicking of the DNA. The gel was rinsed in distilled water to remove any excess of electrophoresis buffer. Strand separation of the DNA was attained by subjecting the gel to an alkaline denaturing solution containing 1M NaCl, 0.5M NaOH. The incubation of the gel was done for 30 minutes with gentle agitation. The gel was rinsed twice with distilled water and then neutralized with a Tris buffer containing 3M NaCl, pH 7.0-7.5 for 30 minutes. The gel was then placed in 20xSSC (Appendix B) for 10 minutes.

The transfer membrane was cut to the exact size of the gel and placed in 2xSSC. Approximately 500mls of 20xSSC was poured 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 20xSSC. This formed a wick through which the salt solution could be drawn. Any air bubbles trapped underneath the paper were rolled out with a glass pipette. The gel was placed on top of the filter paper and again any air bubbles

trapped beneath were rolled out. The filter paper around the gel was covered with plastic wrap to ensure that the salt solution passed only through the gel. Excess liquid was rolled off the top of the gel. The wet nylon membrane was placed on top of the gel taking care to exclude air bubbles. Six sheets of filter paper were cut to the size of the gel, three of these were soaked in 20xSSC and placed on top of the membrane, again excluding any trapped air bubbles. The remaining sheets of dry filter paper were layered on top followed by a stack of absorbant paper towels, a glass plate and a weight of 0.5 to 1kg to distribute the weight evenly over the gel.

Transfer proceeded for 16-20 hours. Afterwards, the paper towels were removed, the well positions of the gel were marked on the membrane which was then peeled off and rinsed in 2xSSC. The membrane was put between 2 sheets of filter paper and baked at 80°C in a vacuum oven for 1 hour. The membranes were then kept at room temperature until used. After transfer the gel was viewed under UV light to check that transfer was complete. Only very rarely did a total transfer not take place and what remained in the gel was not significant.

2.6.4 RADIOLABELLING OF DNA PROBES

DNA can be labelled using radioisotopes ^{32}P, ^{35}S, ^{125}I and ^{3}H. There are 2 methods to label DNA probes that are frequently used. One is the nick translation method (Rigby *et al*, 1977) but when a particular fragment of a probe has to be labelled the procedure of oligolabelling DNA using hexanucleotide primers (Feinberg and Vogelstein, 1983) was used. For both the above-mentioned reactions ^{32}P-deoxycytidine triphosphate (10mCi/ml; 3000Ci/mmol, Amersham) was the labelled nucleotide to be incorporated. The latter method was the one used in this study.

2.6.4.1 OLIGOLABELLING OF DNA FRAGMENTS

This method of labelling DNA, using random hexanucleotide primers is capable of labelling DNA with ^{32}P to specific radioactivities of 10^9 cpm/μg or higher (Feinberg and Vogelstein, 1983). The major advantages of this method over nick translation are:

(i) much higher specific activities are obtained

(ii) utilization of the ^{32}P label is more efficient with up to 90% of the labelled nucleotide being incorporated into the DNA under optimal conditions.

(iii) also, restriction fragments separated from low melting temperature agarose gels may be labelled efficiently and simply, without removal of the agarose and consequent loss of DNA.

The 1μg of probe DNA which was to be labelled was first denatured by boiling for 7 minutes and put on ice for approximately 1 minute. It was then incubated at room temperature for 15 minutes, with the larger fragment of DNA polymerase I (Klenow fragment) in the presence of nucleotides, primers, radioactive deoxynucleoside triphosphate and spermidine trihydrochloride. The latter was found to increase the incorporation of nucleotides. The mechanism is not know, however spermidine has been reported to stimulate the incorporation of gamma-^{32}P ATP in end-labelling reactions by inhibiting nucleases which may be present in some preparations of polynucleotide kinase (Maniatis *et al*, 1982), also polyamines are supposed to unravel the DNA making it more accessible.

The percentage of labelled nucleotide that had been incorporated was checked after 15 minutes. Incorporation of the ^{32}P-dCTP was measured by spotting 0.5μl of the reaction mixture onto a 2.5cm circular Whatman GF/C filter and total counts were noted using a Geiger Counter. The filter was placed onto a vacuum filtering aparatus and was washed once with approximately 15ml of 10% trichloroacetic acid (TCA) containing 0.2M

pyrophosphate followed by three washes with 15ml of 5% TCA. The non-incorporated nucleotides were washed through the filter and the DNA containing the incorporated nucleotides, remained on the filter. The acid precipitable counts were noted and incorporation was calculated as follows:

$$\% \text{ incorporation} = \frac{\text{TCA preciptable counts}}{\text{Total counts}} \times 100$$

More than 50% incorporation was sufficient for labelling whole phage, whereas incorporation above 30% was sufficient for labelling a plasmid or a smaller DNA fragment. The reaction was then incubated at 37°C for a further 15 minutes and then placed on ice.

2.6.4.2 REMOVAL OF NON-INCORPORATED NUCLEOTIDES

Labelled DNA was separated from unincorporated nucleotides by column elution using sephadex G50 (Fine:20 to 80μ), swelled in TE as a matrix and TE pH8.0 as elutor. In the course of this study two methods were used:

Method I

The volume of the reaction mixture was made up to 200μl using TE buffer containing 100μg/ml salmon or herring sperm DNA. An equal volume of phenol saturated with 0.1M Tris pH 8.0 was added to this, mixed well and

centrifuged to separate the 2 phases. The aqueous phase was collected and the phenol phase was back-extracted with 200μl of TE containing the carrier DNA. The aqueous phases were pooled. A 10ml perspex 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 surface of the column and a continuous flow of TE buffer through the column was maintained and three drop fractions were collected. Two peaks of radioactivity were noted, the first one representing the labelled DNA and the second, the free unincorporated nucleotides. The passage of the nucleotide fraction was followed by the Orange G dye which migrates just behind the nucleotides. The fraction of the first peak were pooled so that a 500μl to 1ml volume was obtained and put on ice.

Method II

This method is commonly known as the spin column method. A 1ml syringe clogged at the bottom with glass or nylon wool which was packed by 5 minute centrifugation at full speed with sephadex G-50 until a 0.9ml volume of packed sephadex has been reached. Once this was accomplished the column was calibrated with 100μl volume of 1xTE

until an elution volume of 100μl was collected after centrifugation.

The reaction mixture to be separated was made up to 100μl using 1xTE buffer and loaded on to the top surface of the calibrated column and after the 5 minute spin a 100μl fraction was collected which contained the labelled DNA. The unincorporated nucleotides remained in the column. The volume collected was made up to a ml with 1xTE.

A 10μl aliquot of the labelled DNA (collected by either method) was removed to determine the total specific activity of the DNA using the ^{3}H channel of a scintillation counter. This method of drying counting is known as Cerenkov counting. The counts were corrected for ^{32}P multiplied by three (since counting is only a third efficient) and the aliquot volume multiplied by 100 (since 10μl is out of the total volume of 1ml was counted). The amount of labelled probe to be hybridized to Southern blots was calculated on the basis that 10^{6}cpm/ml of hybridization solution was sufficient to elicit a signal after autoradiography. The exposure time would obviously vary according to the sensitivity of the X-ray film used, as well as the number of blots hybridized together.

Prior to hybridization the labelled DNA fraction was denatured by boiling for 5 to 10 minutes and immediately placed on ice to avoid the re-association of the single strands.

2.6.5 HYBRIDIZATION OF LABELLED DNA TO NYLON SOUTHERN TRANSFERS

2.6.5.1 PREHYBRIDIZATION

Prehybridization of membranes with substances designed to bind to nonspecific nucleic acid binding sites on the solid supports, was drescribed many years ago by Denhardt, (1966), as a means of reducing background hybridization.

Both prehybridization and hybridization solutions were the same and the hybridization was carried out in plastic bags. Fifty percent deionized formamide (Appendix B) was used in the hybridization mix, it has the property of lowering the temperatures required for DNA-DNA duplex formation (Cannon *et al*, 1985). Thus, its inclusion at 50% concentration meant that hybridization could be carried out at 42°C as opposed to 65°C without formamide.

The prehybridization solution (Appendix B) contained the following ingredients: 50% formamide, 10xSSPE, 10xDenhardt's, 1% SDS and 400μg/ml denatured salmon

sperm DNA. The functions of these ingredients are as follows:

The salt solution (SSPE) masks the natural repulsion between the two highly-charged DNA strands thereby enhancing hybridization. Denhardt's solution (Appendix B) blocks non-specific DNA binding sites and acts as a competitor with salmon sperm DNA which is added for the same purpose.

Prior to prehybridization, filters were soaked in 2xSSC. They were placed in a plastic bag, (up to three 20cmx20cm filters could be hybridized together in one bag), and 15ml of prehybridization solution preheated to 42°C was added. The filters were incubated with shaking at 42°C for a minimum period of 1 hour.

2.6.5.2 HYBRIDIZATION

The denatured DNA probe was added to the bags containing the prehybridization solution. Air bubbles were removed and the bag was heat-sealed. Hybridization was allowed to continue at 42°C for a 16-48 hours with gentle agitation after which, the filters were removed from the hybridization solution and washed to remove non-specifically bound substances. The hybridization solution could be re-used on other pre-treated DNA blots if the specific activity of the probe was still high. Before re-using the hybridization mix, it must be heated

at 70°C for 10-30 minutes to ensure denaturation of DNA and then cooling on ice to below 50°C before re-use. The new filters must have been exposed to prehybridization solution for 1 hour. The latter would be removed before adding the already used denatured hybridization solution.

2.6.5.3 POST-HYBRIDIZATION WASHING

The rate of association and/or the stability of DNA duplexes formed during hybridization are markedly affected by temperature and salt concentration (Britten and Davidson, 1985). For a typical DNA association reaction the maximum reassociation rate occurs at 25°C below the melting temperature of the duplexes. At particular salt concentrations, reassociation may cease at temperatures below the optimum. The rate of reassociation rises with the rise of salt concentration but it becomes constant when the concentration exceeds 1.2M NaCl. Under less stringent conditions, sequences which are not perfectly complementary may form duplexes. So, decreasing the salt concentration and increasing the temperature of post-hybridization washes was used after hybridization using probes which detect VNTRs in order to eliminate non-specific hybridization and ensure that only stable DNA duplexes remain on the blot.

After hybridization the membranes were placed into a container with 3xSSC and they were incubated with mild shaking at room temperature for 10 minutes. The filters were then washed for a further 30 minutes in 3xSSC in a water bath at 65°C and then for 2 changes of 0.5xSSC plus 0.1%SDS of 30 minutes each also at 65°C. The decrease radioactivity on the membrane was monitored using a Geiger counter until background activity was reached. The membranes were finally rinsed in 0.5xSSC and then sealed in plastic bags.

2.6.6 AUTORADIOGRAPHY

The membranes were exposed to Kodak XAR5 or Agfa Curix RP1 film in cassettes with cronex quanta fast detail intensifying screens (Protea Electro-Medical), at -70°C for 1 to 10 days depending on the specific activity obtained for the probe. The intensifying screens have a great effect on the detection of ^{32}P and can increase the intensity of the signal by eight to ten-fold (Swanstrom and Shank, 1978). The enhancement is directly related to temperature, there is a great increase as the temperature decreases below 4°C.

The intensifying screens consist of sheets embedded with various inorganic phosphor crystals. The mechanism by which these screens work with ^{32}P are not known but by analogy with X-rays, it appears that the collision of an

emitting particle on the electrons of the phosphor will produce several thousands of photons (Swanstrom and Shank, 1978). Once the film was exposed to the hybridized membranes for a period of time, the film was then removed developed and fixed in a Kodak RPX-O-Mat processor.

If the background was high the filters were re-washed at a higher stringency (65°C in 0.2xSSC and 0.1%SDS) for a further 30 minutes and re-exposed to X-ray film. If on the other hand the bands were very light, re-exposure was carried out and double the amount of time was allowed.

2.6.7 STRIPPING AND RE-USE OF MEMBRANES

Membranes could be re-used several times with relatively low levels of background. The basic principle of stripping 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 membrane to a pH of 7.0. Finally the membranes are rinsed in a 2xSSC solution for 5 minutes and are ready for re-hybridization with a new probe. If not used immediately they can be stored in 2xSSC at room temperature or at 4°C or they can be sealed whilst still moist, in a

plastic bag and stored flat at 4°C.

2.7 ANALYSIS OF HUMAN GENOMIC DNA BY POLYMERASE CHAIN REACTION

Polymerase chain reaction (PCR) was invented by Mullis and was first reported at the American Society of Human Genetics Conference in October 1985 (Saiki *et al*, 1988). PCR is an *in vitro* method for the enzymatic synthesis of specific DNA sequences, using 2 oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the target DNA. A repetitive series of cycles involving template denaturation, primer annealing and the extension of the annealed primers by DNA polymerase results in the exponential accumulation of a specific fragment whose termini are defined by the 5′ ends of the primers. Since 1985, more than 2 000 studies involving research on clinical applications of PCR have been published (Walker and Crisan, 1991) and their numbers appears to be increasing exponentially as PCR itself. The effect of PCR on research in biology and medicine is comparable to that of the discovery of Southern blotting and restriction endonucleases a few years ago. It is an ingenious conceptually simple and rapid technique. For the history of its invention one can be referred to Mullis, (1990) and for reviews on its extensive applications in medicine to Marx, (1988) and Eisenstein, (1990).

One of the most promising approaches appears to be PCR amplification of variable number of tandem repeat (VNTR) loci and electrophoretic detection of polymorphic bands, which does not require Southern blotting and hybridization with specific ^{32}P-labelled or non radioactive probes (Jeffreys *et al*, 1988a; Kasai *et al*, 1990). This method combines the advantage of VNTR loci analysis with those of PCR, that is, high sensitivity and specificity of amplification, simplicity and speed. In the present study, the VNTR locus used in combination with PCR was the one flanking the Apolipoprotein B gene.

2.7.1 VARIABLE NUMBER OF TANDEM REPEAT (VNTR) AT THE APOLIPOPROTEIN B LOCUS

One VNTR is located 0.5kb on the 3′ side of the last aminoacid codon in the apolipoprotein B (APOB) gene (Knot *et al,* 1986; Huang and Breslow, 1987). This gene spans 42kb and maps to the short arm of chromosome 2 (2p24-p23) (Human Gene Mapping 11, 1991). It codes for two forms of the APOB protein, the APOB-48 and APOB-100. The latter is the major apolipoprotein in low density lipoprotein receptor. High levels of LDL and APOB-100 in the plasma correlate with the incidence of premature heart disease. In a pig model of atherosclerosis, particular APOB alleles detectable with alloantisera are associated with susceptibility and elevated plasma cholesterol (Rapacz *et al,* 1986). A number of APOB gene

RFLPs have been reported, some of which have alleles which are associated with plasma LDL level (Berg, 1986). One particular APOB gene point mutation displays a high degree of correlation with hypercholesterolemia (Weisgraber *et al*, 1988; Soria *et al*, 1989).

APOB-100 cDNA and genomic sequences have been cloned and sequenced (Knott *et al*, 1986; Cladaras *et al*, 1986; Law *et al*, 1986; Blackhart *et al*, 1986) and it was found that it is one of the largest monomeric proteins known containing 4 536 aminoacid residues.

Allelic variation at the APOB 3′ hypervariable region is detectable by Southern blot analysis using various restriction enzymes and the 3′ end of the APOB cDNA as a hybridization probe. These Southern blot analysis indicate that this region is complex and the unequivocal assignment of specific alleles to individual DNA bands is difficult because of minor differences in the size of the relatively large restriction fragments. The larger alleles of this hypervariable region are the ones reported to be more common in patients with myocardial infraction (Hegele *et al*, 1986) therefore it would be of utmost importance to devise a strategy that would enable an accurate scoring of these alleles. Boerwinkle *et al*, (1989) and Ludwig *et al*, (1989) analysed this region of the APOB gene by the polymerase chain reaction (Saiki *et al*, 1988). This method provides a higher degree of

resolution among alleles of different sizes thereby enhancing the value of this genetic locus. Boerwinkle *et al*, (1989) in a sample of 125 unrelated individuals scored 12 alleles ranging in size between 570 and 900bp while Ludwig *et al*, (1989) scored 14 alleles in a sample of 318 unrelated individuals ranging in size between 570 and 970bp. In the present study 20 alleles were scored in a sample of 727 unrelated individuals.

2.7.1.1 REACTION MIXTURE

The PCR was carried out in a final volume of 50μl. Each reaction included one microlitre of genomic DNA (500ng-1μg), 100μM of each deoxynucleotide triphosphate (dATP, dCTP, dGTP, dTTP), 1μg of each oligonucleotide primer, 2.5 units of *Taq* Polymerase (Promega) and the reaction buffer recommended by the manufactures. A few drops of mineral oil was added to the reaction to prevent evaporation.

2.7.1.2 PRIMER SELECTION

The oligonucleotide primers were 20 nucleotides long and were synthesized on a Millipore DNA synthesizer. The sequences of the primers flank the target region in the genome on the 3′ side of the APOB gene.

The sequence of the 5′ oligonucleotide used to prime the PCR is:

5′ ATGGAAACGGAGAAATTATG 3′

The sequence of the 3′ PCR primer is:

5′ CCTTCTCACTTGGCAAATAC 3′

2.7.1.3 CYCLING PARAMETERS

The amplification was performed on a DNA thermal cycler (Perkin-Elmer Cetus Instruments) using a step cycle programme set to denature at 94°C for 1 minute, anneal at 60°C for 1 minute and extend at 72°C for 5 minutes for a total of 25 cycles. At the end of the "step cycle" a further extension of 10 minutes at 72°C was allowed.

2.7.1.4 GEL ELECTROPHORESIS

Several composite agarose gels (Seakem HGT Agarose: Nusieve) were tried in different proportions (1:1; 1:3; 1:4) as well as different percentage of HGT agarose gels. The best DNA fragment separation was obtained with 2% agarose (HGT) gels with 0.01% per volume of ethidium bromide run at 120 volts for 3-6 hours. The agarose gels were prepared as in 2.6.2.

2.7.1.5 VISUALIZATION OF FRAGMENTS

DNA in the gel was vilualised over ultraviolet light on a transilluminator. Molecular weight markers (1kb commercial DNA ladder marker and ϕ174 cut with *Hae*III), as well as a mixture of individual samples were run

simultaneously and were used to determine the size of the genomic DNA fragments generated by PCR. The individual samples consisted of DNA from 6 different individuals with fragment sizes ranging from 600bp to 780bp at increments of 30bp.

Gels were photographed under ultraviolet illumination on polaroid type 667 (positive) film using a yellow filter. The first photograph was taken after 2 hours of electrophoresis at 120V and a second one was taken of the same gel 3 hours later after futher electrophoresis.

2.7.2 $(dA\text{-}dC)_n.(dG\text{-}dT)_n$ REPEATS

$(CA)_n$ repeats are a subclass of eukaryote tandemly repeated DNA that contains very short simple sequence repeats such as $(dC\text{-}dA)_n.(dG\text{-}dT)_n$. They are known as microsattelites (Weber and May, 1989). Several examples are know for which the number of repeats within a block of tandemly repeated DNA varies among individuals exhibiting length polymorphism. In the present study 2 (CA) repeats at the APOA2 and D21S168 loci (Weber and May, 1989; Guo *et al*, 1990), were analysed (Table 2.2).

TABLE 2.2 Polymorphic $(CA)_n$ Detecting Markers

Locus	Chromosome Location	Repeat Sequence	PCR Primers
APOA2	1q21-q23	$(CA)_{16}$	GGTCTGGAAGTACTGAGAAA GATTCACTGCTGTGGACCCA
D21S168	21q22.3	$(CA)_{19}$	ATGCAATGTTATGTAGGCTG CGGCATCACAGTCTGATAAA

2.7.2.1 REACTION MIXTURE

The PCR was carried out in a final volume of 50μl. Each reaction included one microliter of genomic DNA (500ng-1μg), 200μM of each deoxynucleotide triphosphate (dATP, dGTP, dTTP), 2.5μM dCTP, 1-2μCi ^{32}P-dCTP at 800 Ci/mmol, 100ng of each oligodeoxynucleotide primer, 1 unit of *Taq* polymerase (Promega) and the reaction buffer recommended by the manufacture. A few drops of mineral oil was added to prevent evaporation.

2.7.2.2 PRIMER SELECTION

The oligonucleotide primers were 20 nucleotides long and were synthesized on a Millipore DNA synthesizer. The sequences of the primers flank the target region in the genome at the APOA2 (Weber and May, 1989) and D21S168 (Guo *et al,* 1990) loci. The sequences of the respective primers are shown in Table 2.2.

2.7.2.3 CYCLING PARAMETERS

The amplification was performed on a DNA thermal cycler (Perkin-Elmer Cetus Instruments) using a step cycle programme set to denature at 94°C for 1 minute, anneal at 55°C for 2 minutes and extend at 72°C for 2.5 minutes for a total of 25 cycles. At the end of the step cycle a futher extension of 10 minutes at 72°C was allowed.

2.7.2.4 ELECTROPHORESIS IN A DENATURING POLYACRYLAMIDE DNA SEQUENCING GEL

Since the expected fragment sizes were smaller than 200bp, standard denaturing polyacrylamide DNA sequencing gels were used (Sambrook *et al*, 1989). Two μl aliquots of the amplified DNA were mixed with 2 μl of formamide sample buffer, denatured at 97°C for 2 minutes and electrophoresed. pBR322 digested with *Hpa*II was used as a molecular weight marker and two heterozygous individuals were used as internal markers for inter gel comparisons of DNA fragment sizes. Gels were not fixed, they were covered with plastic wrap, dried for 20 minutes and prepared for autoradiography.

2.7.2.5 AUTORADIOGRAPHY

The dried gels were exposed to Curix film in cassettes with Tungsten/phosphate calcium intensifying screens at room temperature for one to 24 hours depending on the

activity of the ^{32}P-dCTP, and the set of primers. Once the film had been exposed to the dried gel for a period of time, it was removed, developed and fixed in a Kodak RPX-O-Mat processor.

2.8 STATISTICAL METHODS

2.8.1 ALLELE FREQUENCYS

Allele frequencies were calculated from genotype frequencies using the gene counting method.

A standard error (SE) for each allele frequency was calculated using the following formula: $\sqrt{f(1-f)/2N}$, where f is the calculated observed frequency of the particular allele and N, the number of individuals sampled.

A Chi-square test was used to determine whether the observed distributions of genotypes were significantly different from those expected according to the prediction of Hardy-Weinberg equilibrium and whether populations differ significantly from one another.

2.8.2 HETEROZYGOSITY

Unbiased estimates of expected heterozygote frequencies (heterozygosities or H) were computed as:

$$n(1-\sum_{j=1}^{k}(nj/n)^2)/(n-1)$$

where n1,n2.....nk are the allele counts of k alleles at a locus in a sample of n genes drawn from the population (Nei, 1978). In a randomly mating population in Hardy-Weinberg equilibrium, H is equivalent to the frequency of observed heterozygotes. Heterozygosity (H) can also be used to calculate the expected total number of heterozygotes for a particular locus in a population.

Sample homozygosity, in a random mating population, is obtained by the sum of squares of allele frequencies in the sample.

2.8.3 GENETIC DISTANCES

Genetic distance measurements and cluster analysis using the three VNTR and the two $(CA)_n$ loci frequency data were computed by the methods of Harpending and Jenkins (1973), and the computer program Antana, written and kindly supplied, by Professor H.C.Harpending, Pensylvania State University. Genetic distance, d, is calculated from the formula:

$$d = \frac{\Sigma(p_i - p_j)^2}{p} \text{ where,}$$

pi is the frequency of a particular allele in the first population

pj is the corresponding frequency of the same allele in the second population

The computer program calculates d, for all possible pairs of populations and presents the results in the form of a matrix. To show the relevant genetic similarities (groupings) between the populations a cluster analysis was performed on the genetic distance values and the result presented in a dendrogram. Sixty-six to ninety alleles (binned vs. non-binned alleles) at five loci in 17 populations were used in this distance analysis.

2.8.4 PATERNITY INDEX, COMBINED PATERNITY INDEX, PROBABILITY OF PATERNITY AND AVERAGE POWER OF EXCLUSION

Paternity indices were calculated using data obtained on mother, child and alleged father in conjunction with estimated allele frequencies. The paternity index takes the form of a ratio in which the numerator is the likelihood of the mother and someone with the genotype of the accused father producing a child with the observed genotype and the denominator, the likelihood of someone with the mothers genotype and a random man producing the childs genotype. Paternity indices (PI) for all the genetic systems tested in a particular case may be multiplyed together to produce a combined paternity index (CPI).

Often the combined paternity index will be a number greater than 100 and this indicates that a man with all the same genotypes as the alleged father is over a 100 times more likely to be the biologic father than a man taken at random. However, the courts frequently prefer to have the chance of paternity expressed as a percentage which is also known as the relative chance of paternity or the probability of paternity. In order to convert the combined paternity index into a percentage an additional piece of information is necessary, the probability of paternity based on non-laboratory evidence. This is called the prior probability and may favor paternity or nonpaternity. A prior probability of 0.5 is considered to be neutral and neither favours the mother's nor the alleged father's argument. Since the laboratory is not in a position to evaluate prior probabilities, a prior probability of 0.5 is generally used in the calculation. The formula for calculating the relative chance of paternity (RCP) is:

$$RCP = \frac{CPI}{CPI + 1}$$

In order to access the power that each of the systems used has to exclude falsely accused men, an average power of exclusion (PEX) may be calculated. This is done according to the equation of Garber and Morris (1983):

$$A=\sum_{i=1}^{n}\left[Pj(1-Pi)\right]^{2}+\sum_{i<j}^{n} PiPj\left[(1-Pi)^{1}+(1-Pj)^{3}+(Pi+Pj)(1-Pi+Pj)^{2}\right]$$

where n = number of alleles of the system

Pn = allele frequency

A = PEX

A cummulative PEX value, CPEX, for each of the four major population groups was also calculated as:

$$CPEX=1-\left[(1-PEXa)(1-PEXb)(1-PEXc)....(1-PEXn)\right]$$

where PEXa,b,c,....n = Average power of exclusion of systems a, b, c .... n respectively.

CHAPTER THREE

3 RESULTS

3.1 LYMPHOBLASTOID CELL LINES

Following the described protocol of lymphocyte transformation it has been possible to successfully establish well over 700 lymphoblastoid cell lines from individuals belonging to twelve different southern African population groups.

3.1.1 TIME AND TEMPERATURE CONDITIONS BEFORE TRANSFORMATION

After conditions were optimized, a successful transformation rate above 90% was obtained. There were several parameters however that had to be taken into consideration. An example is the age of the sample, which should not be older than 48 hours, another is the transport conditions (Table 3.1). All samples transformed if they were processed within to two days of collection; thereafter, the transformation rate declined. It is puzzling that there were samples that did not transform three days after collection, whereas others did transform even though they were processed four days after collection.

TABLE 3 . 1 Assessment of lymphocyte transformation on peripheral blood samples affected by parameters of age and temperature (T° C) during transportation prior to transformation.

	DAY 0	DAY 1			DAY 2				DAY 3				DAY 4			
DONORS	CONTROL	RT	4	CB	RT	4	C	CB	RT	4	C	CB	RT	4	C	CB
A	S	S	S	S	S	S	S		U	U	U		U	U	U	
B	S	S	S	S	S	S	S		U	U	U		S	S	U	
C	S	S	S	S	S	S	S		S	S	S		S	U	U	
D	S	S	S	S	S	S	S		U	S	U		S	U	U	
E	S	S	S	S	S	S	S		U	S	S		U	S	S	
F	S	S	S	S	S	S	S		U	U	U		U	U	U	
G	S	S	S	S	S	S	S		U	U	U		S	S	U	
H	S	S	S	S	S	S	S		S	S	S		S	S	S	
I	S	S	S	S	S	S	S		S	S	S		U	S	S	
J	S	S	S	S	S	S	S		S	S	S		U	S	U	

S = Successful
U = Unsuccessful

RT = Room Temperature
CB = Cooler Bag

A-E = Negroid Subjects
F-J = Caucasoid Subjects

3.1.2 FREEZING OF LYMPHOCYTES AFTER SEPARATION

Several freezing methods were experimented with, but a good transformation rate or recovery rate was obtained only when the freezing method described in 2.3.7 was followed. The temperature during the process of freezing may, however fluctuate and have an effect on the viability of the cells. For this reason 10 samples were frozen on different occasions for each donor to ensure a successful recovery rate.

3.2 POLYMORPHISMS IDENTIFIED BY VNTR AND $(CA)_n$ REPEATS

Genotype data for three VNTR loci and two $(CA)_n$ repeats were determined. Two of the VNTR loci (D2S44 and D4S125) were analysed by ^{32}P dCTP labelling and hybridization to Southern blots. The third tandem repeat at the APOB locus was studied by the PCR technique and agarose gel electrophoresis (Table 3.2). The PCR technique and denaturing polyacrylamide gels were used to analyse the two $(CA)_n$ repeat loci (APOA2 and D21S168) (Table 3.2).

The number of alleles observed at 5 loci varied between population groups. The number of alleles and the allele sizes are shown in Table 3.3. These allele numbers do not necessarily correspond to the same allele numbers given by other authors for the same loci since sequencing on individual alleles was not performed.

TABLE 3 . 2 Summary of the five hypervariable loci examined in 17 southern African populations.

LOCUS	PROBE / MARKER	CLASS OF REPEAT	CHROMOSOMAL LOCATION	METHOD OF DETECTION	NO.INDIVIDUALS STUDIED				TOTAL NO.ALLELES OBSERVED			
					NG	KS	CC	CL	NG	KS	CC	CL
D2S44	pYNH24	VNTR	2p	SOUTHERN	293	81	62	28	28	26	13	9
D4S125	pYNZ32	VNTR	4p16.3	SOUTHERN	253	85	64	31	19	13	8	6
APOB	3'beta-L 3'beta-R	VNTR	2p24-p23	PCR	381	81	189	76	19	13	14	14
APOA2	Mfd3	(CA)n	1q21-q23	PCR	402	62	196	60	10	8	10	8
D21S168	334 335	(CA)n	21q22.3	PCR	482	91	217	71	12	11	9	10

NG = NEGROID KS = KHOISAN CC = CAUCASOID CL = "COLOURED"

TABLE 3 . 3 Allele sizes detected at the different loci.

	VNTRS					(CA)n	
	pYNH24		pYNZ32		APOB	APOA2	D21S168
ALLELE	A	B	A	B	A	A	A
	(kb)	(kb)	(kb)	(kb)	(bp)	(bp)	(bp)
1	8.0	8.0	2.6	2.6	420	129	100
2	8.6		2.7		450	131	102
3	8.8		2.8	2.8	480	133	104
4	9.0	9.0	3.0		510	135	106
5	9.4		3.1	3.1	540	137	108
6	9.6		3.2		570	139	110
7	9.8		3.3	3.3	600	141	112
8	10.0	10.0	3.4		630	143	114
9	10.3		3.5		660	145	116
10	10.5		3.6	3.6	690	147	118
11	10.8		3.7		720		120
12	11.0	11.0	3.8	3.8	750		122
13	11.2		3.9		780		
14	11.5		4.0		810		
15	12.0	12.0	4.1	4.1	840		
16	12.3		4.2		870		
17	12.5		4.4	4.4	900		
18	12.8		4.5	4.6	930		
19	13.0	13.0	4.7	4.8	960		
20	14.0	14.0	4.9		990		
21	14.5		5.0	5.0			
22	15.0	15.0					
23	16.0	16.0					
24	16.5						
25	17.0	17.0					
26	18.0	18.0					
27	19.0	19.0					
28	20.0	20.0					

A = All alleles
B = Binned alleles

3.2.1 VNTRs AND THE POLYMORPHISMS THEY IDENTIFY

3.2.1.1 VNTR AT THE D2S44 LOCUS

DNA from 464 unrelated individuals belonging to 17 different southern African population groups was hybridized with the pYNH24 VNTR probe, and 28 different alleles were detected (Table 3.3). Their size range spans 12kb. An example of an autoradiogram showing some allelic variation after digestion and separation of DNA on a 0.7% agarose gel is shown in Fig 3.1.

When dealing with the analysis of tandem repeat sequences one is faced with some limitations with regard to accurate size determination. Band width, electrophoretic resolution and slight variation in electrophoretic mobility of DNA fragments can cause measurement imprecision. The number of samples required to characterize a population increases with increasing complexity or number of alleles at the VNTR locus. To try to minimize measurement error, arbitrarily defined bins were created for alleles of similar size. By binning, 28 alleles were reduced to 13 (Table 3.3). The genotypes of the 464 unrelated individuals belonging to the 17 different population groups and in the four major groups were used to calculate allele frequencies and the corresponding standard errors (Tables 3.4 and 3.5).



FIGURE 3.1 Hybridization pattern revealed by pYNH24 with *Pst*I. There is almost a continuous variation of fragment sizes but only some are indicated on the right hand side

TABLE 3.4 Number and frequencies of binned alleles at the D2S44 locus in 17 southern African populations.

	ZULU N = 54			SWAZI N = 64			XHOSA N = 28			TSONGA N = 62			SOTHO N = 46			TSWANA N = 28		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	1	0.019	0.019	-	-	-	-	-	-	5	0.081	0.035	4	0.087	0.042	-	-	-
2	8	0.149	0.048	10	0.157	0.045	2	0.072	0.049	16	0.258	0.056	6	0.130	0.050	3	0.107	0.058
3	11	0.204	0.055	9	0.140	0.043	8	0.285	0.085	5	0.081	0.035	4	0.087	0.042	9	0.321	0.088
4	7	0.130	0.046	18	0.282	0.056	4	0.143	0.066	4	0.065	0.031	7	0.152	0.053	1	0.036	0.035
5	1	0.019	0.019	1	0.016	0.016	1	0.036	0.035	10	0.162	0.047	6	0.131	0.050	4	0.143	0.066
6	11	0.204	0.055	14	0.219	0.052	1	0.036	0.035	-	-	-	2	0.043	0.030	4	0.143	0.066
7	15	0.278	0.061	-	-	-	6	0.214	0.078	4	0.065	0.031	2	0.043	0.030	4	0.143	0.066
8	-	-	-	2	0.031	0.022	2	0.071	0.049	6	0.087	0.036	10	0.217	0.061	2	0.071	0.049
9	-	-	-	10	0.156	0.045	-	-	-	1	0.016	0.016	4	0.087	0.042	-	-	-
10	-	-	-	-	-	-	-	-	-	8	0.129	0.043	1	0.022	0.022	-	-	-
11	-	-	-	-	-	-	3	0.107	0.058	3	0.048	0.027	-	-	-	-	-	-
12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0.036	0.035
13	-	-	-	-	-	-	1	0.036	0.035	-	-	-	-	-	-	-	-	-

N = Total number of particular alleles identified in each population.

TABLE 3.4 Cont. 2

	PEDI N = 34			LEMBA N = 60			VENDA N = 20			HERERO N = 98			DAMA N = 92			NAMA N = 84		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	5	0.147	0.061	5	0.083	0.036	-	-	-	12	0.122	0.033	6	0.065	0.026	9	0.107	0.034
2	8	0.235	0.073	9	0.150	0.046	2	0.100	0.067	20	0.204	0.041	5	0.055	0.024	12	0.155	0.039
3	4	0.118	0.055	8	0.133	0.044	7	0.350	0.107	8	0.082	0.028	13	0.140	0.036	5	0.060	0.026
4	3	0.089	0.049	6	0.100	0.039	1	0.050	0.049	5	0.051	0.022	10	0.109	0.032	12	0.143	0.038
5	-	-	-	-	-	-	3	0.150	0.080	8	0.082	0.028	13	0.141	0.036	18	0.214	0.045
6	6	0.176	0.065	15	0.250	0.056	2	0.100	0.067	24	0.245	0.043	5	0.054	0.024	8	0.095	0.032
7	3	0.088	0.049	2	0.033	0.023	3	0.150	0.080	14	0.143	0.035	13	0.141	0.036	18	0.214	0.045
8	-	-	-	-	-	-	2	0.100	0.067	1	0.010	0.010	8	0.087	0.029	2	0.024	0.017
9	5	0.147	0.061	10	0.167	0.048	-	-	-	6	0.061	0.024	16	0.174	0.040	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	3	0.050	0.028	-	-	-	-	-	-	2	0.022	0.015	-	-	-
12	-	-	-	2	0.033	0.023	-	-	-	-	-	-	-	-	-	-	-	-
13	-	-	-	-	-	-	-	-	-	-	-	-	1	0.011	0.011	-	-	-

TABLE 3.4 Cont. 3

	SAN N = 78			WHITES N = 52			INDIANS N = 44			JEWS N = 28			"COLOURED" N = 56		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	1	0.013	0.013	1	0.019	0.019	8	0.182	0.058	2	0.071	0.049	5	0.089	0.038
2	6	0.077	0.030	5	0.096	0.041	3	0.068	0.038	3	0.107	0.058	7	0.125	0.044
3	12	0.153	0.041	3	0.058	0.032	2	0.045	0.031	-	-	-	5	0.089	0.038
4	12	0.153	0.041	1	0.019	0.019	11	0.250	0.065	2	0.071	0.049	-	-	-
5	15	0.192	0.045	5	0.096	0.041	12	0.273	0.067	7	0.250	0.082	14	0.250	0.058
6	5	0.064	0.028	8	0.154	0.050	-	-	-	3	0.107	0.058	19	0.339	0.063
7	9	0.116	0.036	21	0.404	0.068	7	0.159	0.055	9	0.321	0.088	6	0.107	0.041
8	1	0.013	0.013	-	-	-	-	-	-	-	-	-	-	-	-
9	8	0.103	0.034	6	0.115	0.044	1	0.023	0.023	2	0.071	0.049	-	-	-
10	4	0.051	0.025	-	-	-	-	-	-	-	-	-	-	-	-
11	2	0.026	0.018	2	0.038	0.027	-	-	-	-	-	-	-	-	-
12	1	0.013	0.013	-	-	-	-	-	-	-	-	-	-	-	-
13	2	0.026	0.018	-	-	-	-	-	-	-	-	-	-	-	-

TABLE 3.5 Number and frequencies of the binned alleles at the D2S44 locus in the four major southern African populations.

	NEGROID N = 586			KHOISAN N = 162			CAUCASOID N = 124			"COLOURED" N = 56		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	38	0.065	0.010	10	0.062	0.019	11	0.089	0.026	5	0.089	0.038
2	89	0.151	0.015	18	0.112	0.025	11	0.088	0.025	7	0.125	0.044
3	86	0.147	0.015	17	0.106	0.024	5	0.040	0.018	5	0.089	0.038
4	66	0.112	0.013	24	0.148	0.028	14	0.113	0.028	-	-	-
5	47	0.080	0.011	33	0.203	0.032	24	0.193	0.035	14	0.250	0.058
6	84	0.143	0.014	13	0.080	0.021	11	0.089	0.026	19	0.339	0.063
7	66	0.112	0.013	27	0.166	0.029	37	0.298	0.041	6	0.107	0.041
8	33	0.056	0.009	3	0.019	0.011	-	-	-	-	-	-
9	52	0.088	0.012	8	0.049	0.017	9	0.073	0.023	-	-	-
10	9	0.015	0.005	4	0.025	0.012	-	-	-	-	-	-
11	11	0.019	0.006	2	0.012	0.009	2	0.016	0.011	-	-	-
12	3	0.005	0.003	1	0.006	0.006	-	-	-	-	-	-
13	2	0.003	0.002	2	0.012	0.009	-	-	-	-	-	-

N = Total number of particular alleles identified in each population.

Binned allele frequencies at the D2S44 locus in the four major groups (NG = Negroid; KS = Khoisan, CC = Caucasoid; CL = "Coloured"), are graphically summarised in Fig 3.2. A similar exercise was done for all the 28 alleles (i.e. before binning) but since the distribution patterns were very similar, these results have not been included. The results are summarized using a format similar to that used by Baird *et al,* (1986).

When this locus was analysed, it was evident that three alleles (*5*, *6* and *7*) were commonly present in each of the four groups although their combined frequencies varied (NG = 34%; KS = 45%; CC = 58%; CL = 70%). Alleles *2* and *3* are the most common in the Negroid (NG = 30%; KS = 22%; CC = 13%; CL = 21%), whereas allele *7* is more prevalent among the Caucasoids (NG = 11%; KS = 17%; CC = 30%; CL = 11%). Allele 5 is more common for the Khoisan, Caucasoid and "Coloured" groups than it is for the Negroids (NG = 8%; KS = 20%; CC = 19%; CL = 25%) and alleles *8* to *13* were not present in the "Coloureds" but are relatively well represented in the other groups (NG = 19%; KS = 12%; CC = 9%; CL = 0%). On the other hand, allele *6* is best represented in the "Coloured" population (NG = 14%; KS = 8%, CC = 9%; CL = 34%). The overall distribution is skew towards the smaller alleles tailing off rather sharply towards the bigger ones.

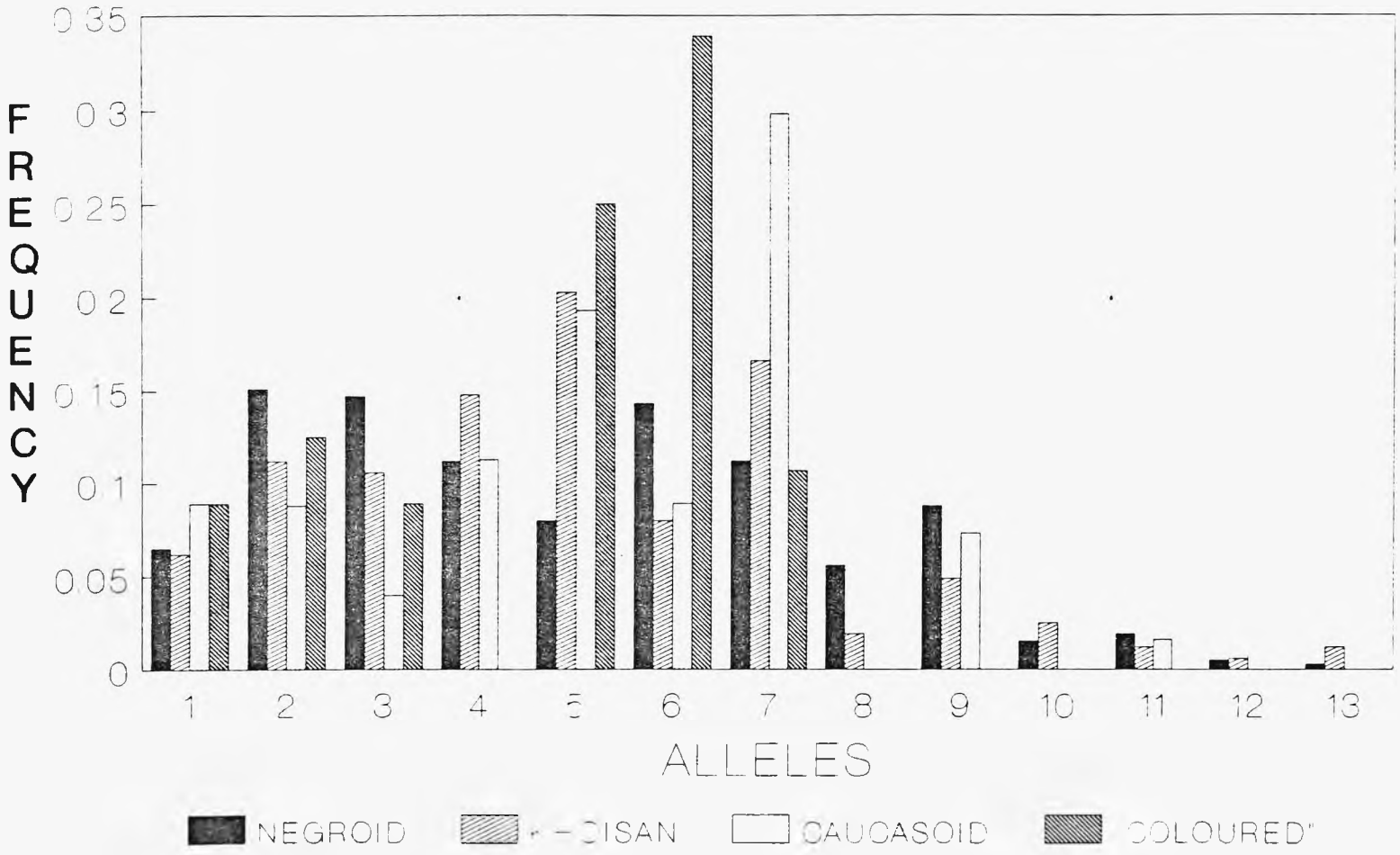


FIG 3.2 Frequency of binned alleles at the D2S44 locus in the four major southern African population groups

Table 3.6 summarizes observed and expected numbers of homozygotes and heterozygotes at the D2S44 locus in the 17 southern African populations considered separately and pooled into larger groups. This comparison, based on Hardy-Weinberg proportions and calculated from the observed and expected observations, showed some heterozygote deficiency in the Negroid, Khoisan ($p<0.05$) and Caucasoid ($0.01<p<0.05$) although this deficiency was greater in the first two groups.

The genotypes of the 464 random individuals were used to study intrapopulation variation, by the calculation of heterozygosity (H), at this locus (Table 3.7). The unbiased heterozygote frequencies calculated from binned genotypes was high in all cases and did not show any remarkable differences between any of the population groups. The lowest value was found to be in the White population group (80.6%) and the highest value in the Sotho group (90.6%). The unbiased heterozygote frequency (H) was calculated when all the 28 alleles were scored and when they were binned. As expected, because of the high number of alleles, the H values in the different populations are consistantly greater in the non-binned samples. When only the four major groups are taken into consideration, the lowest heterozygosity rate for the non-binned data was observed in the "Coloured" group (84.9%) and the highest value was observed in the

TABLE 3.6 Observed and expected numbers of heterozygotes and homozygotes at the D2S44 locus in 17 southern African populations considered separately and pooled into larger groups.

		HOMOZYGOTES		HETEROZYGOTES		
POPULATIONS	**N**	**OBSERVED**	**EXPECTED**	**OBSERVED**	**EXPECTED**	***X***
NEGROIDS	**293**	**59**	**18.7**	**234**	**274.0**	**92.6** *
NGUNI	104	18	7.1	86	97.0	18.2 *
Zulu	*27*	*2*	*3.9*	*25*	*23.3*	*1.0*
Swazi	*32*	*4*	*4.2*	*28*	*27.9*	*>0.1*
Xhosa	*14*	*4*	*2.0*	*10*	*12.0*	*2.5*
Tsonga	*31*	*8*	*3.0*	*23*	*27.6*	*9.4* *
SOTHO-TSWANA	54	9	4.3	45	50.0	5.5 **
Sotho	*23*	*3*	*2.6*	*20*	*20.3*	*0.1*
Tswana	*14*	*2*	*2.0*	*12*	*12.0*	*>0.1*
Pedi	*17*	*4*	*2.5*	*13*	*14.5*	*1.1*
LEMBA	30	9	4.3	21	25.7	6.1 *
VENDA	10	2	1.5	8	8.6	0.2
HERERO	49	15	6.0	34	43.1	15.8 *
DAMA	46	6	3.4	40	42.5	2.2
KHOISAN	**81**	**25**	**6.1**	**56**	**74.9**	**62.7** *
NAMA	42	11	4.9	31	37.1	8.4 *
SAN	39	14	2.8	25	36.2	47.9 *
CAUCASOID	**62**	**17**	**10.0**	**45**	**52.0**	**5.9** **
WHITES	26	7	5.6	19	20.3	0.4
INDIANS	22	7	4.5	15	17.5	1.8
JEWS	14	3	2.8	11	11.2	>0.1
"COLOURED"	**28**	**8**	**5.0**	**20**	**23.0**	**2.1**

N = Number of individuals in each population
* = $p < 0.05$
** = $0.01 < p < 0.05$

TABLE 3.7 Heterozygosity rate at the D2S44 locus when 28 alleles were considered and when these were reduced to 13 bins. Seventeen populations were considered separately and pooled into larger groups.

	ALL ALLELES			BINNED ALLELES	
POPULATIONS	N	K	H (%)	K	H (%)
NEGROIDS	586	**28**	**93.1**	**13**	**88.3**
NGUNI	208	23	94.1	12	88.6
Zulu	*54*	*11*	*89.6*	*7*	*83.6*
Swazi	*64*	*11*	*90.0*	*7*	*83.5*
Xhosa	*28*	*11*	*92.6*	*9*	*89.1*
Tsonga	*62*	*12*	*91.8*	*10*	*87.1*
SOTHO-TSWANA	108	18	93.5	10	90.0
Sotho	*46*	*12*	*92.3*	*11*	*90.6*
Tswana	*28*	*10*	*92.3*	*8*	*87.9*
Pedi	*34*	*9*	*90.6*	*7*	*89.3*
LEMBA	60	10	88.5	9	87.5
VENDA	20	9	94.9	7	88.9
HERERO	98	15	89.7	9	86.1
DAMA	92	20	94.4	11	89.8
KHOISAN	**162**	**26**	**93.6**	**13**	**88.2**
NAMA	84	16	92.8	8	88.4
SAN	78	21	95.2	13	90.1
CAUCASOID	**124**	**13**	**85.1**	**9**	**84.2**
WHITES	52	11	81.2	9	80.5
INDIANS	44	7	83.5	7	83.5
JEWS	28	8	85.9	7	85.3
"COLOURED"	**54**	**9**	**84.9**	**6**	**80.6**

N = Number of alleles sampled
K = Number of alleles in the particular populations
H = Heterozygosity

Khoisan group (93.6%).

Genetic distances between the 17 southern African populations and the four main groups were computed according to the method of Harpending and Jenkins (1973) (Tables 3.8 and 3.9). The binned allele frequency data in Tables 3.4 and 3.5 were used. Populations with greater affinity are separated by smaller genetic distances. In Table 3.8, where all 17 population groups are considered, the lowest values are observed when the Caucasoid groups were compared with one another and with the "Coloured" and Khoisan groups (3 - 7), whilst lower affinity was observed amongst the Negroid groups and between the Negroids and the other groups (2 - 46). In Table 3.9, only the four main groups are considered and, the closest affinity in terms of genetic distance is between the Negroid and Khoisan groups (10) and the largest difference is observed between the Caucasoid and Khoisan groups (28).

Genetic distances were used to generate dendrograms or trees by means of cluster analysis. The choice of a distance measure is not crucial since the clustering procedure depends only on the rank ordering of the pairwise distances. The computer program Antana, written and supplied by H.C. Harpending, Pennsylvania State University, was used to generate dendrograms based on cluster analysis from genetic distances. The resulting

TABLE 3.8 Genetic distances (x100) between 17 southern African populations using the pYNH24 hypervariable probe.

	Zu	Sw	Xh	Tso	Sot	Tsw	Pd	Lem	Ve	He	Da	Na	San	Wht	Ind	Jew	Cl
Zu		25	16	20	17	27	15	20	18	31	26	26	30	27	34	21	31
Sw			15	20	18	23	12	27	21	25	41	19	25	25	29	25	29
Xh				15	6	19	5	10	26	13	22	7	9	9	12	2	12
Tso					26	26	13	13	34	28	19	26	25	30	34	28	29
Sot						20	14	14	30	10	32	8	17	14	16	10	16
Tsw							18	12	22	18	36	30	30	30	32	22	32
Pd								12	16	18	20	12	11	13	12	11	15
Lem									33	11	19	19	20	23	24	16	20
Ve										33	46	36	38	35	36	26	39
He											37	18	26	26	24	21	24
Da												25	24	16	21	20	18
Na													5	6	6	6	4
San														7	6	5	4
Wht															5	5	4
Ind																5	4
Jew																	3
Cl																	

TABLE 3.9 Genetic distances (x100) between four major southern African populations using the pYNH24 hypervariable probe.

	NEGROID	KHOISAN	CAUCASOID	"COLOURED"
NEGROID		10	17	17
KHOISAN			28	21
CAUCASOID				16
"COLOURED"				

dendrograms were used to show the relevant genetic similarities (groupings) between the populations.
Fig 3.3 represents a dendrogram based on the distances shown in Table 3.8 where all populations groups were considered. The Negroid groups show great genetic divergence. The Caucasoid, "Coloured" and Khoisan groups are clustered together and all show greater affinity with one another, away from the Negroid groups. The Nama and San (Khoisan) cluster together. Figure 3.4 shows the dendrogram obtained from the genetic distances in Table 3.9 where only the four main groups are represented. In this case, the Caucasoids diverge the most from the other three groups, with the "Coloureds" clustering between the Negroids/ Khoisan and the Caucasoids.

The effect of binning versus non-binning of alleles was tested. Genetic distances and cluster analyses were carried out when all 28 alleles were considered and when they were binned into 13 alleles. The Zulu, Swazi, Xhosa and Tsonga chiefdoms were amalgamated into the Nguni-speaking Negroids; the Sotho, Tswana and Pedi into the Sotho/Tswana-speaking Negroids; the San and the Nama into the Khoisan, and the Whites, Indians and Jews into the Caucasoid group. The dendrogram in Fig 3.5 is based on binned frequencies. Most of the Negroids are separated from the other groups by a large distance. The Dama did not cluster with the other Negroids but branch

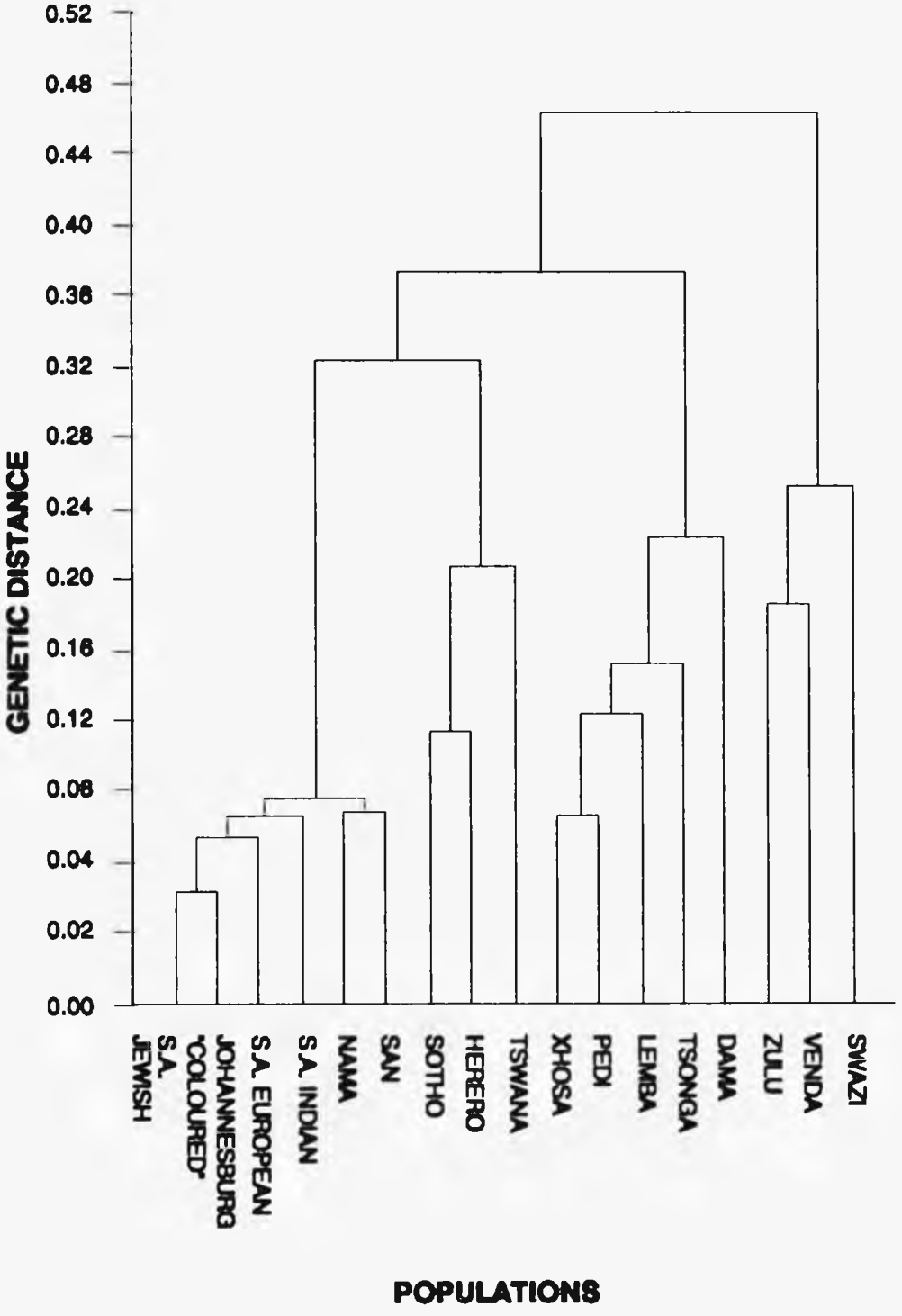


FIGURE 3.3 Dendrogram based on binned frequencies at the D2S44 VNTR locus in 17 southern African populations.

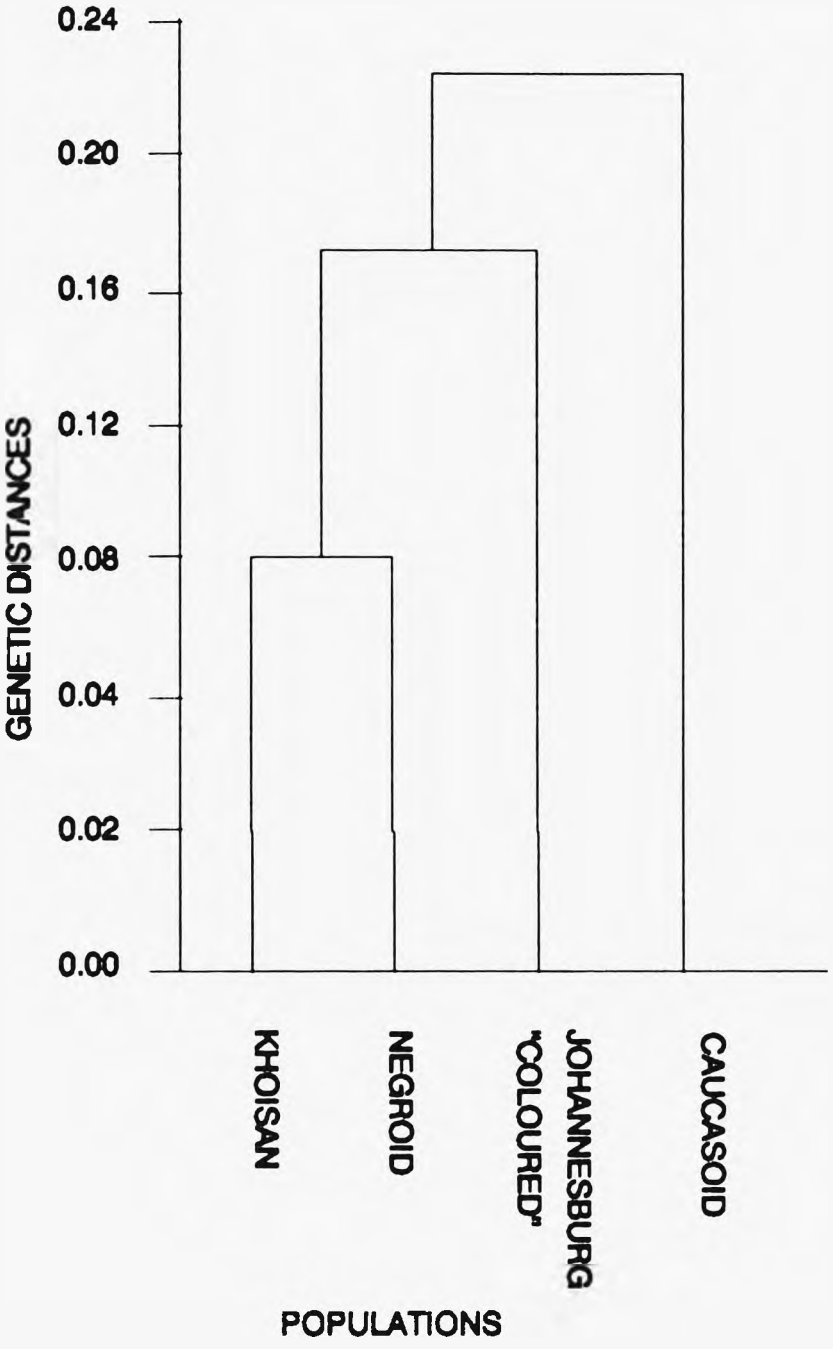


FIGURE 3.4 Clustering implied by binned allele frequencies at the D2S44 VNTR locus in the four major southern African populations.

out from the cluster where the Khoisan is present. It is not entirely unexpected since, although the Dama physically have Negroid features, linguistically they have been grouped with the Khoi because they were slaves of the Nama and speak the same language. They not only aquired their language from the Nama but they may also have received some genes from them. Figure 3.6 is based on the frequencies of all 28 alleles, the dendrogram, unexpectedly, shows the Caucasoid clustering with the Negroid and the Dama and the Khoisan separated in a different cluster by a large distance.

The results in Fig 3.5 based on binned alleles are more concordant with the linguistic classification of populations. The cluster exercise comparing binned and non-binned alleles, performed for this locus will not be repeated individually on the remaining loci. Cluster analyses will have more weight when all the loci are considered together.

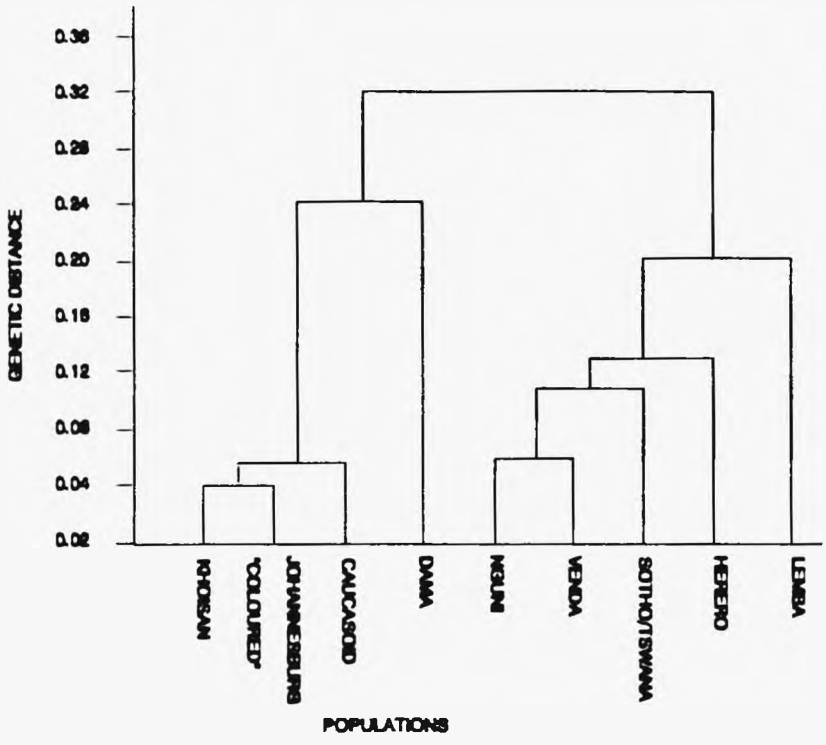


FIGURE 3.5 Dendrogram based on binned frequencies at the D2S44 VNTR locus.

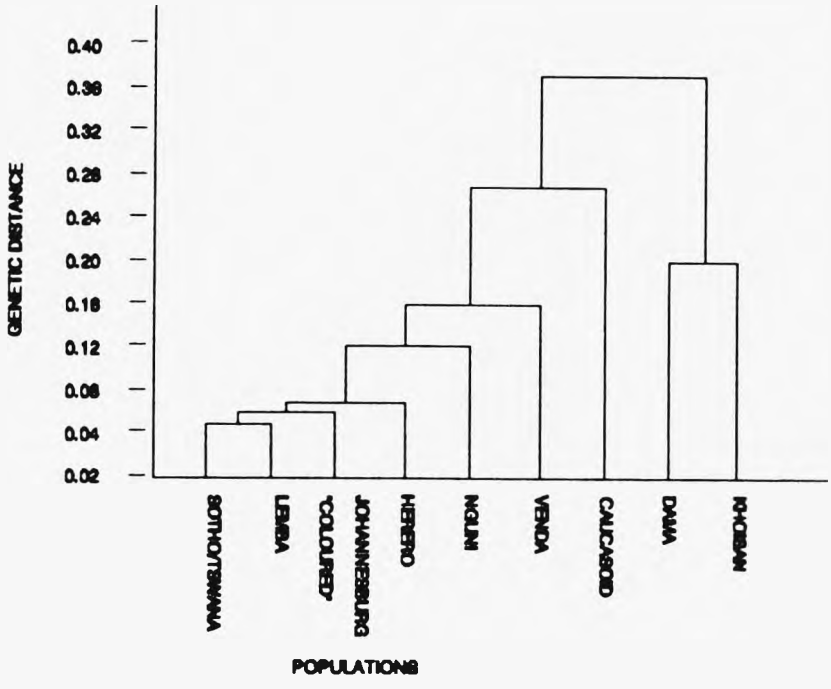


FIGURE 3.6 Dendrogram based on the frequencies of 28 alleles at the D2S44 VNTR locus.

3.2.1.2 VNTR AT THE D4S125 LOCUS

DNA from 433 unrelated individuals belonging to 17 different southern African population groups was hybridized with the pYNZ32 VNTR probe and 21 polymorphic fragments were detected (Table 3.3). Their size range spans 2400 base pairs. An example of an autoradiogram showing allelic variation al this locus is shown in Fig 3.7.

For the reasons mentioned in 3.2.1.1, grouping of alleles in defined bins was done. By binning, 21 alleles were reduced to 11 (Table 3.3) and the genotypes of the 433 unrelated individuals were used to calculated allele frequencies in the 17 different population groups and when pooled into the four major groups, (Tables 3.10 and 3.11).

Binned allele frequencies at the D4S125 locus in the four major groups (NG = Negroid; KS = Khoisan; CC = Caucasoids; CL = "Coloured"), are graphically summarized in Fig 3.8. The overall distribution is skewed to the left, towards the smaller alleles, tailing off in the direction of the larger alleles. A unimodal distribution is observed individually for all populations but whereas the distribution for the Caucasoid and "Coloured" are more skewed to the left, the Negroid and Khoisan distributions are more towards

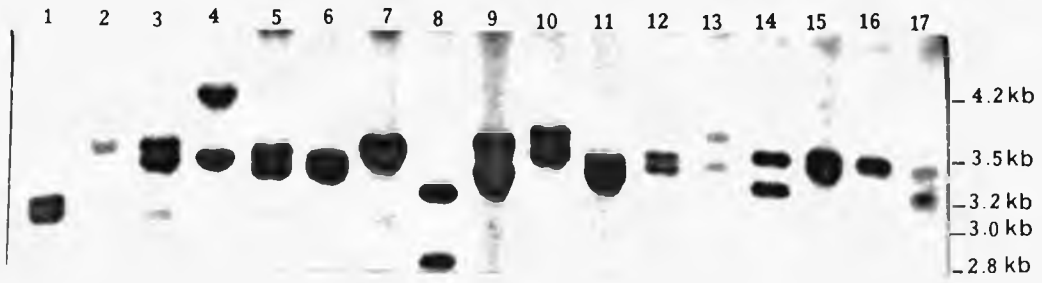


FIGURE 3.7 Hybridization pattern revealed by pYNZ32 with *Pst*I. There is almost a continuous variation of fragment sizes but only some are indicated on the right hand side.

TABLE 3.10 Number and frequencies of binned alleles at the D4S125 locus in 17 southern African populations .

	ZULU N=62			SWAZI N = 62			XHOSA N = 28			TSONGA N = 52			SOTHO N = 46			TSWANA N = 22		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	1	0.016	0.016	2	0.032	0.022	3	0.107	0.058	-	-	-	-	-	-	-	-	-
2	2	0.032	0.022	1	0.016	0.016	-	-	-	1	0.019	0.019	-	-	-	-	-	-
3	13	0.209	0.052	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	28	0.452	0.063	19	0.306	0.059	5	0.179	0.072	12	0.231	0.058	16	0.348	0.070	16	0.727	0.095
5	14	0.226	0.053	28	0.452	0.063	5	0.179	0.072	14	0.269	0.061	23	0.500	0.074	4	0.182	0.082
6	1	0.016	0.016	8	0.129	0.043	14	0.500	0.094	21	0.404	0.068	5	0.109	0.046	2	0.091	0.061
7	1	0.016	0.016	3	0.048	0.027	-	-	-	-	-	-	1	0.022	0.022	-	-	-
8	2	0.032	0.022	1	0.016	0.016	1	0.036	0.035	1	0.019	0.019	1	0.022	0.022	-	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	3	0.058	0.032	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

N = Total number of particular alleles identified in each population.

TABLE 3.10 Cont. 2

	PEDI N = 26			LEMBA N = 54			VENDA N = 36			HERERO N = 40			DAMA N = 78			NAMA N = 74		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	-	-	-	2	0.037	0.026	-	-	-	1	0.025	0.025	2	0.026	0.018	-	-	-
2	-	-	-	-	-	-	1	0.028	0.027	12	0.300	0.072	1	0.013	0.013	-	-	-
3	7	0.269	0.087	-	-	-	16	0.444	0.083	14	0.350	0.075	4	0.051	0.025	6	0.081	0.032
4	15	0.577	0.097	25	0.463	0.068	14	0.389	0.081	13	0.325	0.074	32	0.410	0.056	32	0.432	0.058
5	4	0.154	0.071	21	0.389	0.066	2	0.056	0.038	-	-	-	32	0.410	0.056	20	0.270	0.052
6	-	-	-	5	0.093	0.40	3	0.083	0.046	-	-	-	5	0.064	0.028	12	0.162	0.043
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0.014	0.014
8	-	-	-	1	0.019	0.019	-	-	-	-	-	-	2	0.026	0.018	-	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0.027	0.019
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0.014	0.014

TABLE 3.10 Cont. 3

	SAN N = 96			WHITES N = 40			INDIANS N = 48			JEWS N = 40			"COLOURED" N = 62		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	-	-	-	1	0.025	0.025	4	0.083	0.040	2	0.050	0.034	-	-	-
2	11	0.115	0.033	20	0.500	0.079	28	0.583	0.071	18	0.450	0.079	23	0.371	0.061
3	26	0.271	0.045	9	0.225	0.066	13	0.271	0.064	20	0.500	0.079	33	0.532	0.063
4	39	0.406	0.050	10	0.250	0.068	3	0.063	0.035	-	-	-	5	0.081	0.035
5	12	0.125	0.034	-	-	-	-	-	-	-	-	-	1	0.016	0.016
6	7	0.073	0.027	-	-	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	1	0.010	0.010	-	-	-	-	-	-	-	-	-	-	-	-

TABLE 3.11 Number and frequencies of the binned alleles at the D4S125 locus in the four major southern African populations.

	NEGROID N = 506			KHOISAN N = 170			CAUCASOID N = 128			"COLOURED" N = 62		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	11	0.022	0.007	-	-	-	7	0.055	0.020	-	-	-
2	18	0.036	0.008	11	0.065	0.019	66	0.516	0.044	23	0.371	0.061
3	54	0.107	0.014	32	0.188	0.030	42	0.328	0.041	33	0.532	0.063
4	195	0.385	0.022	71	0.418	0.038	13	0.102	0.027	5	0.081	0.035
5	154	0.304	0.020	32	0.188	0.030	-	-	-	1	0.016	0.016
6	57	0.113	0.014	19	0.112	0.024	-	-	-	-	-	-
7	6	0.012	0.005	1	0.006	0.006	-	-	-	-	-	-
8	8	0.016	0.006	-	-	-	-	-	-	-	-	-
9	-	-	-	2	0.012	0.008	-	-	-	-	-	-
10	3	0.006	0.003	-	-	-	-	-	-	-	-	-
11	-	-	-	2	0.012	0.008	-	-	-	-	-	-

N = Total number of particular alleles identified in each population.

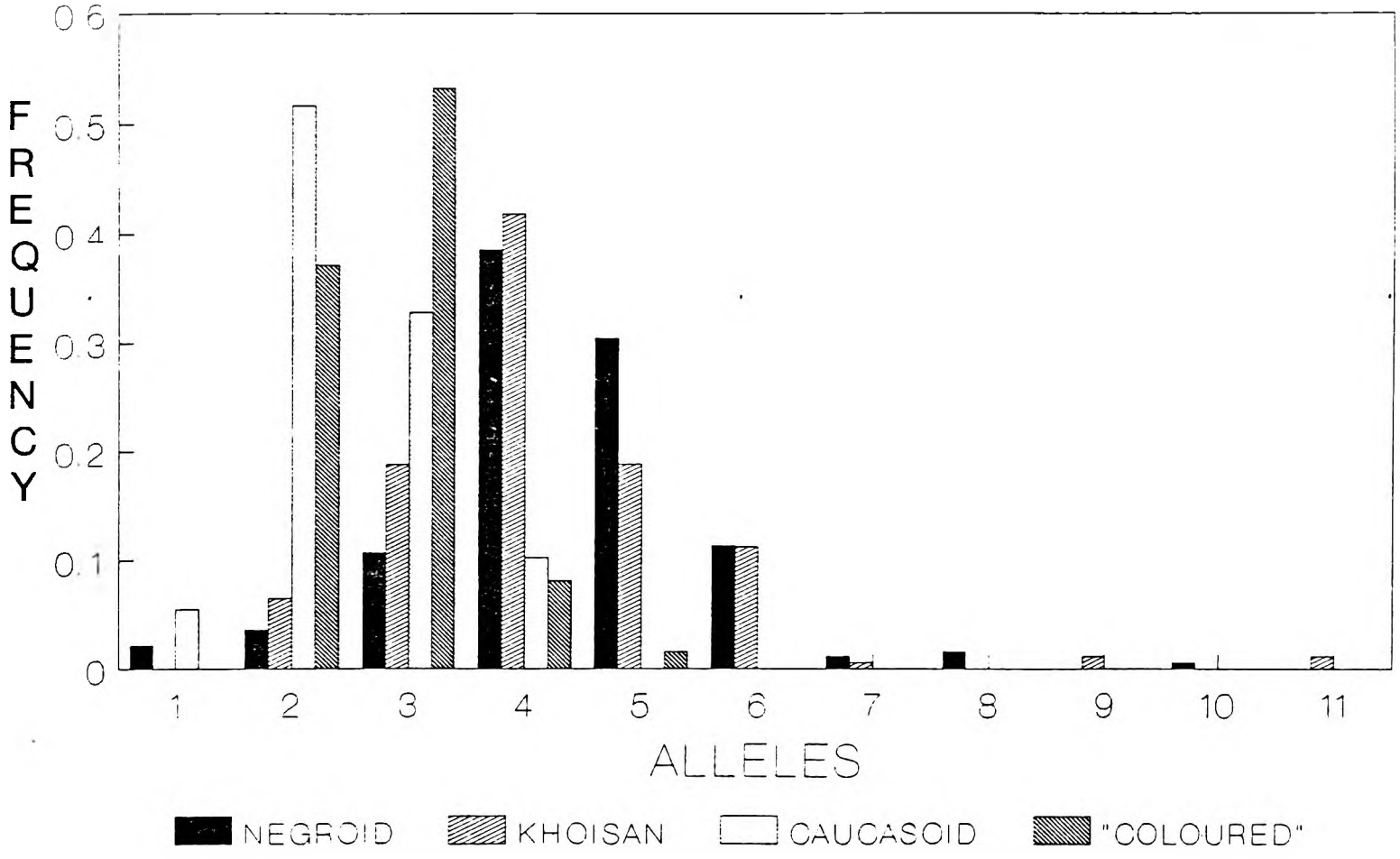


FIG 3.8 Frequency of binned alleles at the D4S125 locus in the four major southern African population groups

the right. Alleles *2*, *3* and *4* account for the bulk of the alleles in the four major groups (GN = 53%; KS = 67%; CC = 95%; CL = 98%). Alleles *2* and *3* are the most common ones in the Caucasoid and "Coloured" populations (NG = 14%; KS = 25%; CC = 84%; CL = 90%), whereas allele *4* is more prominent in the Negroid and Khoisan groups (NG = 39%; KS = 42%; CC = 10%; CL = 8%). Alleles *5* to *11* were not observed in the Caucasoid, are very rare in the "Coloured" and well represented in the Negroid and Khoisan (NG = 45%; KS = 32%; CC = 0%; CL = 2%).

Table 3.12 summarizes observed and expected numbers of homozygotes and heterozygotes in the 17 southern African populations considered separately and pooled into larger groups. The comparison is based on Hardy-Weinberg proportions and shows similarities to the results obtained for the VNTR locus considered in the previous section. Heterozygote deficiency was observed in some of the sub-groups of the Negroid, Khoisan ($p<0.05$) and Caucasoid ($p<0.05$; $0.01<p<0.05$) groups.

The genotypes of the 433 random individuals were used to study intrapopulation variation at this locus (Table 3.13). The unbiased heterozygote frequencies calculated from binned genotypes showed some differences in heterozygosity between the 17 population groups. The lowest values were found to be in the Tswana groups

TABLE 3 . 12 Differences between observed and expected numbers of heterozygotes at the D4S125 locus in 17 southern African populations considered separately and pooled into larger groups.

POPULATIONS	N	HOMOZYGOTES OBSERVED	HOMOZYGOTES EXPECTED	HETEROZYGOTES OBSERVED	HETEROZYGOTES EXPECTED	X	
NEGROIDS	**253**	**107**	**31.4**	**146**	**221.1**	**207.8**	*
NGUNI	102	35	14.0	67	88.3	40.0	*
Zulu	*31*	*7*	*5.4*	*24*	*25.6*	*0.6*	
Swazi	*31*	*9*	*6.1*	*22*	*24.8*	*1.7*	
Xhosa	*14*	*4*	*2.5*	*10*	*11.5*	*1.0*	
Tsonga	*26*	*15*	*5.0*	*11*	*21.0*	*24.6*	*
SOTHO-TSWANA	47	23	8.4	24	38.6	30.6	*
Sotho	*23*	*12*	*5.0*	*11*	*18.0*	*12.6*	*
Tswana	*11*	*6*	*3.4*	*5*	*7.7*	*3.0*	
Pedi	*13*	*5*	*4.4*	*8*	*8.5*	*0.1*	
LEMBA	27	15	4.5	12	22.5	29.2	*
VENDA	18	2	3.9	16	14.1	1.2	
HERERO	20	8	4.8	12	15.2	2.8	
DAMA	39	24	6.9	15	32.1	51.8	*
KHOISAN	**85**	**58**	**10.9**	**27**	**74.3**	**233.9**	*
NAMA	37	27	5.0	10	32.1	113.1	*
SAN	48	31	8.0	17	40.1	78.8	*
CAUCASOID	**64**	**26**	**14.5**	**38**	**49.2**	**11.7**	*
WHITES	20	7	5.4	13	14.6	0.7	
INDIANS	24	12	6.9	12	17.2	5.4	**
JEWS	20	7	5.4	13	14.6	0.6	
"COLOURED"	**31**	**12**	**10.7**	**19**	**20.3**	**0.2**	

N = Number of individuals in each population.
* = $p < 0.05$.
** = $0.01 < p < 0.05$

TABLE 3 .13 Heterozygosity rate at the D4S125 locus when 21 alleles were considered and when these were reduced to 11 bins. Seventeen populations were considered separately and pooled into larger groups.

	ALL ALLELES			BINNED ALLELES	
POPULATIONS	**N**	**K**	**H (%)**	**K**	**H (%)**
NEGROIDS	**506**	**19**	**87.6**	**9**	**73.6**
NGUNI	204	19	87.0	9	75.3
Zulu	*62*	*12*	*84.0*	*8*	*70.8*
Swazi	*62*	*13*	*81.4*	*7*	*69.1*
Xhosa	*28*	*8*	*85.1*	*5*	*70.0*
Tsonga	*52*	*9*	*82.5*	*6*	*72.1*
SOTHO-TSWANA	94	12	83.1	6	63.5
Sotho	*46*	*10*	*80.1*	*5*	*63.2*
Tswana	*22*	*4*	*72.9*	*3*	*45.1*
Pedi	*26*	*6*	*67.9*	*3*	*59.4*
LEMBA	54	9	85.0	5	63.8
VENDA	36	7	80.7	5	65.9
HERERO	40	7	78.0	4	69.9
DAMA	78	11	83.5	7	66.4
KHOISAN	**170**	**13**	**87.9**	**8**	**74.4**
NAMA	74	11	91.9	7	71.6
SAN	96	9	84.3	6	73.5
CAUCASOID	**128**	**8**	**77.4**	**4**	**62.0**
WHITES	40	6	74.9	4	65.3
INDIANS	48	5	73.1	4	58.8
JEWS	40	5	74.7	3	55.9
"COLOURED"	**62**	**6**	**66.5**	**4**	**58.2**

N = Number of alleles sampled
K = Number of alleles in the particular population
H = Heterozygosity

(45.1%) and the highest value for the San (73.5%). When the data on the population groups were pooled into larger groups, the lowest value was seen in the "Coloureds" (58.2%) and the highest in the Khoisan (74.4%). When the unbiased heterozygote frequency was calculated considering all 21 alleles (Table 3.13), as expected, higher values of heterozygosity were shown. The trend was very similar to that obtained when binned alleles were considered. The lowest value was found, as before, among the Sotho-Tswana group, namely the Pedi (67.9%) and the highest value again among the Khoisan, namely the San (91.9%). When only the four major groups were considered, the pattern observed was the same as when all 21 alleles were considered: the lowest value was found in the "Coloured" (66.5%) and the highest for the Khoisan (87.9%).

Genetic distances between the 17 southern African populations and the four main groups were computed as in section 3.2.1.1. The binned frequency data in Tables 3.10 and 3.11 was used. The genetic distances among the 17 populations are shown in Table 3.14. The populations with greater affinities are the Indians and the "Coloureds" (0) and on the other side of the spectrum, the most divergent ones are the Tsonga and the Venda (340). In general greater affinities were seen among the Caucasoid groups themselves and between them and the

TABLE 3 . 14 Genetic distances (x100) between 17 southern African populations using the pYNZ32 hypervariable probe.

	Zu	Sw	Xh	Tso	Sot	Tsw	Pd	Lem	Ve	He	Da	Na	San	Wht	Ind	Jew	Cl
Zu		78	53	57	71	45	79	57	13	75	44	110	30	25	30	29	30
Sw			84	68	86	85	43	92	60	72	92	39	63	64	64	64	64
Xh				120	14	44	67	6	57	37	67	110	29	29	29	28	29
Tso					130	110	150	120	340	140	56	170	91	86	91	91	91
Sot						78	45	10	84	24	90	98	46	47	47	46	47
Tsw							100	59	36	80	50	120	37	38	38	35	38
Pd								71	90	43	98	16	63	67	68	67	68
Lem									65	26	72	120	34	34	34	33	34
Ve										89	95	110	38	33	38	36	38
He											96	75	54	55	55	54	55
Da												130	43	36	43	41	43
Na													87	88	88	87	88
San														>0.1	>0.1	0.1	>0.1
Wht															>0.1	>0.1	>0.1
Ind																0.1	0
Jew																	0.1
Cl																	

"Coloureds", while greater divergence was observed in the majority of the other groups, particularly between the Nama and the Negroid populations. When genetic distances were determined between the four main groups (Table 3.15), the Khoisan shows the greater divergence from the Caucasoid (100) and the Negroid and Caucasoid groups are the ones showing the greatest affinity (11).

TABLE 3 . 15 Genetic distances (x100) between four major southern African populations using the pYNZ32 hypervariable probe.

	NEGROID	KHOISAN	CAUCASOID	"COLOURED"
NEGROID		82	11	30
KHOISAN			100	60
CAUCASOID				42
"COLOURED"				

3.2.1.3 VNTR AT THE APOB LOCUS

The PCR method was used to accurately and rapidly type the hypervariable region associated with the APOB gene. An example of the products of the amplification reactions from 727 random individuals is shown in Fig 3.9. As expected, no individuals exhibited more than two major PCR fragments. Based on the relative migration of fragments with respect to a DNA size standard, the observed fragments were shown to range in size from 420 bp to 990 bp (Table 3.3). Twenty alleles were scored and each one differed in length from its neighbouring allele by 30 bp or two 15 bp repeat units (Boerwinkle *et al*, 1989, Ludwig *et al*, 1989).

The genotypes of the 727 unrelated individuals distributed among the 17 population groups and pooled into the four major groups (NG = Negroid; KS = Khoisan, CC = Caucasoid, CL = "Coloureds") were used to calculate allele frequencies and their standard errors (Table 3.16 and 3.17 respectively).

Allele frequencies at the APOB locus in the four major groups are graphically summarised in Fig 3.10. In some populations the distribution is unimodal and in others it is bimodal, along with differences between the frequencies of specific alleles in different populations.

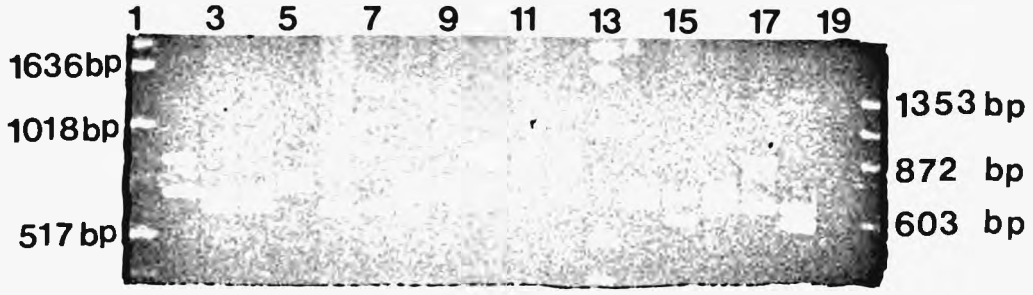


FIGURE 3.9 APOB VNTR amplification products of 15 random individuals after agarose gel electrophoresis and EtBr staining. Lanes 1 and 13 contain DNA size standards from 1kb ladder and lanes 7 and 20 contains DNA standards from ϕX174 digested with *Hae*III.

TABLE 3 . 16 Number and frequencies of the alleles at the APOB locus in 17 southern African populations.

	ZULU N = 102			SWAZI N = 76			XHOSA N = 40			TSONGA N = 80			SOTHO N = 92			TSWANA N = 38		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	-	-	-	-	-	-	1	0.025	0.025	1	0.013	0.012	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	1	0.010	0.010	-	-	-	-	-	-	-	-	-	1	0.011	0.010	-	-	-
4	2	0.020	0.014	2	0.026	0.018	-	-	-	-	-	-	-	-	-	1	0.026	0.026
5	-	-	-	-	-	-	-	-	-	-	-	-	3	0.033	0.019	-	-	-
6	5	0.049	0.021	1	0.013	0.013	-	-	-	3	0.038	0.021	3	0.033	0.019	3	0.079	0.044
7	17	0.167	0.037	7	0.092	0.033	5	0.125	0.052	8	0.100	0.033	13	0.141	0.036	5	0.132	0.055
8	5	0.049	0.021	8	0.105	0.035	2	0.050	0.034	5	0.063	0.027	6	0.065	0.026	2	0.053	0.036
9	12	0.118	0.032	5	0.066	0.028	7	0.175	0.060	15	0.186	0.044	15	0.163	0.039	4	0.105	0.049
10	28	0.274	0.044	23	0.303	0.053	6	0.150	0.056	16	0.200	0.045	19	0.206	0.042	9	0.237	0.069
11	2	0.020	0.014	3	0.039	0.022	1	0.025	0.025	3	0.038	0.021	2	0.021	0.015	-	-	-
12	4	0.039	0.019	4	0.052	0.026	6	0.150	0.056	9	0.113	0.035	3	0.033	0.019	1	0.026	0.026
13	10	0.098	0.029	12	0.158	0.042	4	0.100	0.047	7	0.088	0.031	6	0.065	0.026	6	0.158	0.059
14	8	0.078	0.027	4	0.052	0.026	4	0.100	0.047	6	0.075	0.029	13	0.141	0.036	4	0.105	0.049
15	2	0.020	0.014	6	0.079	0.031	3	0.075	0.042	3	0.038	0.021	4	0.043	0.021	2	0.053	0.036
16	2	0.020	0.014	1	0.013	0.013	-	-	-	2	0.025	0.017	-	-	-	-	-	-
17	1	0.010	0.010	-	-	-	-	-	-	-	-	-	1	0.011	0.011	1	0.026	0.026
18	3	0.029	0.017	-	-	-	-	-	-	2	0.025	0.017	1	0.011	0.011	-	-	-
19	-	-	-	-	-	-	1	0.025	0.025	-	-	-	2	0.021	0.015	-	-	-
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

N = Total number of particular alleles identified in each population.

TABLE 3 . 16 Cont. 2

	PEDI N=40			LEMBA N=84			VENDA N=60			HERERO N=88			DAMA N=62			NAMA N=74		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0.027	0.019
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3	-	-	-	4	0.048	0.023	-	-	-	-	-	-	-	-	-	-	-	-
4	1	0.025	0.025	-	-	-	1	0.017	0.017	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	2	0.050	0.034	-	-	-	4	0.067	0.032	-	-	-	-	-	-	-	-	-
7	5	0.125	0.052	8	0.095	0.032	3	0.050	0.028	11	0.125	0.035	7	0.113	0.040	25	0.337	0.055
8	2	0.050	0.034	3	0.036	0.020	1	0.017	0.017	10	0.114	0.034	15	0.242	0.054	8	0.108	0.036
9	9	0.225	0.066	4	0.048	0.023	1	0.017	0.017	9	0.102	0.032	9	0.145	0.044	6	0.081	0.032
10	5	0.125	0.052	18	0.214	0.044	13	0.217	0.053	16	0.182	0.041	5	0.081	0.035	8	0.108	0.036
11	-	-	-	1	0.012	0.012	1	0.017	0.017	19	0.216	0.044	7	0.113	0.040	8	0.108	0.036
12	6	0.150	0.056	5	0.059	0.026	9	0.150	0.046	7	0.079	0.029	3	0.048	0.027	5	0.068	0.029
13	3	0.075	0.042	14	0.167	0.040	9	0.150	0.046	6	0.068	0.027	5	0.081	0.035	1	0.014	0.013
14	3	0.075	0.042	13	0.155	0.039	12	0.200	0.050	4	0.045	0.022	7	0.113	0.040	2	0.027	0.019
15	2	0.050	0.034	5	0.059	0.026	2	0.033	0.023	4	0.045	0.022	3	0.048	0.027	5	0.070	0.029
16	-	-	-	5	0.059	0.026	1	0.017	0.017	2	0.023	0.016	-	-	-	-	-	-
17	-	-	-	-	-	-	2	0.033	0.023	-	-	-	-	-	-	-	-	-
18	1	0.025	0.025	1	0.012	0.011	-	-	-	-	-	-	-	-	-	2	0.027	0.019
19	1	0.025	0.025	3	0.036	0.020	-	-	-	-	-	-	1	0.016	0.016	2	0.027	0.019
20	-	-	-	-	-	-	1	0.017	0.017	-	-	-	-	-	-	-	-	-

TABLE 3 . 16 Cont. 3

	SAN N = 88			WHITES N = 230			INDIANS N = 78			JEWS N = 70			"COLOURED" N = 152		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	1	0.011	0.011	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	1	0.004	0.004	-	-	-	-	-	-	1	0.006	0.006
3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-	-	-	3	0.020	0.011
5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	-	1	0.004	0.004	1	0.013	0.013	-	-	-	3	0.020	0.011
7	30	0.341	0.051	25	0.109	0.021	7	0.090	0.032	16	0.228	0.050	30	0.197	0.032
8	18	0.205	0.043	14	0.061	0.016	4	0.051	0.025	3	0.043	0.024	5	0.033	0.014
9	8	0.091	0.031	46	0.200	0.026	41	0.526	0.057	13	0.186	0.047	44	0.289	0.037
10	12	0.136	0.037	85	0.369	0.032	12	0.154	0.041	19	0.271	0.053	33	0.217	0.037
11	3	0.034	0.019	9	0.039	0.013	3	0.038	0.022	2	0.029	0.019	9	0.059	0.019
12	5	0.057	0.025	4	0.017	0.009	1	0.013	0.013	-	-	-	5	0.033	0.014
13	2	0.023	0.016	-	-	-	2	0.026	0.018	-	-	-	7	0.046	0.017
14	4	0.045	0.022	1	0.004	0.004	4	0.051	0.025	3	0.043	0.024	3	0.020	0.011
15	4	0.045	0.022	14	0.061	0.016	1	0.013	0.013	9	0.129	0.040	1	0.006	0.006
16	1	0.011	0.011	17	0.074	0.017	2	0.026	0.018	4	0.057	0.028	7	0.046	0.017
17	-	-	-	11	0.048	0.014	-	-	-	1	0.014	0.014	-	-	-
18	-	-	-	2	0.008	0.006	-	-	-	-	-	-	1	0.006	0.006
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

TABLE 3 . 17 Number and alleles frequencies at the APOB locus in the four main southern African populations.

	NEGROID N = 762			KHOISAN N = 162			CAUCASOID N = 378			"COLOURED" N = 152		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	2	0.003	0.002	3	0.018	0.010	-	-	-	-	-	-
2	-	-	-	-	-	-	1	0.003	0.003	1	0.006	0.006
3	6	0.008	0.003	-	-	-	-	-	-	-	-	-
4	7	0.009	0.003	-	-	-	-	-	-	3	0.020	0.011
5	3	0.004	0.002	-	-	-	-	-	-	-	-	-
6	21	0.028	0.006	-	-	-	2	0.005	0.004	3	0.020	0.011
7	89	0.117	0.012	55	0.339	0.037	48	0.127	0.017	30	0.197	0.032
8	59	0.077	0.010	26	0.160	0.029	21	0.056	0.012	5	0.033	0.014
9	90	0.118	0.012	14	0.086	0.022	100	0.265	0.023	44	0.289	0.037
10	158	0.207	0.015	20	0.123	0.026	116	0.307	0.024	33	0.217	0.033
11	39	0.051	0.008	11	0.067	0.020	14	0.037	0.010	9	0.059	0.019
12	57	0.075	0.010	10	0.062	0.019	5	0.013	0.006	5	0.033	0.014
13	82	0.108	0.011	3	0.019	0.011	2	0.005	0.004	7	0.046	0.017
14	78	0.102	0.011	6	0.037	0.015	8	0.021	0.007	3	0.020	0.011
15	36	0.047	0.008	9	0.056	0.018	24	0.063	0.012	1	0.006	0.006
16	13	0.017	0.005	1	0.006	0.006	23	0.061	0.012	7	0.046	0.017
17	5	0.007	0.003	-	-	-	12	0.032	0.009	-	-	-
18	8	0.010	0.004	2	0.012	0.009	2	0.005	0.004	1	0.006	0.006
19	8	0.010	0.004	2	0.012	0.009	-	-	-	-	-	-
20	1	0.001	0.001	-	-	-	-	-	-	-	-	-

N = Total number of particular alleles identified in each population.

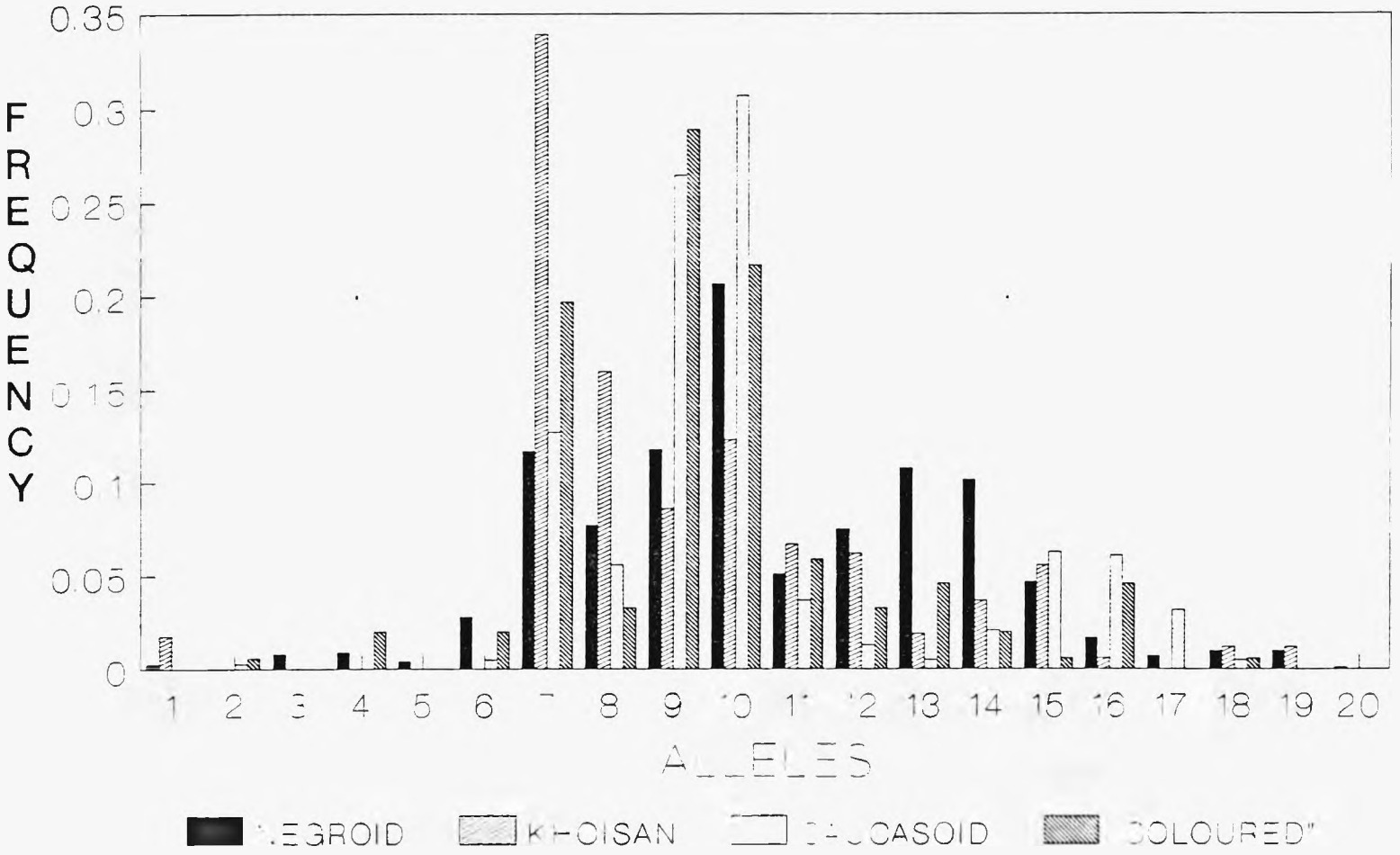


FIG 3.10 Frequency of alleles at the APOB locus in the four major southern African population groups

The bulk of the alleles is accounted for by the alleles *7*, *9* and *10* (NG = 44%; KS = 55%; CC = 70%; CL = 70%). Alleles *9* and *10* are the best represented alleles in Caucasoid and "Coloured" groups. (NG = 33%; KS = 21%; CC = 57%; CL = 51%). Allele *7* is found 2,8 times more frequently in the Khoisan than in the Negroid and Caucasoid groups and 1,7 times more frequently in the "Coloureds" (NG= 12%; KS = 34%, CC = 13%; CL = 20%). Alleles *11* and *12* are more frequent in the Negroids and Khoisan race and nearly twice more in the "Coloureds" than for the Caucasoid (NG = 13%; KS = 13%; CC = 5%; CL = 9%). Alleles *13* and *14* are more common in the Negroid than in any of the other 3 races (NG = 21%; KS = 6%; CC = 3%; CL = 7%). While alleles *1* and *6* are less frequent in the Caucasoid than in any of the other three groups (NG = 5%; KS = 2%; CC = 0.5%; CL = 2%), alleles *15* and *16* are twice more common in the Caucasoid than in any of the other races (NG = 6%; KS = 6%; CC = 12%; CL = 5%). Overall, the distribution is skew to the left.

Table 3.18 shows the observed and expected numbers of homozygotes and heterozygotes in the 17 southern African populations considered separately and pooled into larger groups. This comparison, based on Hardy-Weinberg proportions and calculated from the allele frequencies, showed no significant differences between observed and

TABLE 3 . 18 Observed and expected numbers of heterozygotes and homozygotes at the APOB locus in 17 southern African populations considered separately and pooled into larger groups.

POPULATIONS	N	HOMOZYGOTES OBSERVED	HOMOZYGOTES EXPECTED	HETEROZYGOTES OBSERVED	HETEROZYGOTES EXPECTED	X
NEGROIDS	**381**	**57**	**42.1**	**324**	**338.2**	**5.9** *
NGUNI	149	23	18.7	126	130.0	1.1
Zulu	*51*	*8*	*7.2*	*43*	*43.9*	*0.1*
Swazi	*38*	*3*	*5.9*	*35*	*40.0*	*2.0*
Xhosa	*20*	*3*	*4.9*	*17*	*15.1*	*1.0*
Tsonga	*40*	*9*	*4.8*	*31*	*35.3*	*4.2*
SOTHO-TSWANA	85	14	10.0	71	75.4	1.9
Sotho	*46*	*6*	*5.7*	*40*	*40.1*	*0.02*
Tswana	*19*	*5*	*2.6*	*14*	*16.4*	*2.7*
Pedi	*20*	*3*	*2.5*	*17*	*17.5*	*0.1*
LEMBA	42	4	5.2	38	36.8	0.3
VENDA	30	5	4.3	25	25.8	0.1
HERERO	44	5	5.9	39	38.0	0.2
DAMA	31	6	4.2	25	26.8	0.9
KHOISAN	**81**	**21**	**14.3**	**60**	**66.2**	**3.7**
NAMA	37	6	6.2	31	30.9	0.01
SAN	44	15	8.5	29	35.4	6.0 *
CAUCASOID	**189**	**42**	**36.8**	**147**	**152.2**	**1.0**
WHITES	115	25	23.6	90	90.9	0.1
INDIANS	39	12	12.4	27	26.7	0.01
JEWS	35	5	6.5	30	28.5	0.4
"COLOURED"	**76**	**18**	**13.7**	**58**	**62.0**	**1.6**

N = Number of Individuals in Population

* = $p < 0.02$

expected numbers for all populations except the San. When the data were pooled, a heterozygote deficiency was noted in the Negroid ($p<0.02$).

The genotypes of the 727 random individuals were used to study the intrapopulation variation at this VNTR locus (Table 3.19). The unbiased heterozygote frequencies showed their lowest value for the Indian population (69.2%) and its highest for the Xhosa group (90.1%). When only the four main groups were considered the gap between the lowest and the highest value for H is not as variable as when all the groups were considered separately. The Caucasoid groups showed the lowest value (80.6%) and the Negroid, the highest (89.1%).

Genetic distances at this hypervariable locus were calculated as in section 3.2.1.1. The allele frequency data of Tables 3.16 and 3.17 were used and the resulting distances are shown in Tables 3.20 and 3.21. In the first table, genetic distances were calculated taking in consideration the 17 southern African populations analysed in this study. The populations showing greatest affinity are the Indians and the Jews (0.3) and the ones showing greatest divergence are the Tswana and the San (73). Table 3.21 describes the genetic distances among the four main population groups of southern Africa.

TABLE 3 . 19 Heterozygosity rate at the APOB locus in 17 southern African populations considered separately and pooled into larger groups.

POPULATIONS	ALL ALLELES N	K	H (%)
NEGROIDS	**762**	**19**	**89.1**
NGUNI	298	17	88.0
Zulu	*102*	*15*	*86.7*
Swazi	*76*	*12*	*85.7*
Xhosa	*40*	*11*	*90.1*
Tsonga	*80*	*13*	*89.1*
SOTHO-TSWANA	170	16	88.2
Sotho	*92*	*15*	*88.6*
Tswana	*38*	*11*	*82.0*
Pedi	*40*	*12*	*89.7*
LEMBA	84	13	88.6
VENDA	60	14	87.2
HERERO	88	10	87.6
DAMA	62	10	87.8
KHOISAN	**162**	**13**	**82.7**
NAMA	74	12	84.4
SAN	88	11	82.2
CAUCASOID	**378**	**14**	**80.6**
WHITES	230	13	79.8
INDIANS	78	11	69.2
JEWS	70	9	82.7
"COLOURED"	**152**	**14**	**82.5**

N = Number of alleles sampled
K = Number of alleles sampled in a particular population
H = Heterozygosity

TABLE 3 . 20 Genetic distances (x100) between 17 southern African populations, at the APOB locus.

	Zu	Sw	Xh	Tso	Sot	Tsw	Pd	Lem	Ve	He	Da	Na	San	Wht	Ind	Jew	Cl
Zu		10	1	9	10	22	11	21	28	28	8	61	16	6	5	5	6
Sw			9	2	25	8	21	25	25	23	19	11	59	17	4	4	6
Xh				8	10	21	11	20	26	23	26	6	61	15	5	5	5
Tso					24	10	19	25	25	25	22	11	59	17	4	3	5
Sot						38	17	36	41	35	43	20	70	33	20	19	20
Tsw							35	38	36	37	24	22	73	23	16	16	19
Pd								31	25	28	34	17	55	32	17	17	19
Lem									17	12	22	9	42	19	22	23	24
Ve										11	13	15	21	27	24	24	26
He											16	10	40	17	23	23	23
Da												14	46	26	25	23	27
Na													47	11	8	9	9
San														65	57	54	60
Wht															15	15	13
Ind																0.3	1
Jew																	1
Cl																	

TABLE 3.21 Genetic distances (x100) between the four major southern African populations at the APOB locus.

	NEGROID	KHOISAN	CAUCASOID	"COLOURED"
NEGROID		25	41	16
KHOISAN			30	13
CAUCASOID				32
"COLOURED"				

In this case, greatest affinity is shown between the Khoisan and "Coloured" people (13) and least affinity between the Negroids and Caucasoids (41).

3.2.2 $(CA)_n$ REPEATS AND THE POLYMORPHISMS THEY IDENTIFY

The PCR technique and denaturing polyacrylamide gels were used to analyse the two $(CA)_n$ repeat loci (APOA2 and D21S168). Examples of different fragment sizes are shown for APOA2 and D21S168 in Figs 3.11 and 3.13, respectively. The autoradiograms show darker bands in the position of the allele and some faint artifactual bands differing in length by multiples of the 2bp core repeat.

3.2.2.1 $(CA)_n$ REPEAT AT THE APOA2 LOCUS

Amplified DNA from the APOA2 $(CA)_n$ block from 720 unrelated individuals belonging to 17 different population groups produced allelic fragments ranging in size from 129 bps to 147 bps (Table 3.3). Ten different alleles were scored and each one differs in length from the next one by 2 bp repeat units. An example of an autoradiogram showing allelic variation after amplification and separation of DNA on a standard denaturing polyacrylamide sequencing gel is shown in Fig 3.11.

The genotypes of the 720 unrelated individuals distributed among all 17 populations and samples pooled into the four main groups were used to calculate allele frequen-

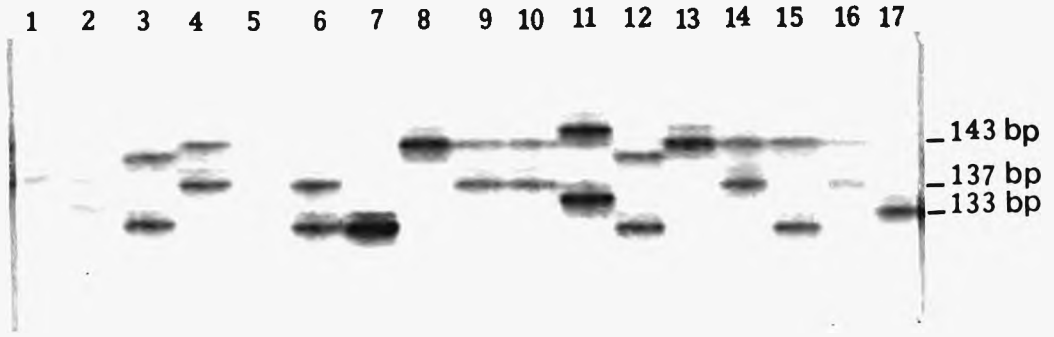


FIGURE 3.11 Autoradiogram of PCR products from 14 unrelated individuals primed with the Mfd3 marker at the APOA2 locus. Lane 3, 11 and 12 contain DNA from two individuals, these were used as internal molecular size markers. There is a continuous variation of fragment sizes but only some sizes in base pairs are indicated on the right hand side.

cies and their standard errors (Tables 3.22. and 3.23).

Allele frequencies at the APOA2 locus in the four main groups (NG = Negroid; KS = Khoisan; CC = Caucasoids; CL = "Coloureds") are graphically summarized in the Fig 3.12. Variation in allele distribution is observed. Alleles *8* and *5* are the commonest alleles in all the four major groups. Allele *8* is commoner in the Caucasoids and "Coloureds" (NG = 20%; KS = 27%; CC = 32%; CL = 31%) and allele *5* is more highly represented in the Negroid and "Coloured" groups (NG = 42%; KS = 19%; CC = 30%; CL = 40%). It is interesting that allele *2* is, on average, twice as frequent in the Khoisan than in any of the others (NG = 9%; KS = 20%; CC = 13%; CL = 10%) and allele *7* is three times less frequent in the Caucasoid and "Coloured" groups than in the Negroid and Khoisan groups (NG = 13%; KS = 15%; CC = 4%; CL = 4%).

Table 3.24 summarizes observed and expected numbers of homozygotes and heterozygotes in the 17 southern African populations considered separately and pooled into larger groups. This comparison, based on Hardy-Weinberg proportions calculated from the observed allele frequencies, shows complete concordance between observed and expected observations for this system.

TABLE 3 .22 Number and frequencies of the alleles at the APOA2 locus in 17 southern African populations.

	ZULU N=76			SWAZI N=64			XHOSA N=44			TSONGA N=38			SOTHO N=90			TSWANA N=64		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	-	-	-	-	-	-	1	0.023	0.023	1	0.026	0.026	-	-	-	-	-	-
2	7	0.092	0.033	6	0.094	0.036	6	0.136	0.052	1	0.026	0.026	11	0.122	0.034	3	0.047	0.026
3	1	0.013	0.013	2	0.031	0.022	2	0.045	0.031	1	0.026	0.026	-	-	-	-	-	-
4	5	0.066	0.028	4	0.063	0.030	6	0.136	0.052	6	0.158	0.059	8	0.089	0.030	4	0.063	0.030
5	34	0.447	0.057	30	0.469	0.062	13	0.295	0.069	11	0.289	0.074	39	0.433	0.052	30	0.469	0.062
6	2	0.026	0.018	5	0.078	0.034	2	0.045	0.031	6	0.158	0.059	-	-	-	2	0.031	0.022
7	9	0.118	0.037	6	0.094	0.036	7	0.159	0.055	6	0.158	0.059	12	0.133	0.036	6	0.094	0.036
8	17	0.224	0.048	10	0.156	0.045	6	0.136	0.052	6	0.158	0.059	16	0.178	0.040	17	0.266	0.055
9	1	0.013	0.013	-	-	-	1	0.023	0.023	-	-	-	4	0.044	0.022	2	0.031	0.022
10	-	-	-	1	0.016	0.016	-	-	-	-	-	-	-	-	-	-	-	-

N = Total number of particular alleles identified in each population.

TABLE 3 . 22 Cont. 2.

	PEDI N = 64			LEMBA N = 88			VENDA N = 64			HERERO N = 120			DAMA N = 92			NAMA N = 78		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	2	0.031	0.022	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	1	0.016	0.016	9	0.102	0.032	4	0.063	0.030	11	0.092	0.026	10	0.109	0.032	9	0.115	0.036
3	2	0.031	0.022	2	0.023	0.016	1	0.016	0.016	-	-	-	-	-	-	4	0.050	0.025
4	9	0.141	0.044	2	0.023	0.016	1	0.016	0.016	4	0.033	0.016	2	0.022	0.015	2	0.026	0.018
5	21	0.328	0.059	42	0.477	0.053	33	0.516	0.062	49	0.408	0.045	35	0.380	0.051	23	0.295	0.052
6	4	0.063	0.030	3	0.034	0.019	3	0.047	0.026	4	0.033	0.016	6	0.065	0.026	3	0.038	0.022
7	13	0.203	0.050	13	0.148	0.038	9	0.141	0.044	19	0.158	0.033	6	0.065	0.026	8	0.103	0.034
8	9	0.141	0.044	15	0.170	0.040	13	0.203	0.050	29	0.242	0.039	25	0.272	0.046	22	0.282	0.051
9	3	0.047	0.026	1	0.011	0.011	-	-	-	4	0.033	0.016	6	0.065	0.026	6	0.077	0.030
10	-	-	-	1	0.011	0.011	-	-	-	-	-	-	2	0.022	0.015	1	0.013	0.013

TABLE 3 . 22 Cont. 3.

	SAN N = 46			WHITES N = 226			INDIANS N = 90			JEWS N = 76			"COLOURED" N = 120		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	-	-	-	1	0.004	0.004	-	-	-	-	-	-	-	-	-
2	16	0.348	0.070	24	0.106	0.020	10	0.111	0.033	15	0.197	0.046	12	0.100	0.027
3	2	0.043	0.030	26	0.115	0.021	16	0.178	0.040	9	0.118	0.037	9	0.075	0.024
4	3	0.065	0.036	1	0.004	0.004	2	0.022	0.015	3	0.039	0.022	1	0.008	0.008
5	1	0.022	0.022	64	0.283	0.030	32	0.356	0.050	22	0.289	0.052	48	0.400	0.045
6	3	0.065	0.036	8	0.035	0.012	3	0.033	0.019	-	-	-	3	0.025	0.014
7	10	0.217	0.061	9	0.040	0.013	5	0.056	0.024	-	-	-	5	0.042	0.018
8	11	0.239	0.063	81	0.358	0.32	18	0.200	0.042	25	0.329	0.054	37	0.308	0.042
9	-	-	-	8	0.035	0.012	2	0.022	0.015	2	0.026	0.018	5	0.042	0.018
10	-	-	-	4	0.017	0.009	2	0.022	0.015	-	-	-	-	-	-

TABLE 3.23 Number and frequencies of the alleles at the APOA2 locus in the four major southern African populations.

	NEGROIDS N = 804			KHOISAN N = 124			CAUCASOID N = 392			"COLOURED" N = 120		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	4	0.005	0.002	-	-	-	1	0.003	0.003	-	-	-
2	69	0.086	0.010	25	0.202	0.036	49	0.125	0.017	12	0.100	0.027
3	11	0.014	0.004	6	0.048	0.019	51	0.130	0.017	9	0.075	0.024
4	51	0.063	0.009	5	0.040	0.018	6	0.015	0.006	1	0.008	0.008
5	337	0.419	0.017	24	0.194	0.036	118	0.301	0.023	48	0.400	0.045
6	37	0.046	0.007	6	0.048	0.019	11	0.028	0.008	3	0.025	0.014
7	106	0.132	0.012	18	0.145	0.032	14	0.036	0.009	5	0.042	0.018
8	163	0.203	0.014	33	0.266	0.040	124	0.316	0.023	37	0.308	0.042
9	22	0.027	0.006	6	0.048	0.019	12	0.031	0.009	5	0.042	0.018
10	4	0.005	0.002	1	0.008	0.008	6	0.015	0.006	-	-	-

N = Total number of particular alleles identified in each population.

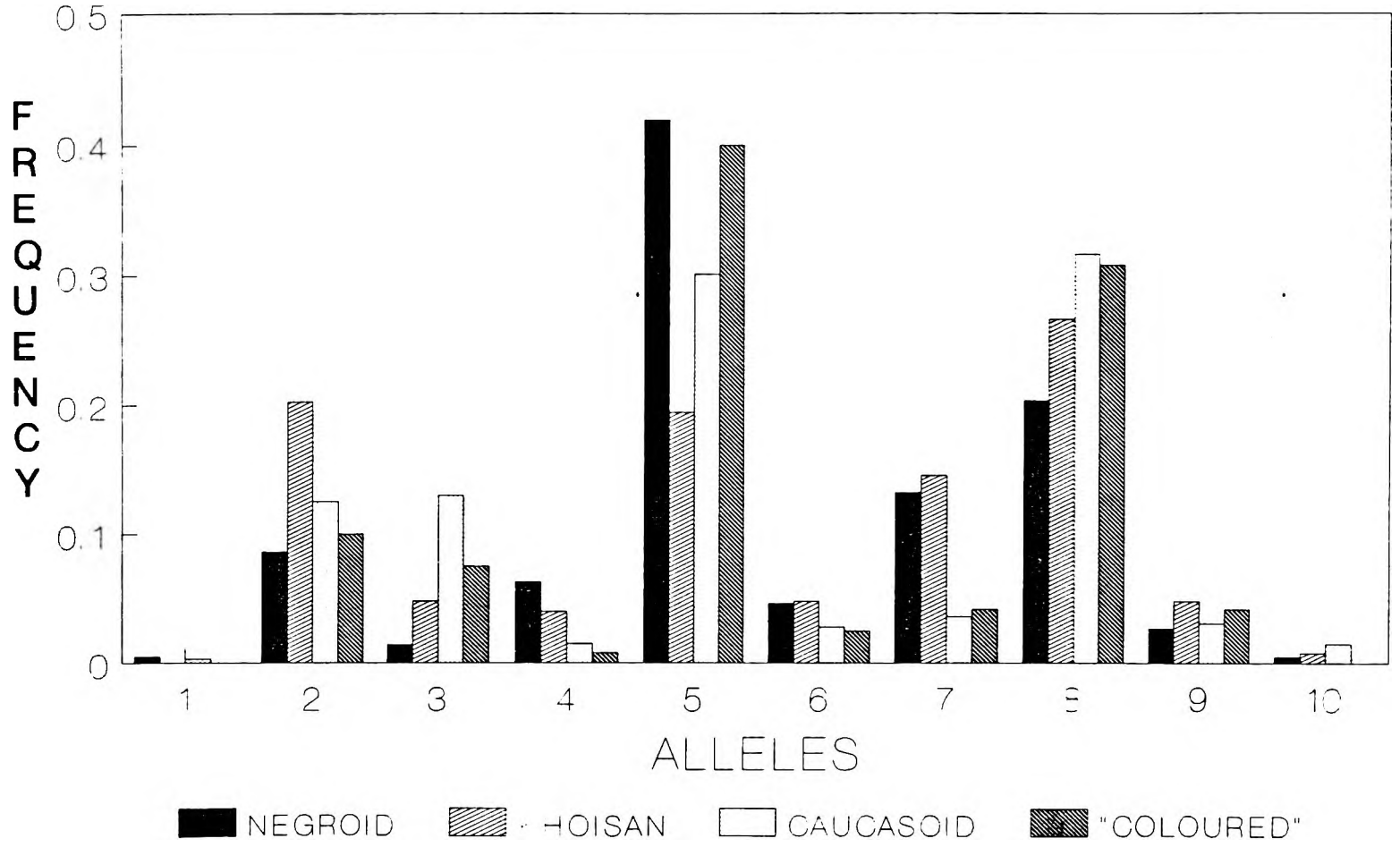


FIG 3.12 Frequency of alleles at the APOA2 Locus in the four major southern African population groups

TABLE 3 . 24 Observed and expected numbers of heterozygotes and homozygotes at the APOA2 locus in 17 southern African populations considered separately and pooled into larger groups.

POPULATIONS	N	HOMOZYGOTES OBSERVED	HOMOZYGOTES EXPECTED	HETEROZYGOTES OBSERVED	HETEROZYGOTES EXPECTED	X
NEGROIDS	**402**	**111**	**100.0**	**291**	**302.0**	**1.6**
NGUNI	111	30	25.1	81	86.1	1.2
Zulu	*38*	*11*	*10.6*	*27*	*27.4*	*0.02*
Swazi	*32*	*8*	*8.6*	*24*	*23.3*	*0.1*
Xhosa	*22*	*5*	*3.8*	*17*	*18.1*	*0.4*
Tsonga	*19*	*6*	*3.5*	*13*	*15.4*	*2.1*
SOTHO-TSWANA	109	29	26.7	80	82.3	0.3
Sotho	*45*	*12*	*11.8*	*33*	*33.1*	*0.01*
Tswana	*32*	*10*	*9.8*	*22*	*22.2*	*0.01*
Pedi	*32*	*7*	*6.3*	*25*	*25.8*	*0.1*
LEMBA	44	14	12.8	30	31.1	0.2
VENDA	32	11	10.7	21	21.4	0.02
HERERO	60	19	15.7	41	44.2	0.9
DAMA	46	10	11.2	36	34.8	0.9
KHOISAN	**62**	**13**	**11.1**	**49**	**50.8**	**0.4**
NAMA	39	7	7.8	32	31.1	0.1
SAN	23	6	5.4	17	17.5	0.1
CAUCASOID	**196**	**55**	**44.4**	**141**	**151.6**	**3.2**
WHITES	113	29	26.8	84	85.5	0.2
INDIANS	45	14	9.7	31	35.3	2.4
JEWS	38	12	9.4	26	28.5	1.0
"COLOURED"	**60**	**17**	**16.5**	**43**	**43.5**	**0.02**

N = Number of Individuals in Population.

The genotypes of the 720 individuals in 17 southern African population groups were determined to study the intrapopulation variation of this $(CA)_n$ repeat (Table 3.25). Unbiased heterozygote frequencies did vary between population groups. Values ranged between 67.7% for the Venda group and 84.6% for the Xhosa chiefdom. When the four major groups were considered, their heterozygosity values varied less dramatically, the lowest heterozygosity was observed in the "Coloured" (73.1%) and the highest heterozygosity in the Khoisan (82.6%).

The allele frequency data of Tables 3.22 and 3.23 were used to calculate genetic distances at this repeat locus (Tables 3.26 and 3.27). In Table 3.26, where all populations were considered, the greatest divergence was shown between the Tswana and the Whites (92) and the least between the Indian, Jewish and "Coloured" peoples (1). Table 3.27, containing the genetic distances between the four main races, shows overall very low values. The least divergent groups were the Caucasoid and "Coloured" (1) and the most divergent ones were the Negroid, Khoisan and "Coloured" (7).

TABLE 3 . 25 Heterozygosity rate at the APOA2 locus in 17 southern African populations considered separately and pooled into larger groups.

POPULATIONS	N	K	H (%)
NEGROIDS	**804**	**10**	**76.1**
NGUNI	222	10	78.8
Zulu	*76*	*8*	*73.2*
Swazi	*64*	*8*	*73.8*
Xhosa	*44*	*9*	*84.6*
Tsonga	*38*	*8*	*83.7*
SOTHO-TSWANA	218	9	76.1
Sotho	*90*	*6*	*74.8*
Tswana	*64*	*9*	*70.3*
Pedi	*64*	*7*	*81.6*
LEMBA	88	7	71.7
VENDA	64	9	67.7
HERERO	120	7	74.5
DAMA	92	8	76.4
KHOISAN	**124**	**9**	**82.6**
NAMA	78	9	80.9
SAN	46	7	77.9
CAUCASOID	**392**	**10**	**77.8**
WHITES	226	10	76.7
INDIANS	90	9	79.2
JEWS	76	6	76.3
"COLOURED"	**120**	**8**	**73.1**

N = Number of alleles sampled
K = Number of alleles in the particular population
H = Heterozygosity

TABLE 3 . 26 Genetic distances (x100) between 17 southern African populations at the APOA2 Locus

	Zu	Sw	Xh	Tso	Sot	Tsw	Pd	Lem	Ve	He	Da	Na	San	Wht	Ind	Jew	Cl
Zu		37	39	41	1	21	68	45	1	30	40	25	7	47	24	27	26
Sw			25	45	41	5	48	56	34	10	19	46	25	7	29	26	26
Xh				28	42	25	9	36	36	27	14	39	31	33	23	20	23
Tso					48	41	43	3	41	50	23	10	45	47	26	23	27
Sot						24	69	54	3	31	43	34	9	51	26	29	29
Tsw							45	5	19	11	18	35	13	92	19	18	17
Pd								55	66	50	25	64	62	46	40	38	41
Lem									46	61	35	9	49	62	37	32	38
Ve										23	40	27	4	43	22	24	23
He											28	52	18	14	29	28	27
Da												32	36	22	19	15	16
Na													29	55	31	29	31
San														39	24	24	24
Wht															26	25	22
Ind																1	1
Jew																	1
Cl																	

TABLE 3 . 27 Genetic distances (x100) between the four major southern African populations at the APOA2 locus.

	NEGROID	KHOISAN	CAUCASOID	"COLOURED"
NEGROID		7	5	3
KHOISAN			4	7
CAUCASOID				1
"COLOURED"				

3.2.2.2 $(CA)_n$ REPEAT AT THE D21S168 LOCUS

DNA from 861 unrelated individuals distributed among 17 different southern African population groups was amplified at the D21S168 $(CA)_n$ repeat block. Twelve different alleles differing from one other by 2 bp repeat units and ranging in size from 100 bp to 122 bp (Table 3.3) were observed. Figure 3.13 shows an autoradiogram of allelic variation after amplification and separation of the DNA from different individuals on a standard denaturing polyacrylamide sequencing gel.

The genotypes of the 861 individuals distributed among the 17 populations and among the four major groups were used to calculate allele frequencies and their standard errors (Tables 3.28 and 3.29).

Allele frequencies at the D21S168 locus among the four major groups are graphically summarized in Fig 3.14. There is striking variation in the distribution of alleles. The Negroid and the "Coloured" groups show a unimodal distribution while the Khoisan and Caucasoid have a bimodal one. In the Negroid group all twelve alleles were observed whereas only nine were observed in the Caucasoids. Alleles 5, 6 and 7 are common alleles in all 4 groups (NG = 47%; KS = 47%; CC = 54%; CL = 44%).

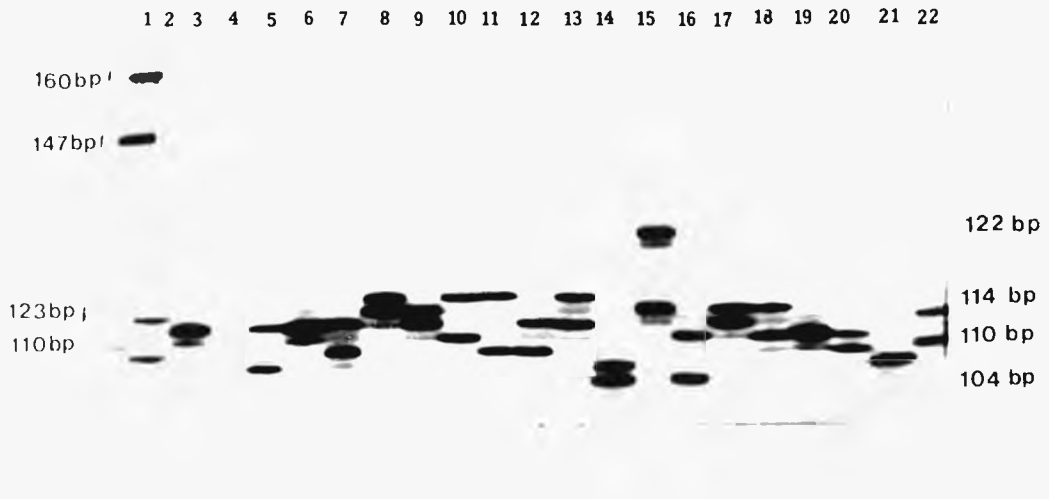


FIGURE 3.13 An autoradiogram showing PCR products at the D21S168 locus from twenty one unrelated individuals. Lane 1 contains a pBR322 molecular size marker. There is a continuous variation of fragment sizes but only some sizes in base pairs (bp) are indicated on both sides of the autoradiograph.

TABLE 3 . 28 Number and frequencies of the alleles at the D21S168 locus in 17 southern African populations.

	ZULU N = 114			SWAZI N = 74			XHOSA N = 50			TSONGA N = 76			SOTHO N = 116			TSWANA N = 84		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	-	-	-	-	-	-	1	0.020	0.020	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	1	0.020	0.020	1	0.013	0.013	2	0.017	0.012	-	-	-
3	5	0.044	0.019	1	0.014	0.014	-	-	-	-	-	-	7	0.060	0.022	1	0.012	0.012
4	4	0.035	0.017	3	0.041	0.023	2	0.040	0.028	2	0.026	0.018	7	0.060	0.022	3	0.036	0.020
5	31	0.272	0.042	17	0.229	0.049	10	0.200	0.057	12	0.158	0.042	18	0.155	0.034	19	0.226	0.046
6	31	0.272	0.042	16	0.216	0.048	11	0.220	0.059	21	0.276	0.051	26	0.224	0.039	16	0.190	0.043
7	18	0.158	0.034	14	0.189	0.046	15	0.300	0.065	23	0.303	0.053	24	0.207	0.038	26	0.309	0.050
8	10	0.088	0.027	12	0.162	0.043	5	0.100	0.042	7	0.092	0.033	16	0.138	0.032	8	0.095	0.032
9	9	0.079	0.025	4	0.054	0.026	2	0.040	0.028	3	0.039	0.022	6	0.052	0.021	5	0.059	0.026
10	4	0.035	0.017	2	0.027	0.019	2	0.040	0.028	5	0.066	0.028	6	0.052	0.021	4	0.048	0.023
11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0.024	0.017
12	2	0.018	0.012	5	0.068	0.029	1	0.020	0.020	2	0.026	0.018	4	0.034	0.017	-	-	-

N = Total number of particular alleles identified in each population.

TABLE 3 . 28 Cont. 2

	PEDI N = 76			LEMBA N = 94			VENDA N = 64			HERERO N = 122			DAMA N = 94			NAMA N = 94		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	2	0.026	0.018	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	2	0.026	0.018	1	0.011	0.011	-	-	-	-	-	-	3	0.032	0.018	3	0.032	0.018
3	3	0.039	0.022	3	0.032	0.018	3	0.047	0.026	10	0.082	0.025	5	0.053	0.023	3	0.032	0.018
4	5	0.066	0.028	2	0.021	0.015	2	0.031	0.022	2	0.016	0.011	3	0.032	0.018	12	0.128	0.034
5	9	0.118	0.037	18	0.191	0.041	25	0.391	0.061	21	0.172	0.034	19	0.202	0.041	8	0.085	0.029
6	18	0.237	0.049	26	0.277	0.046	11	0.172	0.047	27	0.221	0.038	24	0.255	0.045	19	0.202	0.041
7	17	0.224	0.048	19	0.202	0.041	9	0.141	0.044	35	0.287	0.041	23	0.245	0.044	26	0.276	0.046
8	14	0.184	0.044	18	0.191	0.041	5	0.078	0.034	17	0.139	0.031	11	0.117	0.033	13	0.138	0.036
9	5	0.066	0.028	2	0.021	0.015	5	0.078	0.034	8	0.065	0.022	2	0.021	0.015	8	0.085	0.029
10	-	-	-	3	0.032	0.018	-	-	-	1	0.008	0.008	-	-	-	1	0.011	0.011
11	-	-	-	1	0.011	0.011	1	0.016	0.016	-	-	-	1	0.011	0.011	1	0.011	0.011
12	1	0.013	0.013	1	0.011	0.011	3	0.047	0.026	1	0.008	0.008	3	0.032	0.018	-	-	-

TABLE 3 . 28 Cont. 3.

	SAN N = 88			WHITES N = 252			INDIANS N = 96			JEWS N = 86			"COLOURED" N = 142		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	1	0.011	0.011	-	-	-	-	-	-	-	-	-	3	0.021	0.012
3	3	0.034	0.019	36	0.143	0.022	5	0.052	0.023	11	0.128	0.036	11	0.077	0.022
4	18	0.205	0.043	4	0.016	0.008	4	0.042	0.020	-	-	-	12	0.085	0.023
5	7	0.079	0.029	18	0.070	0.016	2	0.021	0.015	5	0.058	0.025	19	0.134	0.029
6	13	0.148	0.038	42	0.166	0.023	19	0.198	0.041	23	0.267	0.048	31	0.220	0.035
7	28	0.318	0.050	89	0.353	0.030	39	0.406	0.050	23	0.267	0.048	32	0.225	0.035
8	2	0.023	0.028	50	0.198	0.025	22	0.229	0.043	16	0.186	0.042	21	0.148	0.030
9	7	0.079	0.029	13	0.050	0.014	2	0.021	0.015	8	0.093	0.031	10	0.070	0.021
10	8	0.091	0.031	-	-	-	2	0.021	0.015	-	-	-	2	0.014	0.010
11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	1	0.011	0.011	-	-	-	1	0.010	0.012	-	-	-	1	0.007	0.007

TABLE 3 . 29 Number and frequencies of the alleles at the D21S168 locus in the four major southern African populations.

	NEGROID N = 964			KHOISAN N = 182			CAUCASOID N = 434			"COLOURED" N = 142		
ALLELE	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.	No	Freq	S.E.
1	3	0.003	0.002	-	-	-	-	-	-	-	-	-
2	10	0.010	0.003	4	0.022	0.011	-	-	-	3	0.021	0.012
3	38	0.039	0.006	6	0.033	0.013	52	0.120	0.016	11	0.077	0.022
4	35	0.036	0.006	30	0.165	0.028	8	0.018	0.006	12	0.085	0.023
5	199	0.206	0.013	15	0.082	0.020	25	0.057	0.011	19	0.134	0.029
6	227	0.235	0.014	32	0.176	0.028	84	0.193	0.019	31	0.220	0.035
7	223	0.231	0.014	54	0.297	0.034	151	0.348	0.023	32	0.225	0.035
8	123	0.128	0.011	15	0.082	0.020	88	0.203	0.019	21	0.148	0.030
9	51	0.053	0.007	15	0.082	0.020	23	0.053	0.011	10	0.070	0.021
10	27	0.028	0.005	9	0.049	0.016	2	0.005	0.003	2	0.014	0.010
11	5	0.005	0.002	1	0.005	0.005	-	-	-	-	-	-
12	23	0.024	0.005	1	0.005	0.005	1	0.002	0.002	1	0.007	0.007

N = Total number of particular alleles identified in each population

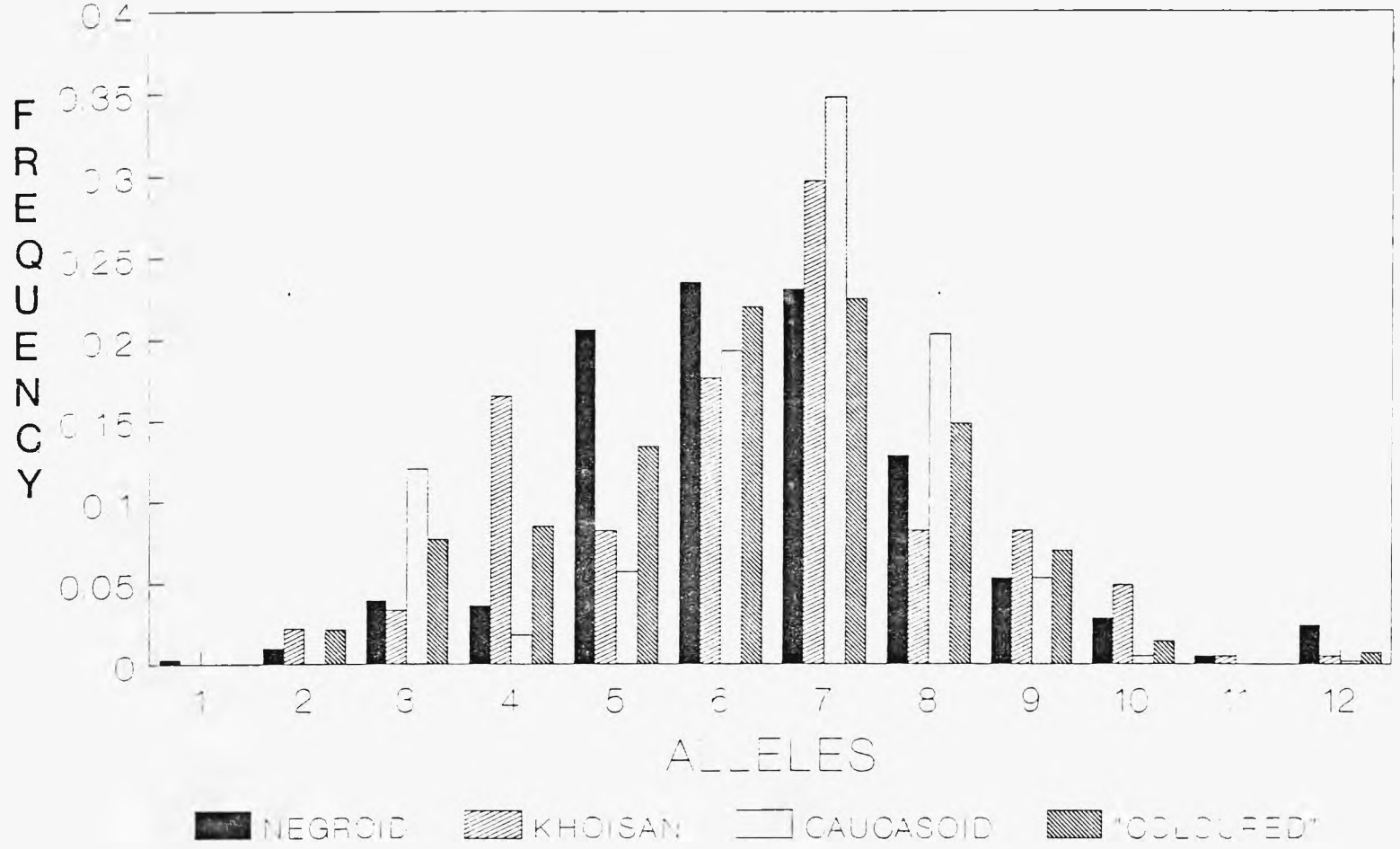


FIG 3.14 Frequency of alleles at the D21S168 locus in four major southern African population groups

Alleles *9, 10, 11* and *12* as a whole, are less represented in the Caucasoid groups, in relation to the Negroid and Khoisan groups (NG = 11%; KS = 14%; CC = 6%; CL = 9%). In contrast allele eight is better represented in the Caucasoid group (NG = 13%; KS = 8%; CC = 20%; CL = 15%).

Table 3.30 summarizes observed and expected numbers of homozygotes and heterozygotes in the 17 southern African populations considered separately and pooled into larger groups. This comparison based on Hardy-Weinberg porportions calculated from the observed allele frequencies, shows complete concordance between observed and expected for all populations except the Lemba ($p<0,05$), San and Jews ($p<0.01$). When the data was pooled, a heterozygote deficiency was noted in the Khoisan only ($p<0.05$). This lack of concordance is probably due to small sample size.

The genotypes of the 861 individuals distributed among the 17 populations were determined to study the intrapopulation variation at this $(CA)_n$ repeat (Table 3.31). The unbiased heterozygote frequency values ranged between (74.5%) in the Indian population and (85.7%) in the Sotho group. When only the 4 major groups were considered, the lowest value was found in the Caucasoid (77.7%) and the highest heterozygosity rate among the "Coloured" (84.8%).

TABLE 3 . 30 Differences between observed and expected numbers of heterozygotes at the D21S168 locus in 17 southern African populations considered separately and pooled into larger groups.

POPULATIONS	N	HOMOZYGOTES OBSERVED	HOMOZYGOTES EXPECTED	HETEROZYGOTES OBSERVED	HETEROZYGOTES EXPECTED	X	
NEGROIDS	**482**	**94**	**84.1**	**388**	**397.396**	**1.4**	
NGUNI	157	33	28.5	124	127.8	0.8	
Zulu	*57*	*12*	*10.9*	*45*	*46.0*	*0.1*	
Swazi	*37*	*4*	*6.3*	*33*	*30.7*	*1.0*	
Xhosa	*25*	*7*	*4.9*	*18*	*20.1*	*1.2*	
Tsonga	*38*	*10*	*7.9*	*28*	*30.0*	*0.7*	
SOTHO-TSWANA	138	24	22.4	114	115.4	0.1	
Sotho	*58*	*9*	*8.7*	*49*	*49.2*	*0.01*	
Tswana	*42*	*7*	*8.4*	*35*	*33.5*	*0.3*	
Pedi	*38*	*8*	*6.3*	*30*	*31.6*	*0.5*	
LEMBA	47	15	9.1	32	37.9	4.7	*
VENDA	32	7	7.0	25	25.0	>0.01	
HERERO	61	7	11.7	54	49.1	2.4	
DAMA	47	8	8.7	39	38.3	0.1	
KHOISAN	**91**	**23**	**15.5**	**68**	**75.3**	**4.297**	*
NAMA	47	7	8.0	40	39.1	0.1	
SAN	44	16	8.3	28	35.7	8.9	**
CAUCASOID	**217**	**54**	**47.8**	**163**	**168.7**	**1.0**	
WHITES	126	26	27.7	100	98.1	0.1	
INDIANS	48	12	12.6	36	35.4	0.04	
JEWS	43	16	8.8	27	34.1	7.3	**
"COLOURED"	**71**	**8**	**11.2**	**63**	**60.0**	**1.1**	

N = Number of individuals in a population

* = $p < 0.05$ ** = $p < 0.01$

TABLE 3 . 31 Heterozygosity rate at the D21S168 locus in 17 southern African populations considered separately and pooled into larger groups.

POPULATION	N	K	H (%)
NEGROIDS	**964**	**12**	**82.5**
NGUNI	314	11	82.0
Zulu	*114*	*9*	*81.6*
Swazi	*74*	*9*	*84.2*
Xhosa	*50*	*10*	*82.2*
Tsonga	*76*	*9*	*80.2*
SOTHO-TSWANA	276	12	83.7
Sotho	*116*	*10*	*85.7*
Tswana	*84*	*9*	*81.0*
Pedi	*76*	*10*	*84.5*
LEMBA	94	11	81.5
VENDA	64	9	79.2
HERERO	122	9	81.5
DAMA	94	10	82.3
KHOISAN	**182**	**11**	**83.9**
NAMA	94	10	83.9
SAN	88	10	83.7
CAUCASOID	**434**	**9**	**77.7**
WHITES	252	7	78.4
INDIANS	96	9	74.5
JEWS	86	6	80.4
"COLOURED"	**142**	**10**	**84.8**

N = Number of alleles sampled
K = Number of alleles in a particular population
H = Heterozygosity

The allele frequency data of Tables 3.28 and 3.29 were used to calculate genetic distances. In Table 3.32 all 17 populations were considered. In general, the greatest affinities are seen among the Caucasoid groups themselves and between the "Coloured" and San, while greater divergence was observed in the majority of the other groups. When only the four major groups were considered (Table 3.33) the greatest divergence was observed between the Khoisan, Negroid and Caucasoid groups (35). The least divergence was between the Caucasoid and the Negroid (3).

TABLE 3 . 32 Genetic distances (x100) between 17 southern African populations at the D21S168 locus.

	Zu	Sw	Xh	Tso	Sot	Tsw	Pd	Lem	Ve	He	Da	Na	San	Wht	Ind	Jew	Cl
Zu		46	4	46	4	53	4	43	13	54	8	24	14	16	15	11	15
Sw			47	0.9	4	7	43	3	36	12	41	16	27	27	28	30	31
Xh				46	2	53	1	42	7	53	4	29	16	19	19	16	19
Tso					39	6	41	4	35	11	40	16	26	26	28	29	29
Sot						45	2	34	8	48	3	22	12	15	14	12	15
Tsw							49	13	42	28	48	17	31	32	34	34	33
Pd								37	8	47	3	25	15	17	17	14	17
Lem									35	8	33	14	27	26	28	30	31
Ve										38	10	24	17	19	17	18	19
He											42	26	41	40	41	46	45
Da												26	17	20	20	17	21
Na													11	10	11	12	13
San														0.9	1	0.9	1
Wht															0.6	1	0.7
Ind																1	0.9
Jew																	1
Cl																	

TABLE 3 . 33 Genetic distances (x100) between the four major southern African populations at the D21S168 locus.

	NEGROID	KHOISAN	CAUCASOID	"COLOURED"
NEGROID		35	3	18
KHOISAN			35	32
CAUCASOID				14
"COLOURED"				

3.2.3 POPULATION AFFINITIES USING DIFFERENT COMBINATIONS OF THE HYPERVARIABLE SYSTEMS

Genetic distance measurements were determined in three different ways: for the three VNTR systems combined, the two $(CA)_n$ repeats combined and the five hypervariable loci used together. The reason for analysing the three VNTR loci independently of the two $(CA)_n$ repeats was to determine whether there is any appreciable difference in the analysis of population affinities when using VNTR loci or CA repeat loci. The genetic distances were calculated between the 17 southern Africa populations, the nine main linguistic groups and the four major racial groups according to the method of Harpending and Jenkins (1973). Dendrograms generated were based on cluster analysis from these genetic distances.

3.2.3.1 AT THE THREE VNTR LOCI

The allele frequency data in Tables 3.4, 3.5, 3.10, 3.11, 3.16, 3.17 were used to generate genetic distances between the 17 southern African population, the nine main linguistic groups and the four major racial groups (Tables 3.34, 3.35 and 3.36). Fig 3.15 represents a dendrogram based on the distances shown in Table 3.34 where all 17 populations were considered.

TABLE 3 . 34 Genetic distances (x100) between 17 southern African populations at three hypervariable loci.

	Zu	Sw	Xh	Tso	Sot	Tsw	Pd	Lem	Ve	He	Da	Na	San	Wht	Ind	Jew	Cl
Zu		25	12	53	59	69	38	60	14	32	53	38	119	236	197	62	11
Sw			48	61	53	75	61	34	41	41	37	68	129	226	158	11	17
Xh				81	83	91	50	86	31	45	77	48	139	266	227	37	37
Tso					43	39	56	71	40	47	55	55	90	191	138	124	56
Sot						56	68	46	61	37	31	59	91	182	134	131	58
Tsw							69	95	60	63	73	67	85	277	123	128	70
Pd								88	36	34	79	53	115	236	201	105	52
Lem									73	56	27	96	146	184	164	132	52
Ve										32	66	39	124	225	176	72	26
He											52	35	112	214	162	102	36
Da												71	114	195	152	124	45
Na													96	252	159	100	56
San														254	200	212	130
Wht															385	310	230
Ind																270	198
Jew																	89
Cl																	

TABLE 3 . 35 Genetic distances (x100) for nine linguistic groups of southern Africa using allele frequencies at three VNTR loci.

	NGUNI	SOT/TSW	LEMBA	VENDA	HERERO	DAMA	KHOISAN	CAUCASOID	"COLOURED"
NGUNI		48	57	52	18	26	100	172	51
SOT/TSW			60	62	48	54	63	214	53
LEMBA				44	46	46	101	178	93
VENDA					42	63	115	202	75
HERERO						24	89	186	65
DAMA							76	169	76
KHOISAN								221	126
CAUCASOID									237
"COLOURED"									

TABLE 3.36 Genetic distance measurements (x100) for the four major southern African populations using allele frequencies at three VNTR loci.

	NEGROID	KHOISAN	CAUCASOID	"COLOURED"
NEGROID		55	55	99
KHOISAN			42	160
CAUCASOID				150
"COLOURED"				

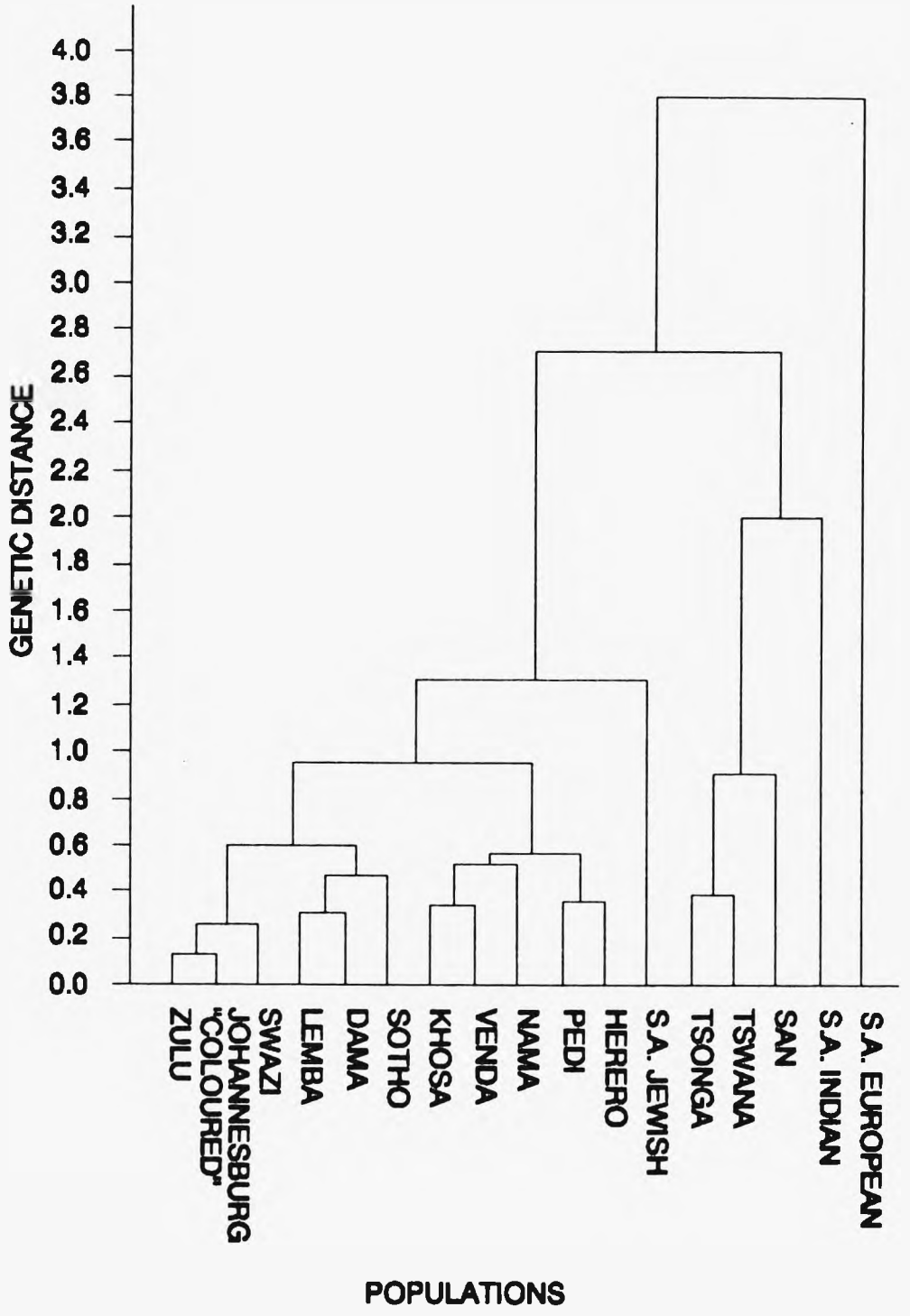


FIGURE 3.15 Dendrogram based on allele frequencies at the three VNTR loci in 17 southern African populations

In Figures 3.16 and 3.17 dendrograms are based on the distances presented in Tables 3.35 and 3.36, respectively where fewer groups are considered.

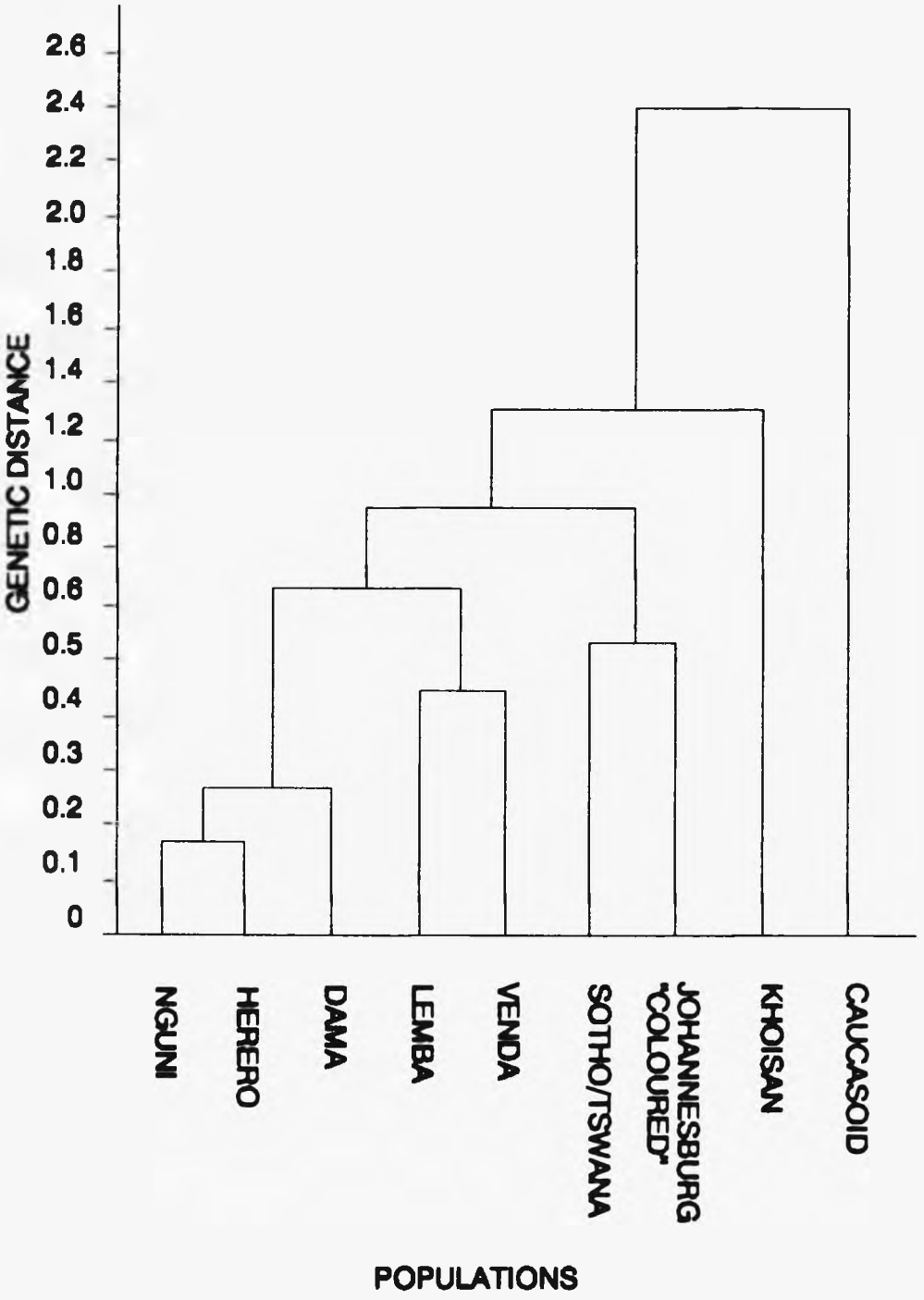


FIGURE 3.16 Dendrogram based on allele frequencies at the three VNTR loci in 9 main linguistic groups of southern Africa

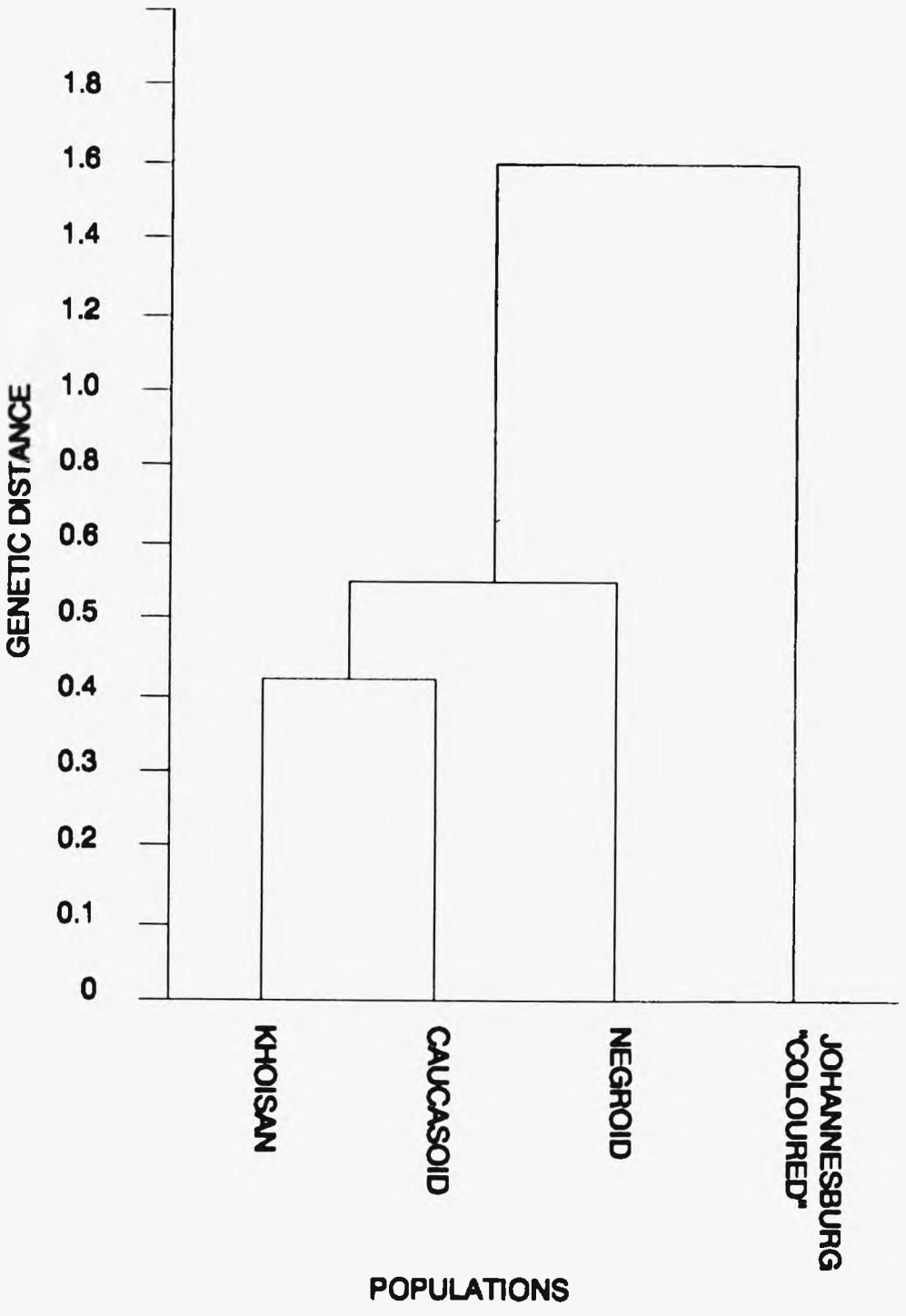


FIGURE 3.17 Dendrogram based on allele frequencies at the three VNTR loci in the four major population groups of southern Africa.

3.2.3.2 AT THE TWO $(CA)_n$ REPEAT LOCI

The frequency data in Tables 3.22, 3.23, 3.28 and 3.29 were used to generate genetic distances between the 17 southern African populations, nine main linguistic groups and the four main races, at the two $(CA)_n$ repeat loci (Tables 3.37, 3.38 and 3.39). Genetic distances were used to generate the dendrograms. Figures 3.18, 3.19 and 3.20 represent the dendrograms based on the distances from tables 3.37, 3.38 and 3.39 where the 17, 9 and 4 population divisions were considered.

TABLE 3 . 37 Genetic distances (x100) between 17 southern African populations at two (CA)n repeat loci.

	Zu	Sw	Xh	Tso	Sot	Tsw	Pd	Lem	Ve	He	Da	Na	San	Wht	Ind	Jew	Cl
Zu		55	82	46	20	73	42	24	32	36	44	27	61	64	71	56	70
Sw			47	82	49	10	40	40	27	35	46	49	45	67	76	58	74
Xh				40	64	48	24	47	46	42	35	52	71	56	89	41	61
Tso					51	88	33	42	51	48	52	32	78	66	69	36	74
Sot						80	39	22	32	34	45	28	47	71	67	58	79
Tsw							48	60	42	52	54	65	56	78	85	61	83
Pd								15	14	13	18	19	68	67	78	43	61
Lem									8	8	18	10	60	58	90	53	57
Ve										2	18	15	53	68	85	51	69
He											16	15	58	65	92	48	67
Da												21	63	67	99	52	63
Na													54	62	82	54	64
San														102	111	44	112
Wht															118	82	17
Ind																105	147
Jew																	96
Cl																	

TABLE 3 . 38 Genetic distances (x100) between nine linguistic groups of southern Africa using allele frequencies at two (CA)n repeat loci.

	NGUNI	SOT/TSW	LEMBA	VENDA	HERERO	DAMA	KHOISAN	CAUCASOID	"COLOURED"
NGUNI		45	78	25	32	40	42	84	65
SOT/TSW			58	45	38	37	67	100	49
LEMBA				45	50	43	49	85	52
VENDA					14	20	46	82	51
HERERO						8	46	85	54
DAMA							41	95	55
KHOISAN								100	84
CAUCASOID									120
"COLOURED"									

TABLE 3.39 Genetic distance measurements (x100) for the four major southern African populations using allele frequencies at two (CA)n repeat loci.

	NEGROID	KHOISAN	CAUCASOID	"COLOURED"
NEGROID		46	51	59
KHOISAN			44	49
CAUCASOID				67
"COLOURED"				

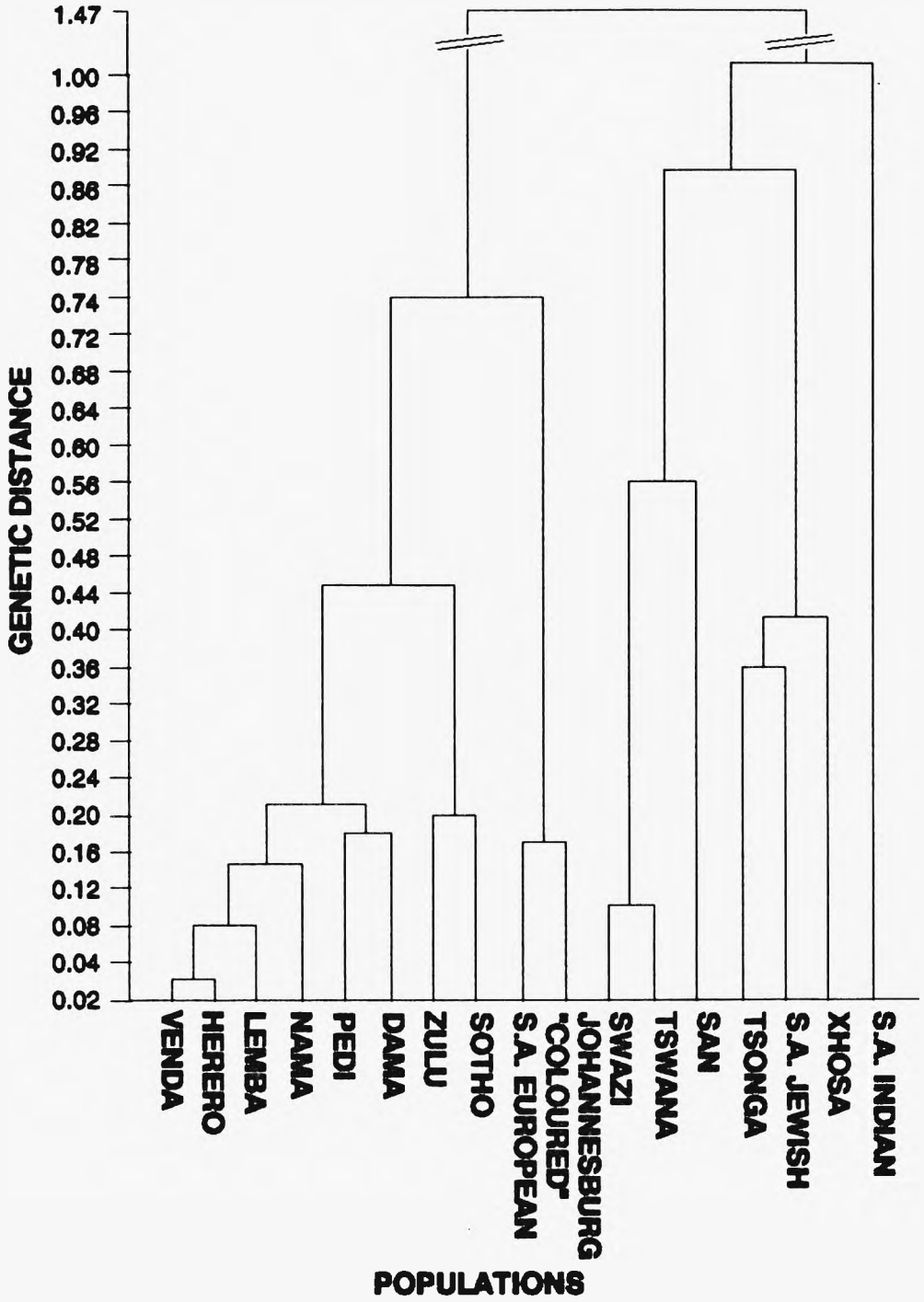


FIGURE 3.18 Dendrogram based on allele frequencies from the two $(CA)_n$ repeat loci in 17 southern African population groups

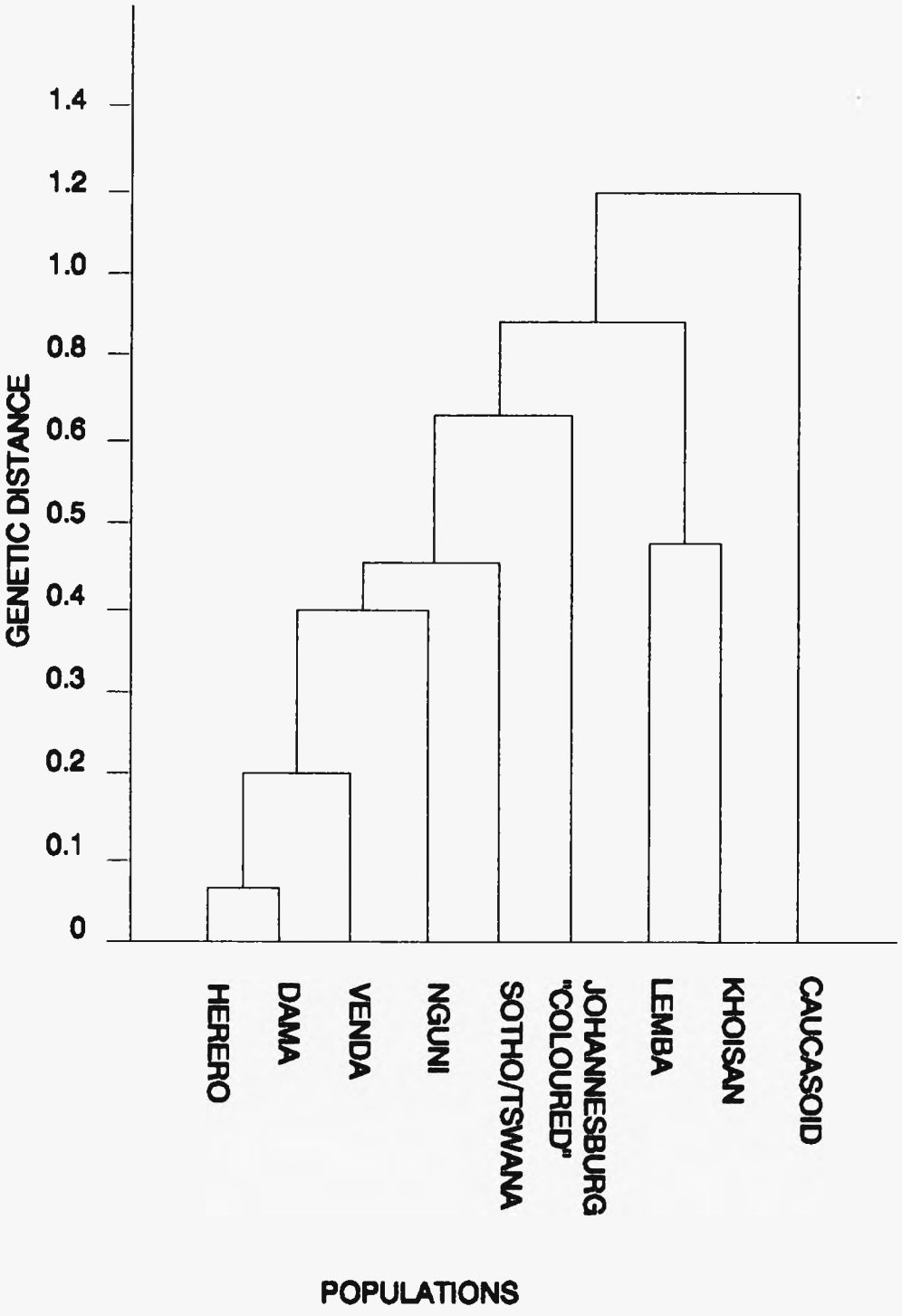


FIGURE 3.19 Dendrogram based on allele frequencies at the two $(CA)_n$ repeat loci for 9 linguistic groups.

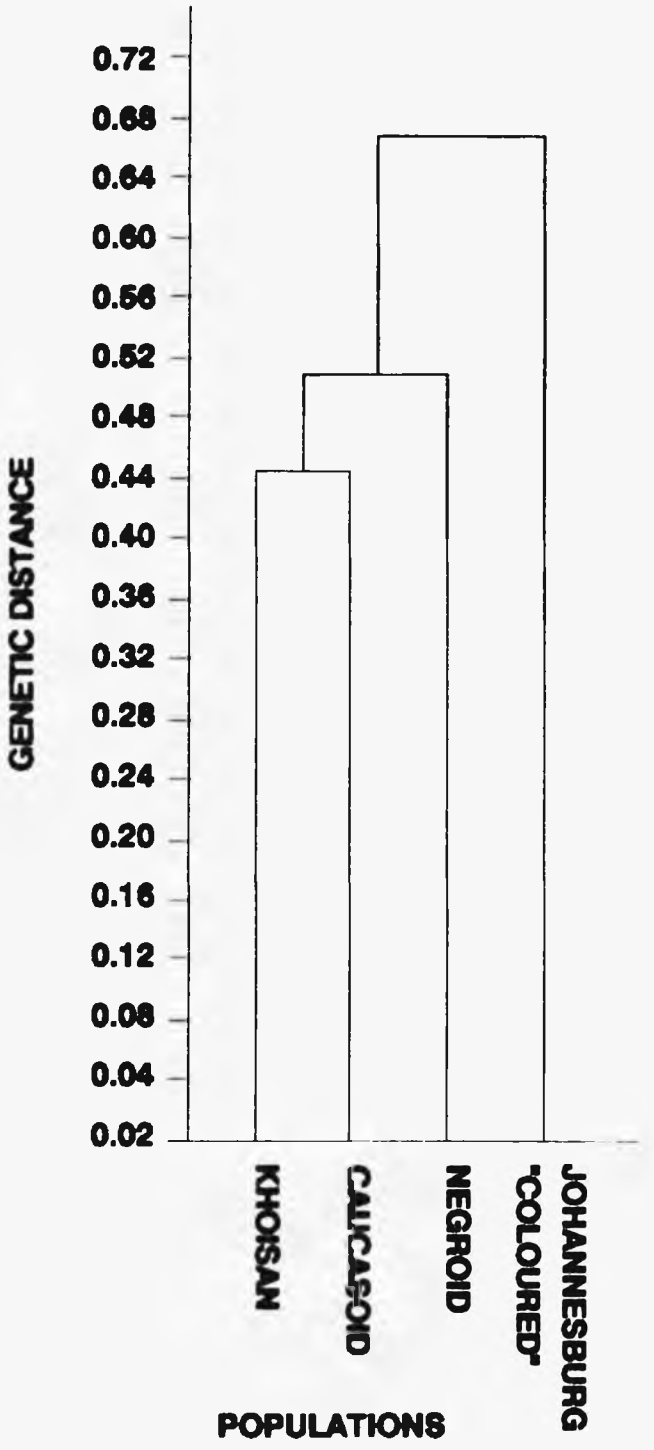


FIGURE 3.20 Dendrogram based on allele frequencies at the two $(CA)_n$ repeat loci in the four major population groups of southern Africa

3.2.3.3 AT THE FIVE HYPERVARIABLE LOCI ANALYSED TOGETHER

The combined frequency data of Tables 3.4, 3.5, 3.10, 3.11, 3.16, 3.17, 3.22, 3.23, 3.28 and 3.29 were used to generate genetic distances between the 17 southern African populations, the 9 main linguistic groups and the four major racial groups (Tables 3.40, 3.41, 3.42). The genetic distances were used to generate the dendrograms. Figures 3.21, 3.22 and 3.23 represent the dendrograms based on the distances from Tables 3.40, 3.41 and 3.42, respectively, where 17, 9 and 4 population groups were considered.

TABLE 3 . 40 Genetic distances (x100) between 17 southern African populations at five hypervariable loci.

	Zu	Sw	Xh	Tso	Sot	Tsw	Pd	Lem	Ve	He	Da	Na	San	Wht	Ind	Jew	Cl
Zu		57	82	93	52	58	47	73	225	320	67	25	104	115	138	146	94
Sw			63	114	83	50	63	70	176	347	132	58	130	191	185	239	94
Xh				111	87	66	65	97	184	265	154	78	99	130	119	194	100
Tso					139	74	117	80	297	211	131	96	84	155	124	202	116
Sot						90	48	117	198	362	131	85	150	123	166	159	117
Tsw							76	52	212	277	130	62	115	182	162	223	86
Pd								81	209	303	122	54	106	121	141	156	119
Lem									205	257	143	87	116	183	179	205	118
Ve										483	331	256	313	305	357	305	251
He											385	275	185	282	305	371	307
Da												93	161	153	147	204	119
Na													73	121	136	174	101
San														116	107	168	163
Wht															95	116	160
Ind																167	194
Jew																	249
Cl																	

TABLE 3 . 41 Genetic distances (x100) for nine linguistic groups of southern Africa using allele frequencies at five hypervariable loci.

	NGUNI	SOT/TSW	LEMBA	VENDA	HERERO	DAMA	KHOISAN	CAUCASOID	"COLOURED"
NGUNI		64	85	59	221	170	68	125	138
SOT/TSW			77	60	219	185	91	114	104
LEMBA				68	21	142	114	138	139
VENDA					137	162	71	129	137
HERERO						227	149	219	230
DAMA							190	222	233
KHOISAN								125	167
CAUCASOID									170
"COLOURED"									

TABLE 3 . 42 Genetic distance measurements (x100) for the four major southern African populations using allele frequencies at five hypervariable loci.

	NEGROID	KHOISAN	CAUCASOID	"COLOURED"
NEGROID		76	142	104
KHOISAN			203	90
CAUCASOID				259
"COLOURED"				

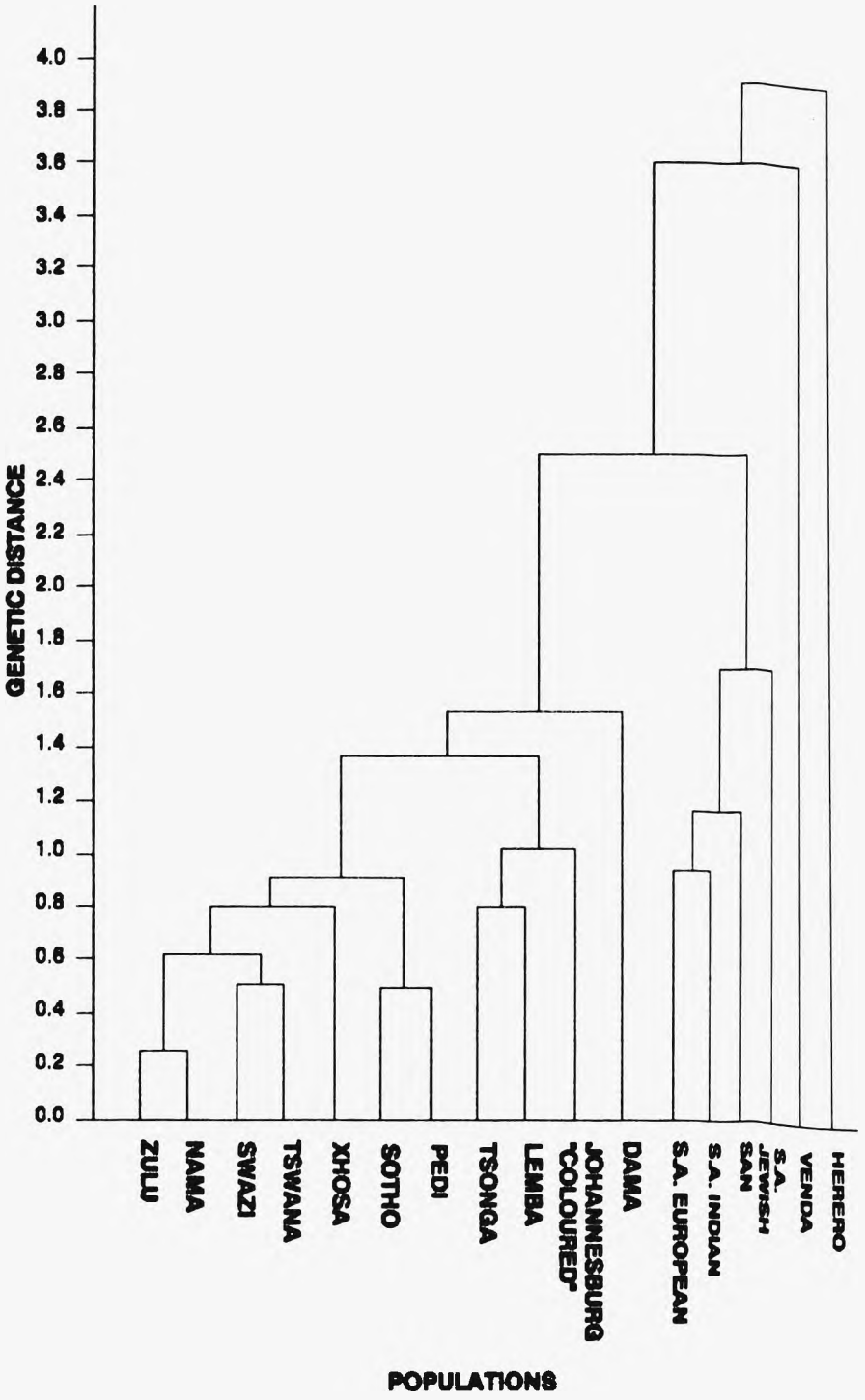


FIGURE 3.21 Dendrogram based on allele frequencies at the five hypervariable loci considered together in 17 southern African populations.

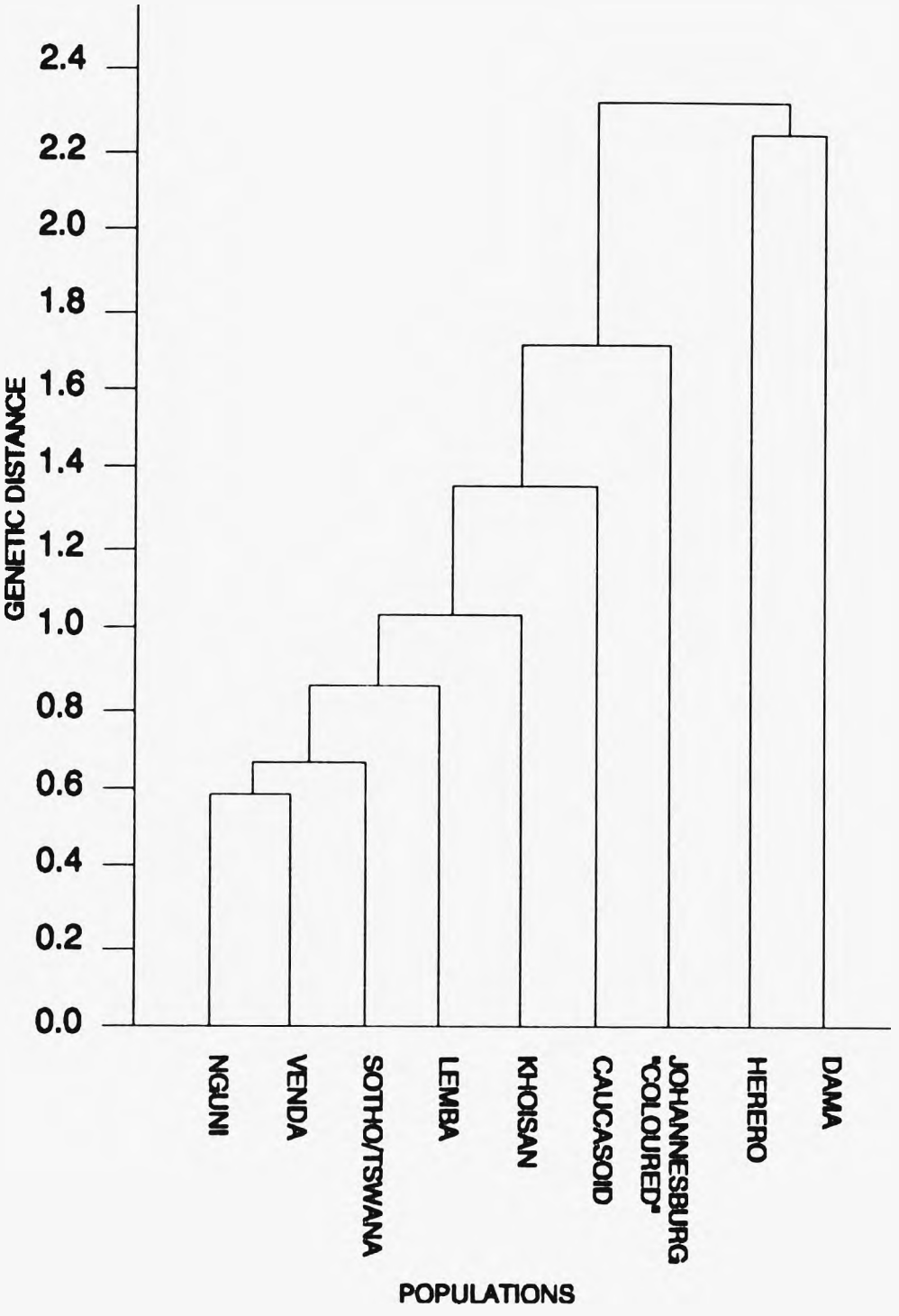


FIGURE 3.22 Dendrogram based on allele frequencies at the five hypervariable loci considered together in 9 main linguistic groups of southern Africa

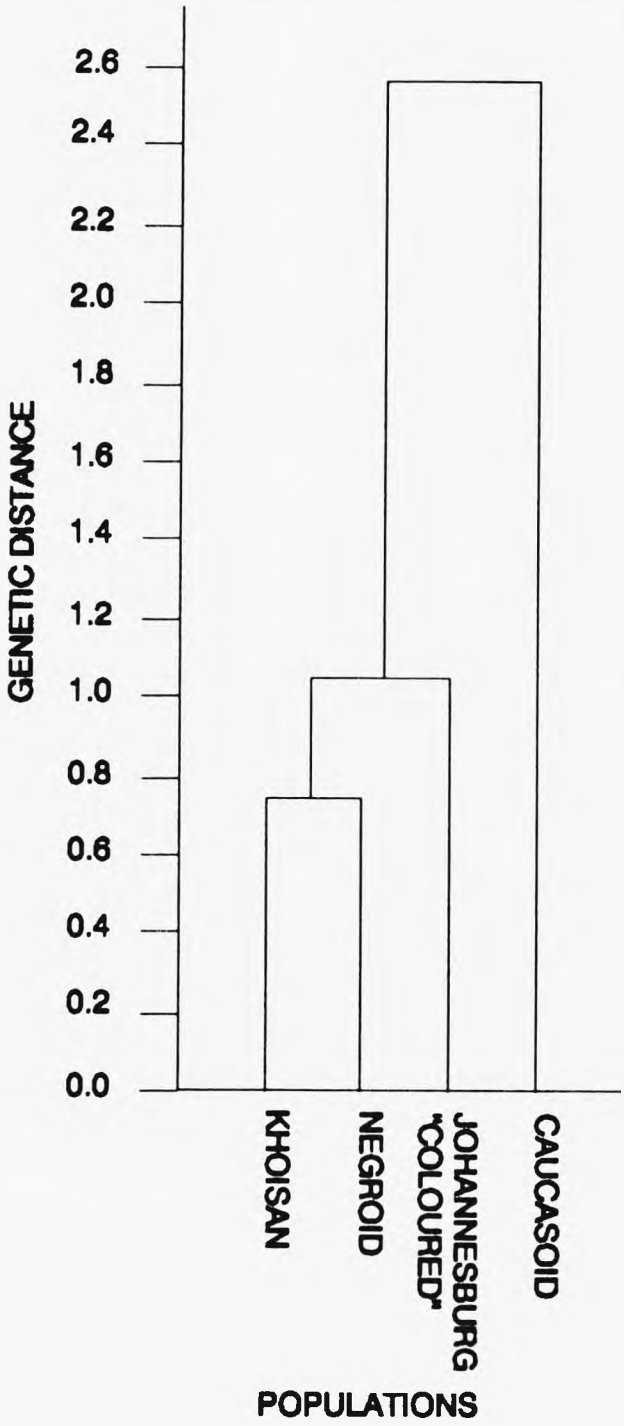


FIGURE 3.23 Dendrogram based on allele frequencies at the five hypervariable loci considered together in the four major population groups of southern Africa.

3.3 APLICATION OF REPEAT LOCI TO PATERNITY DISPUTES

Forty six cases of disputed paternity were tested using the three DNA loci included in this study, APOA2, D21S168 and APOB, that were investigated by the PCR technique. The results obtained from the DNA studies, were compared with those obtained using the HLA system, blood groups and protein polymorphisms (13 loci in total). In 23 cases the probabilities of inclusion calculated for DNA markers were similar to those using the HLA and serogenetic markers. In 13 cases the accused men were excluded from being the biological father of the child on both DNA and serogenetic evidence. In 6 cases the men were excluded by the DNA studies but not by the HLA and serogenetic markers; in two cases the men were excluded by HLA not by the DNA studies. The findings on these 8 cases were similar to those observed in a published study of 102 paternity cases using the D14S1 and HRAS-1 loci (Baird *et al,* 1986).

Two of the paternity dispute cases were studied only at the DNA level because in both cases one of the family members was deceased at the time of the test.

Results obtained on the 23 cases in which the accused man could not be excluded, were used to calculate paternity indices (PI's) and combined paternity indices (CPI's) for DNA, HLA, red cell enzyme and serum protein

polymorphisms (Table 3.43). With the three DNA loci that were used in this study, 9 of the 23 cases, showed the probability of paternity for the DNA markers to be higher than the HLA and serogenetic markers. These results indicate that the use of these three DNA markers alone is not better than the 13 markers used previously. They should therefore not be used instead of the traditional serogenetic and HLA paternity markers, because the overall information content of the new system should equal or be better than the currently used method.

The combined (DNA + Serum) results were used to calculate probability of paternity. There is no doubt that the three DNA loci boosted the probabilities of paternity to ⩾99% in the majority of the cases (17 out of 23 cases) (Table 3.43).

To assess the practical value of the markers at the APOA2, D21S168 and APOB loci in resolving paternity disputes, determination of the so-called average power of exclusion (PEX) of falsely accused man from paternity was calculated according to the formula developed by Garber and Morris (1983). The results of such analysis, the number of individuals used to determine such values and cumulative value of PEX in the four major population groups, are shown in Table 3.44.

TABLE 3.43 Paternity indices (PI's) and combined paternity indices (CPI) for DNA, HLA and serum protein polymorphisms in 23 paternity cases

CASE NO	PI			DNA		SERO AND HLA		CUMULATIVE PROB PAT %
	APOA2	D2IS168	APOB	CPI	PROB PAT %	CPI	PROB OF PAT %	
1		4.20	7.93	33.33	97.0	9.00	90.0	99.00
2	1.19	20.80	4.29	105.95	99.1	383.60	99.7	99.99
3	3.85	2.50	1.89	18.13	94.8	5.90	85.7	99.02
4	1.25	3.38		4.22	80.9	61.50	98.4	96.60
5	1.58	2.50	3.90	15.41	93.9	499.00	99.8	99.98
6		1.42	1.89	2.68	72.9	11.40	91.9	96.87
7	1.19	2.38	4.90	13.88	93.3	20.27	95.3	99.80
8	-	-	4.60	4.60	82.1	11.99	92.3	98.30
9	1.66	2.50	1.89	7.84	88.7	5.07	83.5	97.70
10	3.84	1.40	8.90	47.80	98.0	6.04	85.8	99.86
11	35.70	2.14	55.50	4240	99.9	32.56	97.0	99.99
12	1.58	2.51	15.62	61.95	98.4	11.05	91.7	99.92
13	1.25	2.22		2.78	73.5	51.08	98.1	99.30
14	1.62	2.22	8.47	40.46	96.8	51.08	98.1	99.94
15	3.84	1.42	1.60	8.72	89.8	3.93	79.7	97.30
16	2.46	2.14	4.60	24.20	96.1	36.03	97.3	99.89
17	1.58	1.42	1.62	3.63	78.5	24.00	96.0	98.86
18	3.80	4.20	8.20	130.71	99.3	11.30	91.9	99.94
19	5.88	2.14	6.66	83.80	98.8	2.98	74.9	99.60
20	5.00	2.22	2.53	28.08	96.6	42.48	97.7	99.92
21	3.84	2.50	8.90	85.44	98.8	453.50	99.8	99.83
22	2.46	2.38	4.20	24.77	96.1	393.60	99.7	99.99
23	4.00	4.20	1.60	26.88	96.4	165.70	99.4	99.98

TABLE 3.44 Average power of exclusion (PEX) and cumulative power of exclusion (CPEX) at three hypervariable loci in the four major population groups of southern Africa

POPULATION GROUP	LOCUS	NUMBER OF INDIVIDUALS	PEX	CPEX
NEGROID	APOB	381	.779	
	APOA2	402	.554	97
	D21S168	482	.655	
KHOISAN	APOB	81	.666	
	APOA2	62	.647	96
	D21S168	91	.667	
CAUCASOID	APOB	189	.666	
	APOA2	196	.571	94
	D21	217	.579	
"COLOURED"	APOB	76	.650	
	APOA2	60	.503	94
	D21S168	71	.680	

The results (cumulative PEX values) indicate that the combination of these three loci, will exclude >94 % of falsely accused individuals while empirical values for the 13 HLA and serogenetic markers would exclude 99.05 of wrongly accused men (data obtained in the Department).

In the two paternity disputes that were resolved using DNA markers only, (Case 1 and Case 2 in section 2.1.5), it was possible to exclude one of the men in case two and calculate the probability of paternity for the alleged fathers in case 1 and case 2. In the resolution of these two paternity disputes, other DNA systems besides the ones used in this study were included (Tables 3.45 and 3.46). The allele frequencies used to calculate the PI values in these two cases were obtained from the present study or else from other studies performed in the Department (Denter, 1990; Morris *et al*, 1991; Spurdle, 1992), with the exception of the α- and β- globin frequencies which were obtained from Allen *et al* (1989) and Antonarakis *et al* (1985) respectively.

In case 1 the putative father was not excluded using ten DNA systems from seven loci. The five β- globin RFLPs were used to construct a haplotype which was used to calculate the PI. A CPI of 13215.16, was calculated from seven PI values (Table 3.45). This gave a probability of 99.99% or odds of 9999:1 that the putative father is in

TABLE 3.45 Fragment sizes detected by Southern blotting (S) or PCR (R) at 10 loci, and PI values for case 1

PROBE/ LOCUS	MOTHER	CHILD	PUTATIVE FATHER	PI
pε1.3/ *Hind*III (S)	8.0/3.0	8.0**/3.0*	8.0/8.0	-
Gγ*Hind*III (S)	7.8/7.1	7.8*/7.1**	7.1/7.1	-
Aγ/*Hind*III (S)	3.1/2.8	3.1*/3.1**	3.1/2.8	-
pP3.9/ *Hind*III (S)	7.6/6.0	7.6*/3.0**	6.0/3.0	-
β Globin Haplotype	-++-+ +----	-+-++** +----*	-++-+ -+-++	3.85
49a/*Taq*I (S)		A_3 B_1 C_0 D_1 F_1 G_1 H_1 I_1	A_3 B_1 C_0 D_1 F_1 G_1 H_1 I_1	5.88
pYNH24/ *Pst*I (S)	14.9/9.4	19**/9.4*	19/16	3.4
α3'HVR/ *Pvu*II (S)	4.1/2.5	2.5**/2.5*	3.2/2.5	2.1
APOB (R)	.690/.690	.840**/ .690*	.840/.690	7.9
APOA2 (R)	.143/.131	.143*/ .131**	.131/.131	4.0
D21S168(R)	.112/.108	.112*/ .110**	.112/.110	2.59

* Allele inherited from the mother

** Allele inherited from the biological father

TABLE 3.46 Fragment sizes detected by PCR at nine loci, and PI values for case 2

LOCUS/ MARKER	MOTHER	CHILD	HUSBAND	ACCUSED	PI
ε/*Hind*III	158/ 158	158*/ 72+86	158/ 72+86	155/ 72+86	-
Gγ/*Hind*III	697+91/ 697+91	788/ 697+91*	788/ 697+91	788/ 697+91	-
β *Ava*II	315/214+ 101	214+101/ 214+191	214+101/ 214+101	214+101/ 114+101	-
KM19/*Pst*I	950/950	950*/ 650+300**	950/950	950/650+ 300	1.7
J3.11/*Pst*I	230+150/ 230+150	380/ 230+150	330/ 230+150	380/ 230+150	-
Met/*Msp*I	220+194/ 220+194	414/ 220+194	414/ 220+194	414\ 270+194	-
D21S168	108/112	104**/ 108*	112/114	104/116	4.2
APOA2	137/143	131**/ 143*	137/143	131/141	4.0
APOB	600/720	720*/ 840**	690/870	690/840	7.9

* Allele inherited from the mother

** Allele inherited from the biological father

fact, the biological father of the child.

In case 2, five of the nine systems used were uninformative (Table 3.46), as they failed to distinguish between the two unrelated putative fathers. The husband was excluded from being the biological father of the child with the four remaining systems, while the accused man was not. A CPI of 228.2 was calculated for the accused from four PI values. This gave a probability of paternity of 99.6% or odds of 250:1 that the accused is the biological father of the child.

CHAPTER FOUR

4 DISCUSSION

During the process of their formation and settlement in a given territory, human populations and their genes are subjected to the influences of various evolutionary forces, such as migration, admixture with people from other populations, random drift and natural selection. Despite the complicated interactions of these forces in the different ethnic groups populating southern Africa, there is evidence of the effects of single processes, such as directional patterns of gene flow caused by migration.

The main objective of the present study was to provide genotype and allele frequency data on a set of five hypervariable loci, three VNTR loci and two $(CA)_n$ repeat loci, in a range of southern African populations and to investigate the utility of these hypervariable polymorphisms in population studies.

Southern African populations have already been well characterized in terms of ethnohistory and in terms of genetic markers such as blood groups, protein polymorphisms, autosomal RFLPs, Y-chromosome RFLPs and mitochondrial DNA markers and the aim of this study was to determine whether hypervariable markers could

complement such studies.

4.1 LYMPHOBLASTOID CELL LINES

LCLs play a very important role in population genetics studies in the sense that they can provide an unlimited source of DNA from isolated distinct population groups living in remote areas. Experience has shown that in these areas, suitable local laboratory facilities are usually not available and it is not always possible to find the same donor if a repeat sample should be required. Another problem is that "these stable and isolated human populations" are being encroached upon by, on the one hand, social improvements like transportation and communication which bring them into contact with other people, and on the other hand, by famine and war. Therefore it is of utmost importance to transform and immortalize the genetic material from these unique populations before they disappear, in order to use them in studying and understanding our genetic heritage.

4.1.1 IDEAL TRANSPORTATION CONDITIONS FOR TRANSFORMATION

The transformation rate of blood samples was satisfactory as long as a certain set of conditions were met. All samples transformed succesfully if they were processed within two days of collection and when they

had been stored under cool conditions or at room temperature. Thereafter the transformation success rate declined. It is however, puzzling to find that some samples would not transform after three days but would be successfully immortalized on the fourth day. This was observed in the pilot experiment (Table 3.1) using two donors (one Negroid and one Caucasoid) and it is interesting to note that the samples that transformed on the fourth day were the ones kept at room temperature and at 4°C in both subjects. The failure to transform was seen for the samples kept in the cooler bag. One possible reason for this may be that these samples were in direct contact with the ice packs from day 0, and the cold temperature may have adversely altered the viability of the cells. To date, over 700 samples from twelve different southern African populations have been sucessfully transformed . At least ten aliquots of each have been frozen in four separated liquid Nitrogen containers and they form a valuable resource for future population studies.

4.2 SOUTHERN BLOTTING COMPARED WITH PCR FOR THE DETECTION OF HYPERVARIABLE LOCI

The lower cost, greater sensitivity and increased speed of the PCR technique, offer many advantages over Southern blotting. It has the advantage of clear demonstration of alleles, even if they vary by only a

single repeat and this leads to the more accurate determination of allele frequencies and heterozygosity rates. Another great advantage of the PCR technique is that it can be successfully performed on partially degraded DNA, which is impossible with Southern blot analysis.

Amplification of VNTRs with PCR enables more precise allele determination (Boerwinkle *et al*, 1989; Jeffreys *et al*, 1988a; 1990) but the large size of the products makes them less suitable for general application in comparisson to tandem repeats with amplification products of 100 to 500bp. The small size of dimeric, trimeric and tetrameric STRs facilitates their simultaneous study in multiplex PCR, in which two or more loci can be amplified from a single DNA sample. Simultaneous amplification reduces the cost and the time of determining genotypes and internal size standards provide an additional levels of control to unambiguously score the different alleles. The amplified fragments from several markers can be loaded simultaneously on to a gel so that over a hundred individuals can be typed on a single polyacrylamide gel. Faint bands often occur in the amplification product. These bands are probably due to less specific priming of the synthethic, oligonucleotides, to strand slippage during amplification (Luty *et al*, 1990) or to somatic mutations

(Smeets *et al*, 1989). These faint bands however do not present difficulties when typing alleles at any of the loci studied by PCR. On the contrary, they have been useful in that they can be used on overexposed autoradiograms to visualize a "ladder" of the different allele sizes such that even widely separated alleles differing by, say, eight repeats may accurately be scored.

With Southern blotting it is extremely difficult to size alleles accurately between blots. If alleles differ by one to a few repeat sequences, their resolution is not possible by using conventional agarose-submarine-gel-electrophoresis and Southern blotting (Southern, 1975). Only one band may be apparent because the different bands blur together or coalesce. These heterozygotes are indistinguishable from homozygotes (pseudohomozygosity). This is particularly marked when the overall size of a DNA fragment (or allele) is large and the core repeat sequence is short (Budowle *et al*, 1991a). Not only is it difficult to resolve alleles which differ by a single repeat unit, but the quantity of DNA loaded in each lane affects its relative mobility. Resolution is also related to width of the RFLP band, which with isotopic labels and film-based autoradiographic detection, is difficult to control and will vary because of the quantity of DNA and autoradiographic exposure time. The

quality of the DNA samples is important since bands within a lane that contains degraded DNA will tend to migrate farther than bands in a lane not containing degraded DNA. The size range of the fragments is also important since the smaller the fragment, the more accurately may its size be determined (Gill *et al*, 1990). This favours the choice of allelic systems with small fragment sizes, like the $(CA)_n$ repeats, where closely sized alleles may be more easily distinguished. Larger stretches of dinucleotide repeats show greater polymorphism, but these are still well within the range of accurate detection of PAGE. Budowle *et al*, (1991a) on evaluating the distribution of allelic data from VNTR loci, advocated that a visual evaluation is a necessary first step in their matching criteria method.

When scoring VNTR systems by Southern blotting, where band mobility shifts often occur, it may be most accurate to use the fixed bin approach where fragments of similar size are binned. Arbitrarily defined bins are created to accomodate continuous fragment size data so that alleles can be classified into groups. Since one is "reducing" the number of alleles, the number of samples required to characterize a population may be reduced. Binning of alleles, however, leads to a loss of genetic information in the study groups.

4.3 POPULATION STUDIES

Blood samples from the various population groups were collected in rural and urban areas. Of those collected in the rural areas, the chiefdom affinities were probably more definite than those samples collected in the metropolitan city of Johannesburg. In the latter situation, chiefdom affinities claimed by the donor were accepted but may not necessarily represent the chiefdoms affiliation of relatives from generations past, since marriages between members of different chiefdoms although not common may have occurred. While some individuals may not be accurately classified, or be mixtures of two different language groups, the number of such individuals should comprise only a small fraction of those studied. Broadly defined population groups do not satisfy the premises of the definition of an ideal, non-mixed randomly mating population, but extensive studies of blood groups and isozymes showed that, for statistical computation of genetic probabilities, the binomial (square) law provides a reasonable approximation (Mourant *et al,* 1976). Edwards *et al,* (1992) on studying genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups concluded that the square law may be used to calculate genotype frequencies. Therefore, one can be confident that the allele frequencies derived

from the populations in this study are reliable.

4.3.1 BASIC ASSUMPTIONS FOR THE STUDY OF POPULATION AFFINITIES USING ALLELE FREQUENCIES

When using markers like proteins, blood groups or even single base substitution RFLPs, the polymorphisms are usually limited to two or a few alleles and these show a relatively low mutation rate and have a low heterozygosity rate or polymorphism information content. It is assumed that when the same allele is present in several populations, it has a single origin and therefore came from the same original mutation. It is also often assumed that if an allele attains high frequencies in a number of populations, it is the "oldest" allele and that it arose in the ancestral gene pool, prior to the population split. These assumptions, however, are less likely to hold true for hypervariable polymorphisms, because of their relatively high mutation rate as a result of unequal cross over and slippage mechanisms (Wolff *et al,* 1988; Scholtterer and Tautz, 1992). Such mechanisms of mutation could lead to the multiple origins of alleles of a given size. There is weak statistical evidence suggesting that the mutation rate in VNTR polymorphisms is proportional to allele length (Jeffreys *et al,* 1988b). As the mutation rate increases because the number of alleles at a locus increases, so the heterozygosity at that locus also

increases (Jeffreys *et al*, 1988b).

Another premise which is often necessary for the accurate functioning of algorithms investigating population affinities, is that mutations are cumulative and that back mutations are extremely rare or non-existent. Should back mutation occur at a relatively high rate, as may be the case with VNTRs, then the resulting population affinity calculations would be inaccurate.

4.3.2 ALLELE FREQUENCIES AMONG POPULATIONS

The distribution profile of the alleles for the five hypervariable loci is very variable among the different groups. This is presumably due to unequal recombination resulting from mismatching of the repeat units or from replication slippage. Theoretically, these mechanisms could produce an infinitely large number of alleles. However, structural effects may limit the production of very large or very small alleles, and mutation to previously existing allele sizes probably occasionally occurs. The frequency distribution of the number of repeat units in a population and the maintenace of this distribution, are affected by the rate at which new alleles are generated and the rate at which random genetic drift and natural selection remove alleles from the population. The D4S125 locus is the only one showing

a unimodal distribution in all four major population groups and it is also one of the loci showing a small number of alleles (Fig 3.8). A unimodal distribution is also observed at the D21S168 locus for the Negroid and "Coloured" groups (Fig 3.14). The other loci present with more irregular and variable allele distributions. It appears that the greater the number of alleles, the larger the study population needs to be, in order to get an accurate estimate of allele frequencies. The smallest population group studied was the "Coloured" population (56 to 152 chromosomes) and the largest the Negroid population (506 to 762 chromosomes).

In the case of the D2S44 locus (Fig 3.2), the larger sample sizes (NG, KS and CC) tend to have a more even allele distribution, whereas the CL group (56 chromosomes) has high allele frequencies for alleles 5 and 6. Alleles 5 and 7 are elevated in both the KS and CC groups. A bimodal distribution at this locus was also observed by Gasparini *et al*, (1990), on 102 Italian subjects and by Balaz *et al*, (1989) in 274 American Blacks, 476 Caucasoids and 147 Hispanics.

The unimodal D4S125 allele distribution does show some population differences (Fig 3.8). There is a shift towards smaller allele sizes in the CC and CL groups. It is interesting to note that the two populations of African origin (NG and KS) show very similar allele

distributions whereas the CL group shifts toward the CC group.

The 3'APOB VNTR locus shows a complex multimodal allele distribution in each of the four major groups (Fig 3.10), though alleles *7* to *10* are the most frequent in each. Boerwinkle *et al,* (1989), in studying 125 unrelated individuals observed a bimodal distribution at the 3'APOB VNTR locus. Characteristics of the mispairing and unequal recombination of the VNTRs and the non-identity of the individual repeat units (Boerwinkle *et al*, 1989) are likely to be contributing to these unusual frequency distributions. The arrangement of the repeat units may also be different for the two chromosomes and may affect the process of equal and unequal recombination.

A complex allele distribution pattern is present at the APOA2 locus (Fig 3.12). Although it has only ten alleles, it also shows a complex multimodal allele distribution in each of the four major groups, with two alleles, *8* and 5, being the commonest in all four major groups.

Finally, at the D21S168 locus the intermediate sized alleles predominate though the allele distribution within each group is variable (Fig 3.14). The unimodal allelic distribution in both the Negroid and "Coloured"

groups is very similar as is the bimodal distribution in the Caucasoid and Khoisan groups. Alleles *5*, *6*, and *7* are well represented in all four groups.

Despite the high level of allelic diversity observed between the four major groups, it is striking that the same alleles occur at the highest frequencies in all four populations. This could be the result of different mechanisms: (1) these alleles are substantially older in comparison with others and that they may have been in existence before the geographical dispersal of humans (Deka *et al*, 1991); (2) the common alleles may be the "default" alleles, i.e. they are the ones that are more stable and therefore it will be the other alleles which tend to mutate towards the more stable allele sizes. The latter may not be a valid hypothesis if one accepts a similar mechanism for eukaryotes as the one described for repetitious di- and trinucleotide motifs in prokaryotes (Schlotterer and Tautz, 1992). *In vitro* experiments suggest that slippage occurs within simple sequence stretches and it is estimated that the mutation occurs once in 100 replication events (Levinson and Gutman, 1987b). Mutation rates have not been determined in eukaryotes but it has been suggested that in simple sequence regions in eukaryotes the rate may be similar to that measured in prokatyotes (Schlotterer and Tautz, 1992). In addition, studies in eukaryotes have shown

that the rate of strand slippage for $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences increases when the number of repeat units exceeds about 10 (Weber, 1990). Therefore mutation rates probably differ between microsatellite loci depending on the number of repeats.

4.3.3 HETEROZYGOSITY

Genetic variability is influenced by effective population size, the mutation rate, natural selection and genetic drift. A method for determining the extent of genetic variability was derived by Nei (1978) and is termed genetic diversity or heterozygosity in the case of autosomal markers. The measurement of heterozygosity at different loci may be calculated for different populations. Low values of heterozygosity may be explained by the fact that a population has passed through a recent bottleneck, or that a small population size has presisted for many generations (Flint *et al*, 1989).

VNTR loci usually show high heterozygosity rates and they are thought to have attained such high values because of a high rate of mutation (Jeffreys *et al*, 1988b). If the mutation rate is high at a locus, the rate of approach to mutation-drift equilibrium will also be more rapid (Nei *et al*, 1975).

In comparison with levels observed at protein and blood group loci and even some RFLPs (Nurse *et al*, 1985; Morris *et al*, 1991) heterozygosity values at all five hypervariable loci in all the populations of this study,are very high (Negroid 73.6-89.1%; Khoisan 74.4-88.2%; Caucasoid 62.0-84.2; "Coloured" 58.2-84.8%). The high values found for the Khoisan were rather surprising and in conflict with results obtained from a study using 30 autosomal DNA markers, where the mean heterozygosity was found to be lower in the San than in Caucasoids and Negroid populations (Morris *et al*, 1991). One explanation for this observation was that several of the DNA polymorphisms were originally detected in Caucasoids and by screening a small panel of individuals for polymorphisms, resulting *a priori* in a decreased heterozygosity in other populations (Morris *et al*, 1991). Because of their geographic isolation one expected to find lower heterozygosity values in the San than in the Negroids. In order to explain this finding, one has to combine the suggestion of lack of polygamy in the San (Botha *et al*, 1972) and ancient admixture with Caucasoids (Cavalli-Sforza *et al*, 1988; Cavalli-Sforza, 1989) (explanation in section 4.3.5.1). Polygamy can cause shifts in gene frequency and result in decreased genetic diversity, but this practice is not thought to occur extensively in the San (Botha *et al*, 1972). So the lack of polygamy and possible ancient admixture with

Caucasoids could have resulted in an increased genetic diversity in the San group. High genetic diversity values in the same population groups were also observed by Spurdle, (1992) for some Y-DNA markers. Gasparini *et al*, (1990), used the pYNH24 probe (at the D2S44 locus) to study the frequency distribution of the alleles in 102 Italians, and also obtained a high heterozygosity rate (90%). The method of detection, Southern blotting vs PCR, did not seem to make a significant difference to the level of heterozygosity observed (Tables 3.7; 3.13; 3.19; 3.25; 3.31). The values obtained for non binned alleles were, as expected, higher than those for binned alleles, as were the values for the systems having higher numbers of alleles (D2S44 and APOB).

When the cumulative value for heterozygosity was calculated for the four main population groups for the five hypervariable loci the values were very high in all cases, 99.94-99.98% when binned alleles were considered, and 99.96-99.99% when all alleles were considered. This suggests that cumulative heterozygosities are not useful in distinguishing between populations.

4.3.4 HETEROZYGOTE DEFICIENCY

At the two VNTR loci studied by the method of Southern a number of populations showed heterozygote deficiency. This was probably due to errors in allele sizing since variation in fragment size of alleles scored as being identical would produce "pseudohomozygosity" (Devlin *et al*, 1990) resulting in apparent heterozygote deficiency. In the populations where we observed significant heterozygote deficiency, the number of alleles detected was not significantly lower and would not have resulted in nondetectability of hidden variation. Of course, a definitive answer of the extent of hidden variation can only come from direct sequence analysis.

The third VNTR locus, APOB, was analyzed by the PCR method and the products were visualised directly on ethidium bromide stained agarose gels. This allowed better resolution and more accurate scoring of the alleles. The San is the only group showing heterozygote deficiency at this locus. This is unlikely to be due to sample error since (88 chromosomes were scored). Other possible explanations are inbreeding and genetic drift since the secluded way of life of the San with a limited breeding population may have caused the homozygote excess. The Hardy-Weinberg equilibrium (H-WE) shown in the other populations at this locus is in agreement with

that observed by Budowle *et al*, (1991b); Chakraborty *et al*, (1991); Edwards *et al*, (1992). These authors showed that VNTR genotype distributions scored by PCR, do not exhibit any significant deviation from H-WE.

The two $(CA)_n$ repeat loci in this study were also analysed by the PCR method and the products were separated or polyacrylamide gel electrophoresis (PAGE). Allele sizes were accurately determined from autoradiographs. At the APOA2 locus there was complete agreement between the observed and expected numbers of heterozygotes in all populations. At the D21S168 locus, as for APOB, heterozygote deficiency was seen in the San, but also in the Lemba and the Jews. The latter may have significance since the Lemba claim to be one of the lost tribes of Israel (Van Warmelo, 1974). Both groups may be culturally isolated which may explain the excess of homozygotes.

4.3.5 POPULATION AFFINITIES BY CLUSTER ANALYSIS

Population affinities can be inferred from cluster analysis, generated from genetic distances which are expressed as a function of gene frequencies. The genetic distance between two groups should be small if their gene frequencies are similar. There are a number of reasons why two groups may show similar gene

frequencies: (1) they may share a recent common ancestor, (2) they may exchange genes, (3) they are large, so that little drift has occurred since their separation, and (4) their loci are subject to similar selection pressures, or have been in the past.

When a dendrogram is constructed from genetic distance estimates, the reliability of the topology of the dendrogram depends on the differences in genetic distance among different pairs. If these differences are small, the genetic distances must be estimated accurately by examining many individuals for each locus. Another factor that affects a dendrogram is the level of heterozygosity. As discussed by Nei, (1978), the standard error of genetic distance is large when average heterozygosity is high. Thus, when the average heterozygosity is higher than 0.1, which is the case in hypervariable loci, a relatively large number of individuals should be examined in order to construct a reliable dendrogram.

4.3.5.1 THREE VNTR LOCI

Cluster analysis of the 17 southern African populations at the three VNTR loci (Fig 3.15) revealed an unexpected cluster, since populations with similar linguistic affinities did not cluster together. This phenomenon may not necessarily come as a surprise if it is borne in

mind that languages evolve more rapidly than genes and can undergo rapid replacement, even by an invading minority (Cavalli-Sforza *et al*, 1988). Thus the present language classification of the groups may not reflect their genetic past. On the other hand, it can be argued that the VNTR systems used to study these population affinities, are over sensitive because of their high forward and backward mutation rate, and thus might not allow for the distinction between small population groups or even races.

In this cluster (Fig 3.15) the South African European population is relatively distant from all the other groups including the other Caucasoids, a phenomenon also observed by Spurdle and Jenkins, (1992) and possibly explained by the heterogeneous origins of this population.

The separation of the two Khoisan populations (Nama and San) in two different clusters, is unexpected as is the sharing of a cluster by the San and a Caucasoid group. The Nama and the San have previously been shown to be genetically similar and the San are believed to be free from Caucasoid admixture because of their long isolation (Ramsay and Jenkins, 1988). However, Cavalli-Sforza and coworkers have suggested that admixture with Caucasoids may have occurred 10 000 or more years ago when the Khoisan were widely distributed throughout Africa

(Cavalli-Sforza *et al*, 1988; Cavalli-Sforza, 1989).

The presence of the "Coloured" group deep within the Negroid cluster and distant from the Caucasoid was surprising, making it difficult to justify a Caucasoid contribution to the "Coloured" gene pool. Serogenetic data (Nurse *et al*, 1985), Y-DNA markers (Spurdle and Jenkins, 1992) and mtDNA data (Soodyall, personal communication) support the historical record that the "Coloured" people largely resulted from unions between Khoisan and Negroid females and Caucasoid males.

The cluster analysis of the sub-population groups comprised of smaller sample sizes (Fig 3.15) proved to be sensitive to small changes in gene frequencies, indicating that the data are not very robust. To determine whether larger sample sizes would stabilize the cluster observed, populations were also analyzed as larger groups, classified according to major linguistic affiliations . This analysis (Fig 3.16) showed a much more logical clustering of populations as seen from cluster analysis of smaller groups (Fig 3.15). The Caucasoid exhibit the greatest genetic divergence. All the Negroid groups cluster separately from the Khoisan and Caucasoids. The Western Bantu-speaking group, Herero and the Dama Khoisan-speaking Negroid, unexpectedly share a cluster with the Eastern-Bantu speaking Nguni group which could suggest that the common alleles in

these systems were shared by all the Bantu-speaking Negroids before their Western/Eastern split.

The exact origin of the Dama remains enigmatic, but one favoured theory is that historically the ancestors of the Dama were enslaved by the Khoi, and modern Dama people represent the descendents of runaway slaves who escaped into the desert (Nurse *et al*, 1985). It might be possible that the Dama resulted from hybridization between Western and Eastern Bantu-speaking groups sometime after the Western/Eastern split. Using the Y-specific 49a/*Taq*I RFLP, Spurdle and Jenkins (1992) showed that the Dama cluster with Eastern Bantu-speaking groups. Spurdle (1992) postulated therefore that the male contribution to the Dama were Eastern Bantu-speakers who were enslaved by the Nama in the Eastern Central African region. On the other hand, Soodyall (personal communication), using mtDNA markers, found the Dama clustered with the Western Bantu-speaking Herero population. The results in the present study supports these apparently conflicting results if the following scenario is considered. After the Dama men were enslaved as cattle-herders by the Nama, there was a South eastern movement of these people. It is possible that when they reached the Kalahari region some of the Dama men escaped into the desert. During this movement, they may have taken western Bantu-speaking wives and established

themselves in the region. This would explain the clustering of the Dama with both eastern and western Bantu speaking Negroids when sensitive autosomal marker systems are used. Even if an alternative hypothesis is favoured for the origin of the Khoi, and they are considered to have evolved in the Kalahari from San people (Elphick, 1977), it is not impossible that they enslaved Bantu-speakers of the Eastern-Bantu culture, since the ancestors of the Eastern-Bantu Sotho/Tswana are believed to have stretched much further to the West then they do now (Nurse *et al*, 1985).

Finally, when all samples were classified into only the four major groups, the analysis produced rather perplexing results (Fig 3.17). The "Coloureds" diverged the most from the Khoisan, Negroids and Caucasoids in accordance with the fact that a rich and variable genetic input into this group has come from the Caucasoid, Negroid and Khoisan gene pools. Surprisingly the Khoisan and Caucasoid showed greates affinity with each other. The Khoisan have always been a very distinct group (Nurse et al, 1985) although they have a general African genetic profile (Tobias, 1964; Cavalli-Sforza, 1986). However recent evidence from Cavalli-Sforza (1989) and Spurdle and Jenkins (1992) have been shown between Caucasoid and Khoisan population groups.

4.3.5.2 TWO $(CA)_n$ REPEAT LOCI

As for the three VNTR loci, frequency data for the two $(CA)_n$ repeat loci were analysed in three different ways. The cluster analysis of the frequency data for the 17 different southern African populations revealed unexpected clustering patterns (Fig 3.18). A possible explanation for the data include the rapid evolution of languages and the possibility of slippage within simple sequence stretches like $(CA)_n$ repeats. The latter results in high foward and backward mutation rates.

As observed for similar analysis using data for the three VNTR loci, the Khoisan were separated in two different clusters and clustering was sensitive to small frequency changes.

The cluster analysis for the major linguistic affiliations (Fig 3.19) showed a much more logical clustering of the populations. The Caucasoids exhibit the greatest genetic diversity and the "Coloured" are positioned between the Caucasoid and Negroid. All the Negroid groups with the exception of the Lemba, cluster together away from the Khoisan and the Caucasoids. The position of the Lemba tallies with data from some Y-specific genetic markers (Spurdle, 1992) which indicate that the Lemba are distinctly associated with Caucasoids. Studies using mtDNA markers (Soodyall, 1992)

and autosomal serogenetic markers (Dunn, personal communication) suggested that the Lemba are Negroid in origin but the combined data consolidate the theory that the Lemba population derived from from Arab traders who settled in the area and took Negroid wives (Van Warmelo, 1974).

As for the VNTR systems, the Western- and Eastern-Bantu cannot be distinguished and so the ancestry of the Dama is unclear (see section 4.3.5.1.).

Finally when all the samples were amalgamated into the four major groups (Fig 3.20), the cluster obtained was exactly the same as that obtained for the three VNTR loci. (See section 4.3.5.1. for explanation).

4.5.5.3 FIVE HYPERVARIABLE LOCI

The two different classes of repeats, VNTR and $(CA)_n$ repeats were analysed together with three different population groupings. The analysis of the 17 southern African populations again revealed a weak cluster (Fig. 3.21). In keeping with findings using VNTR (section 4.3.5.1.) and $(CA)_n$ loci, the Caucasoid groups cluster together with the San. The two Khoisan groups (Khoi and San) are separated into different clusters, this consistent pattern may be an indication that the San do indeed share some genetic material with the Caucasoids while the Nama cluster more with the Negroids.

It is known that the Nama took Negroid slaves, particularly Dama. When studying many population groups with systems as polymorphic as the hypervariable loci, it is possible to detect genetic affinities of less significance. The fact that the Venda and the Herero show the greatest divergence from the other populations is perplexing because other studies have not shown them to be so.

The analysis for the nine main linguistic groups (Fig 2.22). resulted in a rather unusual cluster, since the Herero and Dama populations are widely separated from the rest of the Negroid groups which suggest that the Dama may be Western Bantu.

When only the four major groups were analysed (Fig 2.23) the populations clustered according to the linguistic data and hystorical findings. The Caucasoids were the most divergent, and the "Coloureds" occupied an intermediate position between the Caucasoids and the Negroids confirming the contribution from both population groups (see section 4.5.3.1).

The clustering of the Khoisan and the Negroids in this analysis of the combined data (Fig 3.23) strengthens the hypothesis of the common ancestry of these populations which is supported by the analyses using serogenetic and RFLP polymorphisms.

4.4 PATERNITY TESTING

In addition to the evolutionary insights that the hypervariable loci might provide, these loci provide a powerful method for solving paternity and maternity disputes as well as in forensics. Walker and Crisan (1991) defined four generations of paternity testing: the first, blood grouping, which can be expected to exclude 63% of men falsely accused of paternity; the second, serum proteins and red cell enzymes, achieving 85% of exclusions; the third, the HLA system, excluding 93% of men; and finally DNA testing. The latter was initiated in the early 1980's and concentrated mainly on RFLP analysis of single or multiple loci. Hypervariable loci have frequently been used to solve various identification problems. Many cases have been solved by means of DNA fingerprints, and this method is currently considered reliable and highly useful for the identification of individuals (Helminen *et al,* 1991).

During the last five years the isolation and cloning of locus-specific hypervariable DNA areas have produced many single-locus VNTR probes, which have been used increasingly to solve identification problems. These probes and particularly those that can be analysed by PCR, are highly sensitive and can be applied in cases in which the amount and quality, of DNA is a limiting factor if Southern blotting techniques have to be used.

Furthermore, since a maximum of two alleles per person is obtained at each locus, the results are easier to interpret than those obtained by means of multilocus probes. Allele frequencies can also be determined in different specific populations, and this will futher facilitate calculations of PI and CPI.

Justified criticisms of the use of DNA analyses in identification problems have been voiced. Most of this criticism stems from technical artefacts and insufficient intra- and inter-laboratory standards in cases where conclusions based on this analysis are extremely important for the individual (Lander, 1989). As Baird *et al,* (1986) noted, alleles at these hypervariable loci exhibit proper Mendelian segregation obey the Hardy-Weinberg law but their extraordinary allelic diversity and possibly high mutation rates raise some degree of concern.

A number of factors have to be taken into consideration when using Southern blotting and hypervariable loci. These include quality and concentration of the DNA which may affect mobility so that alleles of the same size may appear to be different. On the other hand, alleles that differ by one or very few repeats may not separate well on electrophoresis and heterozygotes may be thought to be homozygotes (Devlin *et al,* 1990; Budowle *et al,* 1991a). Based on these facts, objections to the use of

hypervariable loci and Southern blotting in paternity testing have been raised in courts of law (Green and Lander, 1991). The PCR technique has eliminated a number of these shortcomings. Not only does it allow the precise scoring of alleles but, it can also be used when the quality and quantity of the DNA is poor or extremely small. It is also an inexpensive and rapid technique since two or more loci can be amplified and resolved simultaneously provided the primers are appropriately designed.

Another factor that must be considered when single locus VNTR probes are used is the reported mutation rate, which is much higher than that thought to characterize the simple "classical" RFLPs. The results obtained by using a combination of four or five single locus probes should minimize the possibility of false interpretations due to new, mutated alleles. However, if exclusion is detected with only one of the probes used, the possibility of a new mutation must be kept in mind.

In this experimental trial of the efficacy of VNTRs in paternity testing only three hypervariable loci were analysed. A rate of exclusion of (41%), was observed of 46 cases investigated by DNA analysis which is higher than the 34% exclusion rate observed for 1059 cases analysed using HLA, red cell enzymes and serum protein markers (Dunn *et al*, 1989). Inclusions were observed in

23 out of 46 cases studied by DNA markers and only 9 showed a probability of paternity which was higher than that calculated from the use of standard HLA and serogenetic markers. Although these three hypervariable loci appear to be very useful, their discrimination is no better than the 13 marker loci currently in use (Dunn *et al*, 1989). From these results it can be proposed that at least three other hypervariable loci, scored using the PCR method, be used for a trial period before classical markers are dropped from the laboratory armamentarium. These loci should be well characterized at the population level, so that, a battery of six hypervariable loci could be used to substitute the present HLA/protein marker systems in such a way that the probability of paternity is better than the present system. Since analyses would be done by PCR, conditions could be optimised so that two or more loci are amplified and resolved simultaneously. Paternity testing could then be done accurately, rapidly and inexpensively.

In the present study, DNA markers were able to resolve two difficult paternity cases (Case 1 and Case 2), that could not have been resolved with standard serogenetic markers. High probabilities of paternity were obtained in both cases (99.99% and 99.6% respectively). In Case I, because the child was male, the highly polymorphic

49a *Taq*I system on the Y chromosome was used and this alone provided a PI of 5.88. Markers on the X chromosome have been used in a similar case in which the child was female (Howard *et al*, 1991). In Case 2, one of the men was convincingly excluded with four DNA systems, three of which were multiallelic loci.

Three other cases have been reported in which the putative father was deceased. Since no material was available from them for the testing, paternal grandparents were used in the assessment of paternity (Helminen *et al*, 1991). No exclusions were encountered but this approach could be problematic if an exclusion was encountered as it would be difficult to determine if it was the putative father who was excluded from being the child's father or the paternal grandfather who was excluded from being the father of the putative father.

Although one has to exercise a certain degree of caution when dealing with DNA markers, more particularly hypervariable loci, they do have an indisputable role in individual identification and, together with PCR, they probably constitute the 5th generation of paternity testing.

4.5 HOW USEFUL ARE HYPERVARIABLE LOCI?

Hypervariable loci have not been used extensively in evolutionary studies but, to date, have been mainly used for establishing individual identity. VNTRs are very useful tools for settling parentage dispute issues or answering questions of identification in forensic cases, and they have proved their worth in genetic linkage studies.

Several assumptions which are used in algorithms for the study of population affinity, based on classical markers, may not hold true for hypervariable markers. Their increased mutation rates suggest that the alleles may not have been derived from one another in a chronological sequence, but that multiple alleles of like size have arisen in different populations at different times since the populations diverged. A single allele therefore has probably had many origins and has arisen from different parent alleles. The origin of a given allele may be determined by sequencing the chromosomal background in which it occurs, which may overcome the problem of treating like-sized alleles as having a single origin. This would, however, require an immense amount of work which may be difficult to justify. Standard methods of population analysis may not be suitable for the analysis of hypervariable markers in a population context.

The large number of alleles at each hypervariable locus and the very low frequency of many of them suggest that larger sample sizes are necessary in order to get reasonably accurate allele frequencies. It is apparent that hypervariable loci show distinct allele frequency profiles in different populations, even though there is a tendency for the same alleles to occur at high frequency in different populations. The hypervariable loci may well have a place in the study of population affinities, but the information generated will have to be treated with great care so as not to emphasize insignificant differences. Certain methodological problems have to be solved before they can be used for human evolutionary studies.

A concerted collaborative effort will be required if these hypervariable loci are going to play an important role in our understanding of human origins and evolution.

REFERENCES CITED

AKANE, A., MATSUBARA, K., SHIONO, H., YAMADA, M., NAKAGOME, Y.(1991). Diagnosis of twin zygosity by hypervariable RFLP markers. Am. J. Med. Gen. 41:96-98.

ALDRIDGE, J., KUNKEL, L., BRUNS, G., TANTRAVAHI, U., LALANDE, M., BREWSTER, T., MOREAU, E., WILSON, M., BROMLEY, W., RODERICK, T., LATT, S.A. (1984). A strategy to reveal high-frequency RFLPs along the human X chromosome. Am. J. Hum. Genet. 36:546-564.

ALLEN, R.W., BLISS, B., PEARSON, A. (1989). Characteristics of a DNA probe (pa 3'HVR) when used for paternity testing. Transfusion 29:477-85.

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

BAER, R., BANKIER, A.T., BIGGIN, M.D., DEININGER, P.L., FARRELL, P.J., GIBSON, T.J., HATFULL, G., HUDSON, G.S., SATCHWELL, S.C., SEGUIN, C., TUFFNELL, P.S., BARRELL, B.G. (1984). DNA sequence and expression of the B95-8 Epstein Barr virus genome. Nature 310:207-211.

BAIRD, K.E. (1970). Semantics and Afro-American Liberation. Social Casework 50:265-269.

BAIRD, M., BALAZ, I., GIUSTI, A., MIYAZAKI, L., NICHOLAS, L., WEXLER, K., KANTER, K., GLASSBERG, J., ALLEN, F., RUBINSTEIN, P., SUSSMAN, L. (1986). Allele frequency distribution of two highly polymorphic DNA sequences in three ethnic groups and its application to the determination of paternity. Am. J. Hum. Genet. 39:489-501.

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., BAIRD, M., CLYNE, M., MEADE, E. (1989) Human population genetic studies of five hypervariable DNA loci. Am. J. Hum. Genet. 44:182-190.

BARAN, N., LAPIDOT, A., MANOR, H. (1991). Formation of DNA triplexes accounts for arrests of DNA synthesis at $d(TC)_n$ and $d(GA)_n$ tracts. Proc. Natl. Acad. Sci. USA vol 88:507-511.

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

BELL, G.I., SELBY, M.J., RUTTER, W.J. (1982). The highly polymorphic region near the insulin gene is composed of simply tandemly repeating sequences. Nature 295:31-35.

BERG, K. (1986). DNA polymorphism at the apoliprotein B locus is associated with lipoprotein level. Clin. Genet. 30:515-520.

BIRD, A.G., MCLACHLAN, S.M., BRITTEN, S. (1981). Cyclosporin A promotes spontaneous outgrowth *in vitro* of Epstein-Barr virus-induced B-cell lines. Nature 289: 300-301.

BLACKHART, B.D., LUDWIG, E.M., PIEROTTI, V.R., CAIATI, L., ONASCH, M.A., WALLIS, S.C., POWELL, L., PEASE, R., KNOTT, T.J., CHU, M.L. MAHLEY, R.W., SCOTT, J., MCCARTHY, B.J., LEVY-WILSON, B. (1986). Structure of the human apolipoprotein B gene. J. Biol. Chem. 261:15364-15367.

BODMER, W.F., CAVALLI-SFORZA, L.L. (1976). Genetics, evolution and man. WH Freeman, San Francisco.

BOERWINKLE, E., XIONG, W., FOUREST, E., CHAN, L. (1989). Rapid typing of tandemly repeated hypervariable loci by the polymerase reaction. Applicaton to the apolipoprotein B 3'hypervariable region. Proc. Natl. Acad. Sci. USA 86:212-216.

BOTHA, M.C. (1972). Blood group gene frequencies. An indication of the genetic constitution of population samples in Cape Town. Supplement to S.Afr. Med. J. 46: 1-27.

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

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

BRAATEN, D.C., THOMAS, J.R., LITTLE, R.D., DICKSON, K.R., GOLDBERG, I., SCHLESSINGER, D, CICCODICOLA, A., D'URSO, M. (1988). Locations and contexts of sequences that hybridize to Poly (dG-dT).(dC-dA) in mammalian ribosomal DNAs and two X-linked genes. Nucleic Acids Res. 16(3):865-881.

BRITTEN, R.J., DAVIDSON, E.H. (1985). Hybridization strategy. In:Nucleic Acid Hybridization, Eds Hames BD and Higgins, S.J., Chapter 1:3-15. IRL Press, Oxford.

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

BROWN, T.A. (1986). Cloning vector based on lambda bacteriophage.In: Gene cloning, an introduction. Van Nonstrand Reinhold, UK.

BUDOWLE, B., GIUSTI, A.M., WAYE, J.S., BAECHTEL, F.S., FOURNEY, R.M., ADAMS, D.E., PRESLEY, L.A., DEADMAN, H.A., MONSON, K.L. (1991a). Fixed-bin analysis for statistical evaluation of continuous distributions of allelic data from VNTR Loci, for use in forensic comparisons. Am. J. Hum. Genet. 48:841-855.

BUDOWLE, B., CHAKRABORTY, R., GIUSTI, A.M., EISENBERG, A.J., ALLEN, R.C. (1991b). Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. Am. J. Hum. Genet. 48:137-144.

CAHILL, T.C., BOOKSTEIN, R., AXELROD, F., NGO, K-Y., SCIOSCIA, A.L. (1990). Resolution of DNA linkage discrepancies through analysis of a VNTR locus in a family study of cystic fibrosis. Prenatal Diagnosis 10:795-799.

CANNON, G., HEINHORST, S., WEISSBACH, A. (1985). Quantitative molecular hybridization on nylon membranes. Anal. Biochem. 149:229-237.

CAPON, D.J., CHEN, E.Y., LEVINSON, A.D., SEEBURG, P.H., GOEDDEL, D.V. (1983). Complet nucleotide sequences of the T24 human bladder carcinoma oncogene and its normal homologue. Nature 302:33-37.

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

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

CAVALLI-SFORZA, L.L. (1989). The last 100 000 years of human evolution:the advantage points of genetics and archeology. In HOMINIDAE. Proceedings of the 2nd international congress of human paleontology. Jaca Book.

CAVANEE, W.K., DRYJA, T.P., PHILLIPS, R.A., BENEDICT, W.F., GODBOUT, R., GALLIE, B.L., MURPHREE, A.L., STRONG, L.C., WHITE, R.L. (1983). Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. Nature 305:779-784.

CHAKRABORTY, R., FORNAGE, M., GUEGUEN, R., BOERWINKLE, E. (1991). Population genetics of hypervariable loci: analysis of PCR based VNTR polymorphism within a population. In:Terry B, Gaudenz D., Jeffreys A.J., Wolff R. (Eds) DNA fingerprinting:approaches and applications. Birkhauser, Basel, 127-143.

CHAKRAVARTI, A., PHILLIPS, Y.A.III., 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.

CHANG, J.C., KAN, Y.W. (1980). Beta-thalassemia: a nonsense mutation in man. Proc. Natl. Acad. Sci USA 76:2886-2891.

CHANG, J.C., KAN, Y.W. (1982). A sensitve new prenatal test for sickle cell anemia. N. Engl. J. Med. 307:30-32.

CHENG, H.L., BLATTNER, F.R., FRITZMAURICE, L., MUSHINSKI, J.F., TUCKER, P.W. (1982). Structure of genes for membrane and secreted murine IgD heavy chains. Nature 296:410-415.

CLADARAS, C., HADZOPOULOU-CLADARAS, M., NOLTE, M., ATKINSON, D., ZANNIS, V.I. (1986). The complete sequence and structural analysis of human apolipoprotein B-100: relationship between APOB-100 and APOB-48 forms. EMBO J. 5:3495-3507.

CLARK, A.G. (1987). Neutrality tests of highly polymorphic restriction-fragment-length-polymorphisms. Am. J. Hum. Genet. 41:948-956.

COGGINS, L.W., O'PREY, M. (1989). DNA tertiary structures formed *in vitro* by misaligned hybridization of multiple tandem repeat sequences. Nucleic Acids Res. 17:7417-7426.

COLEMAN, H.C., SILBERBERG, S.J., MACCLAREN, D.C., FLIGNER, C.L., NAKAMURA, Y., O'CONNELL, P., MILNER, E.C.B. (1990). The forensic use of DNA probes to variable number of tandem repeat loci. Proceedings of the IUPAC 3rd International Congress on: Automation and new technology in the clinical laboratory. Ed. by Kiyoshi Okudc. Blackwell Pub.

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

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, Y. (1991). Diagnosis of genetic disease using recombinant DNA. Second edition. Hum. Genet. 87:519-560.

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.

CRABTREE, G.R., KANT, J.A. (1982). Organisation of the rat gamma-fibrogen gene: Alternative mRNA splice patterns produce the gamma A and gamma B (gamma') chains of fibrogen. Cell 31:159-166.

CRAIG, J., FOWLER, S., BURGOYNE, L.A., SCOTT, A.G., HARDING, H.W.J. (1988). Repetitive deoxyribonucleic acid (DNA) and human genome variation:a concise review a relevant to forensic biology. J. Forensic Sci. 33: 1111-1126.

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

CUTTING, G.R., KASCH, L.M., ROSENSTEIN, B.J., ZIELENSKI, J., TSUI, L-C., ANTONARAKIS, S.E., KAZAZIAN, J.H.H. (1990). A cluster of cystic fibrosis mutations in the first nucleoticle binding fold of the cystic fibrosis conductance regulator protein. Nature 346:366-369.

DAVIDSON, E.H., BRITTEN, R.J. (1979). Regulation of gene expression: Possible role of repetitive sequences. Science 204:1052-1059.

DEAN, G. (1963). The Porphyrias. Pitman, London.

DEKA, R., CHAKRABORTY, R., FERRELL, R.E. (1991). A population genetic study of six VNTR loci in three ethnically defined populations. Genomics 11:83-92.

DENHARDT, D.T. (1966). A membrane-filter technique for the detection of complementary DNA. Biochemical and Physical Res. Comm. 23(5):641-646.

DENTER, M. (1990). Haplotype analysis of southern African populations using sex markers loci linked to cystic fibrosis. MSc Thesis, University of the Witwatersrand, Johannesburg.

DEVLIN, B., RISCH, N., ROEDER, K. (1990). No excess of homozygosity at DNA and fingerprint loci. Science 249:1416-1420.

DEVLIN, B., RISCH, N., ROEDER K. (1991). Estimation of allele frequencies for VNTR loci. Am. J. Hum. Genet. 48:662-676.

DONIS-KELLER, H., AND 32 OTHER AUTHORS. (1987). A genetic linkage map of the human genome. Cell 51:319-337.

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

DOVER, G. (1982). Molecular drive: A cohesive mode of species evolution. Nature 299:111-117.

DUNN, D., TESTA, T., MORGAN, C., BARTLEET, S.C., JENKINS, T., BARTLETT, R.G., JOFFE, M. (1989). The usefulness of various polymorphisms in paternity testing. Experience with three southern African populations. SAMJ 76:303-307.

EDWARDS, A., CIUITELLO, A., HAMMOND, H.A., CASKEY, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. 49:746-756.

EDWARDS, A., HAMMOND, H.A., JIN, L., CASKEY, C.T., CHAKRABORTY, R. (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241-253.

EISENSTEIN, B.I. (1990). The polymerase chain reaction. A new method of using molecular genetics for medical diagnosis. N. Engl. J. Med. 322:178-183.

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

EWENS, W.J. (1972). The sampling theory of selectively neutral alleles. Theor. Popul. Biol. 3:87-112.

EWENS, W.J., SPIELMAN, R.S., HARRIS, H. (1981). Estimation of genetic variation of the DNA level from restriction endonuclease data. Proc. Natl. Acad. Sci. USA 78(6):3748-3750.

FEARSON, E.R., VOGELSTEIN, B., FEINSBERG, A.P. (1984). Somatic deletion and duplication of genes on chromosome 11 in Wilm's tumours. Nature 309:176-178.

FEINBERG, A.P., VOGELSTEIN, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Analyt. Biochem 132:6-13.

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

FISHER, R.A. (1930). The distribution of gene ratios for rare mutations. Proc. Roy. Soc. Edinb. 50:204-219.

FLINT, J., BOYCE, A.J., MARTINSON, J.J., CLEGG, J.B. (1989). Population bottlenecks in Polynesia revealed by minisatellites. Hum. Genet. 83:257-263.

FORD, E.B. (1965). Genetic polymorhism. London. Faber & Faber 1-101.

FUCHS, R., BLAKESLEY, R. (1983). Guide to the use of type II restriction endonucleases. Methods in Enzymology. 100:3-38.

GARBER, R.A., MORRIS, J.W. (1983). General equations for the average power of exclusion for genetic systems of n codominant alleles in one-parent and no-parent cases of disputed parentage. In WALKER, R.H. *et al.* (1983). Inclusion probabilities in parentage testing. Virginia: American Association of Blood Banks.

GASPARINI, P., TRABETTI, E., SAVOIA, A., MARIGO, M., PIGNATTI, P.F. (1990). Frequency distribution of the alleles of several variable number of tandem repeat DNA polymorphisms in the Italian population. Hum. Hered. 40:61-68.

GEORGE, M., GUNAWARDANA, A., THREADGILL, D.W., LATHROP, M., OLSAKER, I., MISHRA, A., SARGEANT, L.L., SCHOEBERLEIN, A., STEELE, M.R., TERRY, C., THREADGILL, D.S., ZHAO, X., HOLM, T., FRIES, R., WOMACK, J.E. (1991). Characterization of a set of variable number of tandem repeat markers conserved in Bovidae. Genomics 11:24-32.

GILL, P., JEFFREYS, A.J., WERRET, D.J. (1985). Forensic applications of DNA fingerprints. Nature 318:577-579.

GILL,P., SULLIVAN, K., WERRETT, D.J. (1990). The analysis of hypervariable DNA profiles:problems associated with the objective determination of the probability of a match. Hum. Genet. 85:75-79.

GIUSTI, A., BAIRD, M., PASQUALE, S., BALAZS, I., GLASSBERG, J. (1986). Applicaton of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm. J. Forensic Sci. 31:409-417.

GOODBOURN, S.E.Y., HIGGS, D.R., CLEGG, J.B., WEATHERALL, D.J. (1983). Molecular basis of length polymorphism in the human zetaglobin complex. Proc. Natl. Acad. Sci. USA. 80:5022-5026.

GORDON, W., BOTHA, V. (1969). Lipoid proteinosis in an inbred Namaqualand community. Lancet (i):1032-5.

GORDON, W., BOTHA, V., EDELSTEIN, I. (1971). Lipoid proteinosis. Birth Defects: Original Article Series VII, 8, 164-77.

GREEN, P., LANDER, E.S. (1991). Forensic DNA tests and Hardy-Weinberg equilibrium. Science 253:1038-1039.

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

GROSS, D.S., GARRARD, W.T. (1986). The ubiquitous potential Z-forming sequence of eukaryotes, $(dT\text{-}dG)_n$. $(dC\text{-}dA)_n$ is not detectable in the genomes of eubacteria. Mol. Cell. Biol. 6:3010-3013.

GROSS, D.S., HUANG, S.Y., GARRARD, W.T. (1985). Chromatin structure of the potential Z-forming sequence $(dT\text{-}dG)_n$. $(dC\text{-}dA)_n$. J. Mol. Biol. 183:251-265.

GUO, Z., SHARMA, V., PATTERSON, D., LITT, M. (1990). Dinucleotide repeat polymorphism at the D21S168 locus. Nucleic Acids Res. 18(9):5924.

GUSELLA, J.F., WEXLER, N.S., CONNEALLY, M.P., NAYLOR, S.L., ANDERSON, M.A., TANZI, R.E., WATKINS, P.C., OTTINA, K., WALLACE, M.R., SAKAGUCHI, A.Y., YOUNG, A.B., SHOULSON, I., BONILLA, E., MARTIN, J.B. (1983). A polymorphic DNA marker genetically linked to Hungtington's disease. Nature 206:234-238.

GUTHRIE, M. (1962). Some developments in the pre-history ofthe Bantu-languages. J. Afr. Hist. 3:273-282.

GYLLENSTEN, U.B., ERLICH, H.A. (1988). Generation of single-stranded DNA by polymerase chain reaction and its application to direct sequencing of the HLA-DQ A Locus. Proc. Natl. Acad. Sci USA. 85:7652-6.

HAMADA, H., KAKUNAGA, T. (1982). Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298:396-398.

HAMADA, H., PETRINO, M.G., KAKUNAGA, T. (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA. 79:6465-6469.

HAMADA, H., PETRINO, M.G., KAKUNAGA, T., SEIDMAN, M., STOLLAR, B.D. (1984a). Characterization of genomic poly (dT-dG).poly (dC-dA) sequences, structure, organization, and conformations. Mol. Cell. Biol. 4:2610-2621.

HAMADA, H., SEIDMAN, M., HOWARD, B.H., GORMAN, C.M. (1984b). Enhanced gene expression by the poly (dT-dG).poly(dC-dA) sequence. Mol. Cell. Biol. 4:2622-2630.

HARPENDING, H.C., JENKINS, T. (1973). Genetic distance among southern African populations. In: Crawford, M.H., Workman, P.C., (eds). Methods and Theories of anthropological genetics. University of New Mexico press, Albuquerque. 177-199.

HARRIS, H. (1980). The principles of human biochemical genetics. Third eddition, Elsevier/North Holland, Amesterdam.

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.

HEGELE, R.A., HUANG, L.S., HERBERT, P.N., BLUM, C.B., BURING, J.E., HENNEKENS, C.H., BRESLOW, J.L. (1986). Apolipoprotein B-gene DNA polymorphisms associated with myocardial infarction. N. Engl. J. Med. 315:1509-1515.

HELMINEN, P., JOHNSSON, V., EHNHOLM., C., PELTONEN, L. (1991). Proving paternity of children with deceased fathers. Hum. Genet. 87:657-660.

HELMUTH, R., FILDES, N., BLAKE, E., LUCE, M.C., CHIMERA, J., MADEJ, R., GORODEZKY, C., STONEKING, M., SCHMILL, N., KLITZ, W., HIGUCHI, R. (1990). HLA-DQ2 allele and genotype frequencies in various human populations, determined by using enzymatic amplification and oligonucleotide probes. Am. J. Hum. Genet.47:515-27.

HEYL, T. (1970). Genealogical study of lipoid proteinosis in South Africa. British Journal of Dermatology 83:338-340.

HIGGS, D.R., GOODBOURN, S.E.Y., WAINSCOAT, J.S., CLEGG, J.B., WEATHERALL, D.J. (1981). Highly variable regions of DNA flank the human α-globin genes. Nucleic Acids Res. 9:4213-4224.

HOFKER, M.H., SKRAASTAD, M.I., BERGEN, A.A.B., WAPENAAR, M.C., BAKKER, E., MILLINGTON-WARD, A., VAN OMMEN, G.J.B., PEARSON, P.L. (1986). The X chromosome shows less genetic variation at restriction sites than the autosomes. Am. J. Hum. Genet. 39:438-451.

HOPKINSON, D.A., EDWARDS, Y.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 39:383-411.

HORN, G.T., RICHARDS, B., KLINGER, K.W. (1989). Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction. Nucleic Acids Res. 17:2140-2145.

HOWARD, P.L., COLLINS, C.C., HEINTZ, N.H. (1991). Polymerase chain reaction and allele-specific oligonucleoticles in paternity testing of the deceased. Transfusion 31(5):441-442.

HTUN H., LUND, E., DAHLBERG, J.E. (1984). Human U_1 and RNA genes contain an unusually sensitive nuclease S_1 cleavage site within the conserved 3′ flanking region. Proc. Natl. Acad. Sci USA 81:7288-7292.

HUANG, L.S., BRESLOW, J.L. (1987). A unique AT-rich hypervariable minisatellite 3′ to the apo B gene defines a high information restriction fragment length polymorphism. J. Biol. Chem. 262:8952-8955.

HUFFMAN, T.N. (1982). Archaeology and ethnohistory of the African Iron Age. Ann. Rev. Anthropol 11:133-150.

HUMAN GENE MAPPING 11. (1991). London Conference 1991. Eleven international workshop on human gene mapping. Cytogenet and Cell Genet. 58(1-4).

JARMAN, A.P., NICHOLLS, R.D., WEATHERALL, D.J., CLEGG, J.B., HIGGS, D.R. (1986). Molecular characterisation of a hypervariable region downstream of the human α-globin gene cluster. The EMBO J. 5(8):1857-1863.

JEANG, K.T., HAYWARD, G.S. (1983). A cytomegalovirus DNA sequence containing tracts of tandemly repeated CA dinucleotides hybridizes to highly repetitive dispersed elements in mammalian cell genomes. Mol. Cell. Biol. 3:1389-1402.

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

JEFFREYS, A.J. (1987). Genetic typing and confirmation ofparenthood. Lancet 2:806.

JEFFREYS, A.J., WILSON, V., THEIN, S.L. (1985a). Hypervariable "minisatellite" regions in human DNA. Nature 314:67-73.

JEFFREYS, A.J., WILSON, V., THEIN, S.L. (1985b). Individual specific "fingerprints of human DNA". Nature 316:76-79.

JEFFREYS, A.J. BROOKFIELD, J.F., SEMEONOFF, R. (1985c). Positive identification of an immigration test care using human DNA fingerprints. Nature 317:818-9.

JEFFREYS, A.J., WILSON, G., THEIN, S.L., WEATHERALL, D.J., PONDER, B.A.J. (1986). DNA "Fingerprints" and segregation analysis of multiple markers in human pedigrees. Am. J. Hum. Genet. 39:1I-24.

JEFFREYS, A.J., WILSON, V., NEUMEANN, R., KYTE, J. (1988a). Amplification of human minisatellites by the polymerase chain reaction : Towards DNA fingerprinting of single cells. Nucleic Acids Res. 16:10957-71.

JEFFREYS, A.J., ROYLE, N.J., WILSON, V., WONG, Z. (1988b). Spontaneous mutation rates to new length alleles at tandem repetitive hypervariable loci in human DNA. Nature 332:278-281.

JEFFREYS, A.J., NEUMANN, R., WILSON, V. (1990). Repeat unit sequence variation in minisatellites: A novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60:473-485.

JEFFREYS, M.D.W. (1967). Pre-Columbian maize in southern Africa. Nature 215:695-697.

JENKINS, T., DUNN, D.S. (1981). Haematological genetics in the tropics: Tropical Africa. Clinics in Haematology 10:1029-1050.

JENKINS, T., TOBIAS, P.V. (1977). Nomenclature of population groups in southern Africa. African Studies 36:49-55.

JENKINS, T., STEINBERG, A.G. (1966). Some serum protein polymorphisms in Kalahari Bushmen and Bantu: Gamma globulins, haptoglobins and transferrins. Am. J. Hum. Genet. 18:399-407.

JENKINS T., ZOUTENDYK, A., STEINBERG, A.G. (1970). Gammaglobulin groups (Gm and Inv) of various southern African Populations. Am. J. Physical Anthropology 32:197-218.

JULIER, C., DE GOUYON, B., GEORGES, M., GUENET, J.L., NAKAMURA, J., AVNER, P., LATHROP, G.M. (1990). Minisatellite linkage maps in the mouse by cross-hybridization with human probes containing tandem repeats. Proc. Natl. Acad. Sci. USA 87:4585-4589.

KAISER, L., SEKER G. (1985). Paternity testing 2: Likelihood ratio tests. Am. J. Med. Genet. 20:209-219.

KAN, Y.W., DOZY, A.M. (1978a). Polymorphism of DNA sequences adjacent to human β-globin structural gene: Relationship to sickle cell mutation. Proc. Natl. Acad. Sci. USA 75:5631-5635

KAN, Y.W., DOZY, A.M. (1978b). Antenatal diagnosis of sickle cell anaemic by DNA analysis of amniotic fluid cells. Lancet ii:910-912.

KANTER, E., BAIRD, M., SHALER, R., BALAZS, I. (1986). Analysis of restriction fragment length polymorphisms in deoxyribonucleic acid (DNA) recovered from direct blood stains. J. Forensic Sci. 31:403-408.

KASAI, K., NAKAMURA, Y., WHITE, R. (1990). Amplification of a variable number of tandem repeats (VNTR) locus (pMCT118) by the polymerase chain reaction (PCR) and its application to forensic science. J. Forensic Sci. 35:1196-1200.

KAZAZIAN, H.H. Jr., CHAKRAVARTI, A., ORKIN, S.H., ANTONARAKIS, S.E. (1983). DNA polymorphisms in the human β-globin gene cluster. In: Evolution of Genes and Proteins. Eds. Nci M & Koehn, R.K., Sinauer Associates, Sunderland, Massachussets.

KEMP, D.J., COPPEL, R.L., ANDERS, R.F. (1987). Repetitive proteins and genes of malaria. Ann. Rev. Microbiol. 41:181-208.

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

KIMURA., M. (1983). The neutral theory of molecular evolution. Cambridge University Press. Great Britain.

KIMURA, M. (1989). The neutral theory of molecular evolution and the world view of the neutralists. Genome 31:24-31.

KIMURA, M., CROW, J.F. (1964). The number of alleles that can be maintained in a finite population. Genetics 49:725-738.

KIMURA, M., OHTA, T. (1971). Protein polymorphism as a phase of molecular evolution. Nature 229:467-469.

KIMURA, M., OHTA, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc. Natl. Acad. Sci. USA 75(6):2868-2872.

KNOTT, T.J., WALLIS, S.C., PEASE, R.J., POWELL, L.M., SCOTT, J. (1986). A hypervariable region 3′ to the human apolipoprotein B gene. Nucleic Acids Res. 14:9215-9216.

KONDOLEON, S., VISSING, H., LUO, X., MAGENIS, R., KELLOG, LITT, M. (1987). A Hypervariable RFLP on chromosome 17p13 is defined by an arbitrary single copy probe p144-D. Nucleic Acids. Res. 15: 10605.

KOUFOS, A., HANSEN, M.F., LAMPKIN, B.C., WORKMAN, M.L.,COPELAND, N.G., JENKINS, N.A., CAVANEE, W.K. (1984). Loss of alleles st loci on human chromosome 11 during genesis of Wilm's tumour. Nature 309:170-172.

KWOK, S., HIGUCHI, R. (1989). Avoiding false positives with PCR. Nature 339:237-8.

LALLOZ, M.R., MEVEY, J.H., PATTINSON, J.K., TUDDENHAM, E.G.D. (1991). Haemophilia A diagnosis by analysis of a hypervariable dinucleotide repeat within the factor VIII gene. The Lancet 338:207-211.

LANDER, E.S. (1989). DNA fingerprinting on trial. Nature 339:501-505.

LANDSTEINER, K., LEVINE, P. (1928). On the inheritance of agglutinogens of human blood demonstrable with immune agglutinins. J. Exp. Med. 48:731-749.

LANDSTEINER, K., WIENER, A.S. (1940). An agglutinable factor in human blood recognized by immune sera for rhesus blood. Proc. Soc. Exp. Biol. 43:223.

LANE, A.B., DUNN, D.S., JENKINS. (1985). A genetic profile of South African Ashkenazi Jewish population. S.A. Med. J. 63:935-939.

LATHROP, G.M., O'CONNELL, P., NAKAMURA, Y., LEPPERT, M., CARTWRIGHT, P., LALOUEL, J.M., WHITE, R. (1987). Three genetic linkage groups on chromosome 2. Cytogenet. Cell Genet. 46:644.

LAW, A., WALLIS, S.C., POWELL, L.M., PEASE, R.J., BRUNT, H., PRIESTLEY, L.M., KNOTT, T.J., SCOTT, J., ALTMAN, D.G., RAJPUT, J., MILLER, N.G., *ET AL.* (1986). Common DNA polymorphism within coding sequence of apolipoprotein B gene associated with altered lipid levels. Lancet 1(8493):1301-1303.

LAWN, R.M., FRITSCH, E.F., PARKER, R.C., BLAKE, G., MANIATIS, T. (1978). The isolation and characterization of linked δ- and β- globin genes from a cloned library of human DNA. Cell 15:1157-1174.

LEVINE, P., STETSON, R.E. (1939). An unusual case of intragroup agglutination. J. Am. Med. Assoc. 113:126-127.

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

LEVINSON, G., GUTMAN, G.A. (1987b). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. Nucleic Acids Res. 15:5223-5338.

LEWIS, J.G., WEBER, J.L., PETERSEN, M.B., SLAUNGHENHAUPT, S.A., KWITEK, A., MAY P.E., WARRAN, A.C., *et al.* (1990). Linkage mapping of the highly informative DNA marker D21S156 to human chromosome 21 using a polymorphic GT dinucleotide repeat. Genomics 8:400-402.

LEWONTIN, R.C., HUBBY, J.L. (1966). A molecular approach to the study of genetic heterozygosity in natural populations. II Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595-609.

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

LUDWIG, E.H., FRIEDL, W., McCARTHY, B.J. (1989). High-resolution analysis of a hypervariable region in the human apolipoprotein B Gene. Am. J. Hum. Genet. 45:458-464.

LUTY, J.A., GUO, Z., WILLARD, H.F., LEDBETTER, D.H., LEDBETTER, S., LITT, M. (1990). Five polymorphic microsatellite VNTRs on the human X-chromosome. Am. J. Hum Genet. 46:776-783.

LYNCH., M. (1988). Estimation of relatedness by DNA fingerprinting. Mol. Biol. Evol. 5:584-599.

MACDONALD, M.E., CHENG, S.V., ZIMMER, M., HAINES, J.L., POUSTKA, A., ALLITTO B., SMITH, B., WHALEY, W.L., ROMANO, D.M., JAGADEESH, J., MYERS, R.H., LEHRACH, H., WASHMUTH, J., FRISCHAUF, A.M., GUSELLA, J.F. (1989). Clustering of multiallele DNA markers near the Huntington's disease gene. J. Clin. Invest. 84:1013-1016.

MANDEL, M., HIGA, A. (1970). Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. 53:159:162.

MANIATIS, T., FRITSCH, E.F., SAMBROOK, J. (1982). Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory.

MARAIS, J.S. (1939). The Cape Coloured People. 1652-1937. Johannesburg, 1968. Witwatersrand Univ Press.

MARX, J.L. (1988). Multiplying genes by leaps and bounds (news). Science 240:1408-10.

MATTESON, K.J., OSTRER, H., CHAKRAVARTI, A., BUETOW, K.H., O'BRIEN, W.E., BEAUDET, A.L., PHILLIPS, J.A. (1985). A study of restriction fragment length polymorphisms at the human α1-antitrypsin locus. Hum. Genet. 69:263-267.

McKUSICK, V.A. (1988). Mendelian inheritance in man-catalogs of autosomal dominant, autosomal recessive and X-linked phenotypes 8th edition. The Johns Hopkins University Press.

MIESFELD, R., KRYSTAL, M., ARNHEIM, N. (1981). A member of a new repeated sequence family which is conserved throughout eukaryotic evolution is found between the human δ and β globin genes. Nucl. Acids Res. 22(9): 5931-5947.

MIESLER, M.H. (1983). DNA Polymorphisms. Nature 303:108.

MILLER, C.H., MASON, S.J., GLYDE, D.F., McGINNISS, M.H. (1976). The resistance factor to *Plasmodium vivax* in blacks. The Duffy blood group genotype, *Fy Fy*. New Engl. J. Med. 295:302-304.

MOORE, S.S., SARGEANT, L.L., KING, T.J., MATTICK, J.S., GEORGES, M., HETZEL, D.J.S. (1991). The conservation of dinucleotide microsatellites among mammalion genomes allows the use of hetrologous PCR primer pairs in closely related species. Genomics 10: 654-660.

MORRIS, D.J., HEIM, R.A., VERGA, V., DENTER, M., DUNN, D.S., JENKINS, T. (1991). Study of 30 DNA markers in three southern African populations. Gene Geography 5:1-12.

MOURANT, A.E., KOPEC, A.C., DOMANIEWSKA-SOBAZAK. (1976). The distribution of human blood groups and other polymorphisms. Oxford Univ. Press. New York.

MULLIS, K.B. (1990). The unuusal origin of the polymerase chain reaction. Sci. Am. 262:56-61, 64-65.

MURPHY, G.L., CONNELL, T.D., BARRITT, D.S., KOOMEY, M., CANNON, J.G. (1989). Phase variation of gonococcal protein II: Regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. Cell 56:539-547.

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

MURRAY, J.C., MILLS, K.A., DEMOPULOS, C.M., HORNUNG, S., MOTULSKY, A. (1984). Linkage disequilibrium and evolutionary relationships of DNA variants (restriction enzyme fragment length polymorphisms) at the serum albumin locus. Proc. Natl. Acad. Sci. USA 81:3486-3490.

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

NAKAMURA, Y., LATHROP, P., O'CONNEL, P., LEPPERT, M., BARKER, D., WRIGHT, W., SKOLNICK, M., KONDOLEON, S., LITT, M., LALOUEL, J-M., WHITE, R. (1988a). A mapped set of DNA markers for human chromosome 17. Genomics 2:302-309.

NAKAMURA, Y., CARLSON, M., KRAPCHO, K., KANAMORI, M., WHITE, R. (1988b). New approach for isolation of VNTR markers. Am. J. Hum. Genet. 43:854-859.

NAKAMURA, Y., CULVER, M., O'CONNELL, P.O., LEPPERT, M., LATHROP, G.M., LALOUEL, J-M., WHITE, R. (1988c). Isolation and mapping of a polymorphic DNA sequence (pYNZ32) on chromosome 4p (D4S125). Nucleic Acids. Res. 16(9):4186.

NAYLOR, L.H., CLARK, E.M. (1990). $d(GT)_n.d(CA)_n$ sequences upstream of the rat prolactin gene form Z-DNA and inhibit gene transcription. Nucleic Acids Res. 18:1595-1601.

NEEL, J.V. (1984). A revised estimate of the amount of genetic variation in human proteins. Implications for the distribution of DNA polymorhisms. Am. J. Hum. Genet. 36:1135-1148.

NEI, M. (1975). Molecular population genetics and evolution. North Holland Publishing Company, Amsterdam.

NEI, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics. 89:585-590.

NEWTON, C.R., GRAHAM, A., HEPTINSTALL, L.E., POWELL, S.J., SUMMERS,C., KALSHEKER, N., SMITH,J.C., MARKHAM, A.F. (1989a). Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). Nucleic Acids Res. 17:2503-2416.

NEWTON, C.R., HEPTINSTALL, L.E., SUMMERS, C., *ET AL.* (1989b). Amplification refractory mutation system for prenatal diagnosis and carrier assessment in cystic fibrosis. Lancet ii:1481-1482.

NORDHEIM A., RICH, A. (1983). The sequence $(dC-dA)_n.(dG-dT)_n$ forms left handed Z-DNA in negatively supercoiled plasmids. Proc. Natl. Acad. Sci. USA 80:1821-1825.

NURSE, G.T. (1983). Population movement around the northern Kalahari. African Studies 76:153-163.

NURSE, G.T., WEINER, J.S., JENKINS T. (1985). The peoples of southern Africa and their affinities. Clarendon Press. Oxford.

ODELBERG, S.J., PLAETKE, R., ELDRIDGE, J.R., BALLARD, L., O'CONNELL, P., NAKAMURA, Y., LEPPERT, M., LALOUEL, J-M., WHITE, R. (1989). Characterization of eight VNTR loci by agarose gel electrophoresis. Genomics 5:915-924.

OHNO, S. (1967). Sex chromosomes and sex-linked genes. Springer, Berlin.

OLD, R.W., PRIMROSE, S.B. (1980). Principles of gene manipulation: An introduction to genetic engineering. Oxford. Blackwell Scientific.

OLD, J.M., VARAWALLA, N.Y., WEATHERALL. (1990). Rapid detection and prenatal diagnosis of β-thalassaemia: Studies in Indian and Cypriota populations in the UK. The Lancet 336:834-837.

ONDEK, B., GLOSS, L., HERR, W. (1988). The SV40 enhancer contains two distinct levels of organization. Nature 333:40-45.

ORGEL, L.E., CRICK, F.H.C. (1980). Selfish DNA: The ultimate parasite. Nature 284:604-607.

ORKIN, S.H., LITTLE, P.F.R., KAZAZIAN, H.H., BOEHM, C.D. (1982). Improved detection of the sickle mutation by DNA analysis. N. Engl. J. Med. 307:32-35.

ORKIN, S.H., GOLDMAN, D.S., SALLAN, S.E. (1984). Development of homozygosity for chromosome 11p markers in Wilm's tumour. Nature 309:172-174.

PALACIOS, R. (1981). Cyclosporin A abrogates proliferation of T cells and generation of suppressor and cytotoxic T-cell function induced by Epstein-Barr virus. Immunobiology 160:321-329.

PARDUE, M.L., LOWENHAUPT, K., RICH, A., NORDHEIM, A. (1987). $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ sequences have evolutionarily conserved chromosomal locations in Drosophila with implications for roles in chromosome structure and function. Embo J. 6:1781-1789.

PETERSEN, M.B., ECONOMOU, E.P., SLAUGENHAUPT, S.A., CHAKRAVARTI, A., ANTONARAKIS, S.E. (1990). Linkage analysis of the human HMGIH gene on chromosome 21 using a GT dinucleotide repeat as polymorphic marker. Genomics 7:136-138.

PETERSEN, M.B., SCHINZEL, A.A., BINKERT, F., TRANEBJAERG, L., MIKKELSEN, M., COLLINS, F.A., ECONOMOU, E.P., ANTONARAKIS, S.E. (1991). Use of short sequence repeat DNA polymorphisms after PCR amplifications to detect the parental origin of the additional chromosome 21 in Down syndrome. Am J. Hum. Genet. 48:65-71.

PHILLIPSON, D.W. (1977). The spread of the Bantu language. Scientific Am. 236:106-114.

PROUDFOOT, N.J., GIL, A., MANIATIS, T. (1982). The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. Cell 31:553563.

RAMSAY, M., JENKINS, T. (1988). Alpha-globin gene cluster haplotyoes in the Kalahari San and southern African Bantu-speaking Blacks. Am. J. Hum. Gen. 43:527-533.

RAPACZ, J., HASLER-RAPACZ, J., TAYLOR, K.M., CHECOVICH, W.J., ATTIE, A.D. (1986). Lipoprotein mutations in pigs are associated with elevated plasma cholesterol and atherosclerosis. Science 234: 1573-1577.

RECHSTEINER, M.C. (1991). The Human Genome Project: misguided science policy TIBS 16(12):455-459.

REEDERS, S.T., BREUNING, M.H., DAVIES, K.E., NICHOLLS, R.D., JARMAN, A.P., HIGGS, D.R., PEARSON, P.L., WEATERALL, D.J. (1985). A highly polymorphic DNA marker linked to adult polycystic kidney disease on chrom 16. Nature 317:542-544.

RICHARDS, B., HORN,G.T., MERRILL, J.J., KLINGER, K.W. (1991). Characterization and rapid analysis of the highly polymorphic VNTR locus D4S125 (yNZ32),closely linked to the Huntington disease gene. Genomics 9:235-240.

RIGBY, P.W.J., DICKMANN, M, RHODES, C., BERG P. (1977). Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol Biol 113:237-251.

RISCH, N. (1990). Linkage strategies for genetically complex traits. III The effects of marker polymorphism on analysis of affected relative pairs. Am. J. Hum. Genet. 46:242-253

ROTWEIN, P., POLLACK, K.M., DIDER, D.K., KRIVI, G.G. (1986). Organization and sequence of the human insuline-like growth factor I gene: Alternative RNA processing produces two insulin-like growth factor I precursor peptides. J. Biol. Chem 261:4828-4832.

ROYLE, N.J., CLARKSON, R., WONG, Z., JEFFREYS, A.J. (1988). Preferential localization of hypervariable minisatellites near human telomeres. Cytogenet. Cell Genet. 46:685.

SAIKI, R.K., GELFAND, D.H., STOFFEL, S., SCHARF, S.J., HIGUCHI, R., HORN, G.T., MULLIS, K.B., ERLICH, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

SAIKI, R.K., WALSH, P.S., LEVENSON, C.H., ERLICH, H.A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes. Proc. Natl. Acad. Sci USA 86:6230-4.

SAMBROOK, J., FRITSCH, E.F., MANIATIS, T. (1989). Molecular cloning:a laboratory manual, 2d ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

SCHLOTTERER, C., TAUTZ, D. (1992). Slippage synthesis of simple sequence DNA. Nucleic Acids Res. 20(2):211-215.

SHAH, D.M., ROUSSIS, P., ULM, J., JEANTY, P., BOEHM, F.H. (1990). Cordocentesis for rapid kayrotyping.
Am. J. Obstet. Gynaecol. 162(6):1548-1553.

SHAW, M.W. (1983). Paternity determination: 1921 to 1983 and beyond. JAMA, 250(18):2536-2537.

SHIMATSU, A., ROTWEIN, P. (1987). Mosaic evolution of the insulin-like growth factors. J. Biol. Chem. 262:
7894-7900.

SKOLNICK, M.H., WHITE, R. (1982). Strategies for detecting and characterizing restriction fragment length polymorphisms (RFLPs). Cytogenet. Cell. Genet. 32:58-67.

SLIGHTOM, J.L., BLECHL, A.E., SMITHIES, O. (1980). Human fetal Gγ- and Aγ-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between their duplicated genes. Cell 21:627-638.

SMEETS, H.J.M., BRUNNER, H.G., ROPERS, H.H., WIERINGA, B. (1989). Use of variable simple sequence motifs as genetic markers: Application to study of myotomic dystrophy. Hum. Genet. 83:245-251.

SMITH, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. Science 191:528-535.

SMITH, G.R. (1983). Chi hotspots of generalized recombination. Cell 34(3): 709-710

SMITH, G.R., KUNES, S.M., SCHULTZ, D.W., TAYLOR, A., TRIMAN, K.L. (1981). Structure of chi hotspots of generalized recombination. Cell 24(2):429-436.

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., WALKER, N.F. (1956). Notation for serum protein groups and the genes controlling their inheritance. Nature 178:694-695.

SMITHIES, O., CONNEL, G.E., DIXON, G.M. (1962). Inheritance of haptoglobin subtypes. Am. J. Hum. Genet. 14:14-21

SOODYALL, H. (1992). Mitochondrial-DNA variation in southern African populations. PhD Thesis, University of the Witwatersrand, Johannesburg.

SORIA, L.F., LUDWIG, E.H., CLARKE, H.R.G., VEGA, G.L., GRUNDY, S.M., McCARTHY, B.J. (1989). Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. Proc. Natl. Acad. Sci. USA 86:587-591.

SOUTHERN, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:303-317.

SOUTHERN, E. (1979). Gel electrophoresis of restriction fragments. Methods in enzymology. 68:152-176.

SPANDIDOS, D.A., HOLMES, L. (1987). Transcriptional enhancer activity in the variable tandem repeat DNA sequence downstream of the human Ha-*ras*I. FEBS lett. 218:41-46.

SPURDLE, A. (1992). Y-specific restriction fragment length polymorphisms in southern African populations. PhD Thesis, University of the Witwatersrand, Johannesburg.

SPURDLE, A., JENKINS, T. (1992). Y chromosome probe p49a detects complex *Pvu*II haplotypes and many new *Taq*I haplotypes in southern Africa. Am. J. Hum. Gent. 50:107-125.

STALLINGS, R.L., TORNEY, D.C., HILDEBRAND, C.E., LONGMIRE, J.L., DEAVEN, L.L., JETT, J.H., DOGGETT, N.A., MOYZIS, R.K. (1990). Physical mapping of human chromosomes by repetitive sequence fingerprinting. Proc. Natl. Acad. Sci. USA 87:6218-6222.

STALLINGS, R.L., FORD, A.F., NELSON, D., TORNEY, D.C., HILDEBRAND, C.E., MOYZIS, R.K. (1991). Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. Genomics 10:807-815.

STEINBERG, A.G., JENKINS, T., NURSE, G.T., HARPENDING, H.C. (1975). Gammaglobulin groups of the Khoisan peoples of southern Africa. Evidence for polymorphisms for a $Gm^{1,5,13,14,21}$ haplotype among the San. Am. J. Hum. Genet. 27:528-542.

STEINMETZ, M., STEPHAN, D., LINDAHL, K.F. (1986). Gene organization and recombinational hotspots in the murine major histocompatibility complex. Cell 44:895-904.

STEVENSON, A.C., JOHNSON, H.A., STEWART, M.P.P., GOLDING, D.R. (1966). Congenital malformations: A report of a series of consecutive births in 24 countries. Bulletin of the World Health Organization 34, supplement 1-127.

STREISINGER, G., OWEN, J. (1985). Mechanisms of spontaneous and induced frameshift mutation in bacteriophage T4. Genetics 109:633-659.

SUN, L., PAULSON, K.E., SCHMID, C.W., KADYK, L., LEINWAND, Z. (1984). Non-Alu family interspersed repeats in human DNA and their transcriptional activity. Nucleic Acids Res. 12:2669-2690.

SWANSTROM, R., SHANK, P.R. (1978). X-ray intensifying screens greatly enhance the detection by autoradiography of the radioactive isotopes ^{32}P and ^{125}I. Analyt. Biochem. 86:184-192.

SYKES, B.C. (1983). DNA in heritable disease. Lancet ii:787-788.

TAUTZ, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids. Res. 17:6463-6471.

TAUTZ, D., RENZ, M. (1984). Simple sequences are ubiguitous repetitive components of eukaryotic genomes. Nucleic Acids Res. 12:4127-4138.

TAYLOR, G.R., NOBLE, J.S., HALL, J.L., STEWARD, A.D., MULLER, R.F. (1989). Hypervariable microsatellite for genetic diagnosis. Lancet ii:454.

THEIN, S.L., JEFFREYS, A.J., BLACKLOCK, H.A. (1986). Identification of post-transplant cell population by DNA fingerprinting analysis. Lancet 2:37.

TOBIAS, P.V. (1960). Bushman hunter-gatherers: A study in human ecology. In: Davies D.H.S. (ed). Ecological studies in Southern Africa. Junk: The Hague.

TOBIAS, P.V. (1964). Bushman hunter-gathers: A study in human ecology. In: Davies D.H.S. (ed). Ecological studies in Southern Africa. Junk: The Hague.

TOBIAS, P.V. (1971). The biological invalidity of the term Bantu. S. A. J. Sci. 67:517-520.

TOBIAS, P.V. (1978). *Early Hominoids of Africa*. Jolly C.J. London Duckworth.

TOSATO, G., PIKE, S.E., KOST, I.R., BLAESE, R.M. (1982). Selective inhibition of immunoregulatory cell fucntions by Cyclosporin A. J. Immunol. 128:1986-1991.

TYLER-SMITH, C., BROWN, W.R.A. (1987). Structure of the major block of alphoid satellite DNA on the human Y chromosome. J. Mol. Biol. 195:457-470.

ULLRICH, A., DULL, T.J., GRAY, A., PHILIPS, J.A., PETER, S. (1982). Variation in the sequence and modification state of the human insulin gene flanking regions. Nucleic Acids Res. 10:2225-2240.

ULM, J.E., SHAH, D.M., DEV, V.G., PHILLIPS J.A.III. (1990). Counselling and decision dilemmas associated with fetal blood sampling. Am. J. Med. Genet. 35:75-78.

VAN EEDE, P.H., CUYPERS, T.M., DE LANGE, G.G. (1990). Attempt to calculate the allele frequency distribution of a *Taq*I RFLP detected by the highly polymorphic probe YNH24. Hum. Genet. 84:376-378.

VAN WARMELO, N.J., (1974). The classification of cultural groups In: Hammond-Taske W.D. (ed). The Bantu-speaking peoples of southern Africa. Roatledge and Kagen Paul London and Boston.

WADE, P.T., JENKINS, T., HUEHNS. (1967). Haemoglobin variant in a Bushman: Haemoglobin D Bushman $\alpha_2\beta_2$. Nature 216:688-690.

WAINSCOAT, J.S., HILL, A.V.S., BOYCE, A.L., FLINT, J., HERNANDEZ, M., THEIN, S.L., OLD, J.M., LYNCH, J.R., FALUSI, A.G., WEATHERALL, D.J., CLEGG, J.B. (1986). Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319: 491-493.

WALKER, R.H., CRISAN, D. (1991). DNA Technology: The fourth generation in parentage testing. Transfusion 31(5):383-385.

WALKER, R.H., MEYERS, M.A., PHILLIPS, L.M. (1987). The probability of exclusion of the HLA-A, B system in north American Whites and Blacks in parentage tests. Transfusion 27:75-79.

WAYE, J.S., DURFY, S.J., PINKEL, D., KENWRICK, S., PATTERSON, M., DAVIES, K.E., WILLARD, H.F. (1987). Chromosome-specific alpha satellite DNA from human chromosome 1: Hierarchical structure and genomic organization of a polymorphic domain spanning several hundred kilobase pairs of centromeric DNA. Genomics 1:43-51.

WEATHERALL, D.J. (1991). The new genetics and clinic practice. Nuffield Provicial Hospitals Trust, London.

WEBER, J.L. (1988). Molecular biology malaria parasites. Exp. Parasitol. 66:143-170.

WEBER, J.L. (1990). Informativeness of human $(dC.dA)_n$. $(dG\text{-}dT)_n$ polymorphisms. Genomics 7:524-530.

WEBER, J.L. MAY, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44:388-396.

WEISGRABER, K.H., INNERARITY, T.L., NEWHOUSE, Y.M., YOUNG, S.G., ARNOLD, K.S., KRAUSS, R.M., VEGA, G.L. (1988). Familial defective apolipoprotein B-100: Enhanced binding of monoclonal antibody MB47 to abnominal low density lipoproteins. Proc. Natl. Acad. Sci USA 85: 9758-9762.

WHITE, T.J., ARNHEIM, N., ERLICH, H.A. (1989). The polymerase chain reaction. Trends Genet 5:185-9.

WILSON, M., THOMPSON, L. (1969). The Oxford History of South Africa. Vol I. Clarendon Press, Oxford.

WILSON, J.T., MILNER, P.F., SUMMER, M.E., NALLASETH, F.S., FADEL, H.E., REINDOLLAR, R.H., McDONOUGH, P.G., WILSON, L.B. (1982). Use of restriction endonucleases for mapping the β^s-globin. Proc. Natl. Acad. Sci. USA 79:3628-3631.

WOLFF, R.K., NAKAMURA, Y., WHITE, R. (1988). Molecular characterisation of a spontaneously generated new allele at a VNTR locus: No Exchange of flanking DNA sequences. Genomics 3:347-351.

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

WONG, Z., WILSON, V., JEFFREYS, A.J., THEIN, S.L. (1986). Cloning a selected fragment from a human DNA "fingerprint". Isolation of an extremely polymorphic minisatellite. Nucleic Acids Res. 14:4605-4616.

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

WRIGHT, S. (1931). Evolution in Mendelian populations. Genetics 16:97-159.

WYMAN, A.R., WHITE, R. (1980). A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. USA 77(11): 6754-6758.

YOKOI, T., NATA, M., AOKI, Y., SAGISAKA, K. (1990). An unexcluded paternity case investigated with hypervariable DNA loci. Transfusion 30(9):819-823.

APPENDIX A

SOURCES OF REAGENTS

REAGENT	SOURCE
Agar	Difco
Agarose	FMC
Albumin bovine serum	Sigma
Amberlite	Sigma
Ampicillin	Sigma
Ammonium Persulfate	BioRad
Acrylamide	BioRad
Bacto Tryptone	Difco
BIS	Biorad
BSA	Sigma
Caesium chloride	Boehringer Mannheim
Curix film	Agfa
Cyclosporin A	Sandoz
Dextran sulfate	Sigma
DNA polymerase I	Amersham
DNA polymerase I Klenow fragment	Pharmacia
Dimethyl sulphoxide	Merk
EDTA	BDH
Ethidium bromide	Boehringer Mannheim
Formamide	Merk
Foetal calf serum	Whittaker
Herring sperm DNA	Sigma
Histopaque - 1077	Sigma
B-Hydroxyquinolinol	Sigma

Hybond-N	Amersham
Lysozyme	Sigma
Mineral oil	Sigma
Orange G dye	Sigma
^{32}P-deoxycytidine triphosphate	Amersham
*Pst*I	Promega/Anglian
RPMI 1640	Gibco
Salmon sperm DNA	Sigma
Sephadex G50	Sigma
Sodium dodecyl sulphate	Merck
Spermidine trihydrochloride	Sigma
Taq polymerase	Promega
Temed	BioRad
Urea	Sigma

APPENDIX B

MEDIA AND SOLUTIONS

Distilled water was used to make up all solutions.

Chloroform

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

Deionized Formamide

Formamide	50ml
Amberlite resin	5g

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

Denaturing Solution for DNA Transfer

1.5M	NaCl	87.66 g
0.5M	NaOH	20 g

Denhardt's Solution (100x stock)

Dissolve in H_2O to make 100ml:

Ficoll	2g
PVP	2g
BSA	2g

Dispense in 20ml aliquots and store at -20°C

DNA Markers

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

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

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

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

DNA concurrent with normal hybridization.

Hybridization Solution

Hybond-N

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

LB (Luria Bertani) Medium

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

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

Lysing solution

NH_4CL	8.29	g
$KHCO_3$	1.0	g
EDTA	37.2	mg
H_2O	1.0	l

Phenol

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

Post Hybridization Washing Solutions

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

Prehybridization Solution

Hybond-N

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

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

Neutralizing Solution

Hybond-N

1.5M NaCl
0.5M Tris pH 7.2
0.001M EDTA

20xSSC

3M NaCl

0.3M NaCitrate

0.02M EDTA pH 7.7

Stripping Solution

3M NaCl

0.5M Tris pH 7.0

10XTBE

Tris Base	108g
Boric Acid	55g
Na_2EDTA	9.3g

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

TE

10mM Tris

1mM EDTA

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