Edith Cowan University
Research Online

Theses : Honours Theses

1997

A study of the effects of consanguinity at the genomic level in two Pakistani Bradaris

Sheena Sullivan
Edith Cowan University

Follow this and additional works at: https://ro.ecu.edu.au/theses_hons

Part of the Genetics Commons

Recommended Citation
Sullivan, S. (1997). *A study of the effects of consanguinity at the genomic level in two Pakistani Bradaris*. https://ro.ecu.edu.au/theses_hons/291

This Thesis is posted at Research Online.
https://ro.ecu.edu.au/theses_hons/291

Edith Cowan University

Copyright Warning

You may print or download ONE copy of this document for the purpose of your own research or study.

The University does not authorize you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site.

You are reminded of the following:

- Copyright owners are entitled to take legal action against persons who infringe their copyright.

- A reproduction of material that is protected by copyright may be a copyright infringement. Where the reproduction of such material is done without attribution of authorship, with false attribution of authorship or the authorship is treated in a derogatory manner, this may be a breach of the author's moral rights contained in Part IX of the Copyright Act 1968 (Cth).

- Courts have the power to impose a wide range of civil and criminal sanctions for infringement of copyright, infringement of moral rights and other offences under the Copyright Act 1968 (Cth). Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

USE OF THESIS

The Use of Thesis statement is not included in this version of the thesis.

A STUDY OF THE EFFECTS OF CONSANGUINITY AT THE GENOMIC LEVEL IN TWO PAKISTANI *BRADARIS*

Sheena Sullivan BSc

Faculty of Science, Technology and Engineering

Date of Submission: July 18 1997

ABSTRACT

The purpose of the project was to assess the effects of inbreeding on the genetic constitution of two Pakistani *bradaris* (literally defined as brotherhoods). Both *bradaris* contain children born to consanguineous (first cousin) and non-consanguineous marriages. DNA samples have been supplied by Dr Suhaib Ahmed of the Armed Forces Institute of Pathology in Rawalpindi, for a total of 91 individuals.

The specific allele frequencies, and levels of homozygosity of each *bradari*, were determined using twenty fluorescence-labelled microsatellite markers for chromosomes 13 and 15. Amplification of the DNA was performed using the polymerase chain reaction (PCR). The PCR products were separated electrophoretically on an ABI Prism 310 Genetic Analyzer, with GeneScan software employed to identify the alleles of each individual. Comparisons were made between the two *bradaris*, and between the *bradaris* and previously published data available from the GDB (Genome Data Base) and CEPH (Centre d'Études du Polymorphisme Humain). The level of homozygosity in each *bradari* was also compared to expected levels, calculated assuming random mating and with a correction for the inbreeding coefficient for each pedigree.

The observed allele frequencies differed significantly between the *bradaris* for thirteen of the twenty markers. Allele frequencies in each *bradari* were also compared to the GDB and CEPH data and were found to be significantly different for all loci. The observed levels of homozygosity at each locus varied from 4% to 55% in the Khattar, and 3% to 40% in the Rajpoot. Observed homozygosity in each *bradari* was not statistically different from the GDB or CEPH data. Both the basic and corrected values for expected homozygosity were significantly greater than

observed homozygosity in each *bradari*. An increase in homozygosity in the children of first cousin marriages was observed, however it was less than the predicted 6.25%.

Lower than expected levels of homozygosity in the Pakistani families could indicate that there is preferential early selection against homozygotes in these families. There also appear to be reduced homozygosity levels in some regions of the two chromosomes, which may indicate that the resistance to homozygosity is specific to certain loci.

DECLARATION

I certify that this thesis does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

ACKNOWLEDGEMENTS

First of all I would like to thank my supervisor Professor Alan Bittles for his guidance and patience, and Dr Suhaib Ahmed of the AFIP without whom this project would not have been possible.

Thank you to everyone in the Department of Human Biology at ECU for your support and friendship and to everyone in the Centre for Human Genetics for their help along the way, particularly David Chandler for his patience and expertise, and Lizzie Winchester who did most of the lab work on the Rajpoot pedigree for the first half of the project. Thanks also to Erin Marshall, Eoin Thompson, Janet Cahill, Boyd Salmon, Tamara Rogers and Marie McCluskey for their assistance in the laboratory and with the preparation of this thesis.

Thank you to Prof. L.L. Cavalli-Sforza for the provision of the panel of Human Diversity primers used in the study, and to Dr. Luba Kalaydjieva and Asli Tolún for use of the protocols they developed for these microsatellites.

For their assistance with the statistical analysis my thanks are due to Dr Lincoln Schmitt, Dr Wei Wang and Przemyslaw Radzikowski.

Thanks must also go to Joe Austin-Crowe for his computing expertise, for the map of Pakistan and just generally for his entertainment value.

Finally, thank you to all those people whom I haven't mentioned personally but who gave me support and help along the way.

TABLE OF CONTENTS

ABSTRACT I

DECLARATION III

ACKNOWLEDGEMENTS IV

TABLE OF CONTENTS V

LIST OF FIGURES X

LIST OF TABLES XII

LIST OF EQUATIONS XIII

LIST OF ABBREVIATIONS XIV

I. INTRODUCTION 1

I.1 COEFFICIENT OF INBREEDING 1

I.2 AIMS 3

II. LITERATURE REVIEW 5

II.1 HISTORY OF THE REGION 5

II.2 RAWALPINDI 8

II.3 CONSANGUINITY IN PAKISTAN 9

II.4 MORBIDITY AND MORTALITY ASSOCIATED WITH CONSANGUINITY 12

II.5 METHODS FOR ANALYSING THE GENOME 15

II.5.i Early studies 15

II.5.ii Types of microsatellites 17

II.5.iii Benefits of microsatellite analysis 18

II.5.iv Allele detection methods 19

II.6 EVALUATING POPULATION DIVERSITY USING POLYMORPHIC MARKERS 21

II.7 CONCLUSION 22

III. MATERIALS AND METHODS 24

III.1 SUBJECTS 24

III.1.i Khattar 24

III.1.ii Rajpoot 26

III.2 DNA CONCENTRATION 28

III.3 SELECTION OF PRIMERS 28

III.4 PRIMER WORKING SOLUTIONS 32

III.4.i Stanford Human Diversity primers 32

III.4.ii ABI primers 32

III.5 PCR AMPLIFICATION AND PROTOCOL 32

III.5.i Stanford Human Diversity primers 33

III.5.ii ABI primers 33

III.6 EVALUATION OF MICROSATELLITE PRIMERS 34

III.7 FLUORESCENT DETECTION OF ALLELES 35

III.8 ALLELE IDENTIFICATION AND SIZING 35

III.9 PREPARATION OF THE ABI PRISM 310 GENETIC ANALYZER 35

III.9.i Initial ten primers 36

III.9.ii Second set of ten primers 36

III.10 PREPARATION OF DNA SAMPLES FOR THE ABI PRISM 310 GENETIC ANALYZER 37

III.11 ALLELE ASSIGNMENT 38

III.12 TESTING PRIMERS ON THE ABI PRISM 310 GENETIC ANALYZER 39

III.12.i Initial ten primers 39

III.12.ii Second set of ten primers 40

III.13 STATISTICAL ANALYSIS 40

III.13.i Allele frequency differences 40

III.13.ii Calculation of homozygosity 41

III.13.iii Calculation of the coefficient of inbreeding 42

IV. RESULTS 44

IV.1 POLYMORPHIC MARKERS 44

IV.2 REFERENCE POPULATIONS 47

IV.3 VARIATION IN THE ALLELES 48

IV.3.i Allele size variation 53

IV.3.ii Number of alleles 55

IV.3.iii Allele sharing 56

IV.3.iv Allele frequencies 57

IV.4 LEVEL OF VARIATION IN HOMOZYGOSITY OF LOCI 59

IV.4.i Bradari variation 59

IV.4.ii Variation between the bradaris and reference populations 61

IV.4.iii Expected homozygosity 61

IV.5 HOMOZYGOSITY IN THE CHILDREN OF FIRST COUSIN AND NON-CONSANGUINEOUS MARRIAGES 65

IV.5.i Observed homozygosity 65

IV.5.ii Expected versus observed homozygosity 66

IV.5.iii Individual homozygosity 66

IV.6 THE OBSERVED AND EXPECTED COEFFICIENTS OF INBREEDING 67

V. DISCUSSION 68

V.1 SUITABILITY OF REFERENCE POPULATIONS 68

V.2 SUCCESS OF PCR AMPLIFICATION AND FLUORESCENT DETECTION OF ALLELES 69

V.3 EVALUATION OF DINUCLEOTIDE MARKERS 70

V.4 VARIATION IN THE ALLELES 71

V.5 HOMOZYGOSITY AND THE COEFFICIENT OF INBREEDING 73

V.6 POLYMORPHISM INFORMATION CONTENT OF MARKERS 77

V.7 PROSPECTS FOR FUTURE RESEARCH 80

VI. REFERENCES 84

VII. APPENDIX 96

APPENDIX 1: DEFINITION OF TERMS 96

APPENDIX 2: WWW SITES FOR HUMAN GENOME INFORMATION 98

APPENDIX 3: OBSERVED ALLELE FREQUENCIES 99

LIST OF FIGURES

Figure I.1: Abbreviated diagram of a first cousin marriage 2

Figure II.1: Pakistan 5

Figure III.1: Khattar *bradari* pedigree: each marriage and individual is identified by number 25

Figure III.2: Rajpoot *bradari* pedigree: each marriage and individual is identified by number 27

Figure III.3: GeneScan screen image 39

Figure IV.1: Microsatellite map distances for chromosome 13 45

Figure IV.2: Microsatellite map distances for chromosome 15 46

Figure IV.3: Khattar *bradari* pedigree with genotypes and haplotypes for markers on chromosome 13 49

Figure IV.4: Khattar *bradari* pedigree with genotypes and haplotypes for markers on chromosome 15 50

Figure IV.5: Rajpoot *bradari* pedigree with genotypes and haplotypes for markers on chromosome 13 51

Figure IV.6: Rajpoot *bradari* pedigree with genotypes and haplotypes for markers on chromosome 15 52

Figure IV.7: The mean number of alleles for all loci on each chromosome and for both chromosomes combined 55

Figure IV.8: Allele frequencies for D15S11 58

Figure IV.9: Observed values of homozygosity in the Khattar and Rajpoot *bradaris* 60

Figure IV.10: Mean homozygosity levels ... 61

Figure IV.11: Observed and expected values of homozygosity for the Khattar pedigree ... 62

Figure IV.12: Observed and expected values of homozygosity for the Rajpoot pedigree ... 63

Figure IV.13: Mean differences in homozygosity between children of first cousin and non-consanguineous marriages ... 65

Figure IV.14: The distribution of homozygous loci in the children of first cousin and non-consanguineous marriages ... 66

LIST OF TABLES

Table III.1: Initial set of ten microsatellites 29

Table III.2: Second set of ten microsatellites 31

Table IV.1: Allele size ranges in base pairs 54

Table IV.2: Alleles shared between the two *bradaris* 56

Table IV.3: Values of the *t*-statistic for expected versus observed homozygosity where F=0 (i) and with a correction for F (ii) 64

Table IV.4: Expected and observed inbreeding coefficients 67

Table V.1: PIC, observed heterozygosity and expected heterozygosity for the Khattar pedigree 78

Table V.2: PIC, observed heterozygosity and expected heterozygosity for the Rajpoot pedigree 79

LIST OF EQUATIONS

Equation I.1: Coefficient of inbreeding 2

Equation I.2: Coefficient of inbreeding: multiple generations 3

Equation III.1: Observed homozygosity 41

Equation III.2: Expected homozygosity 41

Equation III.3: Expected homozygosity under inbreeding 42

Equation III.4: Expected pedigree coefficient of inbreeding 42

Equation III.5: Observed coefficient of inbreeding 43

Equation V.1: Polymorphism information content 77

LIST OF ABBREVIATIONS

A·C·T·G	Adenine, Cytosine, Thymine, Guanine
ABI	Applied Biosystems Incorporated
bp	Base pairs
CEPH	Centre d'Études du Polymorphisme Humain
cM	Centimorgan
dH_20	Distilled water
dNTP	Deoxy-N triphosphate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine-tetraacetic acid
EtBr	Ethidium bromide
F	Coefficient of inbreeding
GDB	Genome Data Base
H	Heterozygosity
HLA	Human Leucocyte Antigen
kbp	Kilobase pairs
$MgCl_2$	Magnesium Chloride
μl	Microlitre (10^{-6})
ml	Millilitre (10^{-3})
mM	Millimolar (10^{-3})
nm	Nanometre (10^{-9})
OD	Optical density
PCR	Polymerase chain reaction

PIC	Polymorphism Information Content
r	Correlation coefficient
RFLP	Restriction Fragment Length Polymorphism
SSR	Simple Sequence Repeat
TAE	Tris-acetate electrophoresis buffer
VNTR	Variable Number Tandem Repeat
χ^2	Chi-square statistic

I. INTRODUCTION

Consanguineous marriage, which conventionally is defined as a marital union between a couple related as second cousins or closer, is practised in many parts of the world, particularly in countries within North and Sub-Saharan Africa and West, Central and South Asia (Bittles, 1994). In these regions, the highest incidence of consanguinity is in the poorest and least educated sectors of the population, such as landless families resident in rural areas and lower socioeconomic groups in cities (Bittles, 1994). The socioeconomic status of these sections of society has played an important role in increasing the popularity of consanguineous marriages, since they offer the benefit of significantly decreasing the dowry or bride-wealth that may be payable, and ensure that family property is maintained within the extended family. Marriage with a close biological relative is further believed to facilitate prenuptial arrangements, to increase the likelihood that the bride will have an amicable relationship with her husband and her mother in-law, and to reduce hidden uncertainties regarding the health of the spouse and his/her family (Bittles, Mason, Greene and Rao, 1991).

I.1 Coefficient of inbreeding

Theoretically, inbred populations are expected to exhibit higher levels of homozygosity than non-inbred populations, because of the greater probability that the members of a consanguineous marriage will have inherited identical copies of a gene from a common ancestor (Jorde, 1991). In the case of a first cousin marriage, this shared ancestor is in the grandparental generation. Homozygosity by descent, or autozygosity, can be predicted using the coefficient of inbreeding (*F*), a statistical formula that calculates the increase in homozygosity expected in the children of

consanguineous unions (Jorde, 1991). For example, in Figure I.1, I is the child of a first cousin marriage between D and E with a common ancestor A. *F* is calculated by tracing the paths of the gametes that lead from I's parents back to A, through B and C. The probability of autozygosity of the alleles is one-half because, with Mendelian segregation, the probability that a particular allele present in a parent is transmitted to a child is one-half (Hartl, 1988).

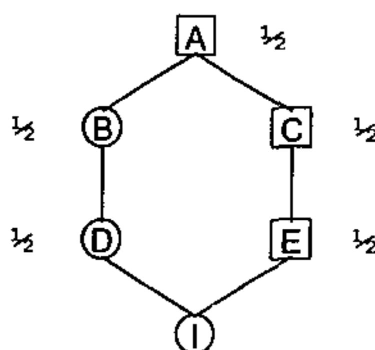


Figure I.1: Abbreviated diagram of a first cousin marriage

Therefore, in the present example, there are five paths between I and A, and so the probability of autozygosity is 1/2 * 1/2 * 1/2 * 1/2 * 1/2 or 1/32. This can be simplified to the equation:

$$F = \sum \left(\frac{1}{2}\right)^{n_1+n_{2-1}}$$

Equation I.1: Coefficient of inbreeding

where n is the number of individuals separating the child and the common ancestor and 1/2 is the probability that the child will inherit the allele of a specific parent. The child of a first cousin marriage will have two such paths, one for each grandparent and so the probability of autozygosity is 1/32 + 1/32 = 1/16. Hence the child of a first cousin union is predicted to display 6.25% greater homozygosity than the child of a non-consanguineous marriage.

In a large pedigree with a history of consanguinity prior to the current generation, the actual coefficient of inbreeding predictably will be higher than can be calculated for a single generation, due to the cumulative effect of inbreeding (Shami, Grant and Bittles, 1994). The effect of prior inbreeding is calculated as the sum of the probability of autozygosity due to each separate path of inheritance of the alleles, and is represented by the equation:

$$F = \sum \left(\tfrac{1}{2}\right)^{n} \left(1 + F_A\right)$$

Equation I.2: Coefficient of inbreeding: multiple generations

where n is the number of individuals in each path connecting the parents and A is the common ancestor in each path (Hartl, 1988).

Although the levels of autozygosity in some human populations have been calculated on a theoretical basis using the above equations, the available literature provides only limited information comparing observed and expected levels of homozygosity in inbreeding communities. Similarly, few data are available comparing the levels of homozygosity in inbred and non-inbred populations. Such an investigation could indicate the extent to which the genome has been affected by inbreeding, and it forms the purpose of the present study.

I.2 Aims

The aim of the project was to use twenty microsatellite markers, ten each for chromosomes 13 and 15, to evaluate how inbreeding may have affected parts of the genome of two *bradaris* from Pakistan. A *bradari*, in Islamic Pakistan, is recognised as the extended family along paternal and/or fraternal lines and frequently includes

consanguineous marriages. The microsatellite analysis had four main aims. These were to:

1. Determine specific allele frequencies for the loci tested in each *bradari*.
2. Compare the observed allele frequencies in each *bradari* with each other and with non-inbred populations.
3. Calculate the observed and expected homozygosity and compare these values in each *bradari*.
4. Compare homozygosity in the *bradaris* with non-inbred populations and at different values of F.

II. LITERATURE REVIEW

II.1 History of the region

Figure II.1: Pakistan

Traces of the earliest inhabitants of modern-day Pakistan exist in the form of stone implements found dating to the second inter-glacial period, from 400,000BC to 250,000BC. These people slowly evolved to form village sites in Baluchistan (the Nal culture), on the Makran Coast to the west of the Indus Delta (the Kulli culture), and along rivers in Punjab and Rajasthan. Later they would found the extensive Indus Valley population which flourished from 2500BC to 1700BC, with its main centres Moenjodaro in Sind and Harappa in Punjab (Thapar, 1990).

The nomadic Aryan-speaking peoples of the Urals and Siberia migrated to the region in about 1500BC, driving the native peoples southward (Wallbank, 1965). By the 9th century BC they had established settlements across the Indian subcontinent, observing a rigid division of labour which was a precursor to the caste system, and practising a Vedic religion which later evolved to Hinduism (Santiago, 1987). At this point there were six main population groups on the subcontinent: the Negrito, Proto-Australoid, Mongoloid (Sino-Tibetan), Mediterranean (Dravidian), Alpine and Aryan (Thapar, 1990).

By the sixth century BC the Persian empire had extended its boundaries to encompass the north-west region of the subcontinent, when Cyrus the Great crossed the Hindu Kush mountains and established a new dynasty (Santiago, 1987). This ended following the conquest of the Persian empire by Alexander of Macedonia in 330BC. Although Alexander returned westwards after reaching the Punjab in 327BC, a number of his soldiers settled in the region where they established and reinforced trade routes between India and Afghanistan, Iran and Asia Minor (Wallbank, 1965). The Greek influence on the cultures subjugated by Alexander and his armies exceeded mere military control, since throughout his conquests Alexander encouraged his soldiers to marry local women (Vollmer, Keall and Nagai-Berthrong, 1983).

The next major political force in the region was the Mauryan empire, established by Chandragupta Maurya who, in the latter years of the fourth century BC, overthrew the Nandan empire in the east of the subcontinent and then successfully campaigned in central and northern India, seizing control of the remnants of the Greek Empire founded by Alexander. The second emperor, Bindushara, extended the area of Mauryan influence by campaigning southward to

Mysore (Allchin *et al.*, 1983). The Mauryan empire is considered to have been one of India's most sophisticated and highly structured political entities, but it lasted for only 100 years, and effectively ceased with the death of its third emperor, Ashoka, in 235BC. Ashoka adopted Buddhism and developed Taxila, in Ghandara, as a centre of religious study. Ghandara was a region that included Peshawar and the Indus, lower Swat and Kabul valleys, and it remained a significant centre of learning for over 700 years. Bactrian Greeks controlled Ghandara and the Indus valley by the second century BC, but they were soon superseded by Scythians from Central Asia, who in turn were displaced from the region by Parthians from Persia (Thapar, 1990).

By the second century AD the Kushans, migrants from Central Asia, had established the Kushan Dynasty, which included Ghandara and extended from eastern Iran to the Chinese frontier and south to the river Ganges. By this time Ghandara had become an important pilgrimage site for Buddhists, and it was a major part of the silk and trade routes which were established between China and India and the Roman Empire (Thapar, 1990). The Kushans gradually lost control of their territories until they held only Ghandara and Kashmir, which eventually fell to the Persian Sassanians in 300AD. The southern areas of Sind and Eastern Punjab became part of the Gupta dynasty of the north-east of the subcontinent. Later, in the fifth century AD, Hepthalites from Central Asia pillaged Ghandara and the Gupta lands, an invasion from which Ghandara never recovered. Eventually, the Hepthalites were driven out by the Turkic dynasty which in turn was replaced by the Hindu Shahi dynasty in the ninth century (Santiago, 1987).

In the tenth century AD, Muslims from Central Asia began their systematic conquest of Indo-Pakistan, and it was the Afghan-Turks who captured the northern region of the subcontinent, bringing it under Islamic rule by the thirteenth century.

Mongolian invaders also exercised some control over the subcontinent, reaching as far as the city of Lahore, until their decline in the fourteenth century. In the sixteenth century, the Moghul Empire was established and its control over most of the subcontinent continued for approximately two centuries. As Moghul power declined, the Sikhs briefly ruled the north-west of the subcontinent, but they were brought under the control of the British Raj in 1849 (Santiago, 1987).

Due to the unrest of Muslim political factions in British India, the concept of a separate Muslim state was suggested in 1930. The name Pakistan, meaning Land of the Pure, was coined and, as part of the negotiations which accompanied the granting of independence, the boundaries between the predominantly Hindu and Muslim states of India and Pakistan were drawn by 1947. The subsequent mass migration of Muslims to West and East Pakistan (now Bangladesh), and of Hindus and Sikhs to India, was accompanied by large-scale rioting and substantial loss of life (Santiago, 1987).

Today, the people of Pakistan represent a mixture of the various ethnic groups which have invaded and settled the Indian subcontinent, and so they include Mongol, Arab, Dravidian and even European influences (King and St Vincent, 1993). They are included in the racial group Caucasian (Stowell, 1996) and in the North, the individuals may be notably fair in complexion. The majority of the population of Pakistan are Punjabi (about 50-60%), with other major ethnic groups being Pathans, Baluch, Sindhis and Mohajirs, the latter being post-Independance migrants from the northern states of India. The predominant religion is Islam, which is followed by 97% of the population, while the remaining 3% are mainly Hindu and Christian (King and St Vincent, 1993).

II.2 Rawalpindi

The two *bradaris* included in this study are from Rawalpindi in the province of Punjab. The city lies on the Potohar plateau occupying the site of an old village inhabited by the Rawals, a group of Yogis. It was part of Ghandara and the Achaemenid Persian Empire, and later became an important centre on the trade route that ran from the Khyber Pass through the Peshawar valley to Lahore (Santiago, 1987). Destroyed during Mongol invasions in the fourteenth century, Rawalpindi was later restored by Jhanda Khan, a Gakhar chief who gave the city its present name. In 1765 a Sikh adventurer named Milka Singh occupied the city and invited people from Jhelum and Shahpur to settle there. It was annexed by the British in 1849, by which time it had become an important commercial centre (Rawalpindi, 1983). According to the 1981 Census of Pakistan, which is the most recent official data source available, the population of Rawalpindi was 764,843. The current population is greatly in excess of that figure.

II.3 Consanguinity in Pakistan

Until recently, the rigid caste system imposed in India did not permit marriages between individuals of different castes or from different regions, and breaches of this law were punished by banishment (Hershman, 1981). Caste endogamy almost certainly has lead to genetic drift among the people of the Indian subcontinent, and it has been suggested to have contributed to differences in genetic traits between the castes (Gadgil and Malhotra, 1983). Perhaps to reduce the adverse effects of inbreeding, permission to marry was only granted if the marriage partners were from different villages, a custom that is still practised today among the Muslim

groups in the Northern Areas (Jammu and Kashmir) in Pakistan (Hershman, 1981; Jamie, 1992).

References to consanguineous marriage appear throughout the history of the subcontinent in the Indo-Aryan texts. The early Aryans associated their god of death with incest, implying that marriage to a close biological relative was unacceptable in their culture (Thapar, 1990). In the [1]*Brahmanas* consanguineous marriage appears to have been permitted to the level of father's brother's daughter (Kapadia, 1958). Later, the [2]*Sutra* writers began imposing restrictions prohibiting marriage between persons related within certain generations of the father and mother. This was known as *sapinda* exogamy and, under these regulations, cross-cousin marriages were proscribed unless justified under specific and peculiar social circumstances. One writer in particular, Gautama, maintained that marriage partners should be related no closer than in the sixth generation on the maternal side, and in the eighth generation on the paternal side, with a lunar penance imposed for unions in the third generation.

An interesting example, which illustrates how views changed at least among the Indo-Aryans, is found in the Royal house of Yadava. Within this family, marriages between partners related within the fourth or fifth generation were frequently contracted, as were first cousin marriages although to a lesser extent. However, it seems probable that by the 7th century AD the attitude towards consanguineous marriage had sufficiently changed to prompt the Yadavas to discredit their ancestors who had contracted marriages in the third generation (Kapadia, 1958).

---

[1] The *Brahmanas* are a Vedic text, written in prose, to guide the four *Vedas* (castes) in their social customs and gives explanations for these customs (Beck, 1996a).

[2] The *Sutras* are various texts believed to have been written by the disciples of Buddha after his death (483BC) that detail appropriate social customs for the castes (Beck, 1996b).

Irrespective of the changing attitudes towards consanguinity in the Indo-Aryans, cousin marriages were customary in the south and in pockets of northern India. Among Dravidians, cross-cousin marriage was a characteristic feature and it is reported in Telugus, Kannadagas and Tamils. Consanguineous marriage prevails among the Muslims throughout the subcontinent and its incidence may have increased in the thirteenth century following the conquest of southern Asia by Muslim invaders from the Middle East. In the Islamic tradition, women can inherit goods and property and so to prevent the dispersal of family wealth, cross-cousin marriages would be favoured (Bittles, 1995). Muslims also practise endogamy within the different sections of Islam. For example, the Ismailis, Sunnis and Shias do not intermarry, and each of the different groups within each of these branches of Islam are also endogamous.

Examination of the 1990-1991 national Pakistan Demographic and Health Survey indicates that consanguinity is strongly favoured in all provinces, and accounts for 62.7% of all marriages (Hussain and Bittles, 1997). Previous studies in urban Punjab found that consanguineous unions constituted approximately 50% of marriages (Bittles, Grant and Shami, 1993; Yaqoob, Gustavson, Jalil, Karlberg and Iselius, 1993). In Pakistan, as in other Muslim countries, the most favoured form of consanguineous marriage is first cousin, especially of the type father's brother's daughter. While this form of marriage results in a coefficient of inbreeding of zero at the X-chromosomal loci ($F_X = 0$), it is more likely to have been adopted for social reasons rather than its genetic benefits - the prophet Mohammed married his cousin, and arranged for his daughter to marry his paternal uncle's son (Bittles, 1994).

In a full first cousin marriage, the couple are predicted to share identical alleles inherited from the same two ancestors (their common grandparents) at one

quarter of all autosomal loci. The marriage partners may actually be even more closely related than can be calculated for a single generation as, in a country such as Pakistan where consanguinity is preferential, it is probable that the couple share more than one common ancestor. *Bradari* marriages are also the norm in most communities, which further increases the probability of homozygosity in the progeny (Shami *et al.*, 1994).

II.4 Morbidity and mortality associated with consanguinity

The available literature indicates that mortality and morbidity are higher in those populations which frequently contract marriages between close biological relatives, because of the increased probability of being autozygous for an otherwise rare recessive disease (Bittles, 1994). Some reviews of the relationship between consanguinity and mortality and morbidity in Pakistan have indicated that mortality and morbidity were higher in the children of consanguineous unions than non-consanguineous marriages (Bittles *et al.*, 1993; Grant, Shami, Milligan and Bittles, 1994). Similar findings have been reported in other spatially distant parts of the world, such as Nigeria (Scott-Emuakpor, 1974), Japan (Schull and Neel, 1972) and Brazil (Freire-Maia, Freire-Maia and Quelce-Salgado, 1963). These studies have not, however, indicated that inbreeding has as much influence on morbidity and mortality as expected.

There are a number of possible explanations for these results. Theoretically, a long history of inbreeding in a community will lead to the elimination of rare lethal genes due to the non-reproduction of persons who are homozygous (Radha Rama Devi, Appaji Rao and Bittles, 1987). A similar picture may hold true for sub-lethal genes, depending upon the severity of the disease. In either case, if inbreeding is

practised over a long period of time then the particular trait may be eliminated, which in turn could mean that the adverse effects of mortality and morbidity in succeeding generations would be substantially reduced.

Mortality may also be recorded as low because deaths go unreported in rural communities, and among groups of low socioeconomic status. Pre-natal deaths may be under-recorded where medical care was not available or deemed necessary, or because the fetus was spontaneously aborted before the mother was aware of the pregnancy. Unrecorded miscarriages are believed to occur quite frequently in all human populations (Roberts and Lowe, 1975), although their actual incidence is difficult to determine (Wilcox *et al.*, 1988). Consanguineous couples may have an elevated frequency of spontaneous abortion, as conceptuses with high autozygosity may have an increased chance of being homozygous for a mutation predisposing to early abortion (Schull and Neel, 1972). A strong argument against this possibility is that sterility appears reduced in consanguineous couples, but again this could be strongly dependent on environmental and age differentials (Bittles *et al.*, 1991). In either case, the inability to determine the actual level of prenatal death in a community with any real degree of confidence means that both theories remain speculative.

Morbidity within an inbreeding population may also be underestimated because it is less precisely defined than mortality. Errors may arise where morbidity within a population is recorded at birth, which severely underestimates the real incidence of disease since only about 40% of congenital abnormalities are diagnosed in the newborn child (Al-Gazali, Dawodu, Sabarinathan and Varghese, 1995; Teebi, 1994). Moreover, morbidity will appear reduced if only severe defects are recorded,

or if the available clinical facilities do not permit the detection of all internal malformations (Teebi, 1994).

Conversely, the apparent risks associated with consanguinity may be exaggerated. Environmental factors associated with the low socioeconomic status of the majority of the couples who contract consanguineous unions predictably will be detrimental to their health (Shami *et al.*, 1994). For example, Yaqoob *et al.* (1993) found that the rate of birth defects increased with decreasing socioeconomic status, but observed no significant association between consanguinity and birth defects.

Secondly, consanguineous marriages are usually contracted between couples who are younger than those not marrying a relative (Satha and Ahmed, 1992). This characteristic may also be linked to socioeconomic status and poor education, as was reported by Mian and Mushtaq (1994), who found that as economic status increased so too did age at marriage. In younger parents, the physical immaturity of the couple also may represent a risk to the health of the developing fetus, particularly the physical and hormonal development of the mother (Bittles *et al.*, 1991).

The presence of recessive lethal genes in inbred and outbred populations is influenced by natural selection, with positive selection for genes which are beneficial. The sickle-cell anaemia gene for example, is a recessive lethal which in the heterozygous state will provide a degree of immunity to malaria (Cummings, 1991). The genes coding for less common disorders may exhibit similar characteristics in inbred communities if they also provide an advantage in coping with specific environmental conditions. Finally, it has also been suggested that some disease-causing genes provide a benefit *in utero*, thereby maintaining their presence in the population (Diamond, 1987).

Despite these caveats, the level of homozygosity in a population may be reflected in the rates of morbidity and mortality observed in inbred communities. Although statistical compensation can be employed to control for socioeconomic and age differentials, the applicability of these adjustments remains largely unproven (Grant and Bittles, 1997) and direct assessment of the level of homozygosity in a population can only be conducted at the genome level. Furthermore, some indication of the genetic isolation of a community may be gained if the degree of genetic variation in the community is less than would be expected with the same allele frequencies, but under conditions of random mating. Several options are available to investigate the effect that inbreeding has exerted on the genome, and these are discussed in the next section.

II.5 Methods for analysing the genome

II.5.i Early studies

The search for the disease loci of inherited diseases and the development of evolutionary trees have contributed to the construction of maps of the human genome. Genetic maps contain fixed reference points known as markers, i.e., nucleotide sequences which display changes in the base pair sequence and are present at polymorphic frequencies. Ideally, a map will contain one marker every 100 kilo base pairs (kbp). These markers are inherited in a codominant Mendelian manner, and for each marker an individual will have two forms, termed alleles, one from each parent (Ott, 1991).

Early studies mapped the human genome using markers such as blood group antigens, serum protein polymorphisms and allozymes, which required a range of biochemical techniques for their analysis and often proved to be uninformative

(Hudson *et al.*, 1992; Queller, Strassmann, and Hughes, 1993). Mapping was revolutionised with the introduction of restriction fragment length polymorphisms (RFLPs) in the early 1980s (Botstein, White, Skolnick and Davis, 1980). RFLPs are identified by employing a DNA restriction enzyme which cleaves the sample DNA at a specific nucleotide recognition sequence, and the resulting products can be detected by agarose gel electrophoresis (Botstein *et al.*, 1980). More than 3000 RFLPs have been described, with uses ranging from the mapping of Mendelian diseases to the investigation of genetic heterogeneity in a given population (Ewens, Spielman and Harris, 1981; Litt, 1991). RFLPs offer simpler identification procedures than previous methods, but are limited in their application since most have a heterozygosity of less than 50% and the results obtained are often difficult to compare between experiments (Donnis-Keller *et al.*, 1987; Kidd *et al.*, 1989).

Mapping was further improved by the description of DNA minisatellites (Jeffreys, Wilson, Thein, Weatherall, and Ponder, 1986) also referred to as variable number tandem repeat sequences (VNTR; Nakamura *et al.*, 1987). Minisatellites are stretches of DNA in which a short nucleotide sequence is repeated some 14 to 100 times (Krontiris, 1995), and the alleles are defined by changes in the number of repeats (Lewontin and Hartl, 1991). They are detected indirectly by excising the region of DNA with a restriction enzyme and estimating the length of the fragment by gel electrophoresis and Southern Blotting (Devlin, Risch and Roeder, 1990). Minisatellites are highly polymorphic, in some cases with heterozygosity of almost 100%, and therefore they are considerably more informative than RFLPs. Their application is, however, limited because they are non-randomly spaced throughout the genome, being concentrated in the telomeric regions of chromosomes (Royle, Clarkson, Wong and Jeffreys, 1988).

The problems associated with RFLPs and VNTRs have been overcome by the use of simple sequence repeats (SSRs) or microsatellites. Microsatellites are DNA sequences that are usually less than 500 base pairs (bp) in length and contain a repeat motif of 1-6bp (Housman, 1995). They are thought to arise due to slippage during replication or unequal meiotic exchanges (Litt and Luty, 1989). Microsatellites are found throughout the genome, within introns, protein coding regions and between genes, and they are highly polymorphic, thereby overcoming the problems associated with VNTRs and RFLPs (Hearne, Ghosh and Todd, 1992).

II.5.ii Types of microsatellites

There are several types of microsatellites, the simplest being the monomeric repeats, such as the A·T multimers which appear to be the most frequent form in the human genome. Dimeric $(C\text{-}A)_n \cdot (G\text{-}T)_n$ repeats are the most common, with one every 30kbp (Weissenbach *et al.*, 1992). Their abundance means that dinucleotides with a high level of heterozygosity are relatively frequent throughout the genome, and for this reason they have been extensively used both for pedigree analysis and to generate high density maps of the human and mouse genomes (Deka *et al.*, 1995).

Trimeric and tetrameric repeats are less common, although they nevertheless are described quite frequently in the human genome, particularly on the X-chromosome where they are found every 300-500kbp (Edwards, Hammond, Jin, Caskey, and Chakraborty 1992). AAAN and AAN repeats are especially abundant (Hearne *et al.*, 1992). They are easier to amplify and to identify than dinucleotides, thus facilitating allele size assignment for DNA typing and genetic mapping (Gill and Evett, 1995).

Trimeric repeats are particularly interesting, because expansions in some of the repeats have been implicated in the aetiology of several important human genetic disorders, including the neurodegenerative disease Huntington's chorea, fragile X syndrome, myotonic dystrophy and Jacobsen syndrome (Hofferbert, Schanen, Chehab and Francke, 1997).

II.5.iii Benefits of microsatellite analysis

The high polymorphism found in microsatellites is due to variation in their number of repeat units (Hearne *et al.*, 1992). Microsatellites have an estimated mutation rate of between $5x10^{-5}$ and $5x10^{-4}$, which is sufficiently large to have allowed the evolution of polymorphisms, yet with a frequency that permits their use in demonstrating inheritance for linkage and population studies, and for forensic studies (Hearne *et al.*, 1992).

A further benefit of microsatellites is that their small size facilitates cloning and synthesis (Yuille, Goudie, Affara and Ferguson-Smith, 1991), and allows them to be reliably amplified using the Polymerase Chain Reaction (PCR; Weber and May, 1989). This method of amplification of a marker is cheaper and quicker than the Southern Blotting technique used for the much larger minisatellites (Boerwinkle, Xiong, Fourest, and Chan, 1989). PCR is also superior because it allows the amplification of more than one locus in a reaction for a single DNA sample, by the technique of "multiplex" PCR (Chamberlain, Gibbs, Ranier, Nguyen and Caskey, 1988). Again, this method is feasible because microsatellites are so small. Multiplex reaction kits are also available, which standardise the reactions, reduce costs and simplify the technical procedures (Edwards, Civitello, Hammond and Caskey, 1991).

Microsatellites provide the additional benefit that alleles at a specific locus can be easily detected using radioactive or non-radioactive labels. The labels are attached to the microsatellite primer sequence during synthesis, or are incorporated into the PCR reaction mixture. Fluorescent labelling, a non-radioactive method of detection, is particularly sensitive and offers a number of advantages over the use of radioactive labels that will be discussed in the next section.

II.5.iv Allele detection methods

The alleles at a locus can be detected in a number of ways however, as previously stated, the two most popular methods are the incorporation of radioactive or fluorescent labels. In both methods, the labels are used to identify the alleles of the target sequence of DNA during PCR amplification, the products of which can then be separated by electrophoresis. Radioactive detection employs X-ray film to visualise the alleles which appear as black bands on the film, while fluorescent labels are detected and interpreted using specialised fluorescence-detecting instruments and computer software. Overall, fluorescent labelling techniques are preferred to radioactive methods because they are more sensitive, offer greater accuracy in sizing, are cheaper and safer.

One advantage of radio-isotope labelling over fluorescent-based methods is that they are simpler to use because the equipment required is not as labour-intensive and technically demanding (Perlin, Burks, Hoop and Hoffman, 1994). They cannot however provide the level of resolution achieved by the instruments used to detect fluorescent labels (Fregeau and Fourney, 1993). This resolution is attained, in part, by the inclusion of a fluorescent standard within each sample to size the alleles. Size standards display a fixed pattern of peaks at certain base pair lengths, which the

instrument can then use as a guide to size the unknown alleles. The method also facilitates comparisons between and within runs, because the software assigns an actual size in base pairs to each peak (Sullivan, Pope, Gill and Robertson, 1992). Knowing the size of each allele aids the researcher further by allowing objective interpretation of the results, unlike autoradiographs where allele assignment is left entirely to the discretion of the researcher (Makino, Yazyu, Kishimoto, Sekiya and Hayashi, 1992).

Although the initial costs of the equipment used with fluorescent detection methods is higher than for radioactive methods, the associated consumable costs are lower (Chehab and Kan, 1989; Ziechler, 1989). The consumable costs can be reduced by several factors, the most significant being the optimisation of conditions. Since the instrument can detect four different coloured dyes, red, blue, green and yellow (one of which is used solely for the size standard), at least three different PCR products can be included in each lane (Makino *et al.*, 1992), and more than three can be included if the size differences between the microsatellites are sufficiently large to avoid overlap (Gill and Evett, 1995).

Amplified microsatellites occasionally will display minor "stutter" bands in front of the allele fragments, particularly in dinucleotide repeats (Love, Knight, McAleer, and Todd, 1990). These bands are thought to arise from impurities in the primers, or from errors in *Taq* polymerase replication during PCR (Litt and Luty, 1989; Perlin *et al.*, 1994). On X-ray film, stutter bands can lead to genotype misclassification because a homozygote may appear to be a heterozygote. The introduction of fluorescence-labelled primers and computer-aided analysis for microsatellite applications has improved allele identification, and so reduced ambiguities associated with stutter bands.

A final important advantage offered by fluorescent labels is that they do not pose the potential health risks that are associated with the use of radio-isotopes, and they also obviate the need for an expensive, high-specification radioisotope laboratory.

II.6 Evaluating population diversity using polymorphic markers

To date, there have been few reported attempts to analyse the differences between inbred and non-inbred communities using microsatellites. Estimates of the coefficient of inbreeding in major populations have been obtained from analysis of marriage registration, and from individual family pedigrees, but these findings have not been supported by genetic analysis (Jorde, 1991). There have been a number of investigations comparing levels of genetic variability in different populations, which have employed a range of markers, including allozymes, RFLPs, VNTRs and microsatellites. This same methodology for investigating differences between major populations can equally be applied to the pedigrees of inbred communities.

Some studies (Chakraborty, Deka, Jin and Ferrell, 1992; Deka *et al.*, 1995) have assessed the heterozygosity in different populations at various polymorphic loci. Although they do not specifically refer to inbreeding, they have investigated various large and small populations and, in the latter case, have effectively addressed the influence of random inbreeding associated with genetic drift. Their findings indicate that heterozygosity decreases with decreasing population size and increasing isolation. Inbred communities, such as *bradaris* in Pakistan, can be relatively small and isolated so, irrespective of the effects of preferential consanguinity, a lower level of heterozygosity would be expected within a *bradari* than in a larger non-divided population.

Other researchers (Murty *et al.*, 1993; Bowcock *et al.*, 1994) investigated the specific allele frequencies present in various populations and reported greater genetic variation within than between populations. However, a comparison of populations separated by large spatial distances displayed greater differences in allele frequency distribution than did populations which were geographically closer together (Hou, Schmitt, Staak, Puers and Prinz, 1994). With these precedents in mind, in the context of the Punjabi *bradaris* it would be expected that there would be greater allele frequency variation within each *bradari* than between them, but that they both would show greater genetic similarity to each other than to non-Punjabi populations.

II.7 Conclusion

There is documentary evidence that consanguineous unions have been strongly favoured in Pakistan for at least the last four generations (Shami *et al.*, 1994), and it seems probable that marriage between close biological kin is a long-standing tradition in the region. Consanguinity has been linked to elevated levels of morbidity and mortality, but whether these findings can be ascribed to an increase in homozygosity, rather than to adverse environmental factors, remains unclear. While morbidity and mortality can indirectly indicate an increase in homozygosity, they cannot provide a true picture of the level of homozygosity in the genome.

Microsatellite analysis of genetic diversity in world populations is revealing much information about genomic differences in ethnic groups, and currently it appears to be the most appropriate method for analysis of the whole genome. While other marker systems exist, including RFLPs and VNTRs, microsatellites present fewer technical problems and they are more economical in large-scale use. Fluorescent labelling of these markers and amplification using PCR appears to be the

best method of detecting the alleles. Therefore, to investigate the effects of inbreeding at the genomic level, the use of fluorescence-labelled microsatellites was selected as the most appropriate option for the present study.

III. MATERIALS AND METHODS

III.1 Subjects

The DNA samples of the *bradari* members under study were initially collected for research into β-thalassaemia, conducted in the Armed Forces Institute of Pathology (AFIP) in Rawalpindi, and University College, London. The samples for microsatellite analysis and the associated pedigrees were provided by Lt. Col. (Dr) Suhaib Ahmed of the AFIP in April 1996.

For the purpose of the study, the definition of consanguineous marriage was limited to couples related as second cousins or closer. In total, 91 individuals were investigated. DNA was not available for deceased individuals, and individuals were not included if they had married into the family and had no children.

III.1.i Khattar

The Khattar *bradari* pedigree represents five generations and comprises 67 individuals. DNA samples were received for 51 individuals from generations III, IV and V, represented in Figure III.1. Of these 51 individuals, there are thirteen marriages with a total of 40 children from marriages 1 to 13. Six of these marriages are non-consanguineous, numbered 1, 2, 3, 5, 6 and 9, with a total of eighteen children. DNA samples were received for only one member in marriages 1, 2, 5 and 6, either because the marriage partner was deceased or their DNA sample had not been obtained. There are six consanguineous marriages in the pedigree, numbered 4, 7, 8, 10, 11, 12 and 13, all of which are between first cousins. The total number of children born to the consanguineous unions is 22.

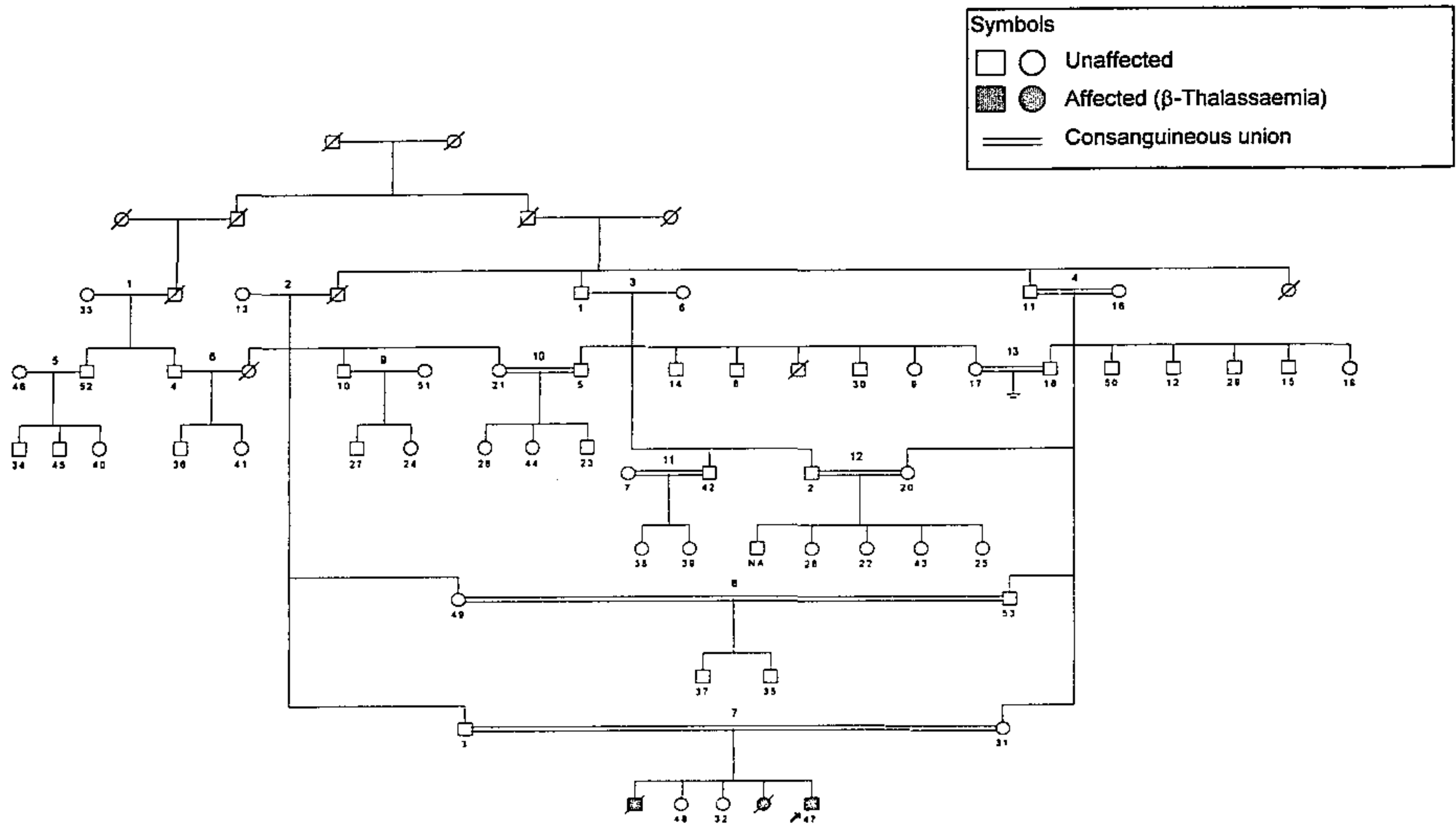


Figure III.1: Khattar *bradari* pedigree: each marriage and individual is identified by number

III.1.ii Rajpoot

The Rajpoot *bradari* pedigree comprises four generations, with individuals identified by number from 1 to 42 for generations II, III and IV (Figure III.2). Only forty DNA samples have been included because individual 1 (the proband) was not received, and individual 13 was excluded because she married into the family and has no children. There are nine marriages within the pedigree, three of which are non-consanguineous and six which are consanguineous. The non-consanguineous marriages, numbered 1 to 3, are all in generation II and have a total of sixteen children. The six consanguineous marriages, numbered 4 to 9, are all in generation III and each is a first cousin union. DNA samples were received for sixteen children from four of the consanguineous marriages.

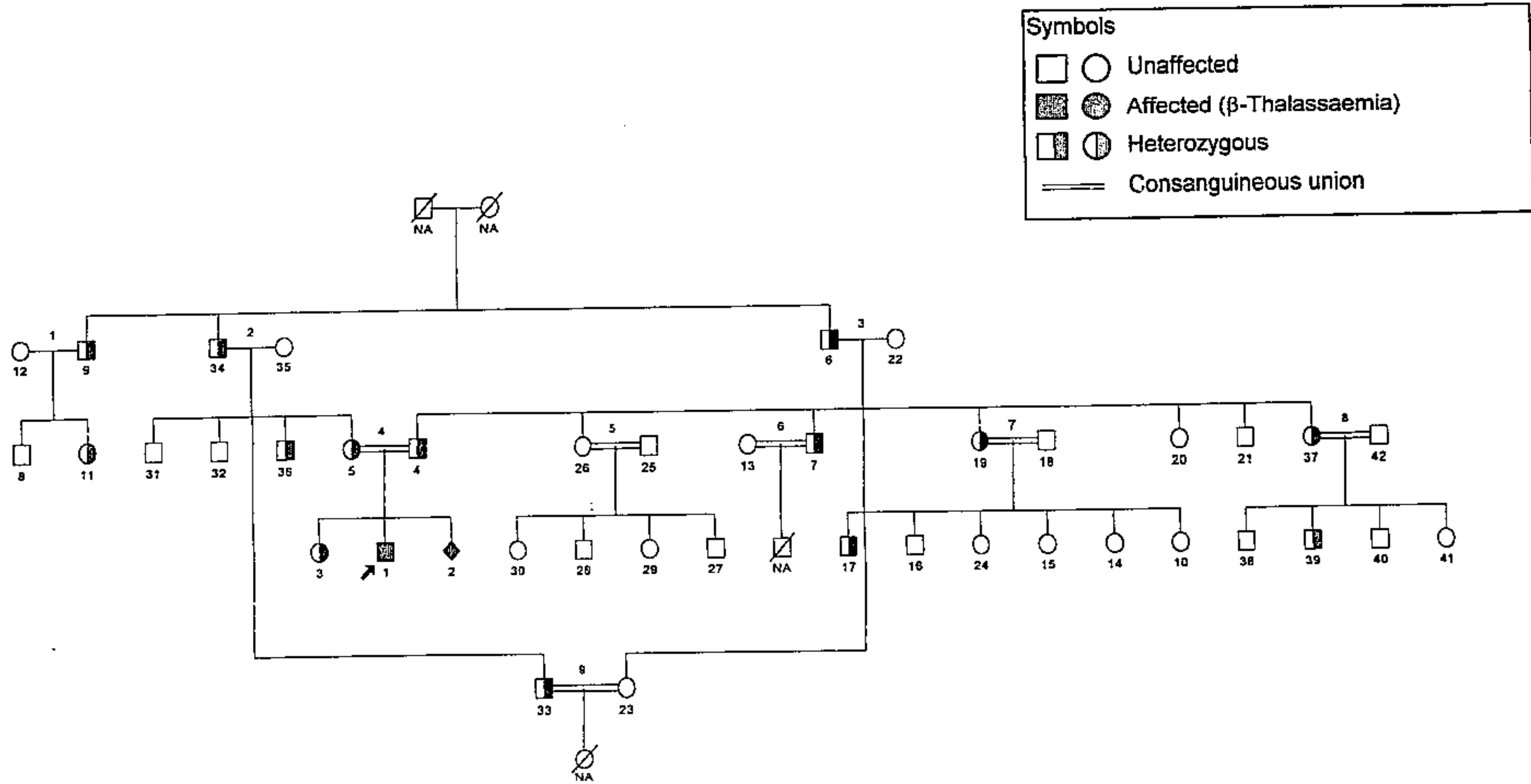


Figure III.2: Rajpoot *bradari* pedigree: each marriage and individual is identified by number

III.2 DNA concentration

The DNA samples were received for each individual in 15-20μg quantities, suspended in approximately 30μl of water. To estimate the concentration of DNA in the samples, two specimens were selected randomly from each pedigree and the DNA concentration determined using a UV/VIS DU640 Spectrophotometer (Beckman). A blank reading on water was recorded before measuring the samples. A fifty-fold dilution of the specimens was read at the 260nm wavelength. DNA concentration was calculated from the 260nm reading, since one optical density (OD) unit at 260nm equates to 50μg/ml of DNA (Sambrook, Fritch and Maniatis, 1989). The average DNA concentration of the four specimens tested was approximately 530ng/μl, but calculations for the preparation of working solutions were based on a 500ng/μl stock concentration to allow for possible overestimation of the DNA concentration in the test samples (this is more easily compensated in the PCR mixture than underestimation). 200μl working solutions were prepared at 5ng/μl concentration using sterile water. Each working sample was gently vortexed and then briefly centrifuged before storage at 4°C. Stock solutions were stored at -80°C to inhibit degradation of the samples.

III.3 Selection of primers

Ten microsatellites were initially selected from the panel of thirty Human Diversity primers made available by Prof. L.L. Cavalli-Sforza, Department of Genetics, Stanford University. Six primers for chromosome 15 and four for chromosome 13 were chosen from twelve primers previously used in the Centre for Human Genetics, Edith Cowan University in population genetics studies. The

conditions for the amplification of these microsatellites had been optimised as part of that research (Kalaydjieva and Tolún, 1996).

Table III.1: Initial set of ten microsatellites

Locus	Label	Oligo name	Sequence
D13S126	FAM	1303L	TCACCAGTAAAATGCTATTGG
		1303R	GTGATTTTCAAATTTGCTCTG
D13S133	TET	CA008L	GGCAACATAGGGAAACCCTAGC
		CA008R	GCTAGGACTACAGGTGCAAACC
D13S192	HEX	HKCA3-1	GGGTAACATAGCAAGACCCC
		HKCA3-2	AGGTATGAGCCATCTCGTCC
D13S270	HEX	084xc5a (CA)	AGTGCCTGGGTATGAACGTG
		084xc5m (GT)	CTGGAAATGCCTTGGAAGGA
D15S101	HEX	MS178L	GAGCCAAGATCATGTTGC
		MS178R	TGCCCACTAGTTTGAGACA
D15S108	HEX	MFD102L	ATTCTTAACAGGAAGTGAGGG
		MFD102R	AACATGAGTTTCAGAGGGG
D15S11	TET	D15S11L	GACATGAACAGAGGTAAATTGGTGG
		D15S11R	GCTCTCTAAGATCACTGGATAGG
D15S97	FAM	MS14L	TCTCCCTCCAATAATGTGAC
		MS14R	TGAGTCAATGATTGAAATTACTG
D15S98	HEX	MS112L	CATGTGAAACTGCAAAAGCTG
		MS112R	AAAAGTCGCATTTGGTCGTT
GABRB3	HEX	L	CTCTTGTTCCTGTTGCTTTCAATACAC
		R	CACTGTGCTAGTAGTTCAGCTC

After the results had been collated for the initial set of ten microsatellites, a further ten were chosen in order to facilitate the development of haplotypes for chromosomes 13 and 15. This second group was composed of six microsatellites for chromosome 13 and four for chromosome 15. Six of the second set were also selected from the Stanford Human Diversity set of primers, three for chromosome 15 and three for chromosome 13. Where more than one primer was located in a region of the chromosome to be investigated, the marker with the highest level of heterozygosity was chosen. The remaining four markers were selected from the ABI Prism Linkage Mapping Set, panels 17, 19 and 22 (Perkin Elmer), for chromosomal locations not covered by any of the Stanford Human Diversity set. There were three such microsatellites selected for chromosome 13 and one for chromosome 15.

Table III.2: Second set of ten microsatellites

Stanford Human Diversity primers

Locus	Label	Oligo name	Sequence
D13S115	HEX	MS34-2	TCTTAGCTGCTGGTGGTGG
		MS34-1	TGTAAGGAGAGAGAGATTTCGACA
D13S144	HEX	1348GT	TCCAAGTATGATTAATCGGAG
		1348R	TCATAATCATGTGAACCAATTC
D13S125	FAM	1320L	GTAACTGGCCAGAATGTCAT
		1320R	GTCCTCCAAAAGAACTCACA
D15S102	FAM	N130-2	TAGGGCCAATGGAGAGAGC
		N130-1	TCAATAACTCCATTGCTCAGTCC
D15S100	TET	MS164-1	CTTTCCAATTCACCCCCAC
		MS164-2	ATCCAGCTCCCCCAAATATT
D15S169	TET	Utsw1591L	CAGGAGAGAGCCTTGGAT
		Utsw1591R	GAGACATCTCTTCTGAAAGCTC

ABI Prism Linkage Mapping Set primers

Locus	Label	Oligo name	Sequence
D13S175	TET	AFM249xb1m	TGCATCACCTCACATAGGTTA
		AFM249xb1a	TATTGGATACTTGAATCTGCTG
D13S173	TET	AFM261yg5a	CCCTGTTCCAGTAATGATGACC
		AFM261yg5m	GTCTCTGGCTGCTCTCAAGACTAT
D13S285	FAM	AFM309va9a	ATATATGCACATCCATCCATG
		AFM309va9m	GGCCAAAGATAGATAGCAAGGTA
D15S120	HEX	AFM164zc9m	GGCTCAAAGTGTTTGCACTG
		AFM164zc9a	TTTGTGATGGTCTTTTATAGGCATA

III.4 Primer working solutions

III.4.i Stanford Human Diversity primers

Fluorescence-labelled forward and reverse primers were received at 8μM concentration and in 200μl quantities. Working solutions of the initial ten primers were prepared in 1.6μM concentrations (20μl primerF + 20μl primerR + 200μl sterile water).

III.4.ii ABI primers

The ABI primers were received from Perkin Elmer in tubes containing 1500pmol of primer at a 10μM concentration. Working solutions were prepared in a 1:1 ratio, in 50μl volumes (25μl primerF and 25μl primerR), to give a 5μM solution.

III.5 PCR amplification and protocol

The dinucleotide tandem repeat sequences were amplified using the polymerase chain reaction (PCR), an *in vitro* method for synthesising defined sequences of DNA using a thermostable DNA polymerase enzyme. The reaction consists of three steps: denaturation, annealing and extension. In the first step, the DNA is separated into single strands that can be used as a template. Step two employs two oligonucleotide primers that anneal to the template DNA at positions flanking the target DNA sequence. Finally, a complementary copy of the region specified by the two primers is synthesised using the enzyme *Taq* polymerase. Repetition of these steps results in exponential amplification of the target sequence (Eeles and Stamps, 1993).

III.5.i Stanford Human Diversity primers

The PCR protocol for the Human Diversity primers developed by Kalaydjieva and Tolún (1996) was used to test the primers. Each PCR was made up to 5μl containing 2μl (10ng) of sample DNA, 1 (0.32μM) of primer, 1μl of 5x buffer (5-7.5mM $MgCl_2$; 1mM dNTPs, 0.5μl 10x polymerase buffer [Biotech]), 0.05-1μl *Taq* DNA polymerase (Biotech) and 0.95μl of dH_20. PCR reaction mixtures were carried out in 0.2ml reaction tubes (Perkin Elmer) in a 96 Well GeneAmp Thermocycler (Perkin Elmer). The thermocycling consisted of four main components. Initially, the samples were denatured for 5 minutes at 94°C. This was followed by fifteen cycles of 20 seconds denaturing at 94°C, one minute of annealing, starting at 63°C and reduced in each cycle by 0.5°C (giving a final temperature of 55.5°C), and a 30 second extension period at 72°C. A further fifteen cycles followed, each consisting of 20 seconds denaturing at 94°C, 20 seconds annealing at 55°C and 30 seconds of extension at 72°C. The cycle concluded with a five minute extension period at 72°C.

III.5.ii ABI primers

A modification of the PCR protocol received with the ABI primers was used to test the primers, employing reduced reaction volumes. PCR reaction mixtures were made to 5μl volumes, consisting of 2μl (10ng) DNA, 0.4μl primer (0.4μM), 1μl 5x buffer (7.5mM $MgCl_2$; 1mM dNTPs; 0.5μl 10x polymerase buffer [Biotech]), 0.05μl of *Taq* polymerase and 1.65μl dH_20. The PCR began with an initial denaturing period lasting five minutes at 95°C. This preceded ten cycles of 15 seconds at 94°C, 15 seconds at 55°C and 30 seconds at 72°C. A further twenty

cycles followed, consisting of 15 seconds at 89°C, 15 seconds at 55°C and 30 seconds at 72°C. The final extension temperature was 72°C and lasted for five minutes.

III.6 Evaluation of microsatellite primers

To ensure that the primers were functioning, PCRs were prepared and tested by agarose gel electrophoresis. This technique employs an electric current to move the negatively charged DNA towards a positively charged electrode through an agarose gel. The larger the allele fragment, the slower it moves through the agarose. Therefore the alleles are differentiated by length fractionation. The PCR products are visualised using Ethidium Bromide (EtBr) under fluorescent light.

A 2% agarose gel solution (0.5g agarose powder [Sigma Chemical Company] in 100ml 1xTAE buffer [0.04M Tris-acetate; 0.001M EDTA]) was prepared and poured on to an 80 ml mini-gel tray, and a small toothed comb was inserted at one end of the tray. The gel was allowed to set for approximately 40 minutes at room temperature, after which the comb was removed and the gel was then placed in the electrophoresis unit. The 5µl PCR products were loaded into the wells with 1µl of 6x Ficoll loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF; 15% Ficoll [Type 400; Pharmacia] in water). pUC19 DNA/Hpa II (0.5mg/ml; Biotech; fragment size range from 26-501bp) was loaded into lane 1 as a size standard. The gel was electrophoresed at 100V for approximately 30 minutes. It was then stained for 10 minutes in EtBr (1.5µl 100% EtBr / 30ml water) and viewed on UV light using a Mighty Bright transilluminator (Hoefer Scientific Instruments). A photograph of the gel was taken using DS34 Direct Screen Instant Camera (Polaroid). All primers produced a good signal in the agarose gel.

III.7 Fluorescent detection of alleles

The forward primer for each of the microsatellites was labelled with one of three fluorescent dyes, each of which appears as a different colour on the computer screen: FAM (blue), TET (green) or HEX (yellow). The standard, GeneScan 500-TAMRA (ABI), is labelled with TAMRA which appears red. The labels emit light at specific wavelengths, which the computer software interprets and outputs as different colours. These fluorescent labels are incorporated into the PCR product during amplification and highlight the alleles.

III.8 Allele identification and sizing

Identification and sizing of the alleles was accomplished using an ABI Prism 310 Genetic Analyzer which employs GeneScan Software (Perkin Elmer, 1995). Alleles are identified by the ABI Prism 310 via a laser-induced fluorescent capillary electrophoresis system. PCR products are loaded on to the autosampler tray where they are individually introduced into the polymer-filled capillary. Each PCR product travels through the capillary to a window where the DNA fragments are illuminated by a laser. The laser excites the fluorescent dyes attached to the fragments, causing each dye to emit light at a specific wavelength (Hearne *et al.*, 1992). A spectrograph then collects the light emitted by the dyes, and separates it according to wavelength. The data are collected by a charged-coupled device camera that allows all four types of fluorescent emissions (FAM, TET, HEX and TAMRA) to be detected simultaneously. The light is stored as electrical signals that can be interpreted by the associated software (ABI Prism 310 Genetic Analyzer Users' Manual, 1995).

III.9 Preparation of the ABI Prism 310 Genetic Analyzer

III.9.i Initial ten primers

A detailed description of the procedures for preparing the ABI Prism 310 can be found in the ABI Prism 310 Genetic Analyzer Chemistry Guide (Research Genetics, 1995). The conditions used were those suggested for short fragments. The capillary was 75μm in diameter, with a total length of 41cm, and length to the window of 30cm. The ABI Prism 310 uses a 2% polymer (5g of 10x Genetic Analyzer Buffer [ABI]; 14.1g of 7% GeneScan Polymer [ABI] in 50ml water) at the anode, and a 3% polymer (6.6M urea, 1x Genetic Analyzer Buffer [ABI]) at the cathode. Buffers are sufficient for up to 96 electrophoretic separations or 24 hours, unless stored at 4°C when it can be used for up to 2 weeks. The instrument also requires washing reagents which are used to clean the capillary between runs. These include a 0.3N NaOH solution, a 1N HCl solution and two tubes of dH_20.

The ABI Prism 310 run module which was routinely selected was GS-Short Denatured C. If the capillary had been used for more than 100 separations the GS-XT-Short Denatured C module was employed, which includes a wash cycle between separations. The matrix file selected was GS C 3 DEN UREA 6.6M, and the size standard file chosen was GS 500 TAMRA 3p.10sec.41cm. For all separations, the injection time was 10 seconds at 7kV, and runs were conducted at 13kV at 30°C for 15 minutes. Analysis parameters were left as default and automatic analysis of each injection was selected. Before the start of each run a wash cycle, GS Wash Capillary and Block, was performed.

III.9.ii Second set of ten primers

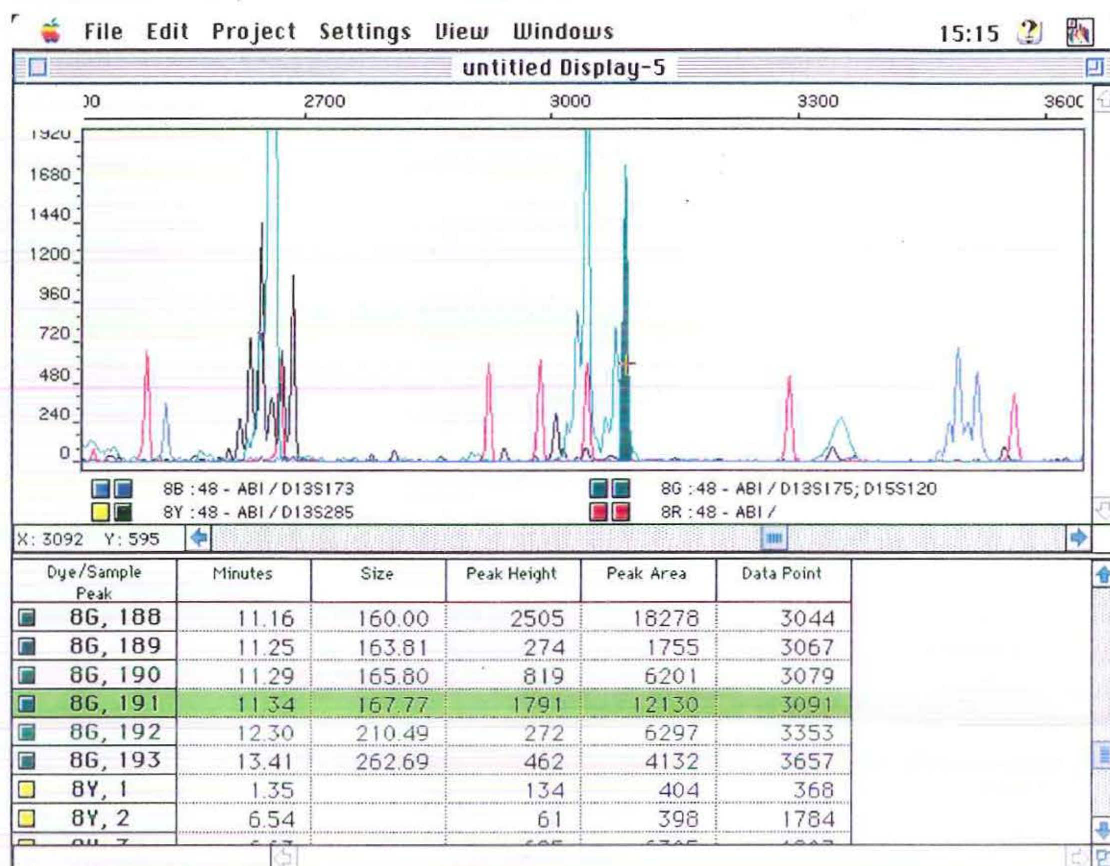
The procedures for preparing the ABI Prism 310 for the second group of primers differed from the initial group because, in the interim, the laboratory had adopted a new, simpler method of analysis. The changes involved the use of a glass syringe to pump Performance Optimised Polymer 4 (POP-4 [ABI]) into the capillary. Instead of 2% and 3% polymer solutions at the anode and cathode respectively, a 1x buffer solution (10-fold dilution of 10x Genetic Analyzer Buffer with EDTA in dH_20) replaced the polymers and the POP-4 was stored in the syringe. The capillary used with POP-4 is 47cm in total length, 36cm in length to the window, and 50μm in diameter. For all separations, the run module selected was GS STR POP4 C, injections were 10 seconds duration at 15kV, and runs lasted 18 minutes at 15kV and 60°C. The matrix file selected was GS STR POP4 C and the size standard file chosen was GS 500 TAMRA POP4 30cm/24min. Again, analysis parameters were left as default and each injection was automatically analysed. Unlike the previous method, this protocol did not require acid-base washing reagents.

III.10 Preparation of DNA samples for the ABI Prism 310 Genetic Analyzer

PCR products were diluted to a 1/20 concentration for markers labelled with FAM and TET, and a 1/10 concentration for markers labelled with HEX. One microlitre portions of the diluted PCR product were pipetted into 0.5ml sample tubes with 12μl of formamide (ABI) and 0.4μl of the standard, TAMRA. Samples were denatured for 2 minutes at 95°C, then chilled on ice before being placed in the ABI Prism 310. Up to four PCR products were included in a single sample tube. This procedure differed slightly for separations using the POP-4 polymer, where the samples required 0.5μl of TAMRA, and were denatured at 95°C for three minutes.

III.11 Allele assignment

The alleles were sized by reference to the standard, GS-500 TAMRA, which was loaded with the samples. The ABI Prism 310 detects the fluorescence emitted by the TAMRA and recognises the size distribution of the fragments, using this as a reference to size the other fragments. Sizes are estimated using GeneScan, with the alleles displayed as peaks and with a corresponding table listing the sizes of all peaks associated with the selected colour. Up to sixteen lanes can be simultaneously examined and the software allows the user to view one to four of the dye colours, so that colours can be examined individually. The individual peaks which represent the alleles can be selected using a mouse, and the corresponding value is highlighted in an adjoining table (see Figure III.3). The results from each run were stored on Cartridge Disks and later on Kodak Writable CD-ROM disks.



Dye/Sample Peak	Minutes	Size	Peak Height	Peak Area	Data Point
8G, 188	11.16	160.00	2505	18278	3044
8G, 189	11.25	163.81	274	1755	3067
8G, 190	11.29	165.80	819	6201	3079
8G, 191	11.34	167.77	1791	12130	3091
8G, 192	12.30	210.49	272	6297	3353
8G, 193	13.41	262.69	462	4132	3657
8Y, 1	1.35		134	404	368
8Y, 2	6.54		61	398	1784

Figure III.3: GeneScan screen image

III.12 Testing primers on the ABI Prism 310 Genetic Analyzer

III.12.i Initial ten primers

Primers were tested on the ABI Prism 310 using the PCR protocol for the Stanford Human Diversity primers described in section III.5.i, and with the conditions specified for the initial set of primers detailed in section III.9.i. The PCR products for D15S101, D15S97, D13S133 and D15S192 gave weak results, and so the concentration of the primers for these markers was increased by 50% in the reaction mixture (i.e., from 1.0μl to 1.5μl, with the volume of water reduced from 0.95μl to 0.45μl to compensate). On re-testing, each of the primers gave good results on the ABI Prism 310.

III.12.ii Second set of ten primers

For the second group of primers, testing was performed using the conditions specified in section III.9.ii. For the Stanford primers, PCRs were amplified using the protocol in section III.5.i. D13S144, D15S102, D15S100 and D15S169 displayed little or no signal, and so their concentration in the reaction mixture was increased by 50%. As D15S169 continued to amplify poorly, it was re-tested with varying concentrations of magnesium chloride ($MgCl_2$; 1.0mM, 1.5mM, 2.0mM and 2.5mM) in the 5x buffer mixture. A concentration of 1.0mM $MgCl_2$ in the buffer gave the best results.

The primers for D13S144 amplified particularly poorly when employed with the entire Khattar pedigree. Upon re-testing, the PCR products produced no results for any of the samples. A new working solution of the primers was prepared on two occasions, but neither produced any results. The lack of signal from the fluorescent marker led to the assumption that the primer had degraded. New primers were ordered and they produced a strong signal in both the agarose gel and on the ABI Prism 310.

The ABI primers were tested using the PCR protocol detailed in section IV.5.ii. All of the primers amplified well on the ABI Prism 310, and so no adjustments were made to the protocol.

III.13 Statistical analysis

III.13.i Allele frequency differences

To calculate the significance of the differences in allele frequencies, a χ^2 test was employed. The criteria for this test are that for each of the two groups being compared, there must be no alleles with an expected frequency of zero, and that at

least 75% of expected allele frequencies must be greater than five. If these criteria are not met, the differences in allele frequencies are significant. Since it is almost inevitable that alleles will not be shared between two groups, and that there may be a frequency of rare alleles (i.e., a frequency < 5), it is possible to compare common alleles with pooled rare alleles, or to group alleles of similar sizes. A significant difference is noted if the results of the χ^2 test calculations are not within the range 0.975>p>0.025, i.e., the 95% confidence interval.

III.13.ii Calculation of homozygosity

Observed homozygosity for each marker was calculated directly from the observed frequency of homozygotes for the marker:

$$Homozygosity = \frac{n_{Homozygous}}{n_{Total}}$$

Equation III.1: Observed homozygosity

where $n_{Homozygous}$ is the number of homozygous individuals and n_{Total} is the total number of individuals.

Expected levels of homozygosity were calculated from the observed allele frequencies, by the formula for predicting heterozygosity under random mating:

$$H = 1 - \sum_{i=1}^{n} p_i^2$$

Equation III.2: Expected homozygosity

where H is heterozygosity, p_i is the frequency of the ith allele, n is the total number of alleles (Murray, 1996). Therefore, H is the probability that a random individual is

heterozygous for any two alleles at a locus with allele frequencies p_i. Since homozygosity is equivalent to 1-H, $\sum p_i^2$ is the homozygosity (Hartl, 1988).

With inbreeding, the allele frequencies will remain the same, however the proportion of heterozygotes will be reduced and so a correction for inbreeding can be included in the equation (Hartl, 1988):

$$H = 1 - \sum_{i=1}^{n} p_i^2(1-F) + p_iF$$

Equation III.3: Expected homozygosity under inbreeding

In large, random mating populations the difference in observed and expected homozygosity would be measured using a χ^2 test, as a means of detecting deviation from Hardy-Weinberg equilibrium. Since this population did not fulfil these requirements, a t-test was used to compare the observed and expected homozygosity at the 5% level of significance.

III.13.iii Calculation of the coefficient of inbreeding

For each of the pedigrees, the mean coefficient of inbreeding was calculated as the average of all individual inbreeding coefficients:

$$F_E = \frac{\sum_{i=1}^{n} F_i}{n}$$

Equation III.4: Expected pedigree coefficient of inbreeding

where F_E is the expected calculated inbreeding coefficient, n is the number of individuals which were investigated, and F_i is the inbreeding coefficient of the ith

individual. Individual inbreeding coefficients were calculated based on parental relationships, that could be determined from the pedigrees.

The observed inbreeding coefficient (F_{IS}) was evaluated by:

$$F_{IS} = \frac{H_E - H_O}{H_E}$$

Equation III.5: Observed coefficient of inbreeding

where F_{IS} is the inbreeding coefficient, H_E is the average expected heterozygosity (including correction for the expected inbreeding coefficient, Equation III.3) and H_O is the average observed heterozygosity (Murray, 1996). The mean expected and observed inbreeding coefficients were compared at the 5% level of significance using a *t*-test.

IV. RESULTS

IV.1 Polymorphic markers

The polymorphic marker information was obtained from the Généthon and the Cooperative Human Linkage Centre (CHLC) microsatellite maps available through the World Wide Web (WWW; see Appendix VII.2). Généthon is an organisation that was initially developed to provide tools for the localisation and cloning of disease-determining genes, and it provides physical and genetic maps of the human genome. CHLC also aims to provide maps of the human genome but has a strong focus on using microsatellites that can be amplified by PCR and which display high heterozygosity.

Sex-averaged map distance information was obtained from both of these sources and used to construct approximate maps of the markers studied from chromosomes 13 and 15. Figures IV.1 and IV.2 show the chromosome maps with approximate distance in centiMorgans (cM) between markers. The distance covered was approximately 112cM for chromosome 13, and 145cM for chromosome 15, i.e., one marker each 10-15cM.

Figure IV.1: Microsatellite map distances for chromosome 13

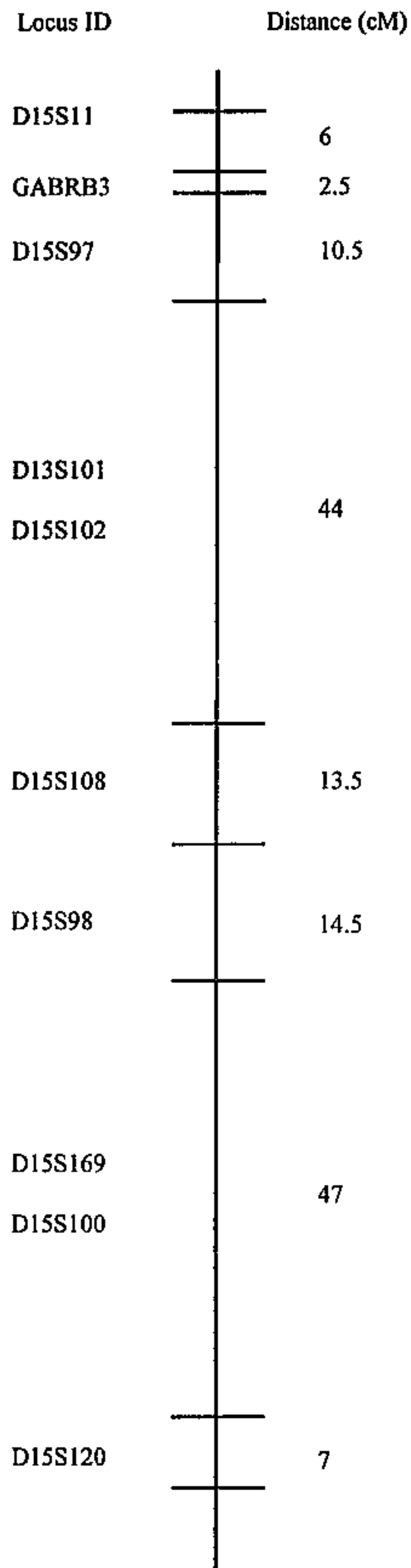


Figure IV.2: Microsatellite map distances for chromosome 15

IV.2 Reference populations

The reference populations chosen were obtained from the Centre d'Études du Polymorphisme Humain (CEPH) and the Genome Data Base (GDB) available on the WWW (see Appendix VII.2). CEPH is a data base that was established in 1984 to provide maps of the human genome, constructed using a reference panel of 61 large families. The GDB is the official repository for genome mapping data for the Human Genome initiative and it includes both CEPH data and results provided by other research laboratories.

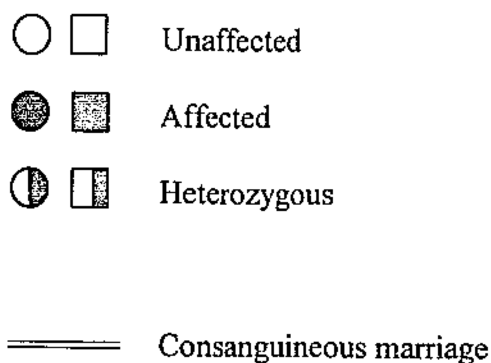
Eighteen of the twenty microsatellites chosen for this study were included in the CEPH data base, the exceptions being D13S133 and D13S192. The information available included average heterozygosity, gene name (where applicable), number of alleles and allele sets, which listed allele sizes and their frequencies. Allele sets were available for five of the chromosome 13 markers, and nine of the chromosome 15 markers.

The GDB contained information on all markers used in the study, listing the maximum heterozygosity, location, primary author, allele sets, populations tested and the number of chromosomes used. This information had not been updated since 1992/1993. The data for seven of the markers were partially or entirely calculated using CEPH families. For five of the microsatellites, D15S97, D15S98, D15S100, D15S101 and D15S102, the allele sets were not listed because the source (Hudson *et al.*, 1992), had used radio-isotope labels which did not give exact allele sizes.

IV.3 Variation in the alleles

As previously stated, allele scoring was performed with the aid of GeneScan. To verify that the alleles for each individual had been inherited in a Mendelian manner, the data in GeneScan were also analysed by GenoTyper (Perkin Elmer, 1995), a program which permits the user to examine each marker in a specific family and includes a function which can check the inheritance of alleles. If an individual did not appear to have inherited one allele from each parent, their sample was reamplified and the products once again analysed by the ABI Prism 310. In all families Mendelian segregation of alleles was observed at all loci.

Figures IV.3 to IV.6 show the genotype and haplotype bars for each chromosome and each *bradari*. For all pedigrees, the disease condition is β-thalassaemia and the symbol definitions are as follows:



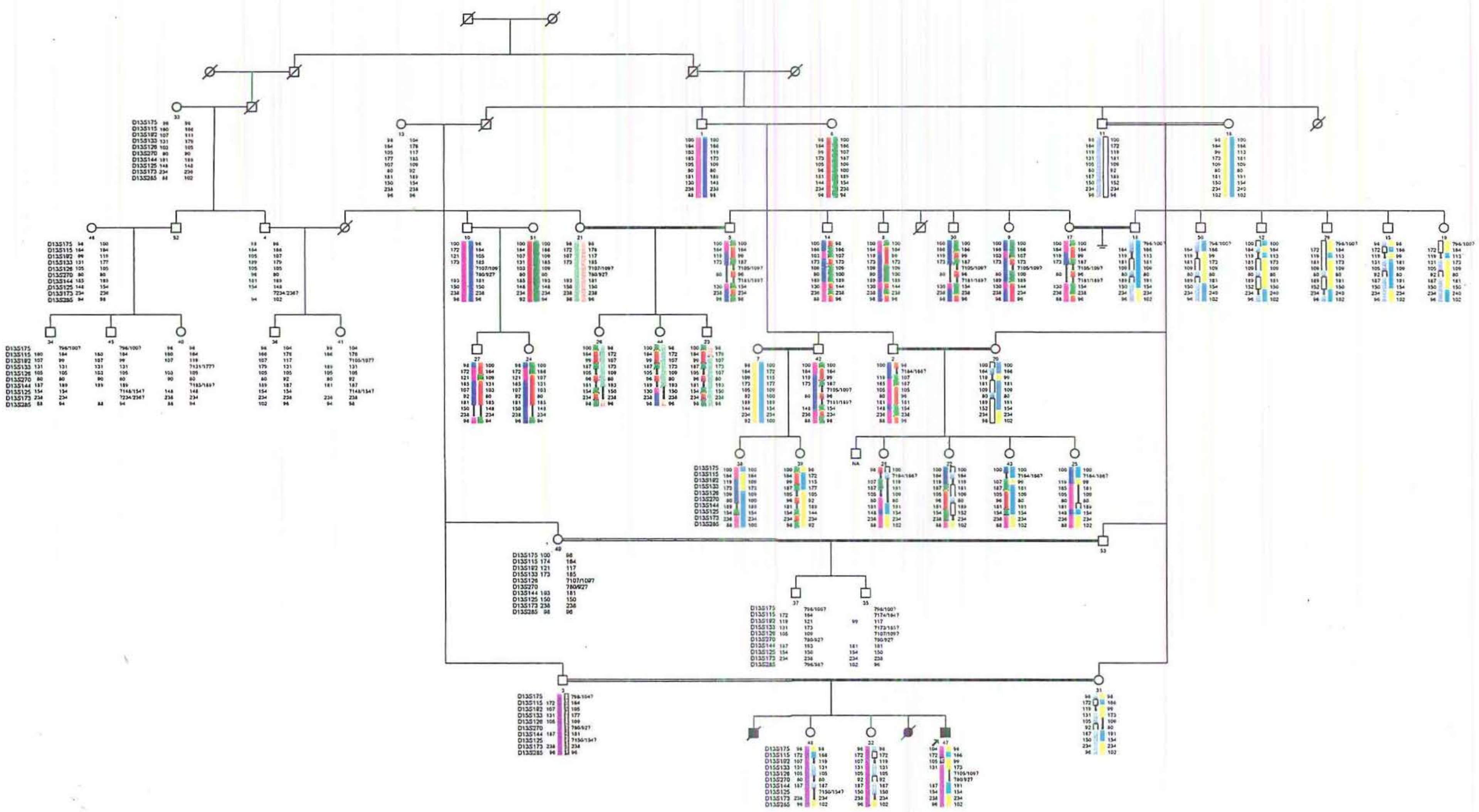


Figure IV.3: Khattar *bradari* pedigree with genotypes and haplotypes for markers on chromosome 13

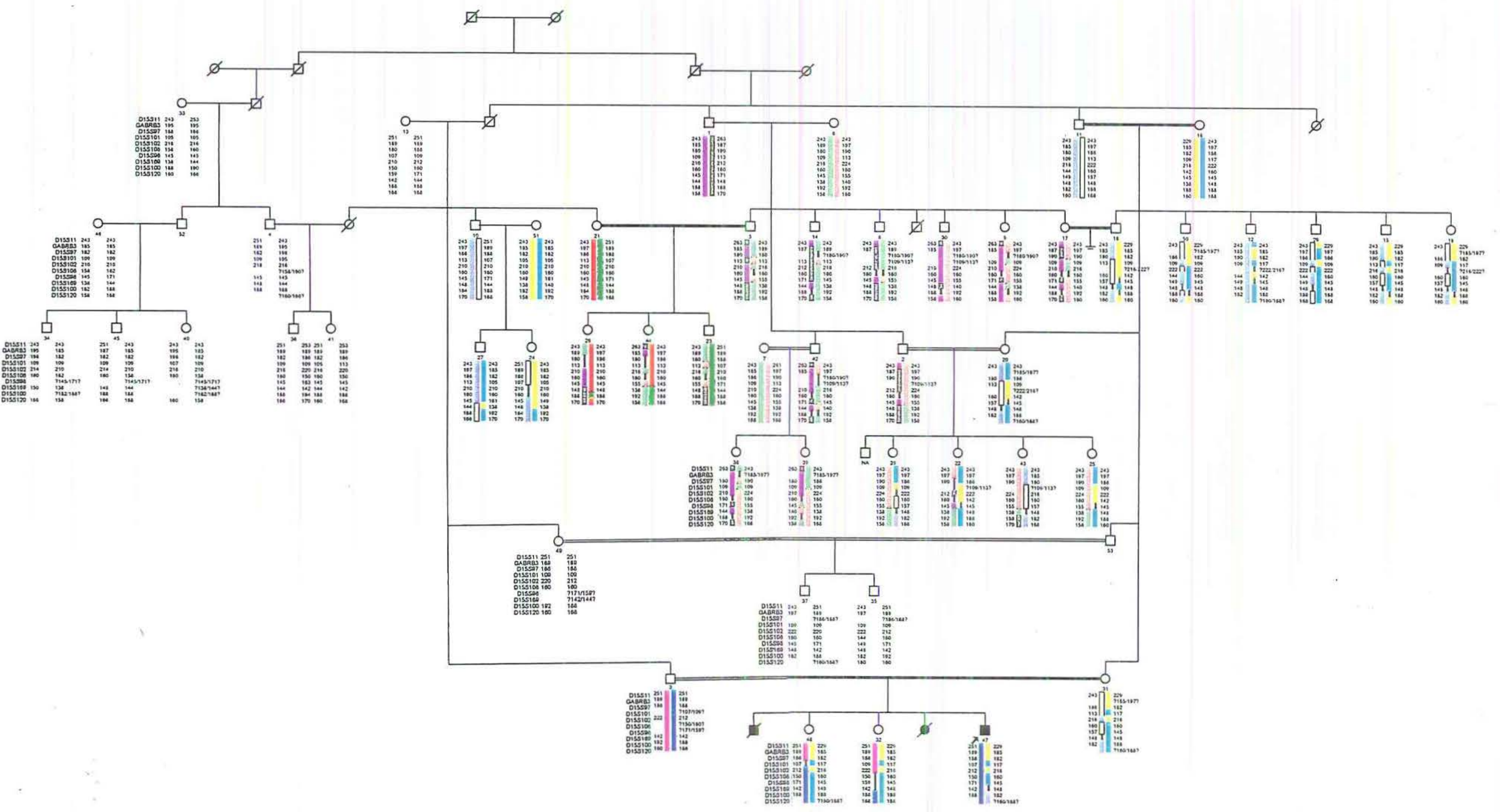


Figure IV.4: Khattar *bradari* pedigree with genotypes and haplotypes for markers on chromosome 15

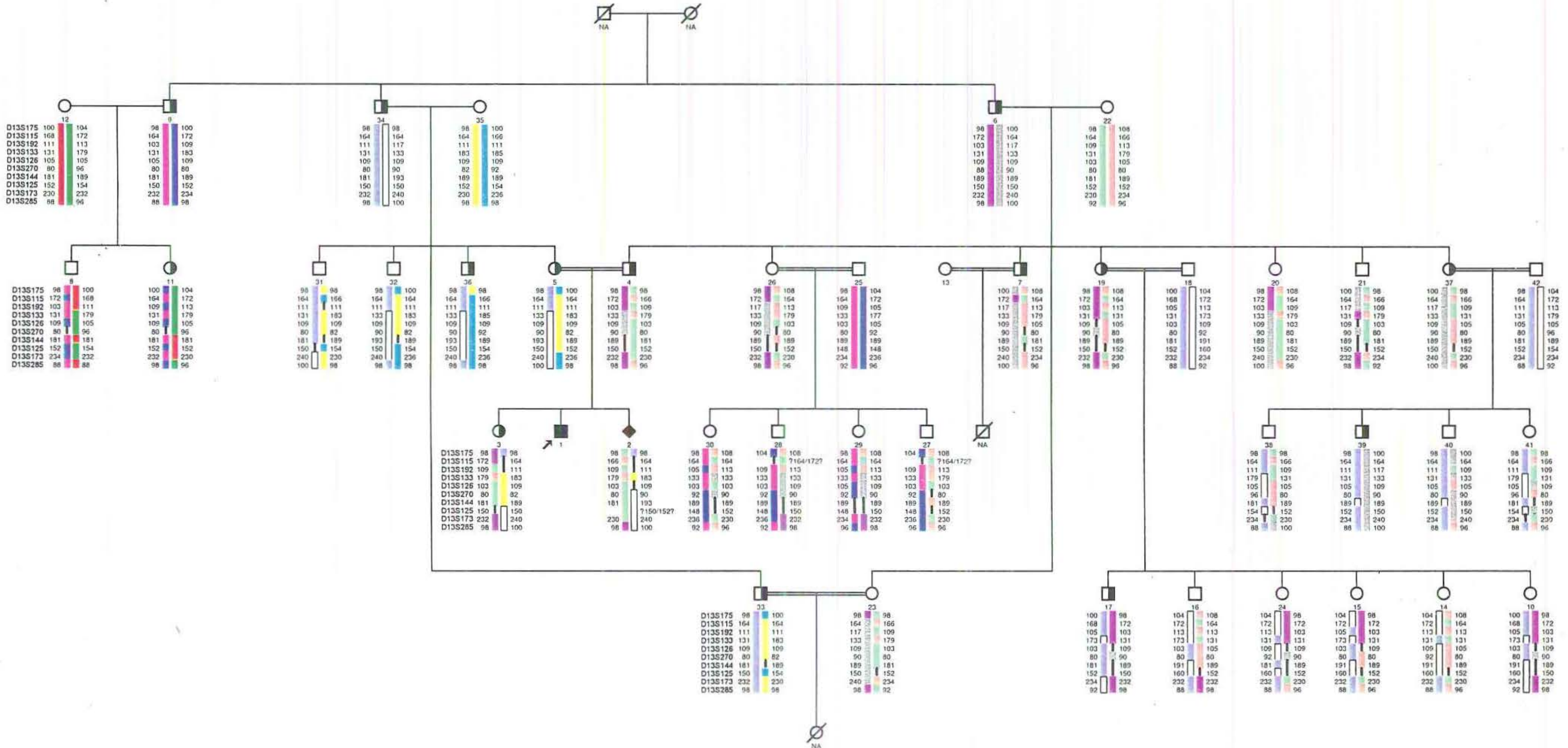


Figure IV.5: Rajpoot *bradari* pedigree with genotypes and haplotypes for markers on chromosome 13

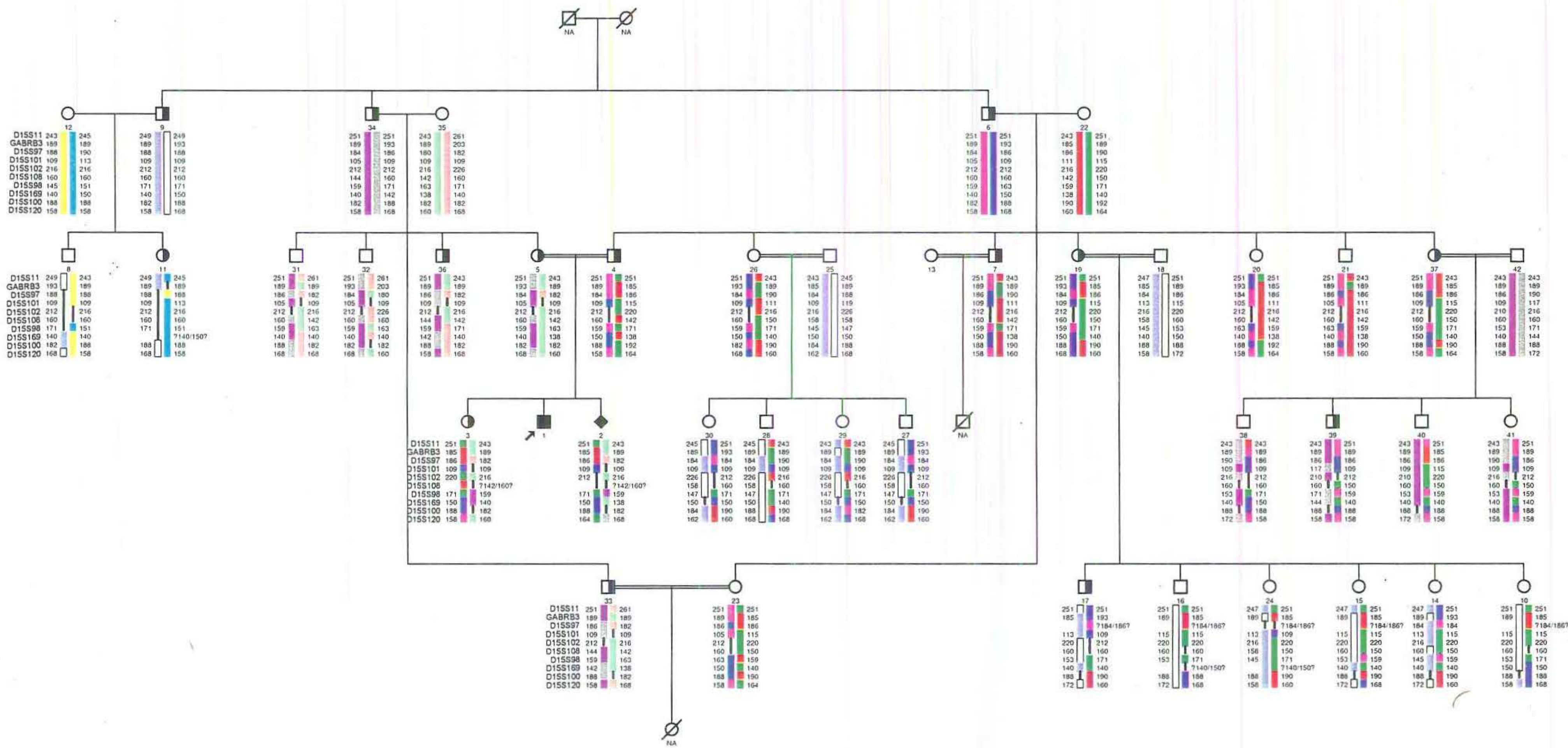


Figure IV.6: Rajpoot *bradari* pedigree with genotypes and haplotypes for markers on chromosome 15

IV.3.i Allele size variation

In general, the observed data displayed similar allele sizes to the reference groups, however there were several instances where allele size ranges in the published data did not match, for example at D13S173, D15S102, D15S108 and D15S100 (see Table IV.1).

For some markers, alleles were identified which had not been previously described by the GDB or CEPH. In general, the new allele was within the ranges described in the published data and, without access to allele sets for all of the published data, it was difficult to determine with confidence whether the allele had not previously been described. However, in one instance (D15S11, allele size 229bp), the allele was not described in any of the published allele sets, nor was it within any of the published ranges.

Table IV.1: Allele size ranges in base pairs

Locus	Khattar	Rajpoot	Total	GDB	CEPH
D13S175	98-104	98-108	98-108	101-113	105-115
D13S115	160-176	164-172	160-176	161-175	161-175
D13S192	99-121	103-117	99-121	88-124	na
D13S133	131-189	131-185	131-189	130-187	na
D13S126	103-109	103-109	103-109	100-112	na
D13S270	80-100	80-96	80-100	79-99	81-95
D13S144	181-193	181-193	181-193	183-199	na
D13S125	130-154	148-160	130-160	131-155	na
D13S173	**234-240**	**230-240**	**230-240**	**166-178**	**166-172**
D13S285	88-102	88-100	88-102	92-106	92-106
D15S11	229-263	243-261	229-263	243-263	238-260
GABRB3	185-197	185-203	185-203	181-201	na
D15S97	180-198	180-190	180-198	159-	168-186
D15S101	105-119	105-119	105-119	104-	110-134
D15S102	**210-224**	**210-226**	**210-226**	**217-**	**98-114**
D15S108	**142-162**	**142-160**	**142-162**	**185-205**	**141-161**
D15S98	145-171	145-171	145-171	152-	141-175
D15S169	138-150	138-150	138-150	142-158	142-158
D15S100	**182-192**	**182-192**	**182-192**	**183-**	**119-131**
D15S120	158-170	158-172	158-172	150-174	150-174

Note: Ranges in bold indicate results which are non-identical in the various reference groups. A value listed as na indicates that no size range was available.

IV.3.ii Number of alleles

The number of alleles which were identified ranged from 3 (D13S175) to 11 (D13S192) among the Khattar samples and from 3 (D13S126) to 7 (D13S133, D15S101, D15S98) in the Rajpoot samples. The mean number of alleles per locus on each chromosome is displayed in Figure IV.7.

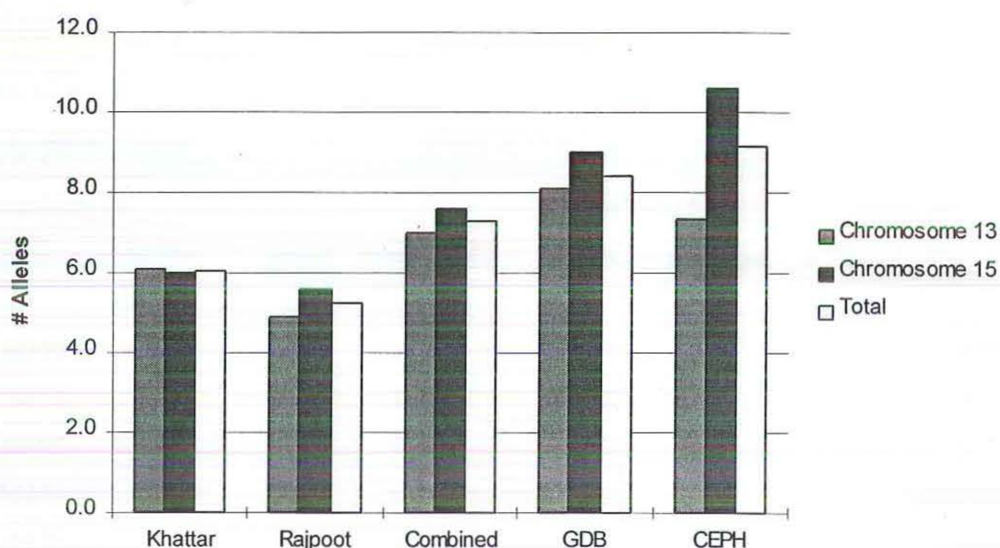


Figure IV.7: The mean number of alleles for all loci on each chromosome and for both chromosomes combined

A paired *t*-test was used to determine if the difference in the number of alleles between each of the *bradaris* was statistically significant. At the 5% level, the test indicated that the difference was significant overall but it did not attain significance for the individual chromosomes. The total number of alleles described for both *bradaris* was combined for comparison with the data from the GDB and CEPH. The results indicated that the number of alleles observed in the combined *bradaris* was significantly fewer than in both of the published data sets. It must be

noted that the number of loci studied differed in the two reference groups employed, and this was taken into consideration when the data were compared.

IV.3.iii Allele sharing

The proportion of alleles shared by the *bradaris* ranged from 0.29 (GABRB3) to 1 (all alleles shared; D15S100). Table IV.2 lists each locus with the total number of alleles and the number shared for that locus.

Table IV.2: Alleles shared between the two *bradaris*

Chromosome 13				Chromosome 15			
Locus	# Alleles	# Shared	Proportion	Locus	# Alleles	# Shared	Proportion
D13S175	4	3	**0.75**	D15S11	9	3	0.33
D13S115	7	3	**0.43**	GABRB3	7	2	**0.29**
D13S192	11	6	0.55	D15S97	7	5	0.71
D13S133	11	5	0.45	D15S101	8	4	0.50
D13S126	4	3	**0.75**	D15S102	9	4	0.44
D13S270	7	4	0.57	D15S108	6	5	0.83
D13S144	6	4	0.67	D15S98	11	4	0.36
D13S125	7	4	0.57	D15S169	6	5	0.83
D13S173	6	3	0.50	D15S100	5	5	**1.00**
D13S285	7	5	0.71	D15S120	8	3	0.38

Note: Proportions in bold represent the lowest and highest for each chromosome.

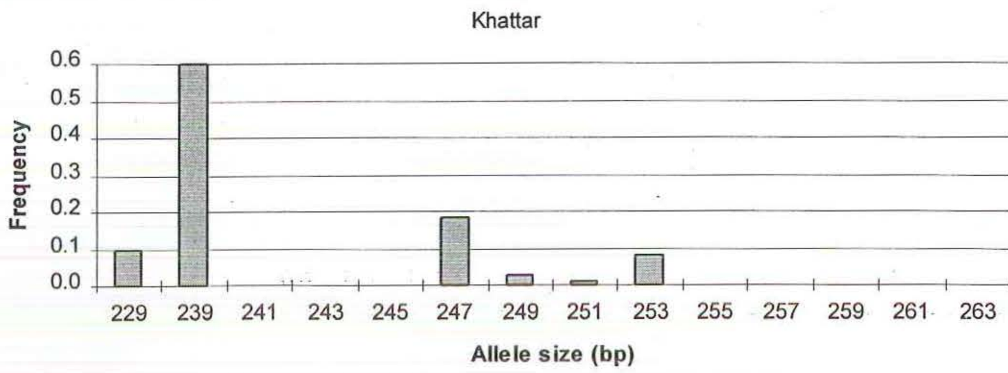
Alleles which were shared by the *bradaris* were observed with a greater frequency than alleles that were not shared, with the overall mean frequency of shared alleles, 0.204, and unshared alleles, 0.055. The most frequently observed allele at a particular locus was found in both *bradaris*, and for eighteen loci the most frequently observed allele in each *bradari* was observed in both pedigrees. The exceptions were at D15S169, where the most common allele identified in the Khattar

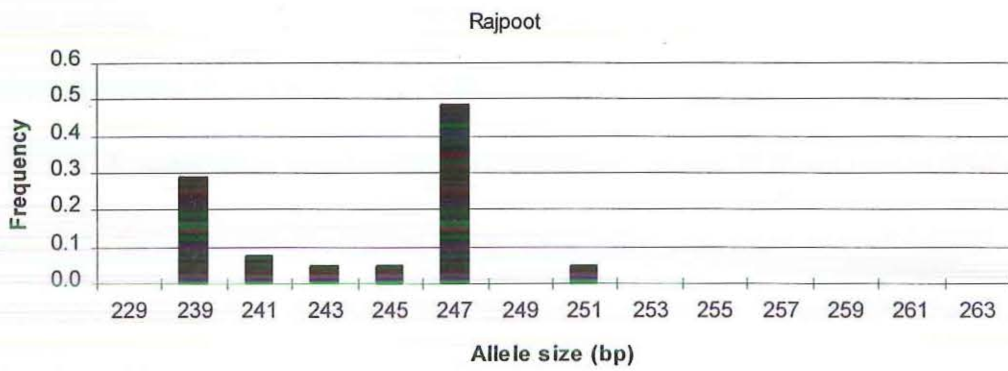
was not observed in the Rajpoot, and at D13S173 where the most common allele in the Rajpoot was not identified in the Khattar.

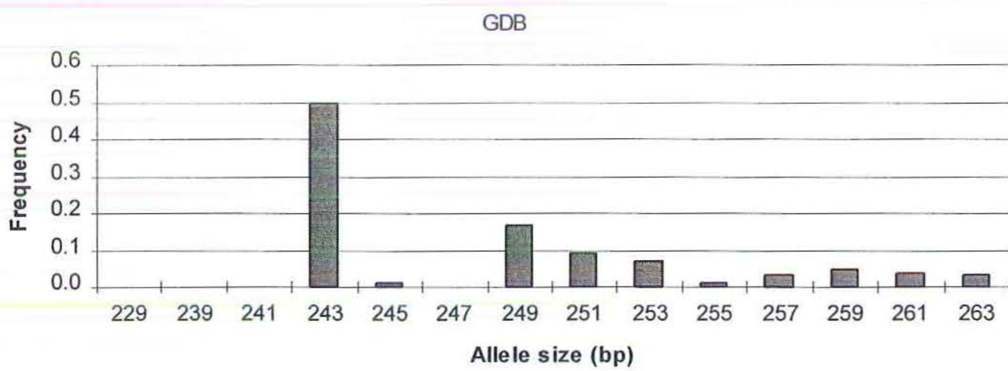
IV.3.iv Allele frequencies

The frequency of alleles ranged from 0 (unobserved) to 0.75, and the most frequent allele for each locus had a frequency of at least 0.25. Allele frequencies were compared as common versus pooled rare alleles or, where necessary, rare alleles were included with alleles of similar size. The χ^2 test indicated that allele frequencies in the two *bradaris* were significantly different at thirteen loci. Figure IV.8 shows the allele frequencies for D15S11, where the allele frequencies were found to be significantly different between the two *bradaris* and the GDB and CEPH data. Specific allele frequency data and results of the χ^2 tests for other loci are listed in Appendix VII.3.

Allele frequencies for each family and for the combined *bradari* data were compared with the GDB and CEPH data, although this could only be accomplished for those markers which had allele sets listed or had similar size ranges. The χ^2 tests indicated that allele frequencies were significantly different at all loci between the *bradaris* and both databases, at the 5% level (data not shown).







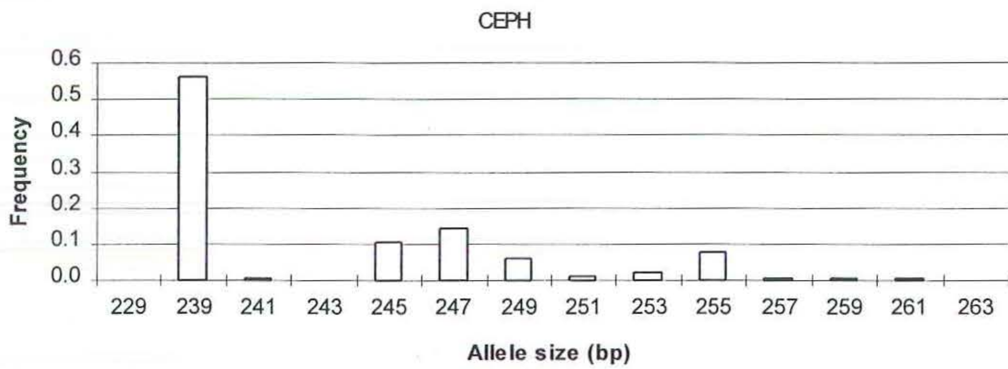


Figure IV.8: Allele frequencies for D15S11

IV.4 Level of variation in homozygosity of loci

IV.4.i *Bradari* variation

Homozygosity varied considerably both within and between the two *bradaris*, with upper and lower limits of 4% (D13S192) and 55% (D15S108) in the Khattar, and 3% (D15S98) and 40% (D15S101) in the Rajpoot. To determine if the levels of homozygosity between the two *bradaris* were statistically different, the data were compared, locus by locus, using a paired *t*-test. No significant difference was observed at the 5% level (Figure IV.9).

A correlation coefficient was calculated to see if there was a relationship between the number of alleles at a locus and observed homozygosity. The correlation coefficient indicated that there was an inverse relationship between the two factors for both chromosomes: r(chr13) = -0.78 and r(chr15) = -0.52.

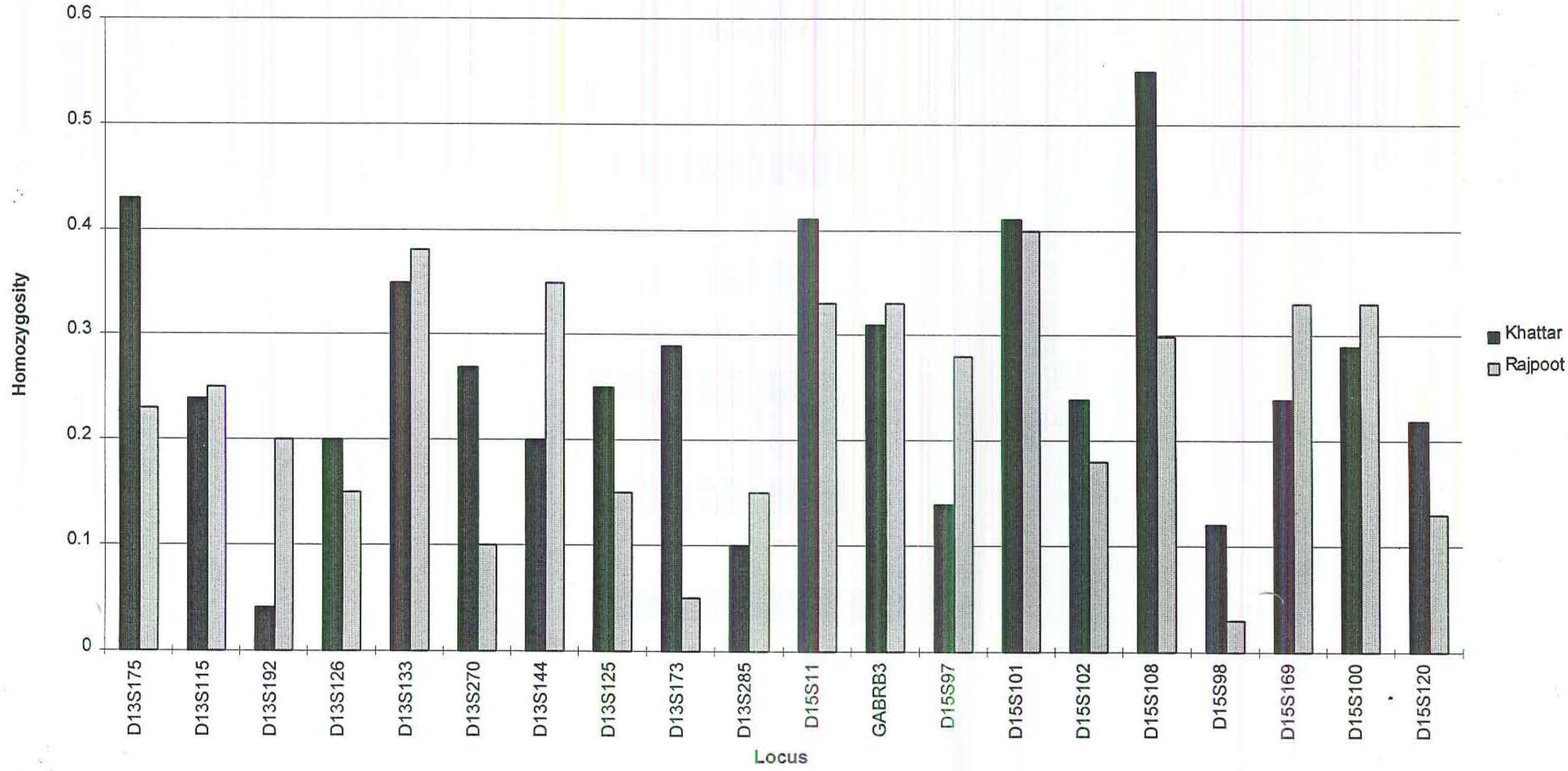


Figure IV.9: Observed values of homozygosity in the Khattar and Rajpoot *bradaris*

IV.4.ii Variation between the *bradaris* and reference populations

The observed levels of homozygosity in the *bradaris* were compared with homozygosity in the published data using a paired *t*-test. At the 5% level, the homozygosity in the Khattar was significantly greater than the CEPH for chromosome 15 (t = 2.489, t-critical = 2.262), and homozygosity in the Rajpoot was significantly less than CEPH for chromosome 13 (t = -4.810, t-critical = 2.262). No significant difference in homozygosity was observed between the *bradaris* and the GDB. Mean homozygosity for each *bradari*, the combined *bradaris* and the published data are shown in Figure IV.10.

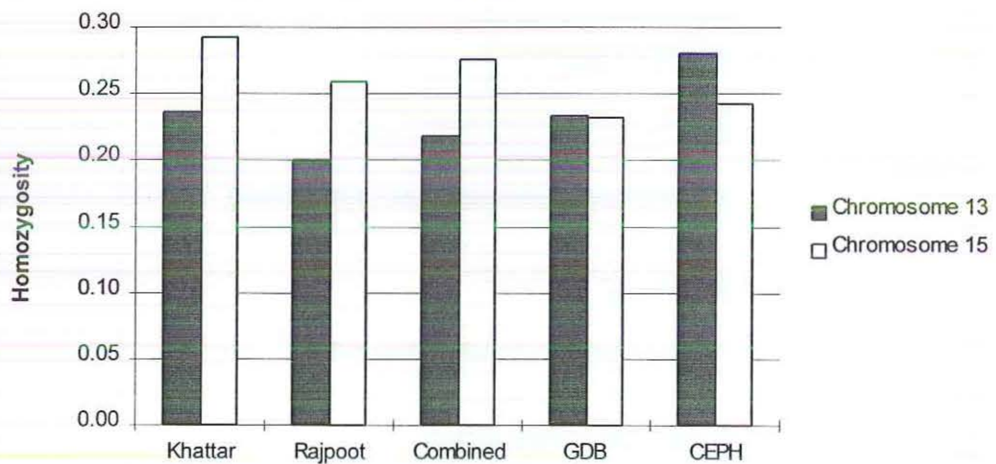


Figure IV.10: Mean homozygosity levels

Note: CEPH does not include D13S133 and D13S192.

IV.4.iii Expected homozygosity

Expected homozygosity was initially calculated using Equation III.2. In general, expected homozygosity exceeded the observed homozygosity, differing by as much as 15% in the Khattar and 22% in the Rajpoot (Figures IV.11 and IV.12).

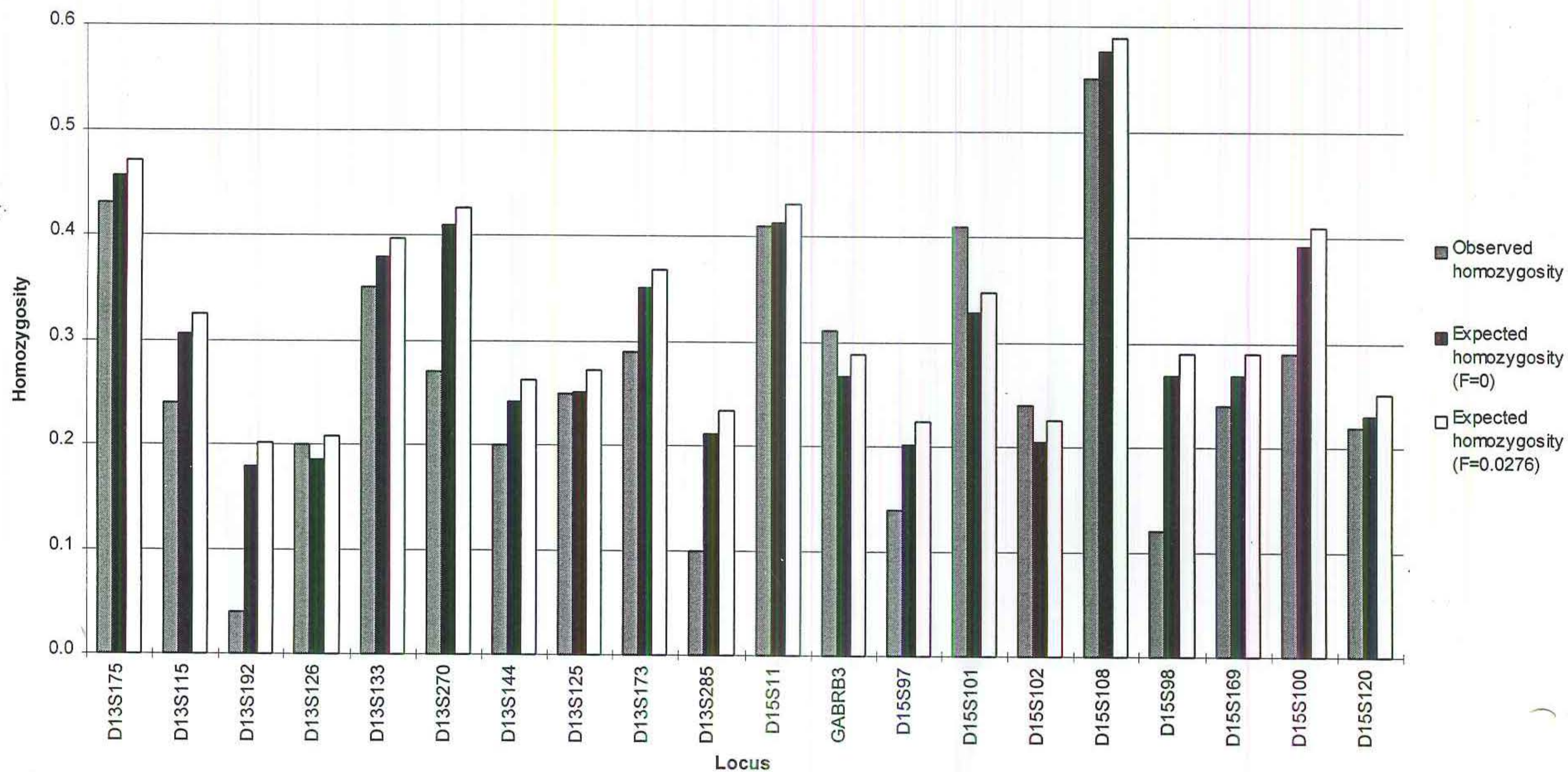


Figure IV.11: Observed and expected values of homozygosity for the Khattar pedigree

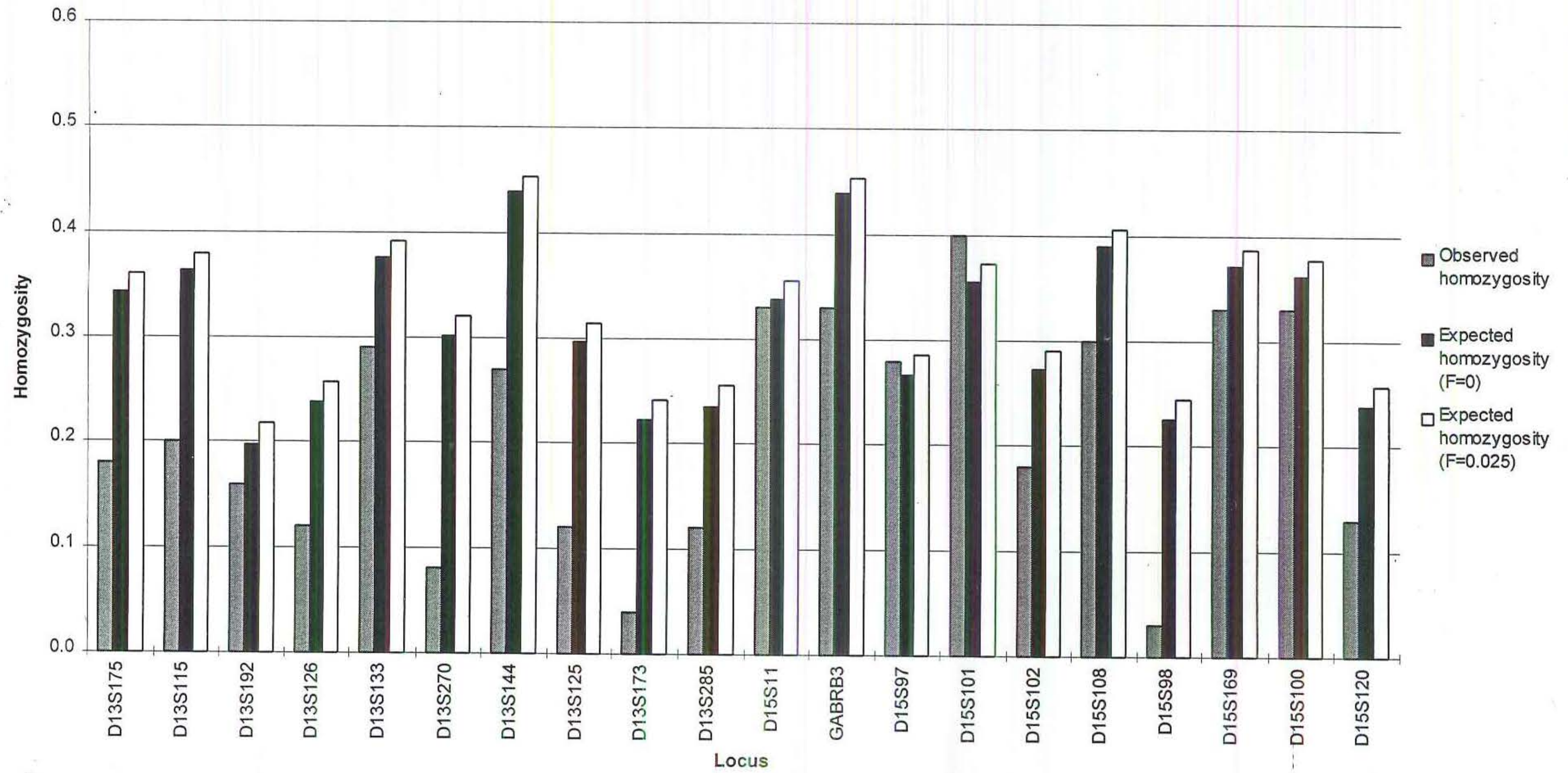


Figure IV.12: Observed and expected values of homozygosity for the Rajpoot pedigree

Observed homozygosity was compared to expected homozygosity using a paired t-test. At the 5% level of significance, a difference between the expected and observed levels of homozygosity was noted in both the Rajpoot and the Khattar pedigrees. However, when the data from both pedigrees were pooled, there was no significant difference in observed and expected homozygosity (Table IV.3.i).

Expected homozygosity was also calculated with a correction for the inbreeding coefficient, using Equation III.3. Again, expected homozygosity generally exceeded observed homozygosity, with a maximum difference of 17% in the Khattar, and 24% in the Rajpoot. A paired t-test indicated that expected homozygosity was statistically greater than observed homozygosity, at the 5% level of significance (Table IV.3.ii).

Table IV.3: Values of the t-statistic for expected versus observed homozygosity where F=0 (i) and with a correction for F (ii)

i

Chromosome	Khattar	Rajpoot	Combined	df	t-critical range
13	3.608*	4.792*	2.2624*	9	$-2.262 < t < 2.262$
15	1.106	2.745*	-0.342	9	$-2.262 < t < 2.262$
Total	3.027*	5.198*	1.225	19	$-2.093 < t < 2.093$

ii

Chromosome	Khattar	Rajpoot	Combined	df	t-critical range
13	4.619*	9.516*	4.026*	9	-2.262 - 2.262
15	1.889	3.4998*	0.572	9	-2.262 - 2.262
Total	4.244*	7.21*	2.692*	19	-2.093 - 2.093

* the value is significant at the 5% level

IV.5 Homozygosity in the children of first cousin and non-consanguineous marriages

IV.5.i Observed homozygosity

The overall increase in the level of homozygosity in the children of first cousin marriages compared with non-consanguineous marriages was 0.0539, i.e., close to the expected 0.0625. In the Khattar, the mean increase in homozygosity in the first cousin progeny was 0.0593 and in the Rajpoot it was 0.0463 (Figure IV.13).

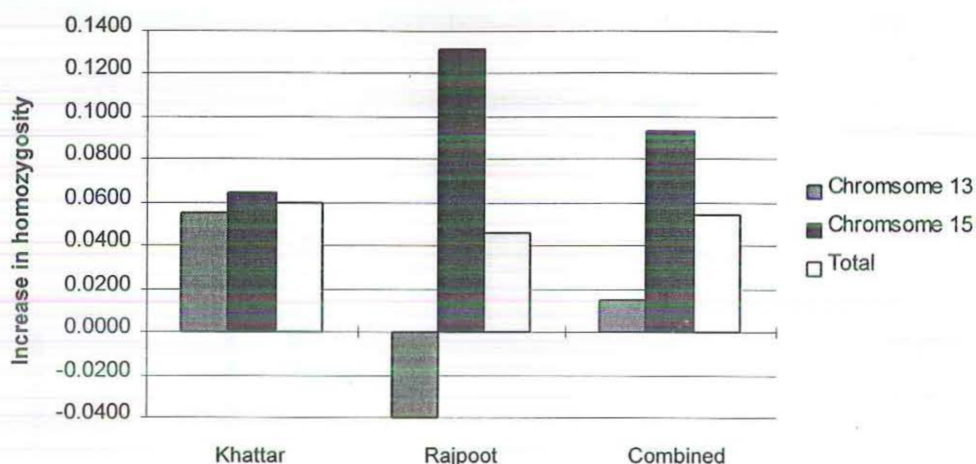


Figure IV.13: Mean differences in homozygosity between children of first cousin and non-consanguineous marriages

A paired *t*-test was used to determine if the increase in homozygosity in the children of first cousin marriages was statistically significant, and indicated that homozygosity was significantly greater in the Rajpoot for chromosome 15 at the 5% level. A *t*-test was also used to compare the increase in homozygosity observed in the two chromosomes. While there was a marked increase in homozygosity on chromosome 15 compared to chromosome 13 this difference was not shown to be statistically significant.

IV.5.ii Expected versus observed homozygosity

The expected homozygosity for each locus was calculated for the children of first cousin (F=0.0625) and non-consanguineous marriages (F=0) using Equation III.3. Expected homozygosity was generally greater than the observed homozygosity, and the two were compared using a paired t-test. At the 5% level, the difference was significant in each pedigree for both types of marriage (data not shown).

IV.5.iii Individual homozygosity

A comparison was also made between the number of homozygous loci in the children of first cousin and non-consanguineous marriages (Figure IV.14).

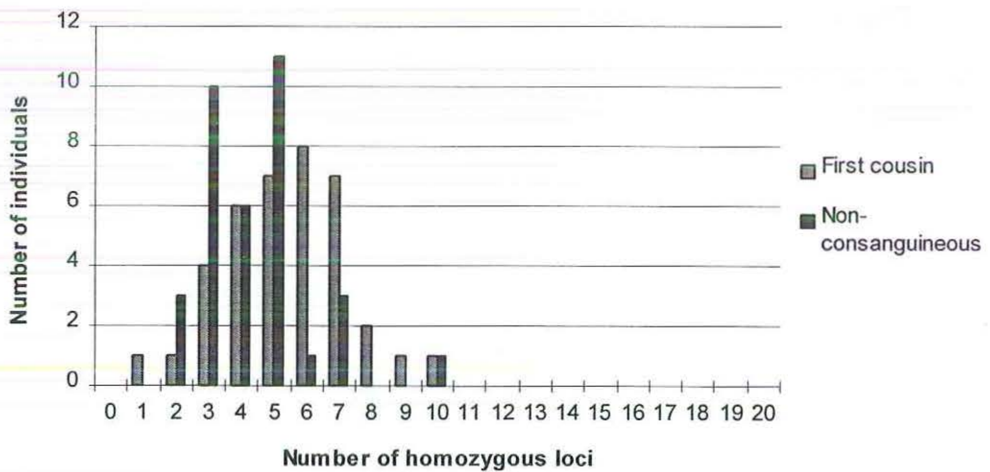


Figure IV.14: The distribution of homozygous loci in the children of first cousin and non-consanguineous marriages

The distribution of homozygous loci in the children of first cousin marriages followed the pattern of a normal distribution with a near bell-shaped curve. Individuals possessed a range of 1 - 10 homozygous loci, with a modal value of 6

homozygous loci. The non-consanguineous progeny exhibited a mixed distribution pattern, with most having 5 homozygous loci.

IV.6 The observed and expected coefficients of inbreeding

For each pedigree, the expected coefficient of inbreeding (F_E) was calculated by Equation III.4. In the Khattar, F_E = 0.0276 for the 51 individuals studied, while in the Rajpoot, F_E = 0.025 for the 40 individuals studied. All consanguineous marriages were first cousin unions, so for the children of these marriages F_E = 0.0625, while F_E = 0 for children of non-consanguineous marriages.

The observed inbreeding coefficient (F_O) was evaluated by Equation III.5. In all cases, the values of the observed inbreeding coefficient were negative and therefore quite different from the expected values (Table IV.4). The negative values of the observed inbreeding coefficient can be attributed to the overall reduction in observed homozygosity compared with expected homozygosity.

Table IV.4: Expected inbreeding coefficients

	Mean expected inbreeding coefficient	Mean observed inbreeding coefficient
Khattar	0.0276	-0.0602
Rajpoot	0.0250	-0.1163
First cousin	0.0625	-0.08787
Non-consanguineous	0.0000	-0.06826

V. DISCUSSION

V.1 Suitability of reference populations

It was reasonable to compare the Pakistani populations with data from predominantly European populations as it is likely that the polymorphisms shared by one group are also present in another. This sharing is expected since the measurement of genetic variability between human populations, F_{ST}, is approximately 0.1, indicating that almost 90% of human genome variability is common to all human sub-populations (Schmitt, 1997). However, the information obtained from the GDB and CEPH are derived from a number of different research groups and, in some instances, has been determined using sample populations which did not contain the same specific individuals. Furthermore, the data may not have been drawn from a randomly selected group, as the individuals usually have been investigated because of specific biological or medical characteristics (Li and Sadler, 1991). Unfortunately, time limitations did not permit multivariate analysis of the observed results to accommodate these variables.

Use of ethnically different sample populations can produce quite variant information about a marker, in particular the information content of a marker. For example, the heterozygosity of a marker must be at least 0.7 for a forensic or linkage study to be cost effective (Ott, 1992). In the present study, the GDB was the first avenue for evaluating the level of heterozygosity that would be expected for a marker, and with the exception of four loci, D13S115, D13S126, D13S125 and D15S108, this value was greater than 0.7. However, the observed levels of heterozygosity for D13S115 and D13S125 were higher than the values published in the GDB and CEPH, and were also greater than 70%. Thus, on the basis of these

observations, the two markers would be considered suitable for linkage or forensic studies.

Despite these minor caveats, it is appropriate to use the GDB and CEPH as reference populations since they have been specifically compiled to provide an indication of the degree of genetic diversity that might be expected in a population, and thus serve as a starting point for genetic studies. Even in linkage studies, these databases are used as reference material, although it appears probable that the allele frequencies observed in the pedigrees under study will differ from the published allele frequencies (Kruglyak, Daly and Lander, 1995).

V.2 Success of PCR amplification and fluorescent detection of alleles

The PCR protocol developed by Kalaydjieva and Tolún required modification for some of the markers, in particular necessitating an increase in the concentration of primers in the reaction mixture. It had been several months since the primers had been tested and their conditions optimised, so it is probable that they had degraded to varying degrees during this time. Degradation of primers is inevitable, as it proceeds at a rate of approximately 0.1% per month because of natural breakdown, but mildly degraded primers are readily accommodated in the PCR reaction mixture. The ease with which the PCR protocol could be modified to suit the degrading primers demonstrates the flexibility of both microsatellite primers and PCR.

After modification of the protocol, PCR amplification of the markers was successful for all microsatellites studied, although it was necessary to rescan some PCR products because the instrument detected little or no signal. Poor detection of alleles by the ABI Prism 310 may have been due to dilution of PCR products in the

multiplex samples, or have been caused by experimental problems such as pipetting errors.

V.3 Evaluation of dinucleotide markers

The effectiveness of the dinucleotide repeat microsatellites used in this study has been confounded by several factors. The most problematic factor was the lack of continuity in allele sizes between published data and the results obtained with the *bradari* DNA samples. Unfortunately, it was not possible to screen samples from the published data to ensure that the allele sizes allocated to DNA from the *bradaris* were in agreement with these sources. Had this opportunity been available, the information could have been used to scale the results from different populations so that allele sizes in the GDB, CEPH and the *bradaris* were congruent with one another (Morell *et al*., 1995). Large differences in allele size could be attributed to the use of differing microsatellite primers. If the primer sequence is longer, then amplification will result in a PCR product with a longer final length. Inconsistent allele sizing can also result from the use of different fluorescent label systems, such as using primers which have been labelled during synthesis, as opposed to incorporating the dyes into the PCR product during amplification.

The allele sizes also occasionally differed by a single base pair. For example, at D13S175 the GDB published sizes begin at 99, but the observed sizes begin at 98. While this can in part be the result of subjective scoring of alleles by the researcher, it can also be due to some measure of incompatibility between the fluorescent detection systems that are employed. The mobility of the DNA fragments may be inconsistent between instruments, resulting in different size allocations for the same allele (Perlin *et al*, 1994). For example, in the Centre for Human Genetics

it was observed that the same PCR product analysed using an ABI Prism 310 and an ABI Prism 373 DNA Sequencer would not always be assigned the same fragment size.

Changes in size between dinucleotide alleles of approximately 2bp did produce some confusion when trying to score alleles. This problem would be overcome by the use of tri-, tetra- or penta-nucleotides, which would produce less ambiguity with scoring alleles because of the greater distinction in fragment size. However, compared to dinucleotides, larger repeats display higher mutation rates and increased selection on repeat length. Hence, different populations display greater similarity in their allele frequencies in these longer nucleotide repeat microsatellites (Wall, Williamson, Petrou, Papiaioannou and Parkin, 1993; Morell *et al.*, 1995). In summary, the differences observed in this study between the two *bradaris,* and between the *bradaris* and the GDB and CEPH, may not have been so great using microsatellites with larger nucleotide repeats. However, it must be remembered that dinucleotides do offer the benefit that they occur more frequently and they are distributed evenly throughout the genome. They also have been tested more extensively than the other types of microsatellite repeat (Moore *et al.*, 1991).

V.4 Variation in the alleles

In general, there were fewer alleles observed in the *bradaris* than in the GDB and CEPH data (refer to Figure IV.7). This finding was expected, since *bradaris* are endogamous communities and therefore they would have a smaller gene pool than a larger, randomly mating population. This factor is further supported by the observation that, for some markers, the reference data (particularly the GDB)

were derived from fewer individuals but displayed greater polymorphism than the equivalent results obtained from the *bradari* members.

Allele sharing among the *bradaris* was quite low at some loci, and allele frequency distributions observed in the Khattar and Rajpoot displayed significant differences at thirteen of the twenty loci. Loci with high proportions of shared alleles in the two *bradaris* tended not to display differences in allele frequencies. The observed allele frequencies indicated that the two Punjabi communities shared greater similarity with each other than with the non-Punjabi reference populations. However, because the *bradaris* were found to be significantly different at more than half the loci studied, this may indicate that they are not as genetically similar as would be expected, given that they are from the same region. This finding may indicate that the *bradaris* have diverged from each other, which could be a result of inbreeding. Alternatively, they may not share the same founding population, which is possible given the history of migration in the region (see section II.1).

The appearance of a previously undescribed allele in the Khattar pedigree (229bp at D15S11), present with a frequency of 0.1, was of interest. The allele was introduced into the *bradari* by individual 16 who married her first cousin, individual 11. The 229bp allele was inherited by six of their seven children, one of whom (31) in turn had passed it on to three of her children. As this allele was not described in any published data, it may be the result of a local mutation. Thompson and Neel (1997) have suggested that private polymorphisms will be observed in relatively undisturbed tribal populations, and this new allele may be an example of such a polymorphism. The high frequency with which the allele has been inherited also demonstrates how quickly the gene pool of a genetically isolated population can be altered by admixture. Furthermore, there appears to be selection in favour of the

heterozygous genotype at this locus. The father's genotype for this marker was 243/243, while the mother was 229/243, so possible genotypes for the offspring are 229/243 or 243/243, but in 6 out of 7 cases the heterozygous genotype was observed. Selection against homozygotes will be discussed in more detail in section V.5.

V.5 Homozygosity and the coefficient of inbreeding

The coefficient of inbreeding provides a mathematical measure of the level of excess homozygosity that results from consanguineous unions. In the pedigrees investigated in this study, the calculated inbreeding coefficient was $F = 0.0276$ in the Khattar and $F = 0.0250$ in the Rajpoot. On the basis of these calculations, which were derived from the five (Khattar) and four (Rajpoot) generation pedigrees available, it was predicted that the level of homozygosity in the *bradaris* would not be significantly different, although the Khattar could be expected to display greater homozygosity. The results of the microsatellite analysis indicated that homozygosity was not significantly greater in the Khattar than in the Rajpoot, although the difference was greater than would be predicted from the pedigrees.

It was also predicted that the observed levels of homozygosity in the *bradaris* would be higher than both the expected homozygosity and the levels of homozygosity in the published data sets. In fact, the reverse proved to be true as the expected homozygosity was actually higher than observed homozygosity, attaining statistical significance for both *bradaris*. Furthermore, the *bradaris* did not display a greater degree of homozygosity than the reference data, as there were a greater, though not statistically significant, number of homozygotes in the CEPH database.

This finding was unexpected, not merely because the CEPH families are not inbred, but also because the number of alleles observed for the CEPH data were

higher than in the Pakistani families. If the number of alleles is inversely proportional to the relative proportion of homozygotes for a locus, then it would have been expected that the CEPH data would have revealed less homozygosity than the *bradari* pedigrees.

All of these observations indicate that the level of observed homozygosity in the *bradaris* is reduced by comparison with the values that would be expected in an inbred community and in the non-inbred reference populations. The most likely explanation would appear to be some selectional mechanism operating against the homozygous genotypes.

In protein and enzyme studies it has previously been proposed that environmental factors can act to positively influence the selection of a heterozygote, as the hybrid form of a gene may confer beneficial intermediary properties not present in either homozygote, i.e., the phenomenon of heterozygote advantage (Fincham, 1975; Clarke, 1979). For example, it has been suggested that several diseases are maintained in the population because of heterozygote advantage. These include cystic fibrosis (Meindl, 1987), Mediterranean fever (Brenner-Ullman, Melzer-Ofir, Daniels and Shohat, 1994) and thalassaemia (Wang and Schilling, 1995), although the only undisputed example is sickle-cell anaemia (Edwards, 1977).

Preferential selection of heterozygotes has also been suggested where a homozygote has specific immunological disadvantages during pregnancy. For example, the HLA region has been linked to recurrent abortion in couples whose level of HLA sharing is high (Beer, Quebbeman, Ayers and Haines, 1981). It is believed that in pregnancy the fetus is maternally detected as a foreign organism and, to prevent its rejection, it stimulates the release of protective antibodies (Schwarzenau, 1990). In couples with high HLA compatibility, the antigenic

differences between mother and fetus are reduced, which could impede the release of these protective antibodies, resulting in spontaneous loss of the pregnancy (Schwarzenau, 1990). However, it has also been suggested that HLA sharing *per se* is not responsible for early embryonic loss, but rather that the HLA may be in tight linkage disequilibrium with the gene or genetic defect responsible for the abortion (Jin, Ho, Speed and Gill, 1995, Brennan, 1997).

Theoretically, if a more general sectional mechanism against high levels of homozygosity is operative at other loci in the genome, then microsatellites that are linked to or are found within these loci are also more likely to be heterozygous. The selective force(s) behind the lower than expected levels of homozygosity described in the present study may be optimally active early in pregnancy, and result in the preferential early abortion of embryos or fetuses which exhibit high homozygosity. In particular, these selectional mechanisms may be expressed at the loci of early developmental genes.

Interestingly, an excess of heterozygosity has also been reported in a number of diverse species undergoing inbreeding, including chickens (Mina, Sheldon, Yoo and Frankham, 1991), *Drosophila* (Rumball. Franklin, Frankham and Sheldon, 1994), and dogs (Rogers, 1995). In particular, Rumball *et al.* (1994) using polymorphic enzyme loci observed that the rate of decline of heterozygotes in full-sib and double first cousin matings was significantly slower than the theoretical expectations. Further, in successive generations of *Drosophila*, the heterozygote was increasingly more fit than the homozygote. If, as believed, the Pakistani families in the present study have been contracting close kin marriages for an extended period of time, then the increased incidence of heterozygosity could perhaps be attributed to a similar increase in their fitness, not present in the homozygotes. However, the family

sizes of close kin unions in Pakistan generally are larger than among non-consanguineous unions, suggesting no decrease in fitness (Bittles *et al.*, 1993).

If prenatal selection against homozygotes is occurring, it would appear to vary in intensity within different regions of the individual chromosomes. While there was no statistically significant difference in the level of homozygosity between the two *bradaris* for either chromosome, differences were noted when data from the individual *bradaris* were compared to the published data. The Khattar were significantly more homozygous than the CEPH data for chromosome 15, and the Rajpoot were significantly less homozygous than the CEPH data for chromosome 13. Similarly, the increase in homozygosity in the children of first cousin marriages and non-consanguineous marriages for the Rajpoot was more pronounced on chromosome 15, and was slightly higher in the Khattar for this chromosome. This effect was also reflected in the differences in the expected homozygosity and observed homozygosity, which were as high as 24% for individual loci on chromosome 15, and 17% for chromosome 13. Overall, these observations support the suggestion that certain regions of the genome may demonstrate resistance to homozygosity (Rogers, 1995).

The level of homozygosity in the children of first cousins was higher than in the children of non-consanguineous marriages, although not by the expected 6.25%. Thus, it would appear that although there may be some degree of preferential selection against homozygotes, the influence of descent from a common ancestor is still present, albeit to a reduced degree. However, as indicated by Figure IV.13, the effects appear to be highly variable across the two chromosomes and in both pedigrees.

V.6 Polymorphism information content of markers

As the observed homozygosity in the *bradaris* was significantly lower than expected, it was decided that the polymorphism information content (PIC) of each marker should also be calculated. The PIC of a marker is the probability that the genotype of a given offspring will permit identification of which marker allele at a locus was inherited from each parent. It is calculated by:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2 p_i^2 p_j^2$$

Equation V.1: Polymorphism information content

where p_i is the frequency of the ith allele, and n is the total number of alleles (Ott, 1991). As the PIC is an adjusted form of heterozygosity, it is conventionally discussed relative to heterozygosity rather than homozygosity.

The PIC, like heterozygosity, decreases with decreasing number of alleles, and so detecting differences in populations is strongly dependent on the number of alleles described for a marker (Taylor, Sherwin and Wayne, 1994). With inbreeding, the PIC may be more informative than heterozygosity, as it gives a more conservative estimate of the possible level of heterozygosity from the observed allele frequencies. For example, if the PIC for a marker is low, it indicates a high probability that the parental haplotypes frequently are identical, with a corresponding increased potential for the homozygous genotype. The PIC would therefore be expected to indicate a greater deviation from expected heterozygosity in a population that is influenced by selection against homozygosity.

With these precedents in mind, the PIC of each marker was calculated according Equation V.1, and compared with the observed and expected levels of heterozygosity using a paired *t*-test (Tables V.1 and V.2).

Table V.1: PIC, observed heterozygosity and expected heterozygosity for the Khattar pedigree

Locus	PIC	Observed heterozygosity	Expected heterozygosity (F=0)	Expected heterozygosity (F=0.0276)
D13S175	0.440	0.570	0.544	0.529
D13S115	0.551	0.760	0.694	0.675
D13S192	0.783	0.960	0.822	0.799
D13S126	0.771	0.800	0.814	0.792
D13S133	0.478	0.650	0.620	0.603
D13S270	0.309	0.730	0.591	0.574
D13S144	0.672	0.800	0.758	0.737
D13S125	0.656	0.750	0.749	0.728
D13S173	0.467	0.710	0.650	0.632
D13S285	0.738	0.900	0.789	0.766
D15S11	0.294	0.590	0.587	0.570
GABRB3	0.625	0.690	0.733	0.713
D15S97	0.752	0.860	0.798	0.776
D15S101	0.498	0.590	0.671	0.653
D15S102	0.737	0.760	0.796	0.774
D15S108	-0.248	0.450	0.424	0.412
D15S98	0.624	0.880	0.731	0.710
D15S169	0.633	0.760	0.731	0.711
D15S100	0.374	0.710	0.608	0.591
D15S120	0.698	0.780	0.770	0.748
t-test		-5.46657	-4.70404	-4.03452

Note: *t*-test values are for comparisons with PIC where df = 19 and *t*-critical = 2.093.

Table V.1: PIC, observed heterozygosity and expected heterozygosity for the Rajpoot pedigree

Locus	PIC	Observed heterozygosity	Expected heterozygosity (F=0)	Expected heterozygosity (F=0.025)
D13S175	0.607	0.820	0.657	0.640
D13S115	0.435	0.800	0.637	0.621
D13S192	0.759	0.840	0.803	0.782
D13S126	0.687	0.880	0.762	0.743
D13S133	0.538	0.710	0.625	0.609
D13S270	0.577	0.920	0.698	0.680
D13S144	0.249	0.730	0.561	0.547
D13S125	0.591	0.880	0.703	0.685
D13S173	0.725	0.960	0.778	0.758
D13S285	0.705	0.880	0.765	0.745
D15S11	0.492	0.670	0.662	0.645
GABRB3	0.219	0.670	0.561	0.547
D15S97	0.637	0.720	0.733	0.714
D15S101	0.429	0.600	0.644	0.628
D15S102	0.624	0.820	0.728	0.709
D15S108	0.359	0.700	0.610	0.594
D15S98	0.732	0.970	0.774	0.755
D15S169	0.456	0.670	0.630	0.614
D15S100	0.439	0.670	0.639	0.622
D15S120	0.699	0.870	0.761	0.742
t-test		-10.0359	-6.83666	-5.86144

Note: *t*-test values are for comparisons with PIC where df = 19 and *t*-critical = 2.093.

At the 5% level, PIC was found to be significantly lower than both observed and expected heterozygosity. Furthermore, the *t*-test indicated that PIC was significantly different from observed heterozygosity to a greater extent than observed versus expected heterozygosity. Therefore, the PIC predicted a reduced level of heterozygosity in the *bradaris* than was observed or would have been predicted by expected heterozygosity, an observation which lends support to the hypothesis of

selection against homozygosity. Under these circumstances, the PIC of a marker may therefore be a better indication that a population is influenced by selection against homozygotes, than the observed homozygosity.

V.7 Prospects for future research

The results of the present study should be treated with caution, due to their preliminary nature, the limited sample size, and the knowledge that the individuals in the pedigrees may carry the recessive disease, β-thalassaemia. There was also the problem of inadequate information on the backgrounds of the individuals studied and the exact relationship between spouses. For example, the inbreeding coefficient of individuals who had married into the pedigree was not known, and a value of $F>0$ could affect the inbreeding coefficient of the whole pedigree. Furthermore, the length of time that inbreeding has been practised in these *bradaris* is uncertain, and from the pedigrees provided only a maximum of two generations include consanguineous marriages. If consanguinity has been a feature of the *bradaris* for more than the last two generations, it predictably would be reflected in their cumulative inbreeding coefficients.

Another major limitation to the study was time, which prevented extensive analysis of the results. For example, analysis of the observed allele frequencies was performed using a χ^2 test, but alternative analyses could be employed, such as the Monte Carlo method which involves random number generation and may be a more appropriate method when dealing with small populations (Scribner, Arntzen and Burke., 1993). Time also prevented thorough analysis of the results to determine if the differences in sample size between the two *bradaris*, and between the *bradaris* and GDB and CEPH were significantly influencing the results. More detailed

analysis of the results should take into consideration the varying population sizes, as it is probable that this factor has been influential when comparing the *bradaris* with published data.

Further analysis of the results also could pay specific attention to the homozygosity and allelic-sharing at specific loci, with respect to the function of any genes that are linked to the microsatellites investigated (e.g. GABRB3). Additionally, the results could be analysed to evaluate any association between homozygosity and certain regions of the chromosomes examined, which may indicate if inbreeding exerts specific regional effects, and may also assess the neutrality of microsatellite markers. The study could also be extended to investigate markers on different chromosomes or specific regions of chromosomes, such as the X-chromosome or the HLA region of chromosome 6, as they may demonstrate greater resistance to homozygosity. Similarly, a variety of markers (e.g. minisatellites, RFLPs) could be compared, as they may each respond to homozygosity and inbreeding in alternative ways, dependent on their mutation rates and positions on the chromosomes.

The mutation rate of the individual microsatellites is another factor which could have been considered in the analysis of the present results. It is accepted that the different types of genetic marker possess different rates of mutation, and these can be predicted by two principle models, the Infinite Allele Model (IAM) and the Stepwise Mutation Model (SMM; Shriver, Jin, Chakraborty and Boerwinkle, 1993). The SMM has been shown to be the most appropriate model for microsatellites (Edwards *et al.*, 1992; Shriver *et al.*, 1993; Valdes, Slatkin, and Freimer, 1993; Di Rienzo *et al.*, 1994), and can be used to detect genetic distance and divergence. This has implications for the results obtained in this study because it provides an

alternative means of comparing the two *bradaris* with each another and with the GDB and CEPH data.

Continuing research would also benefit by increasing the sample size. This proposal is based on the observation that, on a number of occasions, the differences observed between the *bradaris* became less significant as data were combined. Sample size is important from a statistical perspective and, regardless of the relationship between individuals, small number problems will limit the information that can be derived from a study (Cooper, Smith, Cooke, Neimann and Schmidtke, 1985). The sample size could be increased to encompass new families from the Rawalpindi region, and thus investigate how the genomes of separate endogamous communities living in the same geographical area have diverged. Alternatively, an increased sample size could include families from geographically distant regions, and from different ethnic groups, to gain insight into whether inbreeding produces comparable decreases in homozygosity in all human populations.

Further study could also investigate the association between spontaneous abortion and homozygosity. In this investigation, low homozygosity was suggested to be a consequence of the rejection of fetuses with high homozygosity. Early pregnancy loss in humans has recently been found to occur at a rate of between 45% (young women) and 90% (women in their forties) in a study conducted in Bangladeshi women, based on sequential urinary assay of human Chorionic Gonadotrophin levels (Wood, Holman and O'Connor, 1997). Using such a system, the rate of spontaneous abortion could be monitored in both consanguineous and non-consanguineous couples. Moreover, marker analysis could indicate whether homozygosity was higher in abortuses that in live births, although a study of this nature would be faced with major practical and ethical constraints.

In conclusion, the current results could be further investigated to establish more detailed information on the effects of inbreeding in these families, and how they differ from non-inbred populations. Further investigations in other regions of the genome, and in other individuals, may also lead to significant conclusions as to the probable homogeneity of the effects of inbreeding that were observed in this study.

VI. REFERENCES

ABI Prism 310 Genetic Analyzer GeneScan Chemistry Guide. (1995). USA: The Perkin Elmer Corporation.

ABI Prism 310 Genetic Analyzer User's Manual. (1995). USA: The Perkin Elmer Corporation.

Al-Gazali, L.I., Dawodu, A.H., Sabarinathan, K., & Varghese, M. (1995). The profile of major congenital abnormalities in the United Arab Emirates (UAE) population. *Journal of Medical Genetics, 32,* 7-13.

Allchin, F.R., Thapar, R., Calkins, P.B., Srivastava, A.L., Varma, S.P., Percival Spear, T.G., Wolpert, S.A., & Tinker, H.R. (1983). History of the Indian subcontinent. In *Encyclopaedia Britannica* (Vol. 9, pp. 334-430). Chicago: Encyclopaedia Britannica.

Beck, S. (1996a). *The Vedas* [on line]. Available WWW: http://www.west.net/~beck/EC6-Vedas.html.

Beck, S. (1996b). *Buddha and Buddhism* [on line]. Available WWW: http://www.west.net/~beck/EC6-Buddha.html.

Beer, A.E., Quebbeman, J.F., Ayers, J.W.T., & Haines, R.F. (1981). Major histocompatibility complex antigens, maternal and paternal immune responses, and chronic habitual abortions in humans. *American Journal of Obstetrics and Gynecology, 141,* 987-999.

Bittles, A.H. (1994). The role and significance of consanguinity as a demographic variable. *Population and Development Review, 20,* 560-584.

Bittles, A.H. (1995). When cousins marry: a review of consanguinity in the Middle East. *Perspectives in Human Biology, 1,* 71-83.

Bittles, A.H., Grant, J.C., & Shami, S.A. (1993). Consanguinity as a determinant of reproductive behavior and mortality in Pakistan. *International Journal of Epidemiology, 22,* 463-467.

Bittles, A.H., Mason, W.M., Greene, J., & Rao, N.A. (1991). Reproductive behavior and health in consanguineous marriages. *Science, 252,* 789-794.

Boerwinkle, E., Xion, W., Fourest, E., & Chan, L. (1989). Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3' hypervariable region. *Proceedings of the National Academy of Sciences, USA, 86,* 212-216.

Botstein, D., White, R.L., Skolnick, M., & Davis, R.W. (1980). Construction of a human genetic linkage map using restriction fragment length polymorphisms. *American Journal of Human Genetics, 32,* 314-331.

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

Brennan, P. (1997). HLA sharing, and history of miscarriage among women with rheumatoid arthritis. *American Journal of Human Genetics, 60,* 738-740.

Brenner-Ullman, A., Melzer-Ofir, H., Daniels, M., & Shohat, M. (1994). Possible protection against asthma in heterozygotes for familial Mediterranean fever. *American Journal of Medical Genetics, 53,* 172-175.

Chakraborty, R., Deka, R., Jin, L., & Ferrell, R.E. (1992). Allele sharing at six VNTR loci and genetic distances among three ethnically defined human populations. *American Journal of Human Biology, 4,* 387-397.

Chamberlain, J.S., Gibbs, R.A., Ranier, J.E., Nguyen, P.N., & Caskey, C.T. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. *Nucleic Acids Research, 16,* 11141-11156.

Chehab, F.F., & Kan, Y.W. (1989). Detection of specific DNA sequences by fluorescence amplification: a colour complementation assay. *Proceeding of the National Academy of Sciences, USA, 86,* 9178-9182.

Clarke, B.C. (1979). The evolution of genetic diversity. *Proceedings of the Royal Society of London, Series B: Biological Science, 205,* 453-474.

Cooper, D.N., Smith, B.A., Cooke, H.J., Niemann, S., & Schmidtke, J. (1985). An estimate of the unique DNA sequence heterozygosity in the human genome. *Human Genetics, 69,* 201-205.

Cummings, M.R. (1991). *Human Heredity* (2nd ed.). USA: West Publishing Company.

Deka, R., Jin, L., Shriver, M.D., Yu, L.M., DeCroo, S., Hundreisser, J., Bunker, C.H., Ferrell, R.E., & Chakraborty, R. (1995). Population genetics of dinucleotide (dC-dA)*n*·(dG-dT)*n* polymorphisms in world populations. *American Journal of Human Genetics, 56,* 461-474.

Devlin, B., Risch, N., & Roeder, K. (1990). No excess of homozygosity at loci used for DNA fingerprinting. *Science, 249,* 1416-1419.

Diamond, J.M. (1987). Causes of death before birth. *Nature, 329,* 487-488.

Di Rienzo, A., Peterson, A.C., Garza, J.C., Valdes, A.M., Slatkin, M., & Freimer, N.B. (1994). Mutational processes of simple sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences, USA, 91,* 3166-3170.

Donis-Keller, H., Green, P., Helms., Cartinhour, S., Weiffenbach, B., Stephens, K., Keith, T.P., Bowden, D.W., Smith, D.R., Lander, E.S., *et al.* (1987). A genetic linkage map of the human genome. *Cell, 51,* 319-337.

Edwards, A., Civitello, A., Hammond, H.A., & Caskey, C.T. (1991). DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics, 49,* 746-756.

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

Edwards, J.H. (1977). Heterozygote advantage. *Archives of Diseases in Childhood, 52,* 343-344.

Eeles, A., & Stamps, A.C. (1993). *The Polymerase Chain Reaction (PCR): The Technique and Its Applications.* USA: R.G. Landes Company.

Ewens, W.J., Spielman, R.S., & Harris, H. (1981). Estimation of genetic variation at the DNA level from restriction endonuclease data. *Proceedings of the National Academy of Sciences, USA, 78,* 3748-3750.

Fincham, J.R.S. (1975). Heterozygous advantage as a likely general basis for enzyme polymorphisms. *Heredity, 28,* 387-391.

Fregeau, C.J., & Fourney, R.M. (1993). DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. *BioTechniques, 15,* 100-119.

Freire-Maia, N., Freire-Maia, A., & Quelce-Salgado, A. (1963). The load of lethal mutations in White and Negro Brazilian populations: I. First survey. *Acta Genetica, 13,* 185-198.

Gadgil, M., & Malhotra, K.C. (1983). Adaptive significance of the Indian caste system: an ecological perspective. *Annals of Human Biology, 10,* 465-478.

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

Grant, J.C., & Bittles, A.H. (1997). The comparative role of consanguinity in infant and childhood mortality and Pakistan. *Annals of Human Genetics, 61,* 143-149.

Grant, J.C., Shami, S.A., Milligan, P.J., & Bittles, A.H. (1994). Counting the costs of human inbreeding: the application of relative and attributable risks to the assessment of consanguinity-associated mortality. *Perspectives in Human Biology, 4,* 47-54.

Hartl, D.L. (1988). *A Primer of Population Genetics* (2nd ed.). Massachusetts: Sinauer Associates, Inc.

Hearne, C.M., Ghosh, S., & Todd, J.A. (1992). Microsatellites for linkage analysis of genetic traits. *Trends in Genetics, 8,* 288-294.

Hershman, P. (1981). *Punjab Kinship and Marriage.* India: Hindustan Publishing Corporation.

Hofferbert, S., Schanen, C., Chehab, F., & Francke, U. (1997). Trinucleotide repeats in the human genome: size distributions for all possible triplets and detection of expanded disease alleles in a group of Huntington disease individuals by the Repeat Expansion Detection method. *Human Molecular Genetics, 6,* 77-83.

Hou, Y., Schmitt, C., Staak, M., Puers, C., & Prinz, M. (1994). Genetic variation of the amplified VNTR polymorphism COL2A1 in Chinese and German populations. *Human Heredity, 44,* 114-119.

Housman, D. (1995). Molecular medicine: human DNA polymorphism. *The New England Journal of Medicine, 332,* 318-320.

Hudson, T.J., Engelstein, M., Lee, M.K., Ho, E.C., Rubenfield, M.J., Adams, C.P., Housman, D.E., & Dracopoli, N.C. (1992). Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics, 13,* 622-629.

Hussain, R., & Bittles, A.H. (1997). The prevalence and demographic characteristics of consanguineous marriages in Pakistan. *Journal of Biosocial Science*, in press.

Jamie, K. (1992). *The Golden Peak: Travels in Northern Pakistan.* London: Virago.

Jeffreys, A.J., Wilson, V., Thein, S.L., Weatherall, D.J., & Ponder, B.A.J. (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *American Journal of Human Genetics, 39,* 11-24.

Jin, K., Ho, H., Speed, T.P., & Gill, T.J. (1995). Reproductive failure and the major histocompatibility complex. *American Journal of Human Genetics, 56,* 1456-1467.

Jorde, L.B. (1991). Inbreeding in human populations. In R. Dulbecco (Ed.), *Encyclopaedia of Human Biology* (Vol. 4, pp. 431-441). USA: Academic Press, Inc.

Kalaydjieva, L., & Tolún, A. (1996). Personal communication.

Kapadia, K.M. (1958). *Marriage and Family in India* (2nd ed.). India: Oxford University Press.

Kidd, K.K., Bowcock, A.M., Schmidtke, J., Track, R.K., Ricciuti, F., Hutchings, G., Bale, A., Pean, P., & Willard, H.F. with help from Gelernter, J., Giuffra, L., & Kubzdela, K. (1989). Report of the DNA committee and catalogs of cloned and

mapped genes and DNA polymorphisms. *Cytogenetics and Cell Genetics, 51,* 622-947.

King, J., & St Vincent, D. (1993). *Pakistan: A Travel Survival Kit* (4th ed.). Australia: Lonely Planet Publications.

Krontiris, T.G. (1995). Minisatellites and human disease. *Science, 269,* 1682-1683.

Kruglyak, L., Daly, M.J., & Lander, E.S. (1995). Rapid multipoint linkage analysis of recessive traits in nuclear families, including homozygosity mapping. *American Journal of Human Genetics, 56,* 519-527.

Lewontin, R.C., & Hartl, D.L. (1991). Population genetics in forensic DNA typing. *Science, 254,* 1745-1750.

Li, W., & Sadler, L.A. (1991). Low nucleotide diversity in man. *Genetics, 129,* 513-523.

Litt, M. (1991). PCR of TG microsatellites. In M.C. McPherson, P. Quirke & G. Taylor (Eds.), *PCR: A Practical Approach* (pp. 85-99). Oxford; New York: IRL Press at Oxford University Press.

Litt, M., & Luty, J.A. (1989). A hypervariable microsatellite reveaied by *in-vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics, 44,* 397-401.

Love, J.M., Knight, A.M., McAleer, M.A., & Todd, J.A. (1990). Towards construction of high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucleic Acids Research, 18,* 4123-4130.

Makino, R., Yazyu, H., Kishimioto, Y., Sekiya, T., & Hayashi, K. (1992). F-SSCP: fluorescence-based polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP) analysis. *PCR Methods and Applications, 2,* 10-13.

Meindl, R.S. (1987). Hypothesis: a selective advantage for cystic fibrosis heterozygotes. *American Journal of Physical Anthropology, 74,* 39-45.

Mian, A., & Mushtaq, R. (1994). Consanguinity in a population of Quetta (Pakistan): a preliminary study. *Journal of Human Ecology, 5,* 49-53.

Mina, N.S., Sheldon, B.H., Yoo, B.H., & Frankham, R. (1991). Heterozygosity at protein loci in inbred and outbred lines of chickens. *Poultry Science, 70,* 1864-1872.

Moore, S.S., Sardeant, L.L., King, T.J., Mattick, J.S., Georges, M., & Hetzel, D.J.S. (1991). The conservation of dinucleotide microsatellites among mammalian genomes allows the use of heterologous PCR primer pairs in closely related species. *Genomics, 10,* 654-660.

Morell, R., Liang, Y., Asher, J.H., Weber, J.L., Hinnant, J.T., Winata, S., Nyomen, I., Arhya, I.N., & Friedman, T.B. (1995). Analysis of short tandem repeat (STR) allele frequency distributions in a Balinese population. *Human Molecular Genetics, 4,* 85-91.

Murray, B.W. (1996). *The Estimation of Genetic Distance and Population Substructure from Microsatellite Allele Frequency Data* [on line]. Available: http://helix.biology.mcmaster.ca/brent/brent.html

Murty, J.S., Muralidhar, B., Goud, J.D., Mohan, P.J., Babu, B.R., & Rao, V.S. (1993). Hierarchical gene diversity and genetic structure of tribal populations of Andhra Pradesh. *American Journal of Physical Anthropology, 90,* 169-183.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., & White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science, 235,* 1616-1622.

Ott, J. (1991). *Analysis of Human Genetic Linkage*. Baltimore: The Johns Hopkins University Press.

Ott, J. (1992). Strategies for characteristic highly polymorphic markers in human gene mapping. *American Journal Human Genetics, 51,* 283-290.

Perlin, M.W., Burks, M.B., Hoop, R.C., & Hoffman, E.P. (1994). Toward fully automated genotyping: allele assignment, pedigree construction, phase determination, and recombination detection in Duchenne muscular dystrophy. *American Journal of Human Genetics, 55,* 777-787.

Queller, D.C., Strassmann, J.E., & Hughes, C.R. (1993). Microsatellites and kinship. *Trends in Evolution and Ecology, 8,* 285-288.

Radha Rami Devi, A., Appaji Rao, N., & Bittles, A.H. (1987). Inbreeding and the incidence of childhood genetic disorders in Karnataka, South India. *Journal of Medical Genetics, 24,* 362-365.

Rawalpindi. (1983). In *Encyclopaedia Britannica* (Vol. VIII, pp. 346-347). Chicago: Encyclopaedia Britannica.

Roberts, C.J., & Lowe, C.R. (1975). Where have all the conceptions gone? *The Lancet, March 1,* 498-499.

Rogers, T. (1995). An investigation of genetic variation in an inbred group of dogs, using polymorphic DNA markers. *Honours Thesis*. Western Australia: Murdoch University.

Royle, N.J., Clarkson, R.E., Wong, Z., & Jeffreys, A.J. (1988). Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics, 3,* 352-360.

Rumball, W., Franklin, I.R., Frankham, R., & Sheldon, B.L. (1994). Decline in heterozygosity under full sib and double first-cousin inbreeding in *Drosophila melanogaster*. *Genetics, 136,* 1039-1049.

Sambrook, J., Fritch, E.F., & Maniatis, T. (1989). *Molecular Cloning* (2nd ed.). USA: Cold Sprimg Harbour Laboratory.

Santiago, J.R. (1987). *Pakistan: A Travel Survival Kit* (3rd ed.). Australia: Lonely Planet Publications.

Satha, C. & Ahmed, T. (1992). Proximate determinants of fertility. In *Pakistan Demographic and Health Survey* (pp. 85-96). Islamabad: National Institute of Population Studies.

Schmitt, L. (1997). Personal communication.

Schwarzenau, E. (1990). Immunological aspects of abortion. *Triangle, 29*, 143-149.

Schull, W.J., & Neel J.V. (1972). The effects of parental consanguinity and inbreeding in Hirado, Japan: V. Summary and interpretation. *American Journal of Human Genetics, 24,* 425-453.

Scott-Emuakpor, A.B. (1974). The mutation load in an African population: I. Analysis of consanguineous marriages in Nigeria. *American Journal of Human Genetics, 26,* 674-687.

Scribner, K.T., Arntzen, J.W., & Burke, T. (1994). Comparative analysis of intra- and interpopulation genetic diversity in *Bufo bufo*, using allozyme, single-locus microsatellite, minisatellite, and multilocus minisatellite data. *Molecular Biology and Evolution, 11,* 737-748.

Shami, S.A., Grant, J.C., & Bittles, A.H. (1994). Consanguineous marriage within social/occupational class boundaries in Pakistan. *Journal of Biosocial Science, 26,* 91-96.

Shriver, M.D., Jin, L., Chakraborty, R., & Boerwinkle, E. (1993). VNTR allele frequency distribution under the stepwise mutation model: a computer simulation approach. *Genetics, 134,* 983-993.

Stowell, H. (1996). The most beautiful people. *Nature Genetics, 12,* 361-362.

Sullivan, K.M., Pope, S., Gill, P., & Robertson, J.M. (1992). *PCR Methods and Applications, 2,* 34-40.

Taylor, A.C., Sherwin, W.B., & Wayne, R.K. (1994). Genetic variability of microsatellite loci in a bottleneck species: the northern hairy-nosed wombat *Lasiorhinus krefftii*. *Molecular Ecology, 3,* 277-290.

Teebi, A.S. (1994). Autosomal recessive disorders among Arabs: an overview from Kuwait. *Journal of Medical Genetics, 31,* 224-233.

Thapar, R. (1990). *A History of India: Part 1.* USA: Penguin.

Thompson, E.A., & Neel, J.V. (1997). Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. *American Journal of Human Genetics, 60,* 197-204.

Valdes, A.M., Slatkin, M., & Freimer, N.B. (1993). Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics, 133,* 737-749.

Vollmer, J.E., Keall, E.J., & Nagai-Berthrong, E.N. (1983). *Silk Roads, China Ships.* Toronto: Royal Ontario Museum.

Wall, W. J., Williamson, R., Petrou, M., Papaioannou, D., & Parkin, B.H. (1993). Variation of short tandem repeats within and between populations. *Human Molecular Genetics, 2,* 1023-1029.

Wallbank, T.W. (1965). *A Short History of India and Pakistan.* New York: New American Library.

Wang, C.H., & Schilling, R.F. (1995). Myocardial infarction and thalassaemia trait: an example of heterozygote advantage. *American Journal of Hematology, 49,* 73-75.

Weber, J.L., & May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics, 44,* 388-396.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morisette, J., Millasseau, P., Vayssieux, G., & Lathrop, M. (1992). A second-generation linkage map of the human genome. *Nature, 359,* 794-801.

Wilcox, A.J., Weinburg, C.R., O'Connor, J.F., Baird, D.D., Schlatterer, J.P., Canfield, R.E., Armstrong, E.G., & Nisula, B.C. (1988). Incidence of early loss of pregnancy. *The New England Journal of Medicine, 319,* 189-194.

Wood, J.W., Holman, D.J., & O'Connor. (1997, May). *Age and Reproductive Function.* Workshop session presented at: Age - Between Nature and Culture, Max Planck Institute for Demographic Research, Rostock, Germany.

Yaqoob, M., Gustavson, K-H., Jalil, F., Karlberg, J., & Iselius, L. (1993). Early childhood health in Lahore, Pakistan: II. Inbreeding. *Acta Paediatrica Supplement, 390,* 17-26.

Yuille, M.A., Goudie, D.R., Affara, N.A., & Ferguson-Smith, M.A. (1991). Rapid determination of sequences flanking microsatellites. *Nucleic Acids Research, 19,* 1950.

Zeichler, H., Schafer, R., & Epplen, J.T. (1989). Non-radioactive oligonucleotide fingerprinting in the gel. *Nucleic Acids Research, 17,* 4411.

VII. APPENDIX

Appendix 1: Definition of terms

Allele: one of the possible forms of a gene

Allozygosity: homozygosity where the identical alleles are inherited by chance and not by descent from a common ancestor

Autozygosity: homozygosity where the identical alleles are inherited from the same ancestor

***Bradari*:** in Pakistan, the extended family recognised along paternal and/or fraternal lines

Coefficient of inbreeding: the probability that the two alleles at a locus are inherited because of descent from a common ancestor

Consanguinity: mating between two individuals who are related by a common ancestor in the last three or less generations

F_{ST}: a measure of genetic distance between two populations

Haplotype: the alleles received by an individual from one parent

Heterozygosity: alleles at a particular locus are different

Homozygosity: alleles for a particular locus are the same

Human leucocyte antigen: antigens which help the body differentiate between self and non-self

Locus (loci): region where a gene is found

Mendelian inheritance: inheritance where alleles are passed on to the next generation in equal ratios from each parent

Microsatellite: tandem repeats of simple sequences less than 100bp long, occurring randomly in the genome

Minisatellite: repeat sequences where a short nucleotide sequence of DNA is repeated tandemly 20 to 100 times

Morbidity: diseased or disease related

Mortality: death rate in a population

Polymorphism Information Content: probability that the marker genotype of a given offspring will allow deduction of which of the two marker alleles was inherited from each of the parents

RFLP: specific sequences of DNA which can be cleaved by restriction enzymes

Spontaneous abortion: cessation of pregnancy within the first 28 weeks of pregnancy

Appendix 2: WWW sites for human genome information

Généthon

Home Page: www.genethon.fr/genethon-en.html

Map Query: ftp.genethon.fr/pub/Gmap/Nature-1995/data/

Cooperative Human Linkage Centre (CHLC)

Home Page: www.chlc.org/

Map Query: vrww.chlc.or/ChlcMaps.html

Centre d'Études du Polymorphisme Humain (CEPH)

Home Page: www.cephb.fr/cephdb/

Marker Query: www.cephb.fr/cgi-bin/wdb/ceph/systeme/form

Genome Data Base (GDB)

The GDB was upgraded from version 5.6 to version 6.2 in late 1996. Marker information in this study was mainly derived from version 5.6.

Home Page: gdbwww.gdb.org/gdb/ (version 6.2)

gdbwww.gdb.org/gdb/docs/gdbhome.html (version 5.6)

Polymorphism

Query: gdbwww.gdb.org/gdb/ (version 6.2)

gdbwww.gdb.org/gdb-bin/gdb5.6/browser/bin/map?249,19

(version 5.6)

Appendix 3: Observed allele frequencies

χ^2 test values listed are only those that were significant

df = (n - 1)

Chromosome 13

Locus	Allele size (bp)	Khattar	Rajpoot	χ^2
D13S175	98	0.44	0.51	
	100	0.51	0.21	
	104	0.05	0.14	
	108	0.00	0.14	
D13S115	160	0.04	0.00	6.70, df=3
	164	0.45	0.50	
	166	0.25	0.14	
	168	0.00	0.06	
	172	0.19	0.30	
	174	0.02	0.00	
	176	0.05	0.00	
D13S192	99	0.22	0.00	
	103	0.03	0.13	
	105	0.06	0.09	
	107	0.22	0.00	
	109	0.04	0.20	
	111	0.01	0.28	
	113	0.05	0.21	
	115	0.02	0.00	
	117	0.05	0.10	
	119	0.26	0.00	
	121	0.05	0.00	

Chromosome 13 (cont.)

Locus	Allele size (bp)	Khattar	Rajpoot	χ^2
D13S133	131	0.23	0.34	
	133	0.00	0.26	
	165	0.03	0.00	
	173	0.29	0.06	
	177	0.06	0.01	
	179	0.03	0.20	
	181	0.14	0.00	
	183	0.00	0.10	
	185	0.08	0.03	
	187	0.13	0.00	
	189	0.02	0.00	
D13S126	103	0.06	0.24	
	105	0.42	0.26	
	107	0.08	0.00	
	109	0.44	0.50	
D13S270	80	0.59	0.44	
	82	0.00	0.08	
	88	0.00	0.01	
	90	0.03	0.28	
	92	0.20	0.13	
	96	0.13	0.08	
	100	0.06	0.00	
D13S144	181	0.30	0.29	2.12, df=3
	185	0.05	0.00	
	187	0.14	0.00	
	189	0.34	0.59	
	191	0.09	0.06	
	193	0.08	0.06	

Chromosome 13 (cont.)

Locus	Allele size (bp)	Khattar	Rajpoot	χ^2
D13S125	130	0.08	0.00	
	144	0.06	0.00	
	148	0.17	0.08	
	150	0.27	0.35	
	152	0.05	0.38	
	154	0.37	0.13	
	160	0.00	0.08	
D13S175	230	0.00	0.24	
	232	0.00	0.28	
	234	0.45	0.21	
	236	0.15	0.10	
	238	0.35	0.00	
	240	0.05	0.18	
D13S285	88	0.14	0.18	5.04, df=4
	92	0.03	0.13	
	94	0.09	0.00	
	96	0.30	0.24	
	98	0.22	0.33	
	100	0.02	0.14	
	102	0.21	0.00	

Chromosome 15

Locus	Allele size (bp)	Khattar	Rajpoot	χ^2
D15S11	229	0.10	0.00	
	243	0.60	0.29	
	245	0.00	0.08	
	247	0.00	0.05	
	249	0.00	0.05	
	251	0.19	0.49	
	253	0.03	0.00	
	261	0.01	0.05	
	263	0.08	0.00	
GABRB3	185	0.33	0.20	
	187	0.06	0.00	
	189	0.25	0.61	
	193	0.00	0.16	
	195	0.05	0.00	
	197	0.30	0.00	
	203	0.00	0.03	
D15S97	180	0.21	0.03	9.38, df=4
	182	0.23	0.09	
	184	0.00	0.29	
	186	0.25	0.39	
	188	0.11	0.10	
	190	0.17	0.11	
	198	0.04	0.00	
D15S101	105	0.08	0.10	
	107	0.09	0.00	
	109	0.49	0.55	
	111	0.00	0.06	
	113	0.26	0.08	
	115	0.00	0.18	
	117	0.08	0.03	
	119	0.00	0.01	

Chromosome 15 (cont.)

Locus	Allele size (bp)	Khattar	Rajpoot	χ^2
D15S102	210	0.28	0.04	
	212	0.11	0.35	
	214	0.02	0.00	
	216	0.27	0.31	
	218	0.01	0.00	
	220	0.04	0.21	
	222	0.17	0.00	
	224	0.10	0.00	
	226	0.00	0.09	
D15S108	142	0.06	0.15	10.37, df=4
	144	0.05	0.04	
	150	0.07	0.14	
	158	0.05	0.10	
	160	0.75	0.58	
	162	0.02	0.00	
D15S169	145	0.44	0.06	
	147	0.00	0.06	
	149	0.07	0.00	
	151	0.00	0.04	
	153	0.00	0.11	
	155	0.15	0.00	
	157	0.08	0.00	
	159	0.04	0.29	
	161	0.02	0.00	
	163	0.01	0.11	
	171	0.20	0.33	

Chromosome 15 (cont.)

Locus	Allele size (bp)	Khattar	Rajpoot	χ^2
D15S98	138	0.23	0.11	
	140	0.06	0.49	
	142	0.11	0.04	
	144	0.20	0.03	
	148	0.40	0.00	
	150	0.01	0.34	
D15S100	182	0.16	0.21	1.16, df=3
	184	0.04	0.05	
	188	0.55	0.53	
	190	0.01	0.18	
	192	0.25	0.04	
D15S120	158	0.18	0.33	6.89, df=3
	160	0.29	0.18	
	162	0.00	0.05	
	164	0.00	0.08	
	166	0.07	0.00	
	168	0.27	0.28	
	170	0.19	0.00	
	172	0.00	0.10	