1995

Localisation of the gene for a novel form of Charcot-Marie-Tooth disease in an isolated population

Kaite Honeyman
Edith Cowan University

Edith Cowan University

Copyright Warning

LOCALISATION OF THE GENE FOR A NOVEL FORM OF CHARCOT-MARIE-TOOTH DISEASE IN AN ISOLATED POPULATION

BY

Kaite Honeyman

A thesis submitted in partial fulfilment of the requirements

for the award of Bachelor of Science (Human Biology) Honours

at the Faculty of Science, Technology & Engineering,

Edith Cowan University

Date of submission: 6 November, 1995.

USE OF THESIS

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

Declaration

I certify that this thesis does not incorporate without acknowledgement 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.

Signature:

ABSTRACT

Localising the gene for a previously undescribed autosomal recessive form of CMT involved the use of a relatively new approach to rapid genome screening based on the identification of segments which are inherited identical by descent (IBD) from common founding ancestors. It is most feasible for populations which have been founded relatively recently (say less than 25 generations) and which have remained relatively isolated either geographically or culturally. The method is not suitable for highly inbred populations, that is with first and second cousin matings, as many segments will be inherited by chance. It appears to be a suitable screening option for a rare disease trait so that the possibility of allelic heterogeneity is reduced. It is also an option for those pedigrees where a significant result cannot be obtained through traditional linkage analysis, that is, where a large nuclear family with many affected individuals is not available.

Highly informative polymorphic short tandem repeat markers were run on the DNA from an initial selection of ten individuals whose inter-relatedness could be established. In the later stages of the study, DNA from forty-seven members of this randomly inbred Bulgarian Gypsy kindred was used for verification of the IBD finding by lod score analysis.

In the study population where the chosen members are separated from a common ancestor by an average 9.4 meioses, the approximate probability of finding one random segment (not containing the gene) inherited identical by descent (IBD) from this common ancestor was calculated at 9.47 in 1,000 (over 94 markers at an average distance of 12.6 cM). The overall probability of finding an IBD segment (not containing the gene) in three or more chromosomes decreases to 1.7 in 1 million. This study initially identified a 20 cM region on 8q which was shared by three out of six non-sibling chromosomes. This region was extended by saturating with markers at closer intervals and subsequently a 3 cM segment was found to be common among four of six non-sibling chromosomes. Finally, lod score analysis on the forty-seven members of the kindred supported placement of the candidate gene in this 3 cM segment with a maximum lod scores at zero recombination for both marker loci D8S558 and D8S529 being 3.38 and 2.73 respectively. These two markers are mapped at 1 cM apart.

Acknowledgements

Firstly I wish to thank my supervisor, Dr Luba Kalaydjieva, for her contribution of knowledge and expertise to the project, as well as for her patience.

Collaborator on the project, Joachim Hallmayer, has spent a great deal of time on the analysis of the linkage region and has provided much information promptly and enthusiastically. Thanks also to Lodewijk Sandkuijl, whose initial input into the project established its direction. Thank you to the laboratory staff at the University of Sofia who provided the results for chromosome 5.

Thanks to Prof. Alan Bittles for donating his valuable time in editing the final draft of this thesis, and for his continued good humour and support.

Thanks also to the people in the Human Molecular Genetics Laboratory at Joondalup who have been great to work with, especially David Chandler for technical assistance and advice, Jason Williams for drawing the coloured pedigree in appendix C, and Eoin Thompson for being Irish.

Thanks to Alan Honeyman for his advice on mathematical queries as well as for moral support and help with chiidcare.

Finally, I would like to thank my friends in the laboratory at Joondalup for their continued support and encouragement, especially Marie McCluskey and David Chandler who were there when I needed them, and also Lisa Gavranich for making me laugh.

Table of Contents

List of Tables

List of Figures

CHAPTER 1. INTRODUCTION

1.1 Overview

Hereditary Motor and Sensory Neuropathies (HMSN) are disorders involving the peripheral nervous system. With an estimated overall frequency of 1:2500, Charcot-Marie-Tooth (CMT) disease represents the most common HMSN (Lupski *et al.*, 1991). The common clinical symptoms of Charcot-Marie-Tooth disease include weakness and atrophy in the distal muscles of the limbs, skeletal deformities, loss of deep tendon reflexes and a variable degree of sensory loss. However, differences exist between patients with regard to nerve conduction velocity, neuropathological findings and mode of inheritance. With four different genes (located on chromosomes 1, 8, 17 and X) implicated in the development of CMT, the disease is one of the best known examples of locus heterogeneity in genetic disorders in humans (Ben Othmane *et al.*, 1993; Lupski *et al.*, 1991; Gal *et al.*, 1985). Evidence of allelic heterogeneity, that is, different mutations in the same gene, has also been documented (Hayasaka *et al.*, 1993b, 1993c; Roa *et al.*, 1993a, 1993b). In a number of families, linkage to any of the known CMT genes has been excluded which suggests that additional gene loci involved in the development of the CMT phenotype remain to be identified. This is particularly relevant to the autosomal recessive forms of the disease, where only one rare CMT gene has been localised to date.

In most autosomal dominant forms of CMT, neuropathological studies have demonstrated the involvement of myelin, the lipid and protein insulating membrane of peripheral nerves (Dyck *et al.* 1993; Lupski *et al.* 1991). The loss of myelinated fibres, resulting from either faulty production or an increased breakdown of myelin, leads to the reduction in motor nerve conduction velocities (MCV) typical of these forms of the disease. Before the identification of the genes causing this type of CMT, it was not clear whether dysmyelination was a primary phenomenon in the pathogenesis of the disease or if it developed as a consequence of a primary axonopathy. The issue was resolved in the case of the demyelinating autosomal dominant forms of CMT, where mutations in genes encoding major myelin proteins were identified (for example, CMT1B : myelin protein zero; CMT1A : peripheral myelin protein 22). However, the question is still open in the remaining forms of CMT where as yet no genetic lesion has been found. Given the limited knowledge that we have on the interaction between the nerve axon and the myelin-forming Schwann cells, the identification of additional genes involved in different forms of CMT will not only elucidate the pathogenetic mechanisms of those specific disorders but will also contribute to our understanding of the functioning of the peripheral nervous system.

1.2 The clinical presentation of Charcot-Marie-Tooth disease

The Charcot-Marie-Tooth neuropathies were named after the clinicians who, in 1886, first described patients with the typical clinical symptoms of the disease (Lupski *et al.*, 1991). The invariable clinical features of CMT include progressive weakness and atrophy of the distal muscles in the leg, foot deformity, reduced deep tendon reflexes and weakness of intrinsic hand muscles (Roa *et al.*, 1993b; Lupski *et al.*, 1991; Chance & Fischbeck, 1994). Differences in the clinical presentation of the disease, which serve as the basis of the numerous clinical classifications, relate to the age of onset and rate of progression of the above-mentioned symptoms, a variable degree of sensory loss, changes in nerve conduction velocity and the involvement of myelin, as documented by neuropathological investigations.

1.3 Clinical classification of CMT

Although the classification of HMSN has varied, and not all authors have agreed on divisions between types, **Table 1** summarises the widely used subdivisions which were proposed by Dyck (1975). The variable decrease in nerve conduction velocity appears as the most important classification criterion from a pathophysiological point of view. Slow conduction of impulses along the peripheral nerves, as

Table 1 Classification by clinical criteria

CLASSIFICATION	SYMPTOMS	PATHOLOGY/EMG	ONSET	OTHER REFERENCES
HMSN Type I/CMT1	Weakness & wasting of lower limb muscles/foot deformity/hand deformity/sensory loss/ataxia	Symmetrically decreased MCV (< 38 mtrs/sec)/ segmental demyelination & Schwann cell proliferation ("onion bulbs")	1st two decades	Bird & Kraft (1978) Harding & Thomas (1980) Bird et al. (1983)
HMSN Type II/CMT 2	Same as CMT1 but less involvement of intrinsic hand muscles	Axonal neuropathy/normal to slightly reduced MCV	Mostly 2nd decade	Thomas et al. (1974)
HMSN Type III /Dejerine-Sottas	Clubfoot/distal weakness & atrophy in legs & arms/areflexia /sensory loss in all limbs	Hypomyelination/onion bulbs* No myelin* Neuronal/normal to slightly decreased MCV* * These variations in pathology subsequently used for further sub-division of this disease.	Congenital/early childhood	Dyck & Lambert (1974)
HMSN Type IV	Cerebellar ataxia/atypical pigmentary retinopathy/deafness/foot deformity	Increased phytanic acid in body fluids	1st two decades	Barailser (1990) cites Refsum (1945)

documented by electromyography (EMG), correlates with the degree of demyelination found using light and/or electron microscopy. Normal conduction velocities are found in those forms of the disease where axonal loss, rather than dysmyelination, is the characteristic neuropathological feature.

Peripheral neuropathies, and CMT in particular, have been found to occur in association with additional neurological symptoms which are not directly related to the peripheral nervous system (Baraitser, 1990). These associations include deafness (Cornell *et al.*, 1984), neural deafness and optic atrophy (Iwashita *et al.*, 1970), chronic inflammatory demyelinating neuropathy (Gabreels-Festen & Hageman, 1986 cited by Baraitser, 1990), spastic paraplegia (Dyck & Lambert, 1968a, 1968b), optic atrophy (Milhorat, 1943 cited by Baraitser, 1990), retinitis pigmentosa and ataxia (Tuck & McLeod, 1983) and congenital nephritis (Hanson *et al.*, 1970). In a review of 87 Norwegian families with peripheral neuropathies Skre (1974) noted the high incidence of non-specific neurological abnormalities in first degree relatives of CMT patients, and suggested that in the case of autosomal dominant CMT the association may be due to negative selection, in contrast to recessive CMT genes which may exert a pleiotropic effect in heterozygous carriers.

Further subdivision of the above types is summarised in **Table 2** according to mode of inheritance.

Table 2 Classification according to mode of inheritance

Classification	Sub-classification	Distinguishing features of sub-divisions
Autosomal dominant		
HMSN Type I	CMT1A	Intermediate MCV/fewer onion bulbs[1]
	CMT1B	very low MCV/more onion bulbs (Bird *et al.*, 1983)
HMSN Type II	CMT2	
Autosomal recessive		
HMSN Type I	CMT4A	Slow MCV/hypomyelination
	CMT4B	Myelin outfolding[1] /severely decreased MCV
	CMT4C	Axonal loss/normal MCV (Ben Othmane *et al.*, 1993a)
HMSN Type III	Dejerine-Sottas Type A1	Hypomyelination/Onion bulbs
	Dejerine-Sottas Type A2	No myelin
	Dejerine-Sottas Type B	Axonal loss/normal to slightly decreased MCV (Baraitser, 1990)
HMSN Type IV	Refsum's disease	
X-linked		
HMSN Type I	X-linked CMT	Slow MCV/onion bulbs

[1] appearance resulting from demyelination and attempted remyelination

The pattern of inheritance shows correlation with age at onset and the rate of progression of the disease. The most common forms (CMT1A, 1B), inherited in an autosomal dominant fashion, are less severe than the recessive (CMT4) forms (Ben Othmane *et al.*, 1993a). In these dominant forms, the age of onset of clinical symptoms is usually in the first or second decade, however earlier detection of sub-clinical changes is possible by electromyography in those types of CMT in which MCV is reduced (Roa *et al.*,1993b, Chance & Fischbeck, 1994). Although all types of CMT reduce the quality of life of sufferers, they do not seem to result in early death (Dyck *et al.*, 1989).

1.4 The molecular basis of HMSN

1.4.1 The role of myelin in the peripheral nervous system (PNS)

Myelin is the protein and lipid insulating sheath which, in the peripheral nervous system, is formed from the cell membranes of Schwann cells. The importance of myelin in the function of the peripheral nervous system is evident from comparisons made between the speed of transmission of nerve impulses in unmyelinated and myelinated axons. An unmyelinated axon would need to have a diameter approximately 100 times greater than a myelinated axon to achieve equivalent conduction velocities (Berne & Levy, 1988, p.44). This increase in the speed of conduction is due to increased membrane resistance

in the axon regions covered by the myelin which results in the saltatory conduction of action potentials along the nerve axon (Berne & Levy, 1988).

In addition to the role played by myelin as an insulator of the nerve axon, the relationships between axons and Schwann cells may be more complex and interdependent. Evidence that they produce neurotrophic factors suggests that these cells do not merely provide structural support but may directly influence the growth of axons (Meyer *et al.*, 1992 cited by Suter *et al.*, 1993). It is also apparent from studies of the *Trembler* (*Tr*) mouse that Schwann cells affect axonal transport, neurofilament phosphorylation and density (reviewed by Suter *et al.*, 1993). These findings suggest that forms of CMT in which myelin is not affected may still result from mutations in genes expressed in the Schwann cell, but be involved in interactions with axons rather than the process of myelination.

1.4.2 Myelin genes associated with PNS disorders

Genes which have already been identified as being involved in myelin production are Peripheral Myelin Protein 22 (PMP22) Myelin Protein Zero (MPZ or P_0) and Peripheral Myelin Protein 2 (PMP2). Connexin32 (Cx32), although not expressed in

compacted myelin, is expressed at the gap junctions at the ends of the individual Schwann cells (Bergoffen *et al.*, 1993).

Peripheral Myelin Protein-22 (PMP22)/CMT1A/HNPP

The peripheral myelin protein 22 gene (PMP22) was mapped to region 17p11.2-12 (Matsunami *et al.*, 1992), after Lupski and colleagues (1991) found that a 1.5 Mb duplication in this region was associated with CMT1A, and Vance and co-workers localised CMT1A to 17p11.2 (1991). Two point mutations (one recessive), also in the same gene, were found to be associated with CMT1A (Roa *et al.* 1993a, b). A deleted region corresponding to the duplicated area on 17p11.2-12 was found to result in hereditary neuropathy with pressure palsies (HNPP), an autosomal dominant disorder with segmental demyelination and tomaculous or sausage-like formations along the peripheral nerves (Chance *et al.*, 1993).

Expression of this axon-regulated PMP-22 gene is up-regulated during normal development of Schwann cells in the initial stages of the formation of the myelin sheath (Spreyer *et al.*, 1991). It has been suggested that the mechanism of dominance in CMT1A may be a dosage effect of PMP-22, due to the presence of an extra copy of the gene in the duplicated region (Matsunami *et al.*, 1992). Conversely, it has been proposed that

under-expression of the PMP-22 gene is the mechanism responsible for the HNPP phenotype (Chance *et al.*, 1993). Valentijn and colleagues (1995) identified a de novo mutation in the PMP22 gene of a patient with Dejerine-Sottas disease. The patient had an extra exon 1 fragment which was not present in the unaffected parents.

Myelin Protein Zero (MPZ/P_0)/CMT1B

Myelin protein zero (MPZ) is the major structural protein of peripheral nervous system myelin, accounting for more than 50% of the sheath protein (Greenfield *et al.* 1973). Since the extra-cellular domain of MPZ has an immunoglobulin-like structure (Lemke, 1988) and is similar to an adhesive glycoprotein, it has been proposed that it plays a major role in the compaction of myelin (Hayasaka *et al.*, 1993b).

In 1932 Bird and associates demonstrated linkage of an autosomal dominant form of CMT to the Duffy blood group on chromosome 1q. A decade later the MPZ gene was localised to the region 1q21-q23 (Pham-Dinh *et al.*, 1993). This autosomal dominant form of CMT was classified as CMT1B by Bird *et al.* (1983). In 1991 Lebo and colleagues more precisely mapped the CMT1B gene to the same region. Several point mutations in

the MPZ gene have since been identified as being responsible for CMT1B (Hayasaka *et al.*, 1993b, Latour *et al.*, 1995).

Peripheral myelin protein-2 (PMP2)/CMT4A

Peripheral myelin protein 2 (PMP2) is a small, lipid-binding basic protein. The PMP2 gene was mapped to chromosome 8q21.3-22.1 and its cDNA characterised by Hayasaka and colleagues (1993a).

Interestingly, the first autosomal recessive CMT gene (CMT4A) was linked to approximately the same chromosomal region, 8q13-21.1. The linkage study of CMT4A was conducted on four inbred Tunisian families with a total of 13 affected individuals (Ben Othmane *et al.*, 1993a). In a subsequent study Othmane and colleagues (1995) demonstrated the lack of mutations in the PMP2 gene in their CMT4A patients, providing evidence that PMP2 is not the gene involved in this recessive form of Charcot-Marie-Tooth disease.

It thus remains to be determined whether a molecular defect in PMP2 is responsible for another form of CMT. However, together these findings suggest that a cluster of peripheral nervous system genes may be located on the long arm of chromosome 8.

Connexin-32/CMTX

Connexin-32 is expressed in myelinated peripheral nerve at the Nodes of Ranvier, rather than in the compacted myelin sheaths. The Cx32 gene is located on Xq13-21 and an X-linked dominant form of CMT (CMTX) has been mapped to the same region (Gal *et al.*, 1985). A number of different point mutations in the Cx32 gene are now known to cause CMTX (Bergoffen *et al.*, 1993). CMTX is a demyelinating neuropathy characterised by degeneration of the spinal nerve roots, distal muscle atrophy earlier age of onset and increased severity in males. Both dominant and recessive forms have been observed. This variability of expression is possibly the result of allelic heterogeneity (Witkowski and Prokop, 1976 cited by Gal *et al.*, 1985).

1.4.3 Other chromosomal regions linked to CMT

CMT2

CMT Type 2 is an autosomal dominant axonal neuropathy which does not show a dysmyelination pathology. CMT2 has been localised to the chromosomal region 1p35-36 (Ben Othmane *et al.*, 1993a) but as yet a gene has not been identified. Locus heterogeneity in CMT2 is suggested by the fact that

approximately 50% of families with this clinical form of the disease fail to show linkage to the 1p35-36 region.

The underlying genetic defect has been identified in a small proportion of the CMT types. The known genes responsible for the development of CMT encode major myelin proteins and the mutations have a dominant effect, as often is the case with structural proteins. Although one form of autosomal recessive CMT, namely CMT4A, has been mapped, the gene has not been cloned and the general molecular mechanisms of recessive peripheral neuropathies remain unknown. Since loss of function is the general mechanism proposed for autosomal recessive disorders, some of these forms of CMT could be expected to result from mutations in enzymes involved in the complex process of myelin formation and maintenance or in the production of signalling molecules involved in Schwann cell - axon interactions. The identification of such genes lays extremely important ground work in the study of normal function and pathological changes in the peripheral nervous system.

1.5 Kindred studied

A large inter-related group of Gypsy families from the North-West of Bulgaria has provided an opportunity to contribute to this important area of research. This project is part of a major study conducted in

collaboration with the Bulgarian Neurosciences and Brain Research Foundation, the Department of Neurology and the Laboratory of Molecular Pathology at the Medical School in Sofia.

1.5.1 Demographic background of the Gypsy community

The group of Gypsies who form the study population is known to have migrated from Stip, a small town in the present day Republic of Macedonia, in 1886. The community which today numbers about 2000 was founded by a small group of about 50-100 settlers. Although not geographically isolated, the group is endogamous and marriage either with non-Gypsy Bulgarians or with Gypsies from other groups is very rare. These Gypsies have a sedentary lifestyle. Their history and religion prior to settling in the region are not known.

1.5.2. Clinical presentation of CMT in the Gypsy patients

Information on the clinical phenotype has been collected in several different ways:

a) neurological screening of numerous individuals, including all affected persons, performed during field trips;

b) out-patient examinations of six affected individuals independently performed by two neurologists;

c) detailed assessment during hospitalisation of four patients;

d) retrieval of hospital records from previous admissions of several additional patients.

In the majority of cases the disease became manifest during the first decade. The initial symptoms included unsteady gait, weakness in the feet and hands, and numbness. Muscular atrophy first develops in the distal parts of the limbs and gradually progresses to involve proximal muscle groups, leading to severe motor impairment within ten years.

Neurological examination reveals uniform findings in all affected individuals: muscle atrophy, diminished or absent deep tendon reflexes, decreased muscle tone and muscle strength, sensory loss, in particular of the sense of vibration and joint position, and trophic changes in the distal limbs. Hearing loss, especially for high frequencies, becomes manifest by the end of the second decade and progresses to profound deafness in the fourth decade.

Electromyography reveals a profound decrease in motor nerve conduction velocities from an early age, with complete denervation in older patients. Sensory action potentials are unobtainable. The conduction velocities along the facial and the glosso-pharyngeal nerves are diminished.

The neuropathological examination of a sural nerve biopsy, obtained from a 35 year old patient and performed by Prof. P.K. Thomas at the Royal Free Hospital in London, revealed a severe lack of myelinated

fibres with unmyelinated fibres relatively well-preserved. No evidence of hypertrophic changes of the "onion bulb" type was obtained. Electron microscopy demonstrated inclusion bodies of the type reported in Refsum's disease, which in this case were described as non-specific. The conclusion of the neurological report was that the disorder is a severe chronic axonopathy.

The clinical findings in the Gypsy CMT patients reveal interesting features which distinguish the disease from previously described forms of CMT. These include:

a) a marked decrease in nerve conduction velocities which, however, is not combined with the typical neuropathological findings of a demyelinating neuropathy;

b) a hearing loss which is slowly progressive and which has been shown to be sensorineural;

c) the abnormal EMG findings in the facial and other nerves which suggest a more generalised cranial nerve involvement.

1.5.3. Genetic analysis

1.5.3.1. Pattern of inheritance

The overall kindred structure is shown in Appendix B. Although affected individuals appear in the pedigree in all generations, close inspection suggests an autosomal recessive mode of inheritance. Detailed neurological examinations of a number of

the parents of affected individuals, including three couples, has revealed only occasional minor abnormalities (such as insignificantly diminished tendon reflexes). Nerve conduction velocities assessed by EMG in six parental couples were normal, with the exception of one mother who had a nearly 50% decrease in MCV. These findings are consistent with an autosomal recessive transmission of the peripheral neuropathy.

1.5.3.2. Pedigree analysis

The genealogical information has been collected during three field trips and verified through cross-checking of the information obtained from several independent sources. The kindred has a complex structure (Appendix B), with multiple connections between affected branches. The overall level of inbreeding appears to be low, with no consanguineous marriages at a level closer than third cousins. The high frequency of a gene for a rare recessive disorder therefore seems to be due to founder effect and endogamy rather than to immediate preferential inbreeding. The genealogical information obtained suggests the presence of affected individuals at least six generations ago approximately at the time of settling of the group, which indicates that there must have been more than one carrier among the founders of the community.

1.5.3.3. Molecular genetic studies

Linkage analysis using markers located close to known CMT loci has been performed in the affected Gypsy CMT families. The results have demonstrated lack of linkage to the CMT1A, CMT1B, CMT2 and CMT4A loci. Linkage to Cx32 has not been investigated since there is no evidence of X-linked inheritance of the disease.

In conclusion, the neurological findings, as well as the exclusion of linkage to any of the known CMT genes, suggest that the disease segregating in this Gypsy community is a novel type of CMT whose clinical and genetic characteristics can be summed up as follows:

* onset in the first decade initially involving the distal muscles of the legs and gradually progressing proximally, later involving the hands;
* markedly decreased nerve conduction velocities;
* neuropathological findings suggestive of chronic axonopathy;
* cranial nerve involvement, most pronounced in the VII cranial nerve;
* autosomal recessive inheritance;
* high gene frequency in an endogamous community with presumed founder effect.

Given the generalised nature of the defect in this rare form of CMT, identification of the gene mutation causing the disease can be expected to make an important contribution to our understanding of the peripheral nervous system and its disorders.

I.6 Approaches to mapping autosomal recessive disorders

Approaches to mapping disease genes vary depending on a number of variables, such as the pattern of inheritance, the gene frequency, the number of families and affected individuals available for analysis, and the characteristics of the population and families where the disease trait occurs.

These parameters affect the choice of the mathematical approach to estimating the probability of linkage to a specific chromosomal region. However, a common feature of all approaches to mapping is the use of polymorphic markers of known chromosomal location and whose co-segregation with the disease phenotype can be traced. The polymorphic markers which currently are most widely used are a class of DNA repeats known as microsatellites. Microsatellite markers are simple di-, tri- or tetranucleotide units whose polymorphism is related to the variable number of repeats. The availability of such markers is

exponentially increasing with several international consortia working on the identification and mapping of microsatellites (Weber & May 1989; Botstein *et al.*, 1980; Gyapay *et al.*, 1994; Weissenbach *et al.*, 1992; Co-operative Human Linkage Centre (CHLC); NIH/CEPH Collaborative Mapping Group). A map of the human genome indicating the chromosomal location of a large number of microsatellite repeats is already available and is subject to regular updating. The microsatellites on this map can be used as reference points and the position of any disease gene can be established relative to markers of known location. Apart from the large number of alleles and hence the high informativity of this class of polymorphic markers, an additional advantage over classical markers such as the RFLPs is the possibility of utilising PCR technology in the molecular analyses. The unique DNA sequences which precede and follow a microsatellite repeat can be used to design primers for the polymerase chain reaction. Since differences between the alleles are related to the number of repeats, the analysis involves size separation of the PCR products.

A number of strategies have been designed to estimate the probability that a disease gene is located close to a specific polymorphic marker or set of markers:

1.6.1 Linkage analysis

The usual approach to mapping disease genes where no prior knowledge of gene product is available is conventional linkage analysis or positional cloning. The basis of this approach is to trace the segregation of the disease gene and the alleles of a polymorphic marker and, by assessing co-segregation, establish their distance from each other. If the disease gene and the marker investigated are located on different chromosomes they will be inherited independently of each other. If they are located on the same chromosome, the probability of co-inheritance increases as the distance between the two decreases, since the chances of a recombination decline within a small segment of a chromosome. The statistical power of this approach depends on the availability of large pedigrees for analysis. The use of multiple small affected pedigrees is an alternative, however, this is highly dependent on the assumption of locus homogeneity. The results of linkage analysis are presented as LOD scores (logarithm of the odds) in favour of linkage calculated at different recombination frequencies represented by θ. By convention, evidence of linkage is accepted at a lod score of 3 which means odds 1000:1 in favour of linkage (Morton, 1955).

1.6.1.1 Homozygosity mapping

Rare autosomal recessive disorders in closely inbred pedigrees usually occur as the result of the inheritance of the same disease gene from a common ancestor. If the common ancestor is a small number of generations away, the whole chromosomal region containing the disease gene will be transmitted in its original form, i.e. the affected individuals can be expected to be homozygous for both the disease gene and the alleles of the polymorphic markers located in close proximity. This has formed the basis of an approach known as homozygosity mapping (Lander & Botstein, 1987). Unlike conventional linkage analysis, homozygosity mapping allows the localisation of a disease gene through the analysis of a small number of affected individuals, provided that two conditions are fulfilled, namely a high inbreeding coefficient in the family and a low frequency of the disease gene.

1.6.1.2 Segment sharing based on linkage disequilibrium

A recent strategy used in mapping recessive genes in isolated populations is segment sharing (Houwen *et al.*, 1994). This method involves the identification of shared chromosomal segments (haplotypes) which are inherited in a manner that is described as identical by descent (IBD) (Houwen *et al.*, 1994). Mapping by identifying shared segments, or extended regions of

chromosomes which have been inherited from a common founding ancestor, requires three pre-conditions:

1) the disease should be rare,
2) the study population is isolated and
3) the population is derived from a small number of founders.

Unlike homozygosity mapping, segment sharing does not require a high level of inbreeding and so it is believed to be most appropriate in cases when the affected individuals are related via a common ancestor six to ten generations ago. Calculations are made on the probability of several distantly related individuals having inherited the same chromosomal segment from a common ancestor or a set of ancestors, and a set of formulae have been developed which calculate the probability of a false positive finding, i.e. a segment inherited IBD which however does not contain the disease gene (Houwen *et al.*, 1994). A major benefit of this mapping technique is that only a very small number of individuals need to be analysed in order to achieve a statistically meaningful result.

1.7 Aim of the research

The aim of this project was the localisation of the gene for a novel form of Charcot-Marie-Tooth disease (CMT) in an isolated human population.

CHAPTER 2. METHODS

2.1 Kindred studied

2.1.1 Pedigree

Pedigree information has been collected by the Laboratory of Molecular Pathology, Sofia, Bulgaria during three field trips. The information obtained from at least two independent sources has been cross-referenced to verify the connections between separate branches of the pedigree. The structure of the final diagram has been confirmed by several key informants. The full diagram of the kindred is shown in Appendix B. For the purposes of this study, a simplified version of the pedigree (Appendix C), including most of the affected branches, has been compiled by Dr. Luba Kalaydjieva.

The inbreeding coefficient for this pedigree was calculated using the computer program Kinship (Boyce, 1983). The gene frequency for the Bulgarian Gypsy population was calculated using Hardy-Weinberg: $p^2 + 2pq + q^2 = 1$ (where p represents the normal gene frequency and q the disease gene frequency).

2.2 Laboratory procedures

In order to obtain the marker allele segregation data for analysis, the procedures followed were:

- PCR to amplify the different specific regions of DNA
- PAGE to separate the amplified DNA fragments by size
- Visualisation of the separated fragments

Assignment of allele numbers to the fragments

Entry of the alleles on to the pedigrees

Haplotype analysis

DNA had been extracted from peripheral blood lymphocytes by the Laboratory of Molecular Pathology, Medical School, Sofia, Bulgaria. The concentration of each sample had been determined. The range of DNA concentration across the samples varied from 16 μg to 755 μg. The list of DNA concentrations per sample are in Appendix D. A stock bank of DNA at the original concentrations was maintained at -20°C. A working solution was made by diluting with sterile H_2O (Baxter) to a concentration of 2.5μg/mL.

2.2.1 DNA amplification via polymerase chain reaction (PCR)

The method of amplifying specific regions of DNA is the polymerase chain reaction (PCR), devised by Mullis & Faloona (1987). This method utilises an enzyme to exponentially amplify a DNA sequence. The basic requirement is knowledge of part of the sequence at the ends of the target DNA region, so that initiation of the extension of the complement to the template strand can be accomplished. These sequences, referred to as

PCR primers, bind to their complementary base pairs starting at the 5' ends of the single target strands. The target DNA defined by the primers is enzymatically amplified through a series of denaturing, annealing and extension steps. These three steps are performed at different temperatures and form one PCR cycle. The standard PCR protocol usually includes thirty cycles. The number of copies of the target sequence increases exponentially, which allows easy identification and visualisation using any of a number of techniques e.g. autoradiography, fluorescence.

Polymorphic DNA markers (average heterozygosity of 79%) from the Genethon map were used to probe family DNA for evidence of linkage to the disease gene. The PCR primers for these markers were synthesised by Research Genetics Laboratories, California.

Protocols for the amplification of the specific DNA marker sequences using PCR are detailed in Appendix F.

Protocols for using the capillary system of PCR are in Laboratory protocols in Appendix F.

2.2.2 Methods of detection

To visualise the PCR products after electrophoretic separation, the following three labelling methods were used:

- radioactive nucleotide incorporation
- fluorescent deoxyuridine triphosphate ([f]dUTP) incorporation
- fluorescent end-labelled primers.

Radioactive nucleotide and [f]dUTP methods use the same principles, in that the labelled nucleotides are incorporated (along with the other free dNTPs) by complementary base pairing into each new strand of DNA as it is synthesised by the polymerase. For visualisation in the case of dUTP labelling, one of three rhodamine dyes (R110, R6G or Tamra) are added to the PCR master mix. The same thermal cycling conditions are used as in the radioactive labelling. The fluorescent end-labelled primers differ because the label is not incorporated along the strand, it is only present at the 5' end of each primer. These dyes are attached via a linker molecule during synthesis of the primers.

2.2.3 Polyacrylamide gel electrophoresis (PAGE)

Under the influence of an electrical current the negatively charged DNA moves towards the positive electrode. In the case

of PAGE, the gel is set up in a vertical orientation and the amplified PCR products are loaded into wells at the top. Migration of the products is dependent upon the pore size of the gel matrix, the size of the product, and the level of electrical current applied. The larger the product fragment, the slower is its migration through the gel. The choice of 4%, 6% or 8% (W/V) polyacrylamide depended on the sizes of DNA fragments to be separated. Most allele sizes fell into the range of 100 base pairs (bp) to 250 bp, which were suitable for running on 6% (W/V) gels. Alleles smaller than 100 bp gave better separation on 8% (W/V) gels, as the pore sizes resulting from the cross-linked monomers were smaller. Similarly, 4% gels with larger pore sizes were more suited to fragments > 250 bp.

2.2.4 Visualisation

Visualisation of the separated fragments was dependent upon which of the methods of labelling were used (radioactive or fluorescent).

Radioactive method: The alleles are visualised as dark lines on an autoradiograph after the dried gel is exposed overnight at room temperature to X-ray film and developed. Details are listed in "Post-run procedures radioactive labelling" in Appendix E.

Fluorescent methods: The alleles are automatically sized relative to an internal size standard (in number of base pairs) by the Genescan software, which is used in conjunction with the ABI Prism™ 373 DNA Sequencer on which the fluorescent products were run. The principles of operation are:

2.2.4.1 Fluorescent methods (Principles of operation of the ABI Prism™ 373 DNA Sequencer)

A vertical polyacrylamide gel is loaded with the fluorescently-labelled PCR products which migrate downwards to pass a laser-scanning window. Several markers can be run in each lane as differentiation is via the size of the alleles and the different colours of the fluorescent dyes. The alleles are sized relative to an in-lane sizing standard (Perkin-Elmer). Sizes are recorded in base pairs to an accuracy of 2 decimal points. This has the additional benefit of standardising sizes between lanes and gels. An argon laser which moves on a track behind this scanning window excites the fluorescent particles which emit light. These light wavelengths are detected by a photomultiplier tube (PMT) and are converted to an electrical signal which is digitally stored by a Macintosh computer connected to the sequencer.

The visualisation of the varying intensities of fluorescence is via peaks which are graphically displayed on an electrophoretogram.

The raw data collected during scanning are down-loaded during the run to the computer and stored in a collection file on a portable hard disk for later processing. The processed data are available in tabular form as well as the electrophoretogram display.

2.2.5 Assignment of alleles

Radioactive method: The lanes were numbered with the pedigree numbers on the autoradiograph. The allele numbers were then assigned with number one allele designated as the uppermost band and the largest sized fragment.

Fluorescent methods: The sized alleles were re-numbered in the same manner as the radioactively labelled alleles, that is with the largest fragment numbered one followed by consecutive numbers as the sizes reduce. This is not usually necessary but was done for consistency in allocating haplotypes.

The assigned alleles were then entered on to pedigree charts using the computer program Cyrillic. The marker data were

entered in the same order as the markers appear on the chromosome map (Genethon) so that the haplotypes were shown.

2.2.6 Haplotypes

Marker allele data were entered for each chromosome, with marker numbers in consecutive order from the telomeric tip of the p arm to the tip of the q arm. Haplotypes were formed by first setting phase (where possible) for the key affected members 38, 13 and 4 for each set of marker alleles such that the paternal allele was on the left. Where phase was unknown, e.g. where both parents were heterozygous for the same alleles, it was indicated on the pedigree. For the sake of consistency, haplotypes were then formed with the inherited paternal haplotype of each key individual placed in the left position of the father's two chromosomes. Similarly, the inherited maternal haplotype was positioned on the left of the mother's two chromosomes.

CHAPTER 3. RESULTS

3.1 Choice of approach

In order to gain more information on the likely genetic structure of the population and to find the most appropriate method of analysis, the computer program "Kinship" (Boyce, 1983) was used to calculate individual inbreeding coefficients, and the average inbreeding coefficient for the population. The inbreeding coefficients for most individuals were zero, however the values for patients 2 and 4 were $F = 0.003$ (this was expected due to the parents being third cousins). The kinship coefficients for pairs of individuals were also calculated. These are summarised in **Table 3.** Appendix G details the output from the Kinship program including the common ancestor pathways.

Table 3 Kinship coefficients

First individual	Second individual	Kinship coefficient
38	13	0.001
38	4	0.002
13	4	0.016
26	4	0.004
36	26	0.016

The gene frequency for the Bulgarian Gypsy population (approximately two thousand in number with fifteen known affected members) was estimated using the Hardy-Weinberg formula ($p^2 + 2pq + q^2$) with $q^2 = 15/2000$ (15 affected individuals in total population of 2000) therefore,

the disease gene frequency , q = √15/2000 = 0.086. This high disease gene frequency in the population together with the low inbreeding coefficient in the sub-population under investigation makes homozygosity mapping an unsuitable choice of strategy for this study. As previously stated, homozygosity mapping relies on a low gene frequency and a high inbreeding coefficient (Lander & Botstein, 1987). Conventional linkage analysis was not feasible as a screening approach for the pedigree as a whole due to the extensive number of inter-family connections. These relationship loops result in impracticably long computer processing times (e.g. 24 hours for 1 marker). However, this method can be used to a minor extent as an end stage screening on a small number of markers, with only the main inter-family connections included.

3.2 Analysis strategy

3.2.1 Overall design of the analysis

With the above restrictions in mind, it was decided that the rapid genome screening approach based on searching for shared segments, reported by Houwen and associates (1994) and as described in Chapter 1, was appropriate for this population. Collaboration with one of the initiators of the method, Lodewijk Sandkuijl, was obtained. The collaboration involved the selection of the most informative members of the pedigree. This selection was made from a simplified version of the original pedigree. The highlighted individuals are shown in Appendix C.

The analysis followed four stages:

- Preparatory stage : The preparatory stage of the study involved the selection of individuals.
- Stage 1: This stage of the analysis involved identification of "mini-haplotypes", i.e. identical segments (regions bounded by two adjacent markers) shared by at least three of six affected (non-sibling) chromosomes.
- Stage two: These two-marker haplotypes were extended internally by running extra markers at closer intervals (~ 3 cM apart).
- Stage three : The extended pedigree (47 members) was run for those markers involved in shared haplotypes which remained in three or more non-sibling chromosomes.

3.2.1.1 Preparatory stage : Selection of individuals

The individuals were selected according to criteria based on their degree of relatedness being present but low (that is they shared a traceable distant common ancestor). The individuals chosen were related through a total of six possible pathways, however the five pathways with the minimum number of steps were used in the calculation of average separation of 9.4 generations or meioses from the closest common ancestor. This pre-condition

reduced the amount of segment sharing by chance alone and this was determined in the following manner:

Individuals numbered on the pedigree as 38, 13 and 4 were designated as the key members. These individuals are from three nuclear families, the first containing both parents and one affected daughter, the second containing a mother and two affected siblings, and the third containing both parents and two affected siblings. Parents 45, 47, 10, 3 and 1 were included for determination of haplotypes and affected siblings; 49 and 2 were included for comparison of haplotypes (e.g. if an affected sib did not share the same region it probably would not be linked to the disease gene). The relatedness pathways are illustrated by five different colours as shown on the pedigree in Appendix C and are as follows:

Path 1	red	38 via 47 to 13 via 10	10 steps
Path 2	green	38 via 45 to 4 via 1	9 steps
Path 3	blue	38 via 45 to 4 via 3	11 steps
Path 4	yellow	13 via father to 4 via 1	6 steps
Path 5	pink	4 via 1 to 4 via 3	11 steps

The average number of steps (each step represents a meiotic event or generation) 47/5 = 9.4.

DNA was used from affected siblings to obtain additional information on haplotypes. For example, in regions where three or more of six chromosomes were identical in affected persons but where siblings showed different haplotypes in these areas, it was an indicator that it was unlikely to contain the disease gene. Conversely, where the affected siblings showed the same haplotypes, it indicated that the region may have contained the disease gene. DNA from the parents of affected members was used where available, in order to determine the parental origin of alleles (to set phase) to assist in tracing inheritance routes and for the construction of haplotypes.

3.2.1.2 Probability of false positive findings

Several formulae (Houwen *et al.*, 1995) were used to arrive at an overall probability of finding a false positive anywhere in the genome i.e. an IBD region which does not contain the disease gene. The initial formula

$$p_1(x) = 0.5^g(1-0.01x)^g$$

was used to calculate the approximate probability of finding a random chromosomal segment between two markers x cM apart that was identical to that of an ancestral chromosome g generations ago (Houwen *et al.*, 1995). For example, the probability (p_1) of one random

11 cM segment (not containing the disease gene) being inherited identical by descent (IBD) from an ancestor 9.4 generations ago is 1 in 2000 (**Table** 4). A second calculation p_2 estimated the probability that k **of** n chromosomes transmitted via a common ancestor g generations ago would be IBD between two markers x cM apart where

$$p_2 (k, x) = (nCk) . (p_1(x))^k . (1 - p_1(x))^{n-k}.$$

Thus the probability of **three out of six** chromosomes sharing an 11 cM IBD segment (not containing the disease gene) from 9.4 generations ago was calculated at 2.42^{-9} (**Table** 4).

However, as the IBD segment could have come from **any one of four ancestral chromosomes,** the following formula was used to evaluate the probability of its occurrence in **three or more** chromosomes:

$$p_3(k,x) = 4. \sum p_2 (i,x).$$

This gave a total probability of an 11 cM segment (not containing the disease gene) being inherited IBD in three to six chromosomes (i.e. the sum of the probabilities for k = 3 to k = 6) as 9.69×10^{-9} **(Table 4).**

Table 4 PROBABILITY OF FALSE POSITIVE IBD FINDINGS - SEGMENT SHARING METHOD

Distance between markers x cM	Marker Frequency	p1	p2(n=6;k=3)	p2(n=6;k=4)	p2(n=6;k=5)	p2(n=6;k=6)	Sum of p2 for k = 3..6	p3 4*sum p2 for k=3..6	p3 x frequency
2	2	.22E-03	3.66E-08	3.36E-11	1.65E-14	3.37E-18	3.66E-08	1.46E-07	**2.93E-07**
4	3	1.01E-03	2.05E-08	1.55E-11	6.25E-15	1.05E-18	2.05E-08	8.19E-08	**2.46E-07**
6	5	8.27E-04	1.13E-08	7.02E-12	2.32E-15	3.21E-19	1.13E-08	4.52E-08	**2.26E-07**
7	7	7.48E-04	8.36E-09	4.70E-12	1.41E-15	1.76E-19	8.37E-09	3.35E-08	**2.34E-07**
8	4	6.76E-04	6.16E-09	3.13E-12	8.46E-16	9.54E-20	6.17E-09	2.47E-08	**9.87E-08**
9	8	7.48E-04	8.36E-09	4.70E-12	1.41E-15	1.76E-19	8.37E-09	3.35E-08	**2.68E-07**
10	7	5.50E-04	3.32E-09	1.37E-12	3.01E-16	2.76E-20	3.32E-09	1.33E-08	**9.29E-08**
11	6	4.95E-04	2.42E-09	8.99E-13	1.78E-16	1.47E-20	2.42E-09	9.69E-09	**5.81E-08**
12	7	4.45E-04	1.76E-09	5.88E-13	1.05E-16	7.78E-21	1.76E-09	7.05E-09	**4.93E-08**
13	8	4.00E-04	1.28E-09	3.83E-13	6.12E-17	4.08E-21	1.28E-09	5.11E-09	**4.08E-08**
14	11	3.59E-04	9.21E-10	2.48E-13	3.56E-17	2.13E-21	9.21E-10	3.69E-09	**4.05E-08**
15	3	3.21E-04	6.62E-10	1.60E-13	2.05E-17	1.10E-21	6.63E-10	2.65E-09	**7.95E-09**
16	4	2.87E-04	4.75E-10	1.02E-13	1.18E-17	5.64E-22	4.75E-10	1.90E-09	**7.59E-09**
17	3	2.57E-04	3.39E-10	6.52E-14	6.70E-18	2.87E-22	3.39E-10	1.35E-09	**4.06E-09**
18	5	2.29E-04	2.41E-10	4.14E-14	3.79E-18	1.45E-22	2.41E-10	9.62E-10	**4.81E-09**
20	2	1.82E-04	1.20E-10	1.63E-14	1.19E-18	3.60E-23	1.20E-10	4.80E-10	**9.59E-10**
21	1	1.61E-04	8.41E-11	1.02E-14	6.58E-19	1.77E-23	8.41E-11	3.36E-10	**3.36E-10**
22	2	1.43E-04	5.87E-11	6.31E-15	3.62E-19	8.63E-24	5.87E-11	2.35E-10	**4.70E-10**
24	1	1.12E-04	2.82E-11	2.38E-15	1.07E-19	1.99E-24	2.82E-11	1.13E-10	**1.13E-10**
25	1	9.91E-05	1.94E-11	1.44E-15	5.72E-20	9.45E-25	1.94E-11	7.77E-11	**7.77E-11**
26	2	8.73E-05	1.33E-11	8.72E-16	3.05E-20	4.43E-25	1.33E-11	5.32E-11	**1.06E-10**
29	1	5.92E-05	4.14E-12	1.84E-16	4.35E-21	4.29E-26	4.14E-12	1.66E-11	**1.66E-11**
30	1	5.18E-05	2.78E-12	1.08E-16	2.24E-21	1.93E-26	2.78E-12	1.11E-11	**1.11E-11**
	94	9.47E-03	1.65E-05	1.18E-07	4.53E-10	7.22E-13	1.03E-07	4.12E-07	**1.67E-06**

Average no. generations (g) = 9.4; k = No. of shared segments; n = Total No. of chromosomes (6)

Table 5 PROBABILITY OF FALSE POSITIVE IBD FINDINGS - SEGMENT SHARING METHOD

Distance between markers x cM	Marker Frequency	p1	p2(n=6;k=3)	p2(n=6;k=4)	p2(n=6;k=5)	p2(n=6;k=6)	Sum of p2 for k = 3..6	p3 4*sum p2 for k = 3..6	p3 x frequency
1	3	1.35E-03	4.87E-08	4.92E-11	2.65E-14	5.97E-18	4.87E-08	1.95E-07	5.84E-07
2	4	1.22E-03	3.66E-08	3.36E-11	1.65E-14	3.37E-18	3.66E-08	1.46E-07	5.85E-07
3	5	1.11E-03	2.74E-08	2.29E-11	1.02E-14	1.89E-18	2.74E-08	1.10E-07	5.48E-07
4	6	1.01E-03	2.05E-08	1.55E-11	6.25E-15	1.05E-18	2.05E-08	8.19E-08	4.91E-07
5	2	9.14E-04	1.52E-08	1.04E-11	3.82E-15	5.83E-19	1.52E-08	6.09E-08	1.22E-07
6	6	8.27E-04	1.13E-08	7.02E-12	2.32E-15	3.21E-19	1.13E-08	4.52E-08	2.71E-07
7	10	7.48E-04	8.36E-09	4.70E-12	1.41E-15	1.76E-19	8.37E-09	3.35E-08	3.35E-07
8	4	6.76E-04	6.16E-09	3.13E-12	8.46E-16	9.54E-20	6.17E-09	2.47E-08	9.87E-08
9	8	7.48E-04	8.36E-09	4.70E-12	1.41E-15	1.76E-19	8.37E-09	3.35E-08	2.68E-07
10	6	5.50E-04	3.32E-09	1.37E-12	3.01E-16	2.76E-20	3.32E-09	1.33E-08	7.97E-08
11	6	4.95E-04	2.42E-09	8.99E-13	1.78E-16	1.47E-20	2.42E-09	9.69E-09	5.81E-08
12	7	4.45E-04	1.76E-09	5.88E-13	1.05E-16	7.78E-21	1.76E-09	7.05E-09	4.93E-08
13	8	4.00E-04	1.28E-09	3.83E-13	6.12E-17	4.08E-21	1.28E-09	5.11E-09	4.09E-08
14	11	3.59E-04	9.21E-10	2.48E-13	3.56E-17	2.13E-21	9.21E-10	3.69E-09	4.05E-08
15	4	3.21E-04	6.62E-10	1.60E-13	2.05E-17	1.10E-21	6.63E-10	2.65E-09	1.06E-08
16	2	2.87E-04	4.75E-10	1.02E-13	1.18E-17	5.64E-22	4.75E-10	1.90E-09	3.80E-09
17	3	2.57E-04	3.39E-10	6.52E-14	6.70E-18	2.87E-22	3.39E-10	1.35E-09	4.06E-09
18	5	2.29E-04	2.41E-10	4.14E-14	3.79E-18	1.45E-22	2.41E-10	9.62E-10	4.81E-09
20	1	1.82E-04	1.20E-10	1.63E-14	1.19E-18	3.60E-23	1.20E-10	4.80E-10	4.80E-10
21	1	1.61E-04	8.41E-11	1.02E-14	6.58E-19	1.77E-23	8.41E-11	3.36E-10	3.36E-10
22	2	1.43E-04	5.87E-11	6.31E-15	3.62E-19	8.63E-24	5.87E-11	2.35E-10	4.70E-10
25	1	9.91E-05	1.94E-11	1.44E-15	5.72E-20	9.45E-25	1.94E-11	7.77E-11	7.77E-11
26	2	8.73E-05	1.33E-11	8.72E-16	3.05E-20	4.43E-25	1.33E-11	5.32E-11	1.06E-10
29	1	5.92E-05	4.14E-12	1.84E-16	4.35E-21	4.29E-26	4.14E-12	1.66E-11	1.66E-11
30	1	5.18E-05	2.78E-12	1.08E-16	2.24E-21	1.93E-26	2.78E-12	1.11E-11	1.11E-11
	109	1.27E-02	1.94E-07	1.55E-10	7.00E-14	1.37E-17	1.94E-07	7.77E-07	3.60E-06

Average no. generations (g) = 9.4; k = No. of shared segments; n = Total No. of chromosomes (6)

Finally, the overall probability of finding a segment which is shared IBD, but does not contain the gene, having run 89 markers covering 94 intervals at varying lengths and frequencies, was calculated using p3 x marker frequency for each interval. Houwen & colleagues noted that this formula does not take into account the probability of a false IBD finding in one interval not being independent of a similar finding in an adjacent interval. However, the summary of the p3 values over all the intervals covered in Stage 1 of this study resulted in an overall probability of a false positive IBD finding being approximately 1 in 600,000 (1.67×10^{-6})(Table 4). When the extra markers for Stage 2 are taken into account (109 intervals in total), the probability increases to 1 in 278,000. This is due to the increased number of smaller intervals and reduction in larger segments (Table 5). That is, there is a greater chance of false positive findings with smaller intervals between markers.

3.2.2 Feasibility of the study

3.2.2.1 Informativity of polymorphic markers

The polymorphic marker loci information was taken from the Genethon microsatellite map catalogue (Gyapay *et al.*, 1994) and map distances were sex-averaged. The

average distance between the marker loci was 11 cM and the average published heterozygosity was 0.79.

In order to obtain an estimation of the heterozygosity index in this population, and therefore to determine whether the markers used were appropriate, ten unlinked markers were analysed. These markers were run on extended versions of the pedigree as follows: three on chromosome 1 (D1S233, D1S186, D1S228), three on chromosome 8 (not linked to the gene region) (D8S88, D8S279, D8S257), three on chromosome 17 (D17S578, D17S849, D17S796) and one on chromosome 5 (D5S819). The number of alleles per marker versus the number of alleles published for the CEPH families and obtained from the Genome Data Base is listed in **Table 6.**

Table 6 Number of alleles in a sample of 10 markers run for the extended pedigree

Locus name	Actual # alleles	Published # alleles
D1S233	6	9
D1S186	9	10
D1S228	5	7
D5S819	5	Not listed
D8S88	10	10
D8S279	10	10
D8S257	5	9
D17S578	4	11
D17S849	4	6
D17S796	5	8

Allele frequencies were calculated and are shown in **Table 7.** Frequencies could not be compared to the published frequencies as allele numbers were allocated on a relative basis for the radioactively labelled markers.

Table 7 Allele frequencies for 10 unlinked marker loci

n = no. of chromosomes/marker

Allele #		D1S233	D1S186	D1S228	D5S819	D8S88	D8S279	D8S257	D17S578	D17S849	D17S796
	n =	70	90	66	66	98	90	98	62	62	68
1		0.04	0.04	0.07	0.27	0.16	0.15	0.04	0.04	0.24	0.14
2		0.17	0.07	0.04	0.17	0.12	0.15	0.30	0.07	0.60	0.07
3		0.17	0.07	0.26	0.23	0.02	0.05	0.14	0.58	0.04	0.14
4		0.35	0.03	0.52	0.17	0.02	0.05	0.47	0.31	0.12	0.24
5		0.17	0.14	0.11	0.17	0.19	0.02	0.05			0.41
6		0.10	0.14			0.28	0.02				
7			0.31			0.07	0.32				
8			0.17			0.07	0.02				
9			0.03			0.05	0.05				
10						0.02	0.17				

As a measure of the degree of polymorphism of this sample of ten markers **Table 8** shows the heterozygosity calculations. This population is very heterozygous with four of the markers giving a higher observed heterozygosity than the published figures, an indication that these markers are highly polymorphic for this pedigree.

Table 8 Heterozygosity calculations

Marker #	# Hetero-zygotes	Total # individuals	Observed Heterozygosity	Published Heterozygosity
D1S233	29	35	0.83	0.84
D1S186	31	45	0.89	0.84
D1S228	19	33	0.57	0.76
D5S819	30	35	0.86	0.83
D8S88	43	49	0.88	0.82
D8S279	35	45	0.78	0.86
D8S257	33	49	0.67	0.71
D17S578	21	31	0.68	0.63
D17S849	18	31	0.58	0.67
D17S796	21	34	0.62	0.81

Table 9 shows the calculation of the inbreeding coefficient of the pedigree by firstly finding the expected heterozygosity using the

formula $H = 1 - \Sigma p_i^2$ where H is the actual heterozygosity observed and p_i is the allele frequency. Then $F = (H_o - H) / H_o$ where F is the inbreeding coefficient and H_o is the expected heterozygosity for a randomly mating population.

Table 9 Calculation of coefficient of inbreeding for 10 unlinked marker loci

Marker #	Σp_i^2	H_0 (Expected)	H (Actual)	F
D1S233	0.22	0.78	0.83	0.00
D1S186	0.18	0.82	0.89	0.12
D1S228	0.35	0.65	0.57	0.12
D5S819	0.21	0.79	0.86	0.00
D8S88	0.17	0.83	0.88	0.00
D8S279	0.18	0.82	0.78	0.05
D8S257	0.33	0.67	0.67	0.00
D17S578	0.43	0.57	0.68	0.00
D17S849	0.43	0.57	0.58	0.00
D17S796	0.27	0.73	0.62	0.15

This inbreeding coefficient can be interpreted in terms of the probability that any two alleles chosen randomly from this population are identical by descent from a non-inbred ancestral founder i.e. at the point where F was zero (Hartl, 1988). The level of heterozygosity observed was greater than expected in six out of ten of the markers. These inbreeding coefficients are recorded as zero.

3.2.2.2 Homozygosity in the ten selected members

Homozygosity levels were recorded for all the markers run for stages 1, 2 and 3. There were five individuals in each of the two categories, affected and non-affected. A summary of the results comparing these two categories for each chromosome is given in **Table 10.** The markers for chromosome 8 show the highest level of homozygosity in the affected individuals with 20.87% compared to 12.17% homozygosity in non-affected persons.

Table 10 Homozygosity levels in the five affected and five non-affected individuals chosen for the segment sharing method

Chrom.#	Total # markers x total # indivs.	Homozygotes/chromosome			
		Affected		Non-affected	
		No.	%	No.	%
1	160	27	16.9	31	19.4
5	170	15	8.8	15	8.8
8	230	48	20.9	28	12.2
9	160	15	9.4	25	15.6
11	170	22	12.9	15	8.8
17	110	12	10.9	15	13.6
	1000	139	27.8	129	25.8

3.3 Segment sharing analysis of candidate chromosomes

Stage 1 : Segment sharing - identification of shared haplotypes

At this stage haplotypes shared by three or more non-sibling chromosomes were identified. These haplotypes were located over two adjacent markers at an average distance of 11 cM apart. It was also noted at this stage whether or not the affected siblings carried the same alleles for these markers.

Stage 2 : Extension of shared haplotypes

The purpose of this step was to exclude areas in which some of the marker alleles coincidentally were the same size in base pairs (identical by state - IBS) rather than because they were co-inherited (identical by descent - IBD) from a common founding ancestor. This was achieved by running additional polymorphic markers internally to the two-marker haplotypes identified at stage 1.

Stage 3 : Extended pedigree analysis of identified IBD regions

If the stage 2 markers also formed identical, extended haplotypes over three or more non-sibling chromosomes, data were then obtained for the extended pedigree. This was achieved by running the markers which comprised the identified IBD region for

the forty-seven members of the pedigree for whom DNA was available.

The allele assignments were sent to our statistical collaborator Joachim Hallmayer at Stanford University, for calculation of the LOD scores in order to verify or exclude the proposed linkage region.

3.3.1 Criteria for exclusion

The criteria for exclusion of regions as identical by state (IBS) rather than IBD are as follows:

* Extended haplotypes different
* IBD index is 0 (i.e. zero identical segments shared by affected sibs)
* Recombinations within the extended region

3.3.2 Order of analyses

Chromosomes known to contain regions or genes associated with a form of CMT or myelin protein production were first processed. These were chromosomes 1 (CMT1B, CMT2), 8 (CMT4A), 9 (Ciliary Neurotrophic Factor Receptor), 11 (Ciliary Neurotrophic Factor) (Lev *et al.*, 1993; Masu *et al.*, 1993), and 17 (CMT1A). In the

event that these chromosomes were excluded, a search of the remaining autosomes would be undertaken.

3.3.3 Markers analysed

The initial screening process involved the running of 89 markers on chromosomes 1, 5, 8, 9, 11 and 17, the chromosomes which contained candidate genes for CMT or PNS proteins. The fraction of the genome which was covered totalled 1192 centiMorgans (cM) out of a total length of ~ 3817 cM, that is 31.2%. The average distance between markers run was 12.6 cM. The breakdown of distances per chromosome covered is shown in **Table 11**.

Table 11 Distance of genome covered in centiMorgans

Chromosome	Total distance cM	No. of intervals	Average distance cM
1	305	23	13.2
5	233	17	13.7
8	178	17	10.5
9	173	14	12.3
11	155	13	11.9
17	148	10	14.8
	1192	94	12.6

Total genome distance 3817 cM (approximated from Human Screening Set Genethon map catalogue : 347 markers at average spacing of 11 cM)

Percentage of genome covered: 1192/3817 x 100 = 31.2%

Table 12 summarises the details of markers run by giving the number of markers which amplified successfully, the number which failed to amplify, the number which were uninformative (allele origin could not be traced), and the average distance between those markers which amplified successfully (including those which were uninformative).

Table 12 Analysis of markers used for Stage 1 screening

Chrom.#	# which gave results	# which did not amplify	# uninformative results	Average distance* cM
1	23	1	5	13.2
5	18	5	4	13.7
8	15	1	0	10.5
9	13	1	2	12.3
11	14	0	0	11.9
17	11	1	1	14.8

* This is between markers for which results were obtained

Figure 1 Map distances

Chromosome 1

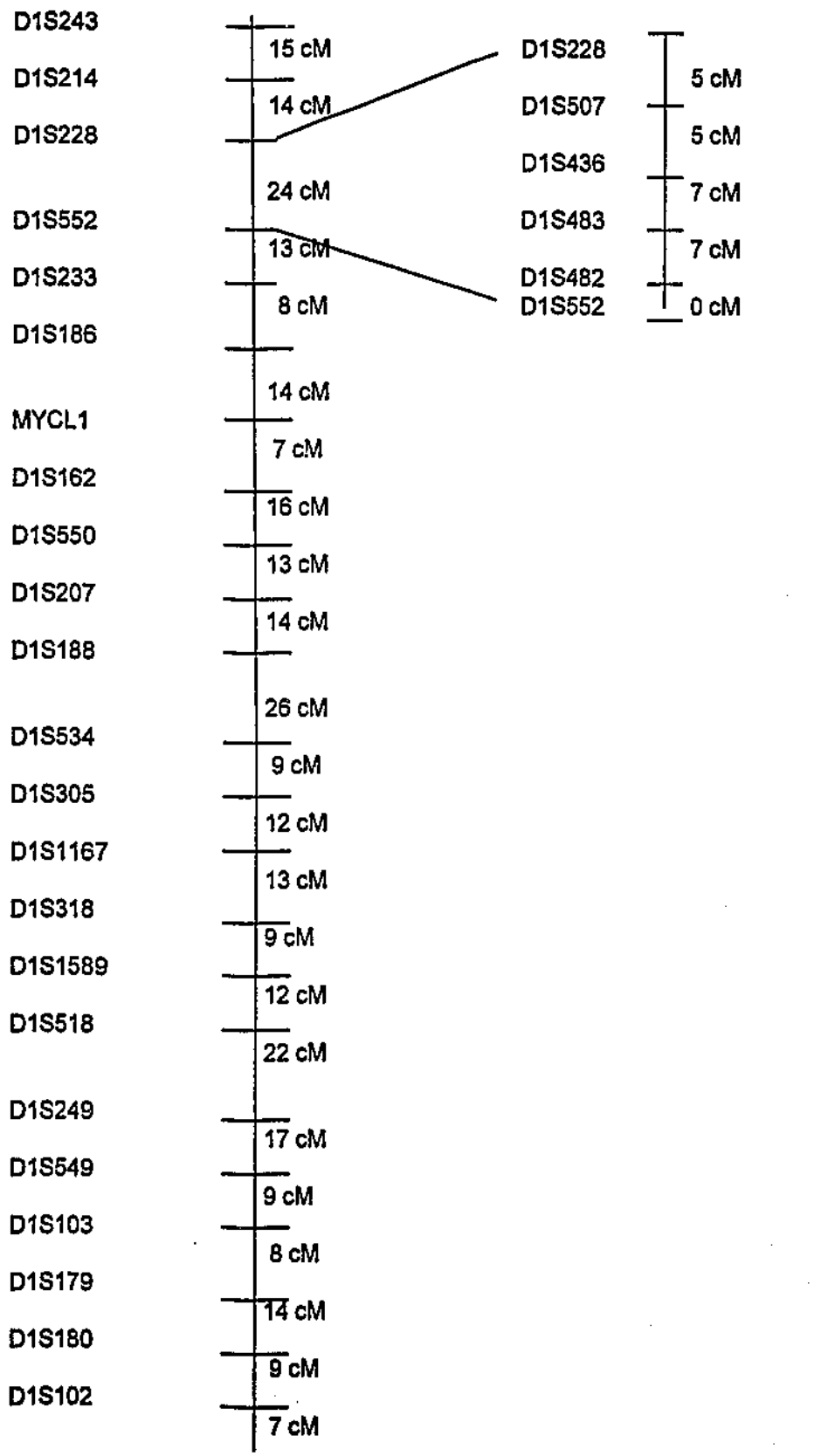


Figure 2 Map distances

Chromosome 5

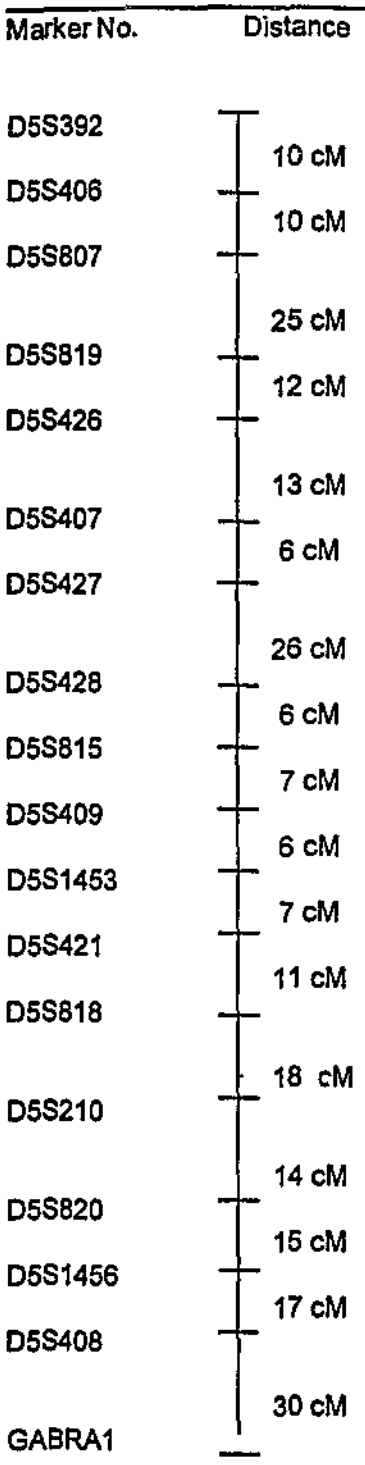


Figure 3 Map distances

Chromosome 8

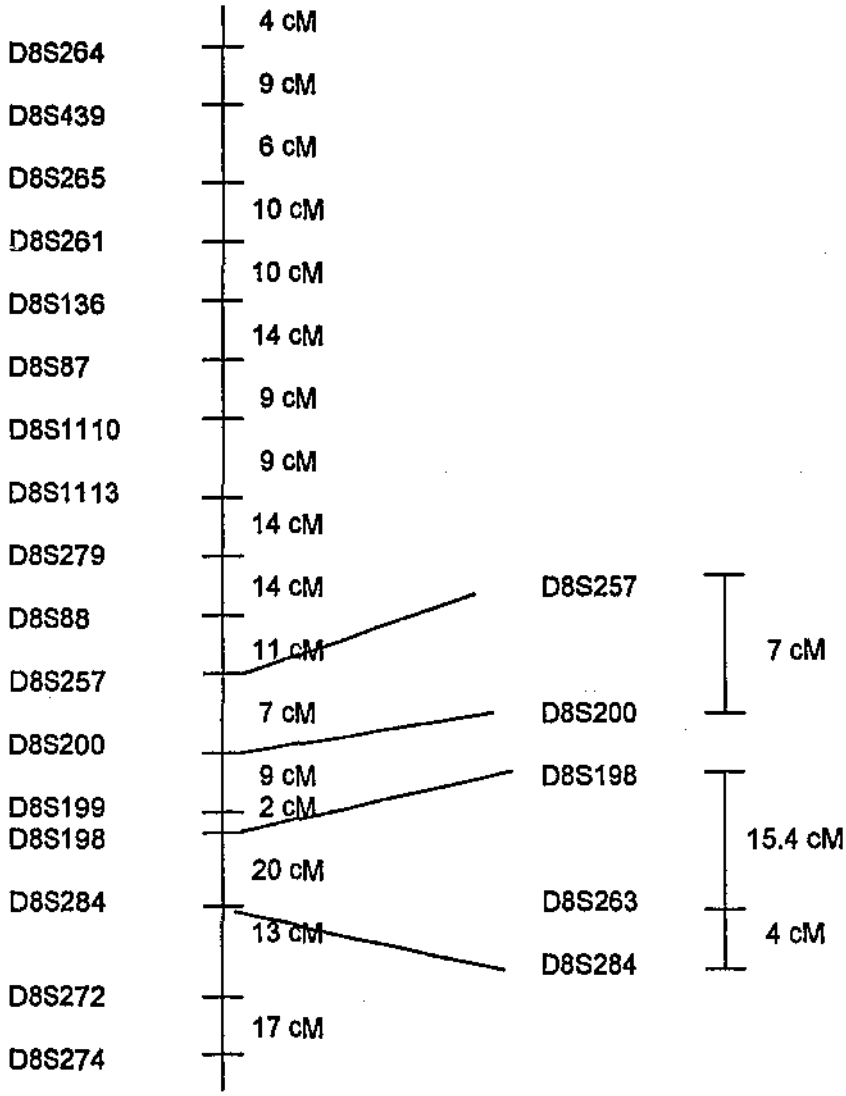


Figure 4 Map distances

Chromosome 9

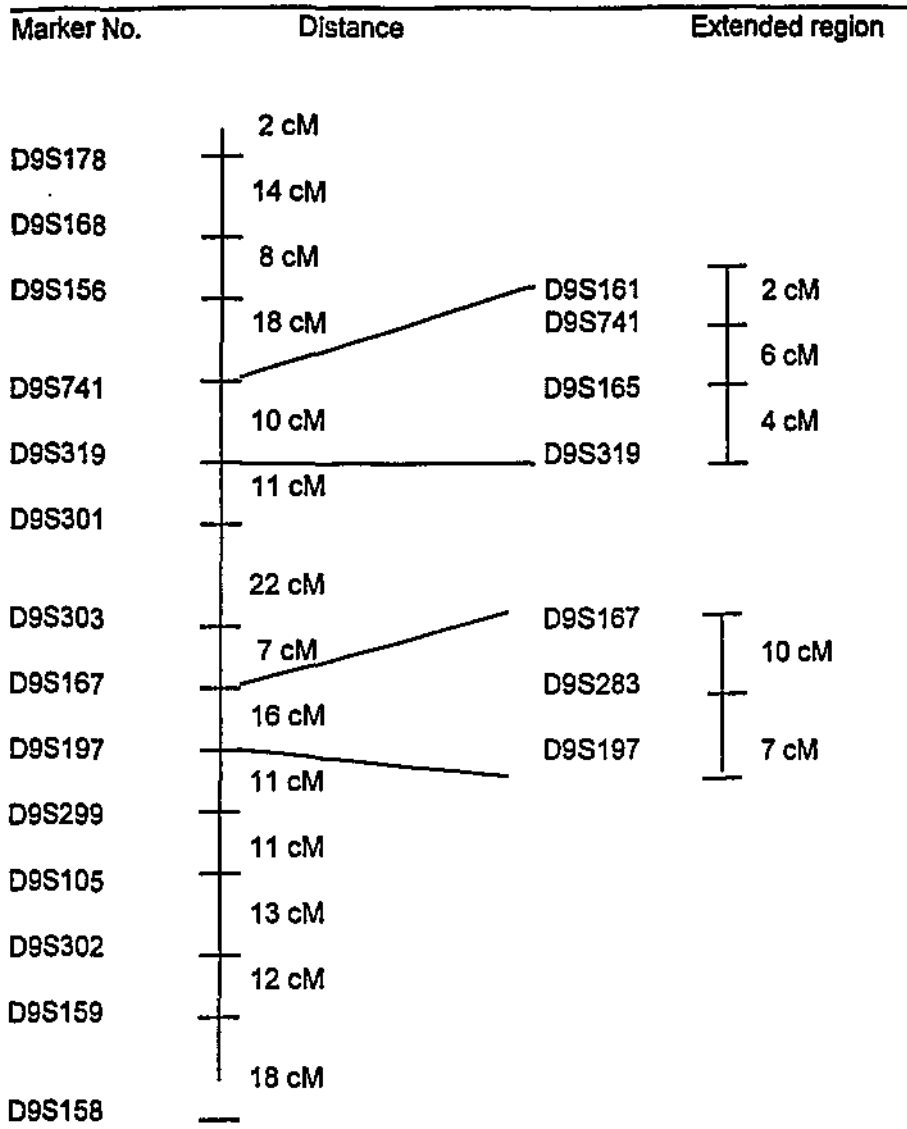


Figure 5 Map distances

Chromosome 11

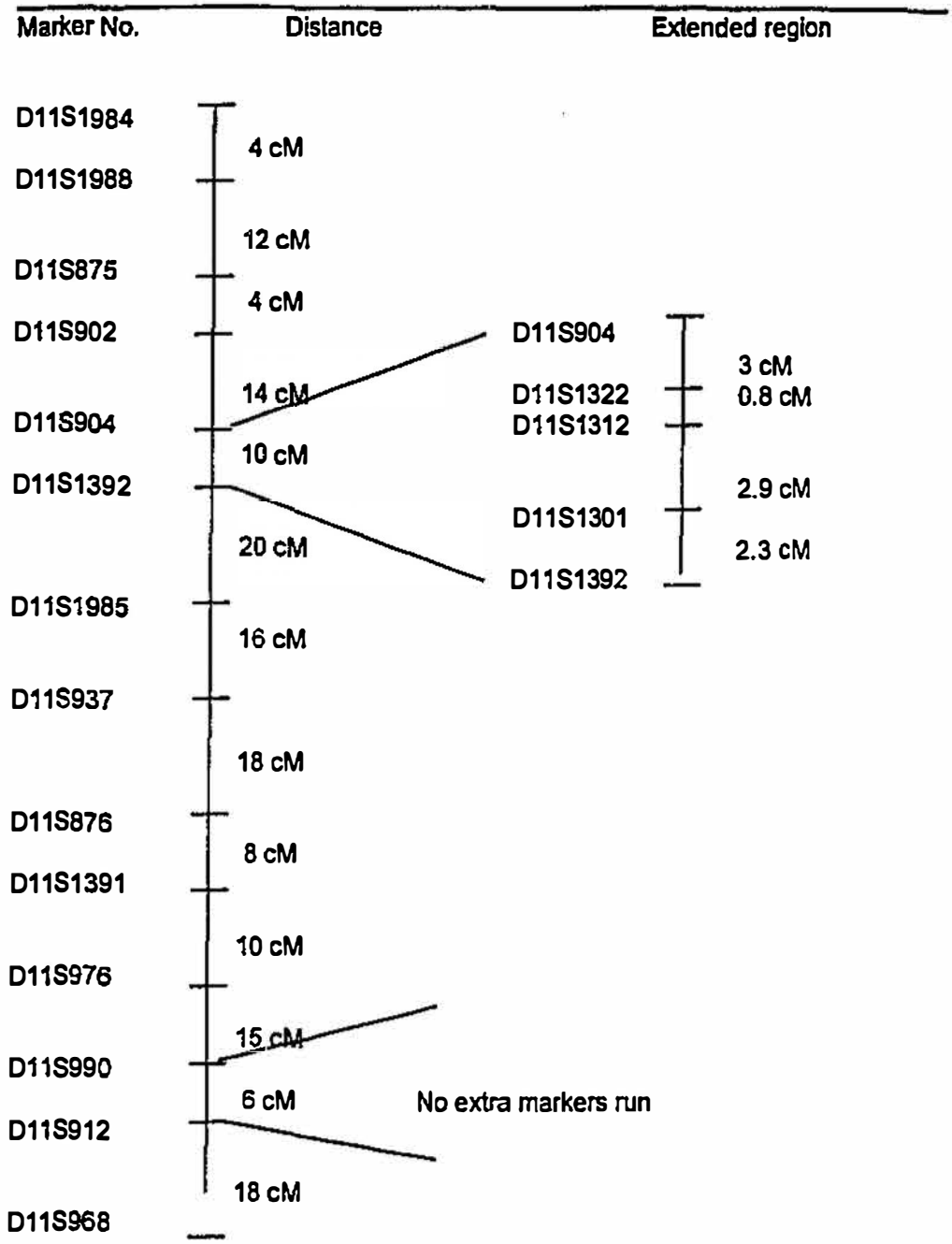
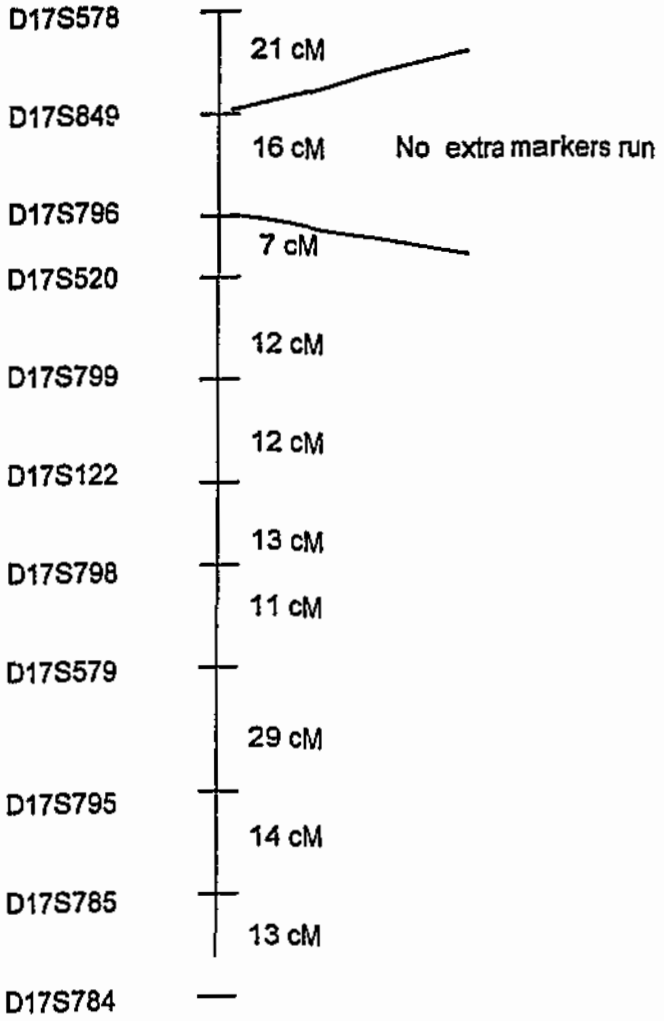


Figure 6 Map distances

Chromosome 17



From the twenty-four markers run on **chromosome 1 (Figure 1)**, one could not be successfully amplified by PCR (D1S248). Five markers were uninformative for these families. In most cases this was caused, by homozygosity in both parents. The average distance between the successfully run markers was 13.2 cM.

Twenty-three **chromosome 5 (Figure 2)** markers were run for the forty-seven family members by our collaborators in Sofia. Of these, five were unsuccessful and of the remaining markers four were uninformative. The average distance between markers was 13.7 cM.

On **chromosome 8 (Figure 3)**, sixteen markers were initially run with fifteen successful (D8S261 did not amplify). The average distance between markers was 10.5 cM.

On **chromosome 9 (Figure 4)**, fourteen markers were run at an average distance of 12.3 cM. One (D9S175) was unable to be amplified and two were uninformative.

Fourteen markers were successfully run on **chromosome 11 (Figure 5)** at an average distance of 11.9 cM apart. Three markers were not run as they were very close to neighbouring markers and six were uninformative.

Eleven successful markers were run on **chromosome 17 (Figure 6)** at an average distance of 14.8 cM. One marker did not amplify (D17S809) and one was situated between two close markers so was not run.

3.3.4 Stage 1 : Identification of shared haplotypes

Nine regions shared by three or more non-sibling chromosomes were identified as possible IBD regions. In order to check whether these regions were inherited IBD, they were saturated with a total of eighteen additional markers. Figures 1 to 6 illustrate the total chromosome, intervals and markers as well as those regions where the number of markers has been extended. These regions occurred on chromosomes 1p, 8q, 9p, 9q, 11p and 17p.

Table 13 summarises the segments identified at stage 1 of the screening process as those which are shared by three or more affected non-sibling chromosomes. It also records whether there is sharing by a sibling and how many copies of each haplotype the sibling has inherited. The lack of sharing of the segment between siblings is an indication that it is unlikely to be inherited IBD.

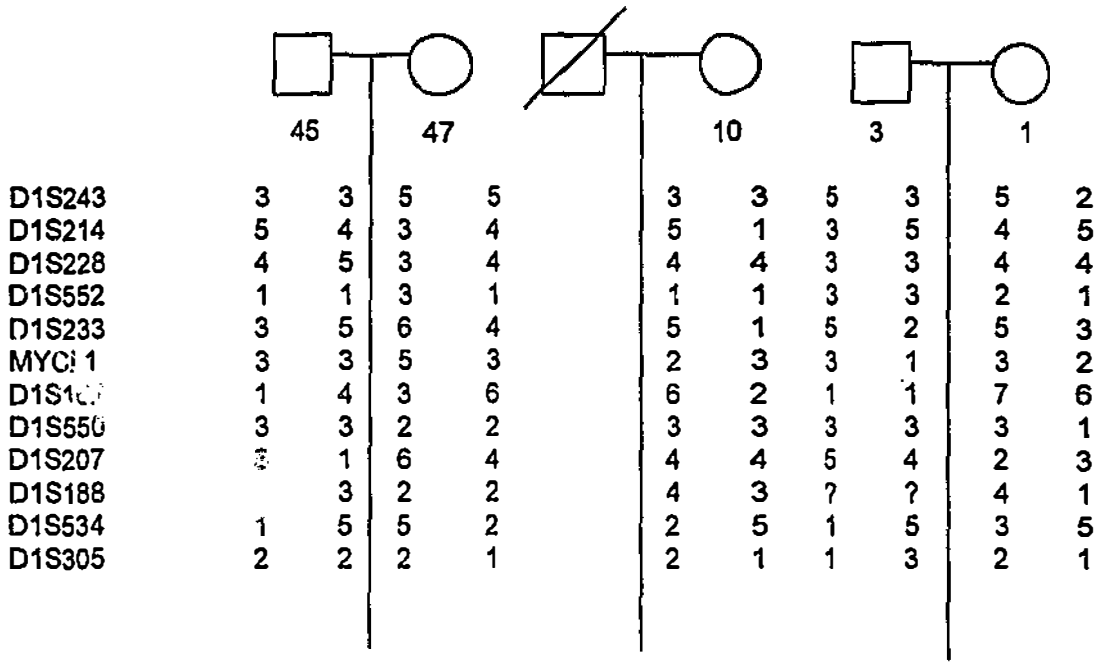
Table 13 Regions which are shared by at least 3 of 6 (non-sibling) chromosomes

Chrom #	First Marker #	Distance cM	Second Marker #	No. of haplo-types shared (non-sib)	Haplotypes shared by siblings Y/N	Ped.#	Chroms each *
1	D1S228	24	D1S552	4/6	Y	13, 49	2
8	D8S257	7	D8S200	3/6	Y	4, 2	2
8	D8S198	13	D8S284	4/6	Y	13,49/4, 2	2/1
9	D9S741	10	D9S319	3/6	N		
9	D9S167	16	D9S197	3/6	Y	13, 49	1
11	D11S904	10	D11S1392	3/6	Y	13, 49	1
11	D11S990	6	D11S912	3/6	Y	13, 49	1
17	D17S849	16	D17S796	3/6	Y	4, 2	2

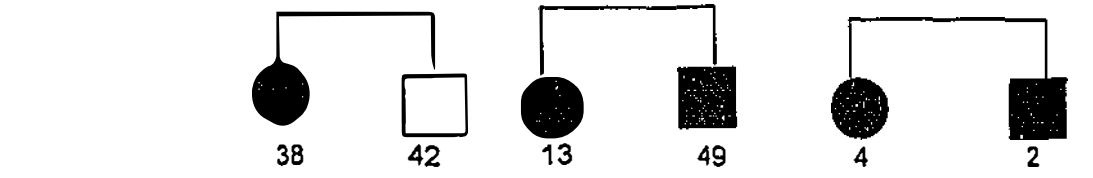
* This indicates whether the affected sibling is homozygous or heterozygous for the alleles of this haplotype.

Figure 7 Stage 1: Segment sharing results
Regions shared by three or more non-sibling chromosomes

Chromosome 1p



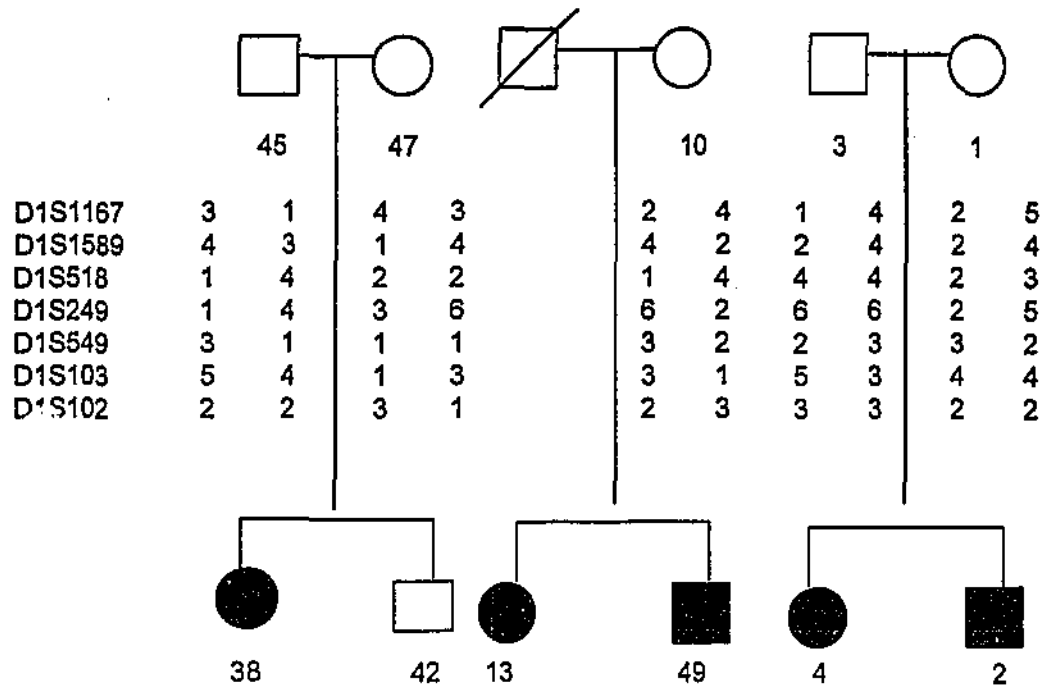
D1S243	3	3	5	5		3	3	5	3	5	2
D1S214	5	4	3	4		5	1	3	5	4	5
D1S228	4	5	3	4		4	4	3	3	4	4
D1S552	1	1	3	1		1	1	3	3	2	1
D1S233	3	5	6	4		5	1	5	2	5	3
MYCL1	3	3	5	3		2	3	3	1	3	2
D1S162	1	4	3	6		6	2	1	1	7	6
D1S550	3	3	2	2		3	3	3	3	3	1
D1S207	[illegible]	1	6	4		4	4	5	4	2	3
D1S188		3	2	2		4	3	?	?	4	1
D1S534	1	5	5	2		2	5	1	5	3	5
D1S305	2	2	2	1		2	1	1	3	2	1



D1S243	3	5		3	3	3	3	5	5	5	5
D1S214	5	3		3	5	3	1	3	4	3	4
D1S228	4	3		4	4	4	4	3	4	3	4
D1S552	1	3		1	1	1	1	3	2	3	1
D1S233	3	6		3	5	4	1	5	5	5	5
MYCL1	3	5		3	2	4	3	3	3	3	3
D1S162	1	3		4	6	5	2	1	7	1	[illegible]
D1S550	3	2		4	3	4	3	3	3	3	[illegible]
D1S207	5	6		2	4	6	4	5	2	4	2
D1S188	3	2		4	4	4	4	?	?	4	4
D1S534	1	5		4	2	4	5	1	3	5	3
D1S305	2	2		1	2	2	1	1	2	3	2

Figure 8 Stage 1 : Segment sharing results

Chromosome 1q

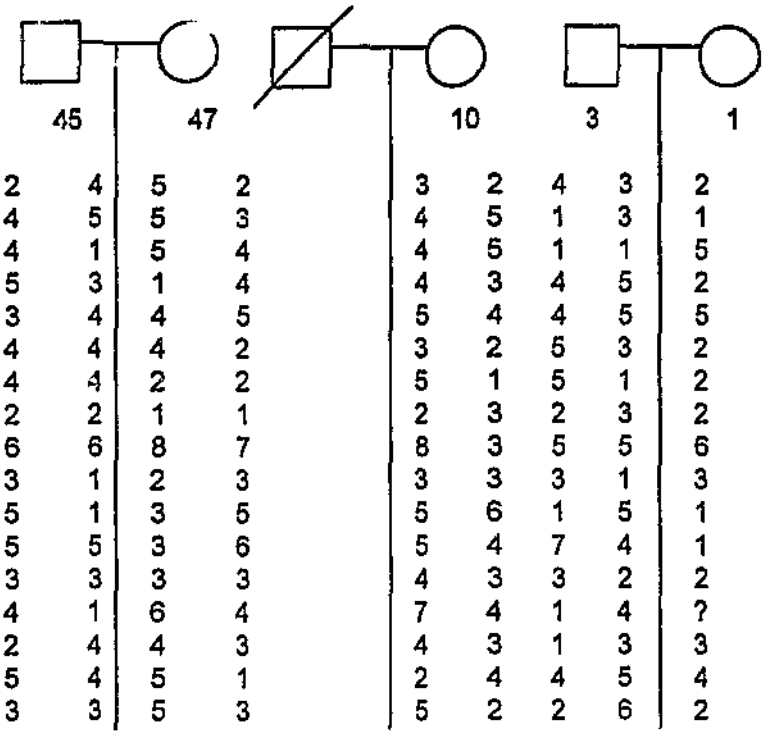


D1S1167	3	1	4	3		2	4	1	4	2	5
D1S1589	4	3	1	4		4	2	2	4	2	4
D1S518	1	4	2	2		1	4	4	4	2	3
D1S249	1	4	3	6		6	2	6	6	2	5
D1S549	3	1	1	1		3	2	2	3	3	2
D1S103	5	4	1	3		3	1	5	3	4	4
D1S102	2	2	3	1		2	3	3	3	2	2

D1S1167	3	4		2	2	4	2	1	2	4	5
D1S1589	4	1		4	4	4	2	2	2	2	4
D1S518	1	2		4	1	1	1	4	2	4	3
D1S249	1	3		1	6	1	6	6	2	6	2
D1S549	3	1		1	3	1	3	2	3	2	2
D1S103	5	1		2	3	2	3	5	4	5	4
D1S102	2	3		3	2	3	2	3	2	3	2

Figure 9 Stage 1 : Segment sharing results

Chromosome 5



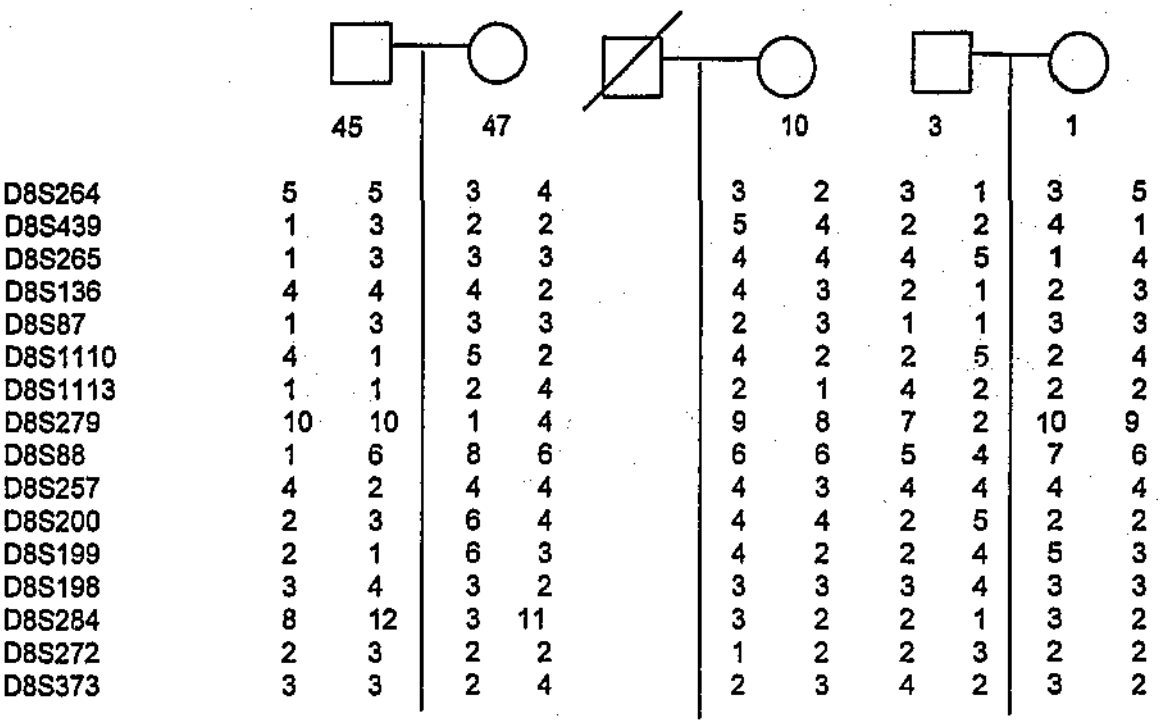
D5S392	2	4	5	2		3	2	4	3	2	3	
D5S406	4	5	5	3		4	5	1	3	1	5	
D5S807	4	1	5	4		4	5	1	1	5	1	
D5S819	5	3	1	4		4	3	4	5	2	4	
D5S426	3	4	4	5		5	4	4	5	5	1	
D5S407	4	4	4	2		3	2	5	3	2	3	
D5S427	4	4	2	2		5	1	5	1	2	1	
D5S428	2	2	1	1		2	3	2	3	2	1	
D5S815	6	6	8	7		8	3	5	5	6	4	
D5S409	3	1	2	3		3	3	3	1	3	2	
D5S1453	5	1	3	5		5	6	1	5	1	1	
D5S421	5	5	3	6		5	4	7	4	1	6	
D5S818	3	3	3	3		4	3	3	2	2	2	
D5S210	4	1	6	4		7	4	1	4	?	?	
D5S820	2	4	4	3		4	3	1	3	3	4	
D5S1456	5	4	5	1		2	4	4	5	4	3	
GABRA1	3	3	5	3		5	2	2	6	2	3	

D5S392	4	2		2	3	5	3	4	3	4	3	
D5S406	5	3		3	4	3	4	1	5	1	5	
D5S807	1	4		3	4	3	4	1	1	1	1	
D5S819	3	4		5	4	5	3	4	2	4	4	
D5S426	4	5		5	5	5	5	4	5	4	1	
D5S407	4	2		4	3	4	3	5	2	5	3	
D5S427	4	2		*1	5	5	5	?	?	5	1	
D5S428	2	1		1	2	1	3	2	2	2	1	
D5S815	6	7		5	8	5	3	?	?	5	4	
D5S409	1	3		2	3	2	3	3	3	3	2	
D5S1453	1	5		?	?	1	5	1	1	1	1	
D5S421	5	6		5	5	4	5	7	1	7	6	
D5S818	3	3		2	4	2	4	3	2	2	2	
D5S210	1	4		4	7	4	7	1	4	1	4	
D5S820	4	3		2	4	2	4	?	?	1	4	
D5S1456	4	1		3	2	3	4	4	3	5	3	
GABRA1	3	3		2	5	2	2	2	3	6	3	

* Phase unknown

Figure 10 Stage 1 : Segment sharing results
Regions shared by three or more non-sibling chromosomes

Chromosome 8



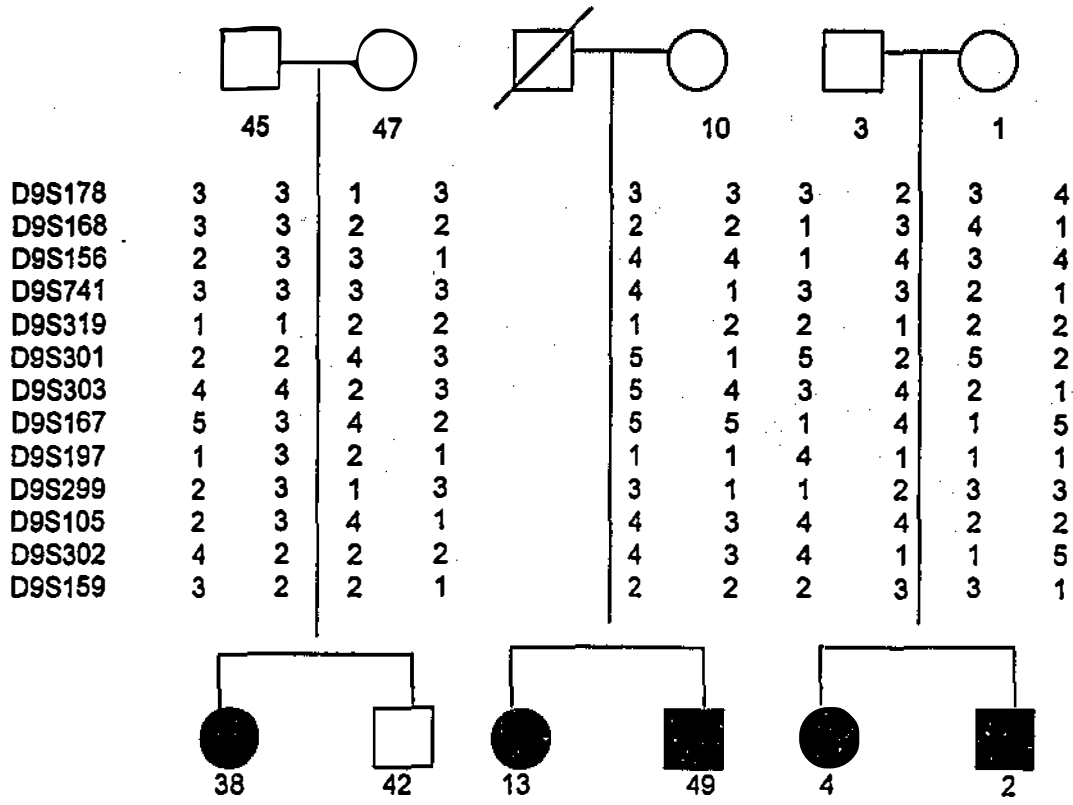
D8S264	5	5	3	4		3	2	3	1	3	5	
D8S439	1	3	2	2		5	4	2	2	4	1	
D8S265	1	3	3	3		4	4	4	5	1	4	
D8S136	4	4	4	2		4	3	2	1	2	3	
D8S87	1	3	3	3		2	3	1	1	3	3	
D8S1110	4	1	5	2		4	2	2	5	2	4	
D8S1113	1	1	2	4		2	1	4	2	2	2	
D8S279	10	10	1	4		9	8	7	2	10	9	
D8S88	1	6	8	6		6	6	5	4	7	6	
D8S257	4	2	4	4		4	3	4	4	4	4	
D8S200	2	3	6	4		4	4	2	5	2	2	
D8S199	2	1	6	3		4	2	2	4	5	3	
D8S198	3	4	3	2		3	3	3	4	3	3	
D8S284	8	12	3	11		3	2	2	1	3	2	
D8S272	2	3	2	2		1	2	2	3	2	2	
D8S373	3	3	2	4		2	3	4	2	3	2	

D8S264	5	3		2	3	2	3	3	3	3	3	
D8S439	1	2		1	5	3	5	2	4	2	4	
D8S265	1	3		2	4	3	4	4	1	5	1	
D8S136	4	4		2	4	2	4	2	2	1	3	
D8S87	1	3		2	2	2	*3	1	3	1	3	
D8S1110	4	5		3	4	3	2	2	2	5	4	
D8S1113	1	2		*1	2	3	1	4	2	2	2	
D8S279	10	1		7	9	3	9	7	10	7	10	
D8S88	1	8		1	6	5	6	5	7	5	7	
D8S257	4	4		4	4	4	4	4	4	4	4	
D8S200	2	6		4	4	5	4	2	2	2	2	
D8S199	2	6		7	4	3	4	2	5	2	5	
D8S198	3	3		3	3	3	3	3	3	3	3	
D8S284	8	3		3	3	3	3	2	3	2	3	
D8S272	2	2		1	1	1	1	2	2	2	2	
D8S373	3	2		1	2	1	2	4	3	4	3	

* Phase unknown

Figure 11 Stage 1 : Segment sharing results
Regions shared by three or more non-sibling chromosomes

Chromosome 9



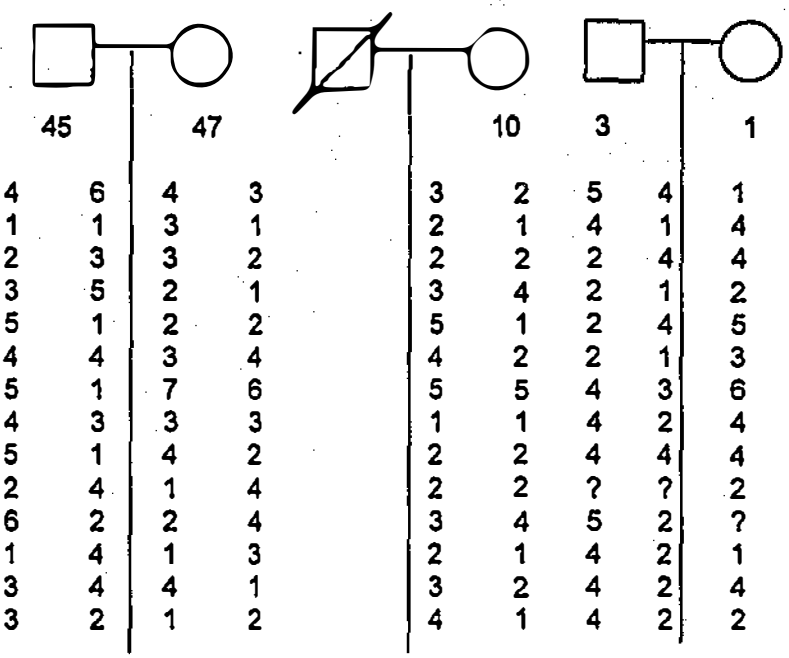
D9S178	3	3	1	3	3	3	3	2	3	4		
D9S168	3	3	2	2	2	2	1	3	4	1		
D9S156	2	3	3	1	4	4	1	4	3	4		
D9S741	3	3	3	3	4	1	3	3	2	1		
D9S319	1	1	2	2	1	2	2	1	2	2		
D9S301	2	2	4	3	5	1	5	2	5	2		
D9S303	4	4	2	3	5	4	3	4	2	1		
D9S167	5	3	4	2	5	5	1	4	1	5		
D9S197	1	3	2	1	1	1	4	1	1	1		
D9S299	2	3	1	3	3	1	1	2	3	3		
D9S105	2	3	4	1	4	3	4	4	2	2		
D9S302	4	2	2	2	4	3	4	1	1	5		
D9S159	3	2	2	1	2	2	2	3	3	1		

D9S178	3	1	3	3	3	3	3	3	2	3		
D9S168	3	2	2	2	3	2	1	4	3	4		
D9S156	2	3	3	4	1	4	1	3	4	4		
D9S741	3	[3]	[3]	4	2	4	[3]	2	3	1		
D9S319	1	[2]	*[2]	1	1	1	[2]	2	1	2		
D9S301	2	4	3	5	3	5	5	5	2	2		
D9S303	4	2	3	5	3	4	3	2	4	1		
D9S167	[5]	4	1	[5]	1	5	1	1	4	[5]		
D9S197	[1]	2	4	[1]	4	1	4	1	4	[1]		
D9S299	2	1	*1	3	3	3	1	3	1	3		
D9S105	2	4	*3	4	3	4	4	2	4	2		
D9S302	4	2	*3	4	2	3	4	1	1	5		
D9S159	3	2	3	2	3	2	?	?	3	1		

* Phase unknown

Figure 12 Stage 1 : Segment sharing results
Regions shared by three or more non-sibling chromosomes

Chromosome 11



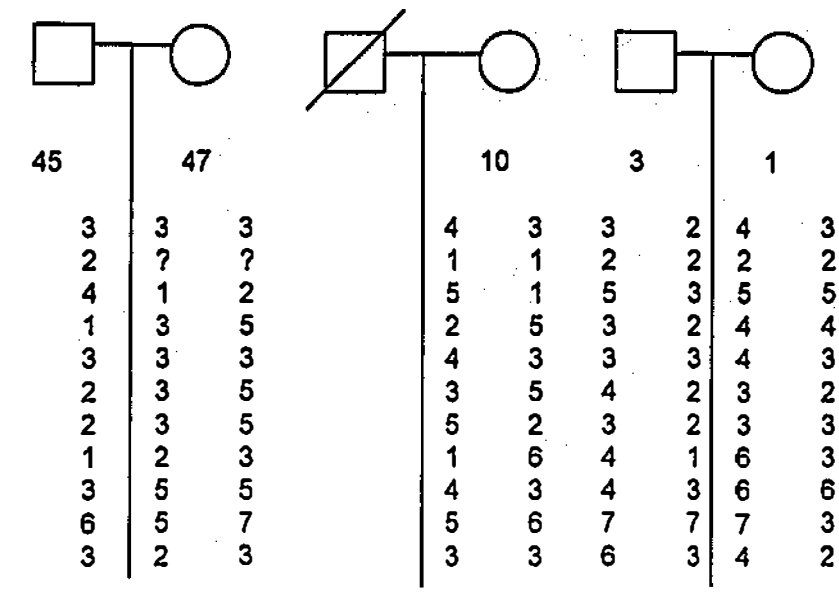
D11S1984	4	6	4	3		3	2	5	4	1	4
D11S988	1	1	3	1		2	1	4	1	4	4
D11S875	2	3	3	2		2	2	2	4	4	1
D11S902	3	5	2	1		3	4	2	1	2	4
D11S904	5	1	2	2		5	1	2	4	5	5
D11S1392	4	4	3	4		4	2	2	1	3	4
D11S1985	5	1	7	6		5	5	4	3	6	1
D11S937	4	3	3	3		1	1	4	2	4	4
D11S876	5	1	4	2		2	2	4	4	4	3
D11S1391	2	4	1	4		2	2	?	?	2	2
D11S976	6	2	2	4		3	4	5	2	?	?
D11S990	1	4	1	3		2	1	4	2	1	5
D11S912	3	4	4	1		3	2	4	2	4	4
D11S968	3	2	1	2		4	1	4	2	2	2

D11S1984	4	4		2	3	3	3	5	1	4	1
D11S988	1	3		*1	2	1	2	4	4	1	4
D11S875	*2	3		4	2	4	2	2	4	4	4
D11S902	3	2		1	3	1	3	2	2	1	4
D11S904	5	2		*1	5	1	5	2	5	4	5
D11S1392	4	3		2	4	2	4	2	3	1	4
D11S1985	5	7		8	5	2	5	4	6	3	1
D11S937	4	3		1	1	1	1	4	4	2	4
D11S876	5	4		2	2	4	2	4	4	4	3
D11S1391	2	1		3	2	4	2	1	2	1	2
D11S976	6	2		1	3	2	3	5	5	5	5
D11S990	1	1		*1	2	1	2	4	1	4	5
D11S912	3	4		4	3	4	3	4	4	2	4
D11S968	3	1		4	4	2	4	4	2	2	2

* Phase unknown

Figure 13 Stage 1 : Segment sharing results
Regions shared by three or more non-sibling chromosomes

Chromosome 17



D17S578	4	3	3	3		4	3	3	2	4	3	
D17S849	2	2	?	?		1	1	2	2	2	2	
D17S796	5	4	1	2		5	1	5	3	5	5	
D17S520	3	1	3	5		2	5	3	2	4	4	
D17S799	4	3	3	3		4	3	3	3	4	3	
D17S122	4	2	3	5		3	5	4	2	3	2	
D17S798	2	2	3	5		5	2	3	2	3	3	
D17S579	4	1	2	3		1	6	4	1	6	3	
D17S795	5	3	5	5		4	3	4	3	6	6	
D17S785	7	6	5	7		5	6	7	7	7	3	
D17S784	4	3	2	3		3	3	6	3	4	2	

D17S578	4	3		*3	4	3	3	3	4	3	4	
D17S849	*2	1		1	1	4	1	2	2	2	2	
D17S796	5	1		3	5	3	5	5	5	5	5	
D17S520	3	3		4	2	4	5	2	4	2	4	
D17S799	4	3		*3	4	3	3	3	4	3	4	
D17S122	2	3		4	3	4	5	2	2	4	2	
D17S798	2	5		*2	5	2	2	3	3	3	3	
D17S579	1	2		1	1	4	6	1	3	1	3	
D17S795	3	5		5	4	6	4	4	6	3	6	
D17S785	7	5		3	5	5	*6	7	3	7	3	
D17S784	4	3		4	3	4	3	3	4	6	4	

* Phase unknown

Figure 7 shows the alleles for the thirteen markers run on **chromosome 1p**. It highlights the 24 cM region between D1S228 and D1S552 which is common to four out of six non-sibling chromosomes. D1S228 is uninformative in two of the families and D1S552 is uninformative in the family where two of the three identical flanking alleles were identified. Affected sibling 49, although showing the same haplotype for these two markers, has inherited a different maternal chromosome. Number 2 has a double recombinant maternal chromosome in the region (24 cM).

Figure 8 shows the results for **chromosome 1q**. There were no more than 2/6 shared haplotypes on this arm of chromosome 1.

The data for **chromosome 5** is presented in **figure 9**. However, there were no segments shared by more than two out of six non-sibling chromosomes.

Figure 10 gives the results of the initial set of marker loci run the length of **chromosome 8**. Two regions of sharing were found, one with flanking markers D8S257 and D8S200 (a distance of 7 cM) and another between D8S198 and D8S284 (a distance of 20 cM). These two segments are separated by 11 cM.

Figure 11 gives the results for markers run on **chromosome 9** where two separate regions are shared between three of six non-sibling chromosomes. The first region, on 9p, is a 10 cM segment between markers D9S741 and D9S319. Both markers are uninformative. No affected siblings share this region.

The second region on 9q between markers D9S167 and D9S197 covers 17 cM. Only one affected sibling (number 49) has this haplotype from the maternal triple recombinant chromosome. The segment may be from the same chromosome as number 13, however it is not clear at which point the second crossover occurred (i.e. either between D9S303 and D9S167 or between D9S197 and D9S299).

Figure 12 gives the results for the original fourteen markers run on **Chromosome 11.** It highlights two regions which showed sharing of haplotypes in three of six non-sibling chromosomes. The first on 11p was between markers D11S904 and D11S1392, a 10 cM region. Affected sibling 49 has inherited the same maternal chromosome as his affected sister. Affected sibling 4 has inherited a different maternal chromosomal segment in this region.

Chromosome 17 results are shown in **figure 13.** A 16 cM segment is shared by 3/6 non-sibling chromosomes and numbers 4 and 2 are homozygous for both these flanking marker loci. The first marker, D17S849, is uninformative.

3.3.5 Stage 2 : Detailed analysis of shared haplotypes

Exclusion of stage 1 haplotypes is summarised in **Table 14.** The IBD index refers to the number of identical haplotypes shared by affected siblings, such that a 0 indicates that no sibs have the same segment, and a 1 or 2 indicates that a sib has the same haplotype on either one or two chromosomes. Exclusion results when either one of the following criteria is met:

- the extended haplotype is different
- the IBD index is zero or one

Table 14 Exclusion criteria for Stage 1 segment sharing

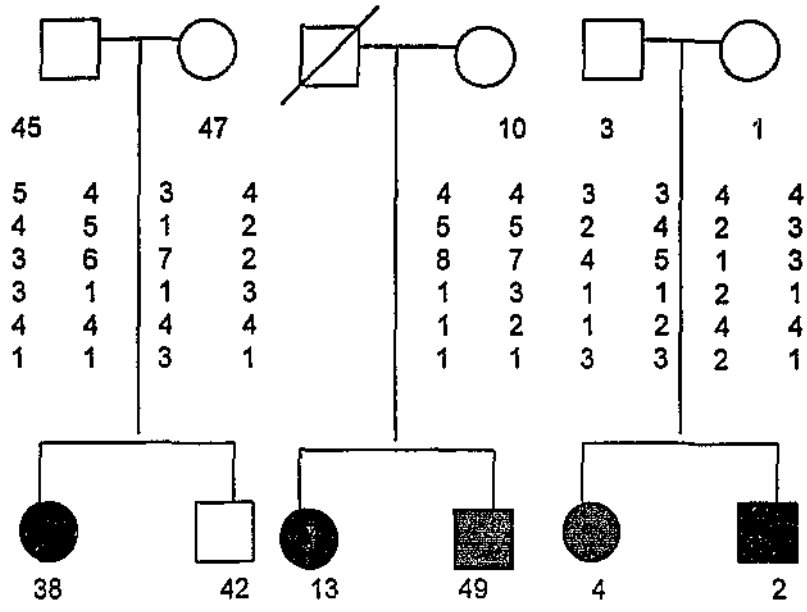
Chrom.#	IBD index 0-2	Extended haplotype identical	Excluded y/n
1	2	n	y
8	2	y	n
9	0	n	y
9	1	n	y
11	1	n	y
17	1		y *

* The region on chromosome 17 was excluded by LOD score analysis prior to commencement of the segment sharing method, as it is a candidate gene region. Extra markers were not run.

The results of the stage 2 detailed screening follow with figures 14 to 18 showing the haplotypes for the ten selected members.

Figure 14 Stage 2: Segment sharing results
Detailed analysis of shared haplotypes

Chromosome 1p



D1S228	5	4	3	4		4	4	3	3	4	4	
D1S507	4	5	1	2		5	5	2	4	2	3	
D1S436	3	6	7	2		8	7	4	5	1	3	
D1S483	3	1	1	3		1	3	1	1	2	1	
D1S482	4	4	4	4		1	2	1	2	4	4	
D1S552	1	1	3	1		1	1	3	3	2	1	

D1S228	4	3	4	4	4	4	3	4	3	4
D1S507	5	1	1	5	1	5	4	2	4	3
D1S436	6	7	*7	8	7	7	5	1	5	3
D1S483	1	1	1	1*	1	3	1	2	1	1
D1S482	4	4	3	1	4	2	2	4	2	4
D1S552	1	3	1	1	1	1	3	2	3	1

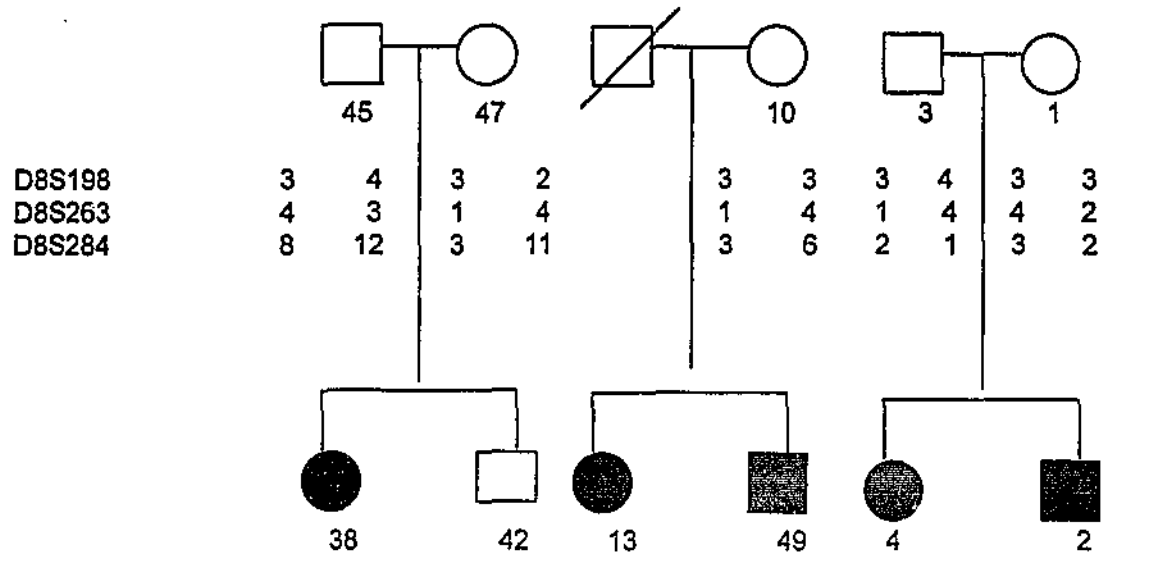
* Phase unknown

** Exclusion of IBD because the haplotypes within the original (highlighted in bold) markers are not identical.

*** Original marker D1S228 is uninformative in the families (10,13 & 49) and (1,3,4,2), and D1S552 is uninformative in family (10,13 & 49) which is where 2/6 chromosomes were first identified.

Figure 15 Stage 2 Detailed analysis of shared haplotypes

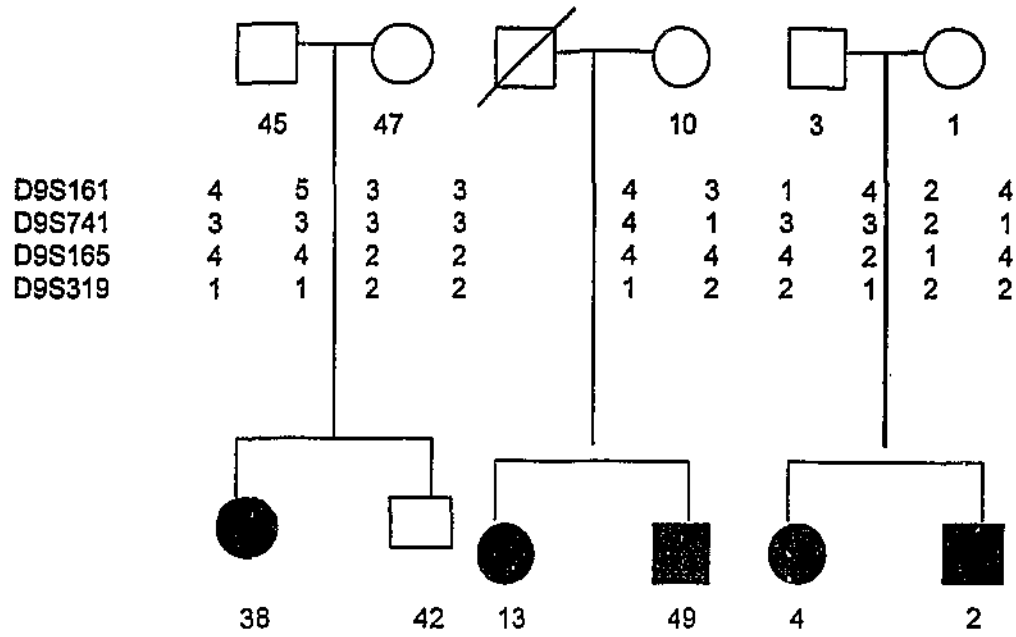
Chromosome 8q



Extended region still shared by 3/6 non-sib chromosomes. Identical region shared by sibs 13 & 49.

Figure 16 Stage 2 : Detailed analysis of shared segments

Chromosome 9p



D9S161	4	5	3	3		4	3	1	4	2	4	
D9S741	3	3	3	3		4	1	3	3	2	1	
D9S165	4	4	2	2		4	4	4	2	1	4	
D9S319	1	1	2	2		1	2	2	1	2	2	

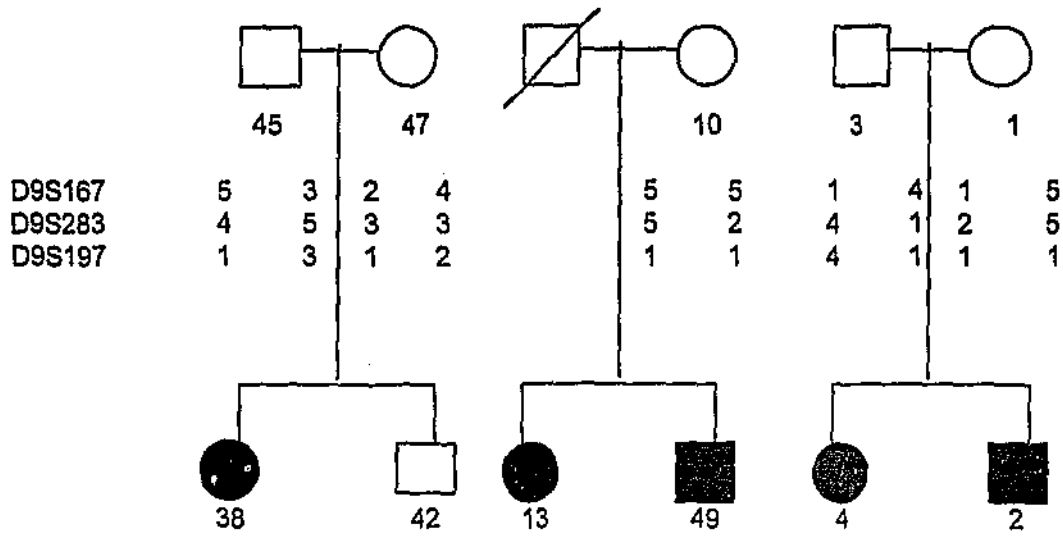
D9S161	4	3		1	4	4	4	1	2	4		4
D9S741	3	3		3	4	2	4	3	2	3		1
D9S165	4	2		3	4	2	4	4	1	2		4
D9S319	1	2		*2	1	1	1	2	2	1		2

* Phase unknown

Extended region reduces to below threshold with 0/6 identical non-sibling haplotypes.

Figure 17 Stage 2 : Detailed analysis of shared haplotypes

Chromosome 9q



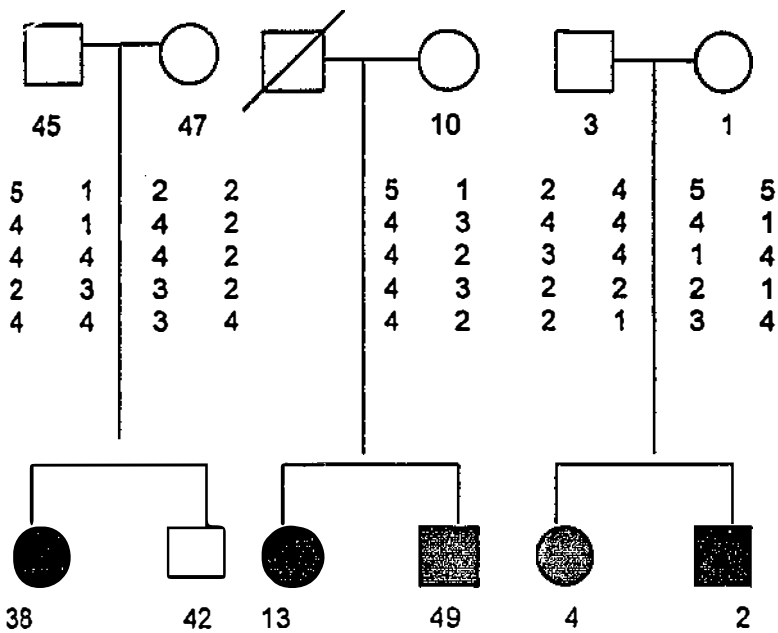
D9S167	5	3	2	4			5	5		1	4	1	5	
D9S283	4	5	3	3			5	2		4	1	2	5	
D9S197	1	3	1	2			1	1		4	1	1	1	
D9S167	5	4			1		5	1	5	1	1	4	5	
D9S283	4	3			*2		5	2	5	4	2	4	5	
D9S197	1	2			4		1	4	1	4	1	4	1	

* Phase unknown

Extended region reduces to below threshold with 2/6 identical non-sibling haplotypes

Figure 18 Stage 2 : Detailed analysis of shared haplotypes

Chromosome 11



D11S904	5	1	2	2		5	1	2	4	5	5
D11S1322	4	1	4	2		4	3	4	4	4	1
D11S1312	4	4	4	2		4	2	3	4	1	4
D11S1301	2	3	3	2		4	3	2	2	2	1
D11S1392	4	4	3	4		4	2	2	1	3	4

D11S904	5	2		1	5	1	5	2	5	4	5
D11S1322	4	4		1	4	1	4	4	4	4	1
D11S1312	4	4		4	4	4	4	3	1	4	4
D11S1301	2	3		2	4	2	4	2	2	2	1
D11S1392	4	3		2	4	2	4	2	3	1	4

The three original haplotypes when extended each show a different internal set of alleles.

Sibs have different haplotypes except for 13 & 49.

Figure 14 shows an expanded view of this 24 cM shared region on chromosome 1p with four extra markers run between D1S228 and D1S552 at an average distance of 6 cM. Each one of the four expanded haplotypes is different over the 24 cM. However, three of six common haplotypes over three markers D1S483, D1S482 and D1S552 are shown (7 cM region). Results from the region D1S483 to D1S482 (also 7 cM as D1S482 and D1S552 are shown at the same map locus) show four out of six non-sibling chromosomes with the 1,4 and 3/6 with the 1,4,1 haplotype in pedigree members 38, 49 and 2. These two 7 cM segments (1,4 or 1,4,1) are not shared by sibling number 4 who has inherited the other maternal chromosome with a recombination outside the region. They are not shared by affected sibling 13 as 49 has inherited a recombinant paternal chromosome where the crossing-over occurred in the middle of this 7 cM segment. Phase is unknown for 13 at marker locus D1S436 and for 49 at marker D1S483 (one of the 7 cM shared 1,4 haplotypes). There are only three alleles for each of the flanking markers D1S483 and D1S552 of the 7 cM shared haplotype.

Figure 15 shows the extended haplotypes for D8S198 to D8S284. The extended region still showed sharing by 3/6 non-sibling chromosomes with an identical region shared by

siblings 13 and 49 on both chromosomes. Numbers 4 and 2 also have this region IBD, although with different haplotypes.

Figure 16 gives the results for the 9p region between D9S161 and D9S301. The only available marker between D9S741 and D9S319 was D9S165 at 6 cM telomeric to D9S741. However, D9S161 was close to the region at 2 cM centromeric to D9S741 so this was run also. These markers did not show more than two of three common haplotypes.

On 9q two extra markers, D9S283 and D9S318, were run between D9S167 and D9S197 but D9S318 did not amplify. This left D9S283 at a distance of 10 cM telomeric to D9S167. The results from these investigations are shown in **figure 17.** Even though phase is unknown for members 13 and 49 for marker D9S283, reversing the allele order still would not result in more than two common haplotypes.

The second region was on 11q between D11S990 and D11S912, a distance of 6 cM. One affected sibling (number 49) has also inherited the same alleles for these flanking markers. The results of three extra markers, D11S1322, D11S1312 and D11S1301, run between markers D11S904

and D11S1392, at an average distance of 2.5 cM, are shown in **figure 18**. Two identical haplotypes over a distance of 3.8 cM were seen between D11S904 and D11S1312. Four non-sibling chromosomes show the 4,4 haplotype between D11S1322 and D11S1312 which are 0.8 cM apart. No extra marker was run in the second region between D11S990 and D11S912 covering 6 cM on 11q. Siblings 13 and 49 share the same haplotype although it may be from the other paternal chromosome.

Table 15 summarises the sharing patterns in the extended regions. It shows that only the two chromosome 8 regions (separated by 11 cM) remain with 3/6 identical non-sibling chromosomes.

Table 15 Sharing patterns in extended regions

Chrom No.	First Marker #.	Distance cM	Second Marker #	No. of haplotypes shared by non-sibs	Shared by siblings Y/N	Ped.#	No.Chroms each
1	D1S228	5	D1S507	2/6	Y	13, 49	1
	D1S507	5	D1S436	2/6	Y	13, 49	1
	D1S436	7	D1S483	2/6	Y	13, 49	1
	D1S483	7	D1S482	4/6	N		
	D1S482	0	D1S552	3/6	N		
	Total extended haplotype			**0/6**			
8	D8S257	7	D8S200	3/6	Y	4, 2	2
	D8S198	15.4	D8S263	3/6	Y	13, 49	2
	D8S263	4	D8S284	3/6	Y	13, 49	2
	Total extended haplotype			**3/6**			
9	D9S161	2	D9S741	2/6	N		
	D9S741	6	D9S165	2/6	N		
	D9S165	4	D9S319	2/6	N		
	Total extended haplotype			**0/6**			
9	D9S167	10	D9S283	2/6	Y	13, 49	1
	D9S283	7	D9S197	2/6	Y	13, 49	1
	Total extended haplotype			**2/6**			
11	D11S904	3	D11S1322	3/6	Y	13, 49	1
	D11S1322	0.8	D11S1312	4/6	Y	13, 49	1
	D11S1312	2.9	D11S1301	3/6	Y	13, 49	1
	D11S1301	2.3	D11S1392	2/6	Y	13, 49	1
	D11S990	6	D11S912	3/6	Y	13, 49	1
	Total extended haplotype			**0/6**			
17	D17S849	16	D17S796	3/6	Y	4, 2	2

** Two different haplotypes

3.3.6 Stage 3: Extended pedigree analysis of IBD region

The next stage involved running the whole family against the three markers in the region. These data were transmitted to Joachim Hallmayer at Stanford for analysis using the computer program Fastlink. Nuclear families were treated as separate and the results combined. These results are shown in **Table 16.** Theta (θ) represents the recombination fraction or the genetic distance between the marker locus and the disease gene locus.

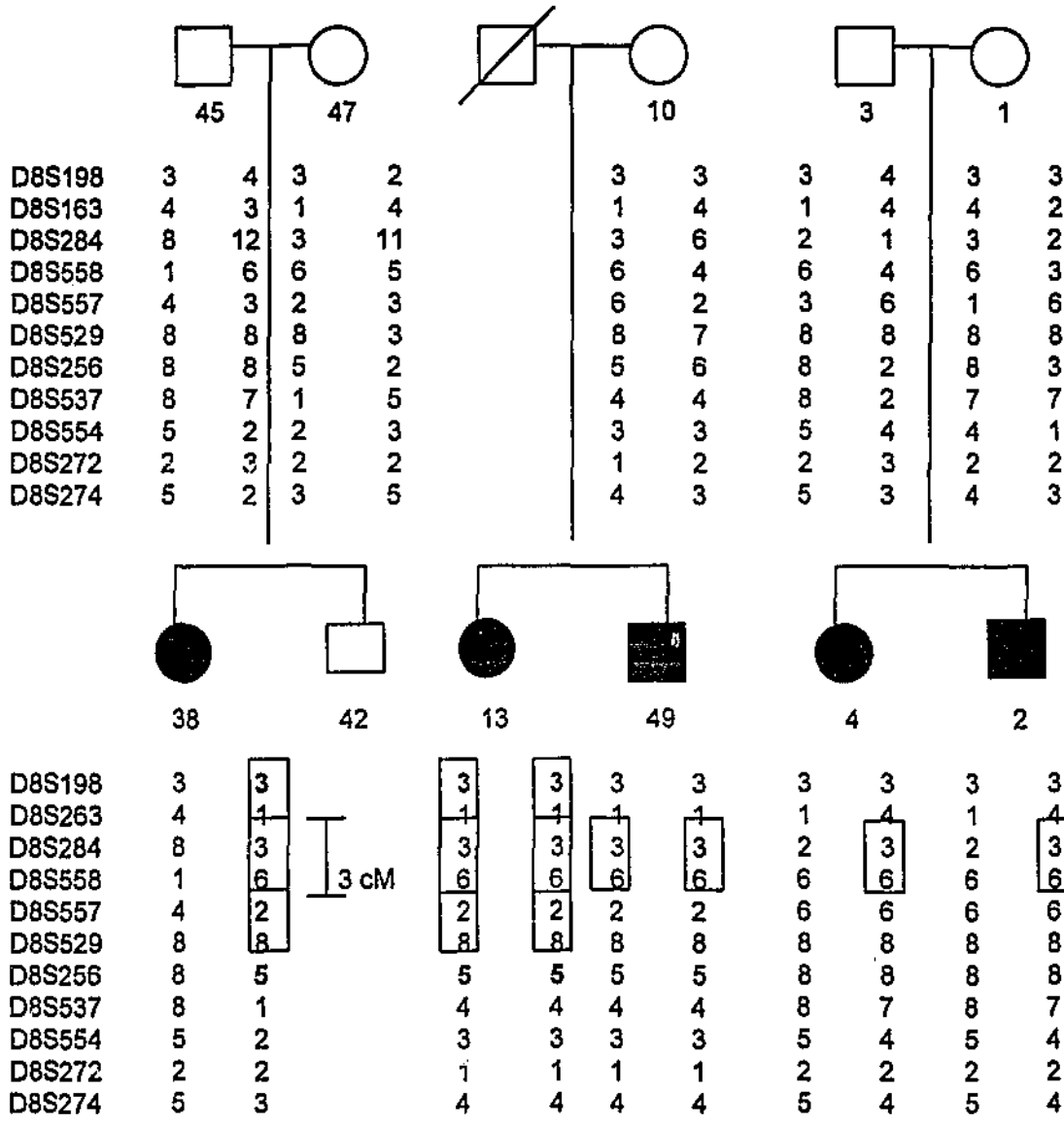
Table 16 Two point LOD scores for IBD region on 8q

	Recombination fraction (θ)				
Locus	0.0	0.05	0.1	0.15	0.2
Published allele frequencies (GDB)					
D8S198	-3.09	-0.38	-0.03	-0.02	-0.04
D8S263	-0.85	0.57	0.70	0.66	0.55
D8S284	3.41	2.93	2.45	1.98	1.52
Equal allele frequencies					
D8S198	-3.09	-0.38	-0.12	-0.03	-0.02
D8S263	-0.60	1.02	1.06	0.94	0.77
D8S284	3.44	2.96	2.48	2.00	1.54

Due to the increasing lod scores approaching marker D8S284, and the significant score at this locus, additional markers were obtained from Research Genetics for the region between D8S284 and D8S274. The extended haplotypes for the segment sharing members between marker loci D8S198 and D8S274 are shown in **figure 19.**

Figure 19 Stage 4 : Fine mapping the region of interest by further extension of haplotypes

Chromosome 8q



D8S198	3	4	3	2			3	3	3	4	3	3		
D8S163	4	3	1	4			1	4	1	4	4	2		
D8S284	8	12	3	11			3	6	2	1	3	2		
D8S558	1	6	6	5			6	4	6	4	6	3		
D8S557	4	3	2	3			6	2	3	6	1	6		
D8S529	8	8	8	3			8	7	8	8	8	8		
D8S256	8	8	5	2			5	6	8	2	8	3		
D8S537	8	7	1	5			4	4	8	2	7	7		
D8S554	5	2	2	3			3	3	5	4	4	1		
D8S272	2	3	2	2			1	2	2	3	2	2		
D8S274	5	2	3	5			4	3	5	3	4	3		
D8S198	3	3			3	3	3	3	3	3	3	3		
D8S263	4	1			1	1	1	1	1	4	1	4		
D8S284	8	3			3	3	3	3	2	3	2	3		
D8S558	1	6			6	6	6	6	6	6	6	6		
D8S557	4	2			2	2	2	2	6	6	6	6		
D8S529	8	8			8	8	8	8	8	8	8	8		
D8S256	8	5			5	5	5	5	8	8	8	8		
D8S537	8	1			4	4	4	4	8	7	8	7		
D8S554	5	2			3	3	3	3	5	4	5	4		
D8S272	2	2			1	1	1	1	2	2	2	2		
D8S274	5	3			4	4	4	4	5	4	5	4		

Extended haplotype 3,1,3,6,2,8 (over 27 cM) still shared by 3/6 non sib chromosomes

Extended haplotype shared by sibs 13 & 49

Haplotype 3,6 between D8S284 and D8S558 is shared by 4/6 non-sib chromosomes,

The two point lod scores for these additional markers are listed in **Table 17.** These results are based on a pedigree structure with connections between families 2 and 3 as well as the branch containing numbers 14 and 26.

Table 17 Two point LOD scores for fine mapping of IBD region on 8q

	Recombination fraction (θ)				
Locus	0.0	0.05	0.1	0.15	0.2
Equal allele frequency; 0.05 disease gene frequency					
D8S284	-infinity	2.23	1.95	1.57	1.19
D8S557	1.83	1.93	1.69	1.39	1.09
D8S558	3.38	2.81	2.29	1.83	1.41
D8S529	2.73	2.30	1.88	1.47	1.10
D8S256	2.29	2.41	2.09	1.69	1.28
D8S537	1.88	1.94	1.69	1.38	1.07

The different lod scores at locus D8S284 between the initial calculations and these later results reflects the different treatment of the pedigree within the Fastlink program in terms of inter-family connections. In Table 16 the nuclear families were treated as separate (apart from linked branches family numbers 2 and 3) and the results summed. In the later calculations, additional links were included.

Haplotypes formed from marker loci D8S263, D8S284, D8S558, D8S557, D8S529 and D8S256 for the extended pedigree follow in **figures 20 to 23.** Map distances according to the CEPH map (GDB) are:

D8S263	to	D8S284	2 cM
D8S284	to	D8S558	3 cM
D8S558	to	D8S557	0 cM
D8S557	to	D8S529	1 cM
D8S529	to	D8S256	2 cM

The full pedigree was split into separate nuclear families numbered 1 to 6. This is reflected in the original numbering system where the family number precedes the individual number e.g. 2-5 (Appendix D shows both sets of numbers). A total of six different mutant haplotypes were identified in this region and these are highlighted by colours so that the inheritance pathways can be traced.

Figure 20

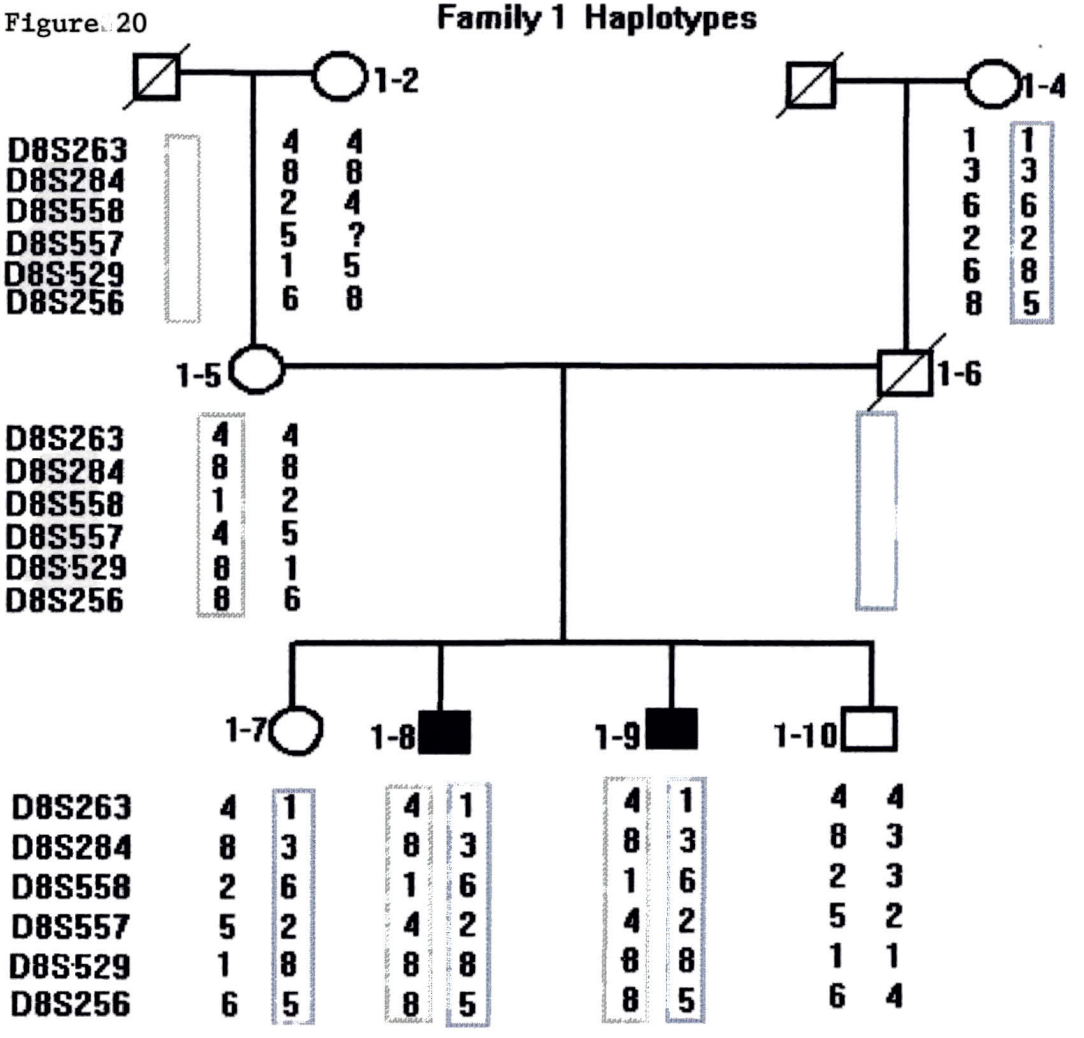

Family 1 Haplotypes
1-2
1-4
D8S263
D8S284
D8S558
D8S557
D8S529
D8S256
4 4
8 8
2 4
5 ?
1 5
6 8
1 1
3 3
6 6
2 2
6 8
8 5
1-5
1-6
D8S263
D8S284
D8S558
D8S557
D8S529
D8S256
4 4
8 8
1 2
4 5
8 1
8 6
1-7
1-8
1-9
1-10
D8S263
D8S284
D8S558
D8S557
D8S529
D8S256
4 1
8 3
2 6
5 2
1 8
6 5
4 1
8 3
1 6
4 2
8 8
8 5
4 1
8 3
1 6
4 2
8 8
8 5
4 4
8 3
2 3
5 2
1 1
6 4

Figure 21

Family 3 + Family 2 Haplotypes

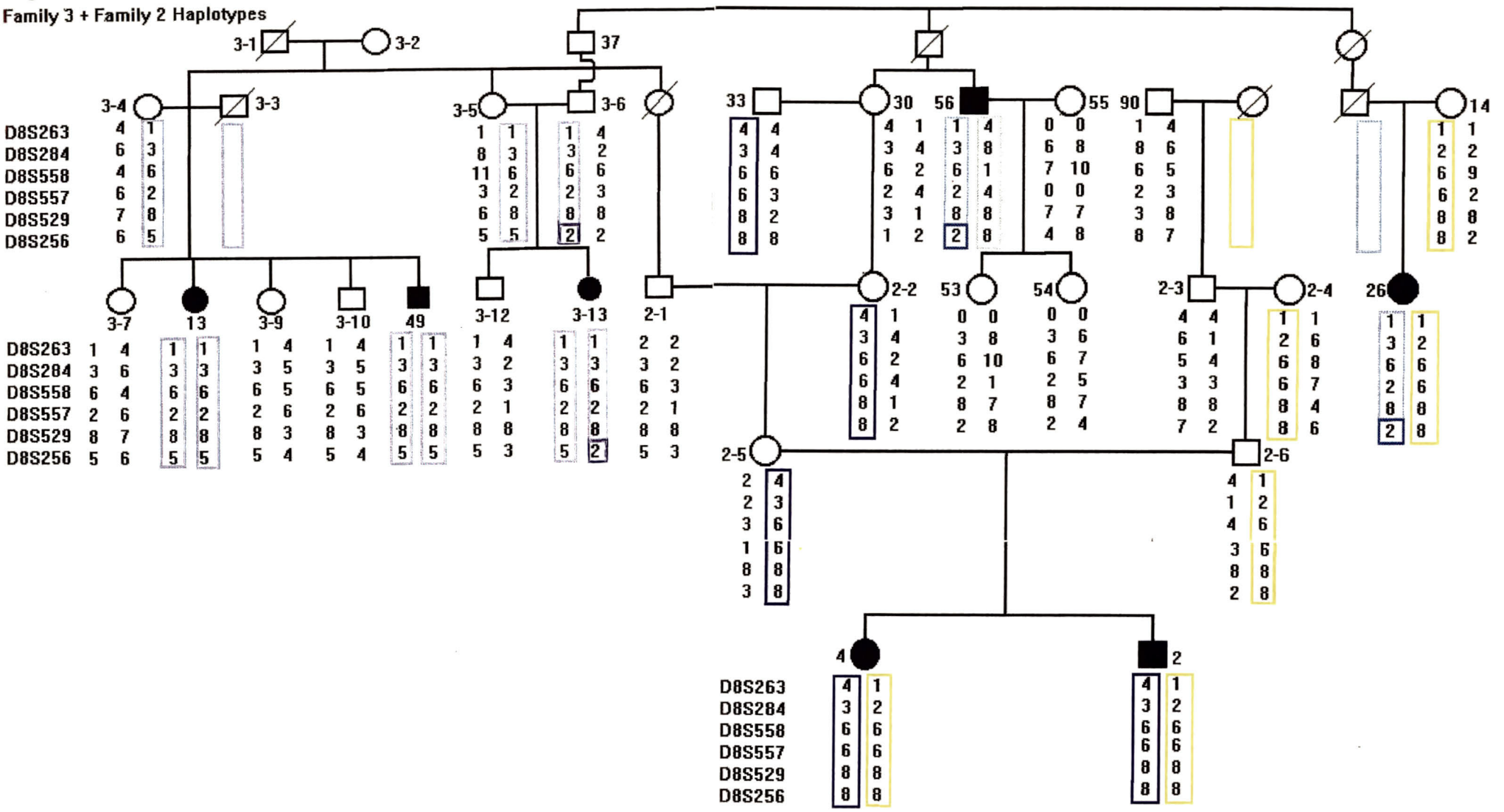


Figure 22

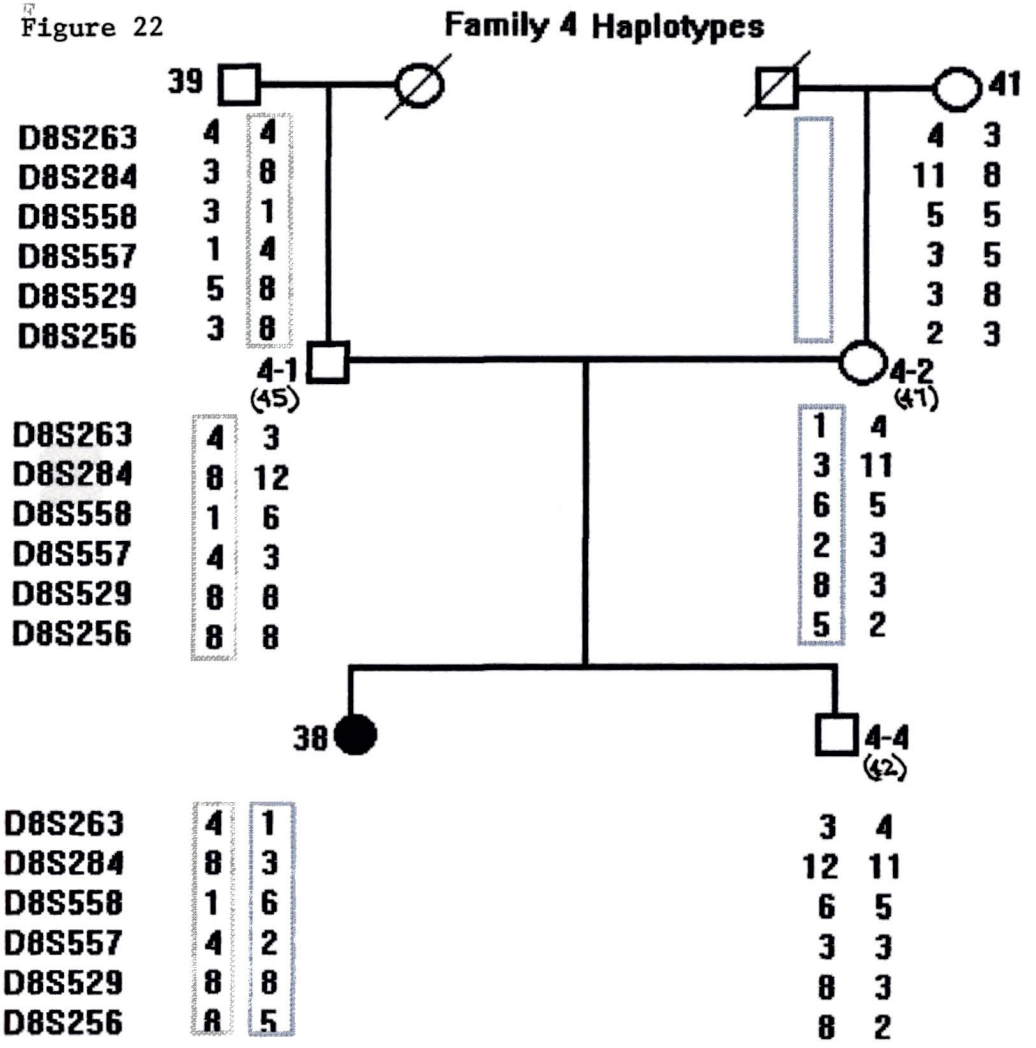
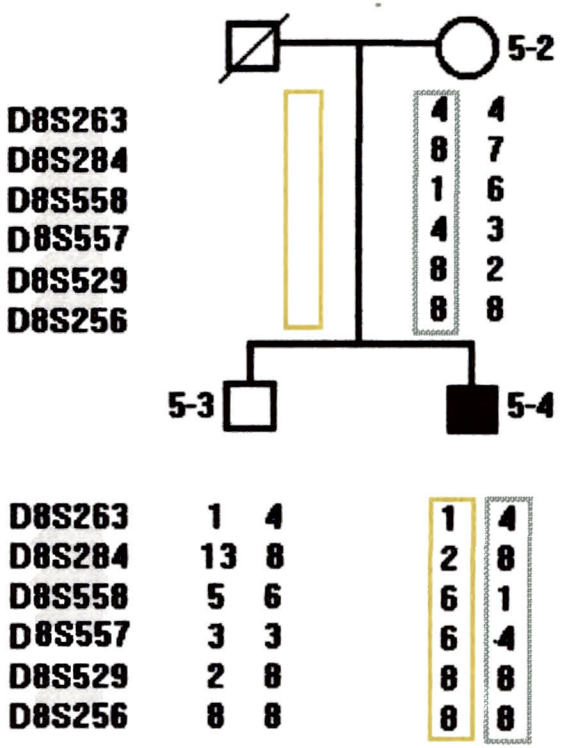
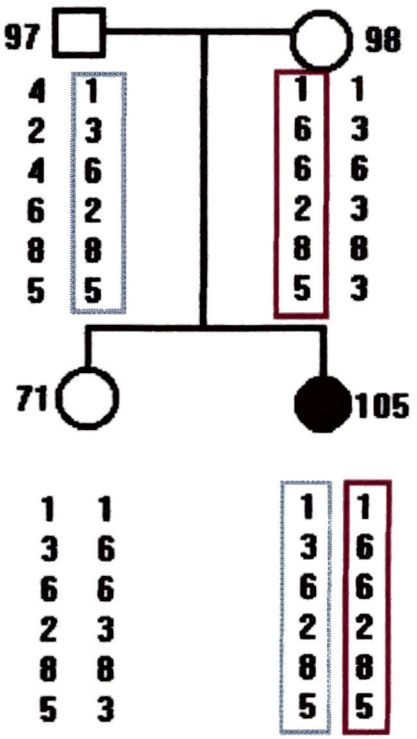

Family 4 Haplotypes
39
41
D8S263 4 4
D8S284 3 8
D8S558 3 1
D8S557 1 4
D8S529 5 8
D8S256 3 8
4 3
11 8
5 5
3 5
3 8
2 3
4-1
(45)
4-2
(41)
D8S263 4 3
D8S284 8 12
D8S558 1 6
D8S557 4 3
D8S529 8 8
D8S256 8 8
1 4
3 11
6 5
2 3
8 3
5 2
38
4-4
(42)
D8S263 4 1
D8S284 8 3
D8S558 1 6
D8S557 4 2
D8S529 8 8
D8S256 8 5
3 4
12 11
6 5
3 3
8 3
8 2

Figure 23

Family 5 Haplotypes



Family 6 Haplotypes



CHAPTER 4. DISCUSSION & CONCLUSION

A locus for a previously undescribed autosomal recessive form of CMT has been mapped to a 27 cM region 8q24.13-qter using a relatively new approach to rapid genome screening, namely segment sharing. The method is based on the identification of identical chromosomal segments shared among distantly related members of a population exhibiting random inbreeding. Subsequent lod score analysis places the candidate gene in the region between markers D8S557 and D8S256, a distance of 3 cM.

Stage one of this rapid genome screening method resulted in the identification of eight regions flanked by markers at an average of 13.7 cM apart. These shared regions were on chromosomes 1, 8, 9, 11 and 17. Stage two, the detailed analysis of these haplotypes, excluded identity by descent (IBD) in all but one region. For this 20 cM region, located on 8q, sharing of alleles IBD was confirmed over the extended region for three out of six non-sibling chromosomes. Using the results of the formula for p_2 in Table 4, the probability of a false positive result over a segment of this length being found in three out of six chromosomes had been estimated at 1.2×10^{-10}. The validity of these findings was confirmed by conventional linkage analysis of the extended pedigree. The maximum lod score for the region was obtained for D8S558, 3.38 at zero recombination. As time

permits, more markers will be run in the interval D8S557 to D8S526 to reduce the linkage region.

In order to gain an indication of the underlying genetic structure of this population, in particular the level of genetic heterogeneity and inbreeding, a sample of ten unlinked markers was examined. These markers were on four chromosomes and they were run for the extended pedigree. The first aspect examined was the actual number of alleles compared to those published for CEPH families. The total number of alleles over 9 of the markers (for one the published number of alleles was not listed) in the actual sample was 32.6% less than the published number. This seemed to be in keeping with the expectations for an inbred, isolated population (i.e. a restricted gene pool with increased homozygosity). To follow through on this hypothesis the heterozygosity indices were calculated and compared to those published for the CEPH families. In four of the ten markers the observed heterozygosity was higher than published and for another two markers the difference was within 0.04. Inbreeding coefficients were positive in only three of the ten markers. However, two of these were higher than would be expected at 0.12 and 0.15. For one of these markers (D17S796) this is due to a high number of homozygotes for the highest frequency allele.

Overall the level of inbreeding for the population is extremely low (at eight decimal places it was still zero), even though there are many

connections between affected branches of the kindred. This low level of inbreeding, as well as the calculated disease gene frequency of 0.08 in the population of 2000 Bulgarian Gypsies, directed the study away from homozygosity mapping and towards the segment sharing approach.

The outcome of this study, that is linkage of the gene for this autosomal recessive CMT, was successful. The region identified as IBD was linked to the disease gene. For the 31% of the genome covered, the segment sharing method did not produce any false positive IBD regions. The screening of only ten members was an efficient and time-saving facet of the method. However, further reduction in the second stage screening may be possible. For the purposes of a rapid screening method, the question arises of whether it is efficient to continue to stage two for those regions which can be excluded without extending the haplotypes. One criterion which could be applied for exclusion at Stage 1 is:

* affected sibs do not share the same haplotype.

If it is assumed that the segment is IBD, these regions would be unlikely to be linked to the disease gene. This would not be the case if two or more different mutations existed at the same locus or more than one disease locus was present. However, in these cases the power to detect IBD is reduced and so the choice of this method may be inappropriate. For this reason, the choice of method is influenced

by the structure and age (since founding) of the population under investigation, in addition to the disease gene frequency.

The correct choice of individuals and knowledge of possible relationships within the population is one pre-requisite to the success of the method. In our case, it might perhaps be suggested that good fortune was another, since alternative choices of individuals and/or polymorphic markers in this study could have resulted in the 8q region being overlooked at Stage 1 screening. As an example, the individuals selected could have been numbers 26, 36 and 4. Number 36 is related to 26 over a 7 step pathway, 36 related to 4 over two pathways one 7 step and one 9 step (minimum is used), and 26 is related to 4 via 9 steps. The choice of these members would have resulted in an average of 7.6 meiotic events separating the individuals by comparison with 9.4 for the actual individuals selected. Numbers 13 and 36 were not considered as they are first cousins (a five step pathway) and as such, share 1/16 of their genes. The kinship coefficients for various combinations of individuals showed that the degree of relationship between the alternatives 36 and 26 was 0.016, the same as that between 13 and 4. Although 4 and 26 were closer than those selected the difference between coefficients was only 0.002.

In theory, the more distantly related individuals were better choices due to the reduced chance of IBS sharing. However, for this project

the selection of individuals proved to be invalid since, when the full results were obtained for the forty-seven members in the 8q "IBD" region, it was discovered that some of the shared haplotypes were not transmitted along the expected common ancestor pathways.

The shared region between marker loci D8S198 and D8S256 (Figure 19) identified in three of the six non-sibling chromosomes (in numbers 38 and 13), appeared to be IBD. However, one mutant haplotype can only be traced from 38 to maternal grandfather and it is not known whether it was transmitted via the expected route of great-grandmother with links to the other branches, or if it was transmitted from the maternal great-grandfather with no known connections. This was the case for number 4 where one mutant haplotype was not transmitted via the expected route (through individual number 90) but instead came from the paternal great-grandmother who has no known links with 38 and 13. The second mutant haplotype was transmitted to number 4 by maternal great-grandfather number 33 rather than via number 30 who also has no known links with 38 and 13.

These examples highlight the problems with using this method for such a randomly inbred pedigree. A total of six different mutant haplotypes were identified in the twelve affected individuals. To have this number of recombinants the mutation/s must have been introduced into the population a long time ago and most probably by a number of carriers.

Would the alternative choices have given a truly IBD region? Numbers 36 and 26 share one identical haplotype - 26 via the paternal grandmother and 36 via the paternal grandfather. However, 36 and 4 do not share any identical haplotypes, as the mutant chromosomes inherited by 4 have come from individuals (2-4 and 33) whose connections within the pedigree are unknown. With hindsight it is apparent that there was no ideal choice of individuals for the segment sharing method in this pedigree. Although the desired result was obtained, i.e. localising the disease gene for this form of CMT, it was not as a result of the applicability of the segment sharing method.

The linkage analysis performed for the extended pedigree served to reinforce the IBD finding on 8q with the 3/6 shared haplotypes giving maximum lod scores for D8S198 of 0.02 at 15% recombination, and for D8S284 of 3.41 at zero recombination. This positive indication of linkage for D8S284 was strengthened by running more markers telomeric to this marker. This second group of lod scores gave a different result for D8S284 at zero recombination, minus infinity, because of extra inter-family connections being included in the calculations. Originally the results were calculated and summed for the nuclear families as though they were unrelated. The results of analysis of the pedigree including the links between families 2 and 3 and individuals 14 and 26 gives a maximum lod score of 3.16 at 5% recombination for marker D8S284. A maximum lod score of 2.73 at

zero recombination was obtained for the marker D8S529, 3.5 cM telomeric to D8S284. The next three markers telomeric to D8S529 show decreasing lod scores, an expected trend. Therefore the most likely placement of this new CMT disease gene is between markers D8S284 and D8S529. This is only a relative genetic map position as placement of these markers has not been physically confirmed.

REFERENCES

Baraitser M. (1990). The Genetics of Neurological Disorders (2nd ed.). Oxford: Oxford University Press.

Ben Othmane K., Hentati F., Lennon F., Ben Hamida C. *et al.* (1993a). Linkage of a locus (CMT4A) for autosomal recessive Charcot-Marie-Tooth disease to chromosome 8q. *Human Molecular Genetics* 2(10): 1625-1628.

Ben Othmane K., Middleton L.T., Loprest L.J., Wilkinson C.M., Lennon F. *et al.* (1993b). Localization of a gene (CMT2A) for autosomal dominant Charcot-Marie-Tooth disease type 2 to chromosome 1p and evidence of genetic heterogeneity. *Genomics* 17: 370-375.

Ben Othmane K., Loeb D. Hayworth-Hodgte R., Hentati F., Rao N. *et al.* (1995). Physical and genetic mapping of the CMT4A locus and exclusion of PMP-2 as the defect in CMT4A. *Genomics* 28: 286-290.

Bergoffen J., Scherer S.S., Wang S., Oronzi Scott M., Bone L.J. *et al.* (1993). Connexin mutations in X-linked Charcot-Marie-Tooth disease. *Science* 262: 2039-2042.

Berne R.M. and Levy M.N. (Eds.). (1988). *Physiology* (2nd ed.). St. Louis:C.V. Mosby Company.

Bird T.D. and Kraft G.H. (1978). Charcot-Marie-Tooth disease: Data for genetic counselling relating age to risk. *Clinical Genetics* 14: 43-49.

Bird T.D., Ott J. and Giblett E.R. (1982). Evidence for linkage of Charcot-Marie-Tooth neuropathy to the Duffy locus on chromosome 1. *American Journal of Human Genetics* 34: 388-394.

Bird T.D., Ott J. and Giblett E.R. (1983). Genetic linkage evidence for heterogeneity in Charcot-Marie-Tooth neuropathy (HMSN type I). *Annals of Neurology* 14: 679.

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

Boyce A.J. (1983). Computation of inbreeding and kinship coefficients on extended pedigrees. *The Journal of Heredity* 74: 400-404.

Chance P.F., Alderson M.K., Leppig K.A., Lensch M.W., Matsunami N. *et al.* (1993). DNA deletion associated with hereditary neuropathy with liability to pressure palsies. *Cell* 72: 143-151.

Chance P.F. and Fischbeck K.H. (1994). Molecular genetics of Charcot-Marie-Tooth disease and related neuropathies. *Human Molecular Genetics* 3: 1503-1507.

Cornell J., Sellars S. and Beighton P. (1984). Autosomal recessive inheritance of Charcot-Marie-Tooth disease associated with sensorineural deafness. *Clinical Genetics* 25: 163-165.

Dyck P.J. (1975). Inherited neuronal degeneration and atrophy affecting peripheral motor, sensory and autonomic neurons in *Peripheral Neuropathy* Vol. 2 (3rd ed.). (Eds. Dyck P.J. and Thomas, P.K.). Philadelphia: Saunders.

Dyck P.J., Chance P., Lebo R. and Carney J.A. (1993) Hereditary motor and sensory neuropathies in *Peripheral Neuropathy* Vol. 2 (3rd ed.). (Eds. Dyck P.J. and Thomas, P.K.). Philadelphia: Saunders.

Dyck P.J., Karnes J.L., Lambert E.H. (1989). Longitudinal study of neuropathy deficits and nerve conduction abnormalities in hereditary motor and sensory neuropathy type I. *Neurology* 39: 1302.

Dyck P.J. and Lambert E.H. (1968a). Lower motor and primary sensory neuron diseases with peroneal muscular atrophy:I. Neurologic, genetic and electrophysiologic findings in hereditary polyneuropathies. *Archives Neurology* 18: 603-618.

Dyck P.J. and Lambert E.H. (1968b). Lower motor and primary sensory neuron diseases with peroneal muscular atrophy:II. Neurologic, genetic and electrophysiologic findings in hereditary polyneuropathies. *Archives Neurology* 18: 619-625.

Gal A., Mucke J., Theile H., Wieacker P.F., Ropers H.H. and Weinker T.F. (1985). X-linked dominant Charcot-Marie-Tooth disease: Suggestion of linkage with a cloned DNA sequence from the proximal Xq. *Human Genetics* 70: 38-42.

Greenfield S., Brostoff S., Eylar E.H. and Morell P. (1973). Protein composition of myelin of the peripheral nervous system. *Journal of Neurochemistry.* 20: 1207-1216.

Gyapay G., Morissette J., Vignal A., Dib C., Fizames C. *et al.* (1994). 1993-1994 Genethon human genetic linkage map. *Nature Genetics* 7: 246-249.

Hanson P.A., Farber R.E. and Armstrong R.A. (1970). Distal muscle wasting, nephritis, and deafness. *Neurology* 20: 426-434.

Harding A.E. and Thomas P.K. (1980). The clinical features of hereditary motor and sensory neuropathy types I and II. *Journal of Medical Genetics* 17: 329-336.

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

Hayasaka K., Himoro M., Takada G., Takahashi E., Minoshima S, Shimizu N. (1993a). Structure and localization of the gene encoding human peripheral myelin protein 2 (PMP2). *Genomics* 18: 244-248.

Hayasaka, K., Ohnishi, A., Takada, G., Fukushima, Y., Murai, Y. (1993b). Mutation of the myelin P_0 gene in Charcot-Marie-Tooth neuropathy type 1. *Biochemical and Biophysical Research Communications* 194(3): 1317-1322.

Hayasaka, K., Himoro, M., Sawaishi, Y., Nanao, K., Takahashi, T. *et al.* (1993c). De novo mutation of the myelin P_0 gene in Dejerine-Sottas disease (hereditary motor and sensory neuropathy type III). *Nature Genetics* 5: 266-268.

Houwen, R.H.J., Baharloo, S., Blankenship, K., Raeymaekers, P., Juyn, J. *et al.* (1994). Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nature Genetics* 8: 380-386.

Iwashita H., Inoue N., Araki S. and Kuroiwa Y. (1970). Optic atrophy, neural deafness and distal neurogenic amyotrophy. Report of a family with two affected siblings. *Archives of Neurology* 22: 35-64.

Lander E.S. and Botstein D. (1987). Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 236: 1567-1570.

Latour P., Blanquet F., Nelis E., Bonnebouche C., Chapon F. *et al.* (1995). Mutations in the myelin protein zero gene associated with Charcot-Marie-Tooth disease type 1b. *Human Mutation* 6: 50-54.

Lemke G., Lamar E. and Patterson J. (1988). Isolation and analysis of the gene encoding peripheral myelin protein zero. *Neuron* 1: 73-83.

Lev A.A., Rosen D.R., Kos C., Clifford E., Landes G. *et al.* (1993). Human ciliary neurotrophic factor: localization to the proximal region of the long arm of chromosome 11 and association with CA/GT dinucleotide repeat. *Genomics* 16: 539-541.

Lebo R.V., Chance P.F., Dyck P.J., Bird T.D. (1991). Chromosome 1 Charcot-Marie-Tooth disease (CMT 1b) locus in the Fcγ receptor gene region. *Human Genetics* 88: 1-12.

Lemke G. (1988). Unwrapping the genes of myelin. *Neuron* 1: 535-543.

Lupski J. R., Montes de Oca-Luna R., Slaugenhaupt S., Pentao L., Guzzetta V. *et al.* (1991). DNA duplication associated with Charcot-Marie-Tooth Disease Type 1A. *Cell* 66: 219-232.

Masu Y., Wolf E., Holtmann B., Sendtner M., Brem G., Thoenen H. (1993). Disruption of the CNTF gene results in motor neuron degeneration. *Nature* 365: 27-32.

Matsunami N., Smith B., Ballard L., Lensch M.W., Robertson M. *et al.* (1992). Peripheral myelin protein-22 gene maps in the duplication in chromosome 17p11.2 associated with Charcot-Marie-Tooth 1A. *Nature Genetics* 1: 176-179.

Morton N.E. (1955). Sequential test for the detection of linkage. *American Journal of Human Genetics* 7: 266-318.

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

Pham-Dinh D., Fourbil Y. Blanquet F., Mattei M.G., Roeckel N. *et al.* (1993). The major peripheral myelin protein zero gene: Structure and localization in the cluster of Fcγ receptor genes on human chromosome 1q21.3-q23. *Human Molecular Genetics.* 2: 2051-2054.

Roa B.B., Dyck P.J., Marks H.G., Chance P.F., Lupski J.R. (1993a). Dejerine-Sottas syndrome associated with point mutation the peripheral myelin protein 22 (PMP22) gene. *Nature Genetics* 5: 269-273.

Roa B.B., Garcia C.A., Pentao L., Killian J.M., Trask B.J. *et al.* (1993b). Evidence for a recessive PMP22 point mutation in Charcot-Marie-Tooth disease type 1A. *Nature Genetics* 5: 189-194.

Robinson G.C., Jan J.E. and Miller J.R. (1977). A new variety of hereditary sensory neuropathy. *Human Genetics* 11: 53-56.

Saiki R.K. (1990). Amplification of genomic DNA in *PCR Protocols: A Guide to Methods and Applications.* Academic Press Inc.

Sambrook J., Fritsch E.F. and Maniatis T. (1989). Molecular Cloning: A Laboratory Manual (2nd Ed.). Cold Spring Harbor Laboratory Press: New York.

Skre H. (1974) Genetic and clinical aspects of Charcot-Marie-Tooth's disease. *Clinical Genetics* 6: 98-118.

Spreyer P., Kuhn G., Hanemann C.O., Gillen C., Schaal H. *et al.* (1991). Axon-regulated expression of a Schwann cell transcript that is homologous to a "growth arrest-specific" gene. *EMBO Journal* 10: 3661-3668.

Suter U., Welcher A., Snipes G. (1993). Progress in the molecular understanding of hereditary peripheral neuropathies reveals new insights into the biology of the peripheral nervous system. *Trends in Genetics* 16(2): 50-56.

Tuck R.R. and Mcleod J.G. (1983). Retinitis pigmentosa, ataxia, and peripheral neuropathy. *Journal of Neurology Neurosurgery & Psychiatry* 46: 206-213.

Valentijn L.J., Ouvrier R.A., van den Bosch N.H., Bolhuis P.A., Baas F. and Nicholson G.A. (1995). Dejerine-Sottas neuropathy is associated with a de novo PMP22 mutation. *Human Mutation* 5: 76-80.

Vance J., Barker D.F., Yamoaka L.H., Stajich J., Speer C.S. *et al.* (1991). Localization of Charcot-Marie-Tooth disease type 1A (CMT 1A) to chromosome 17p11.2. *Genomics* 9: 623-628.

Weber J.L and 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., Morissette J. *et al.* (1992). A second-generation linkage map of the human genome. *Nature* 359: 794-801.

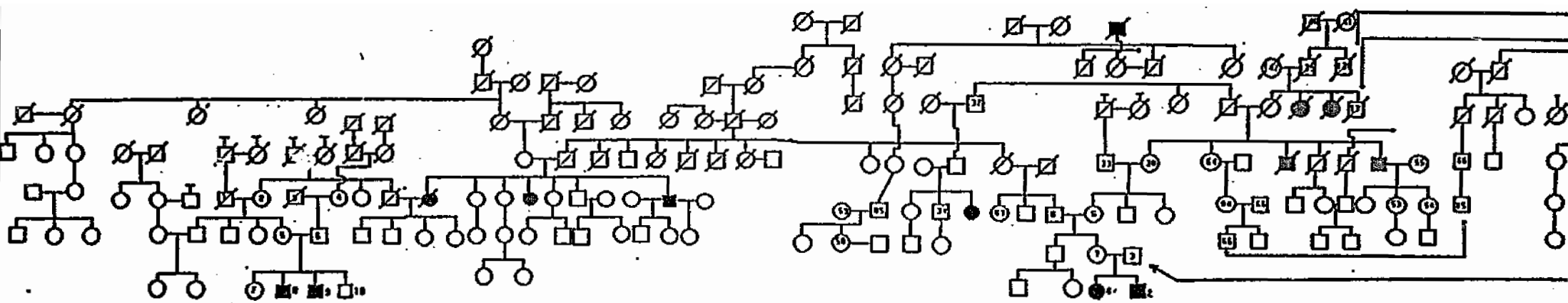
APPENDIX A

LIST OF ABBREVIATIONS

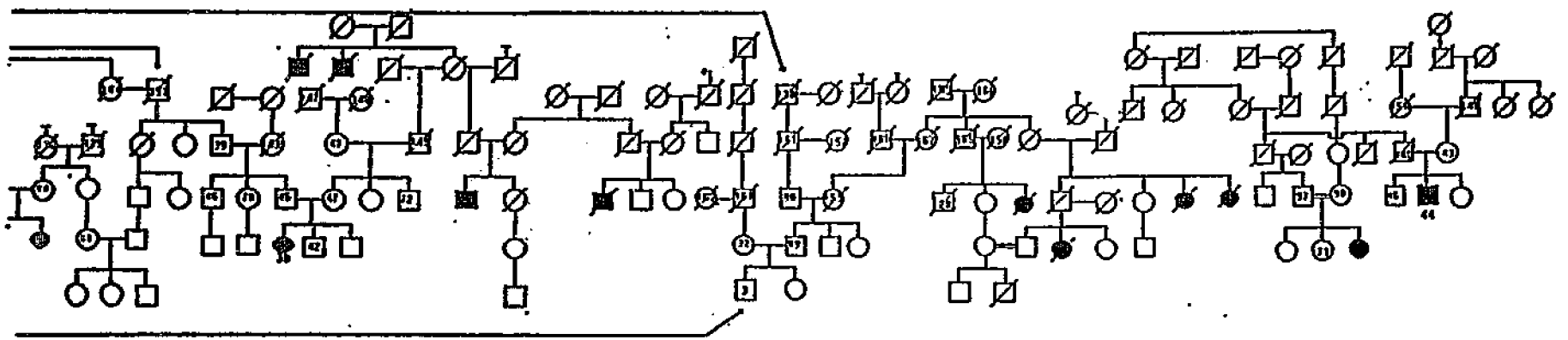
cM	Centi-Morgan
CMT	Charcot-Marie-Tooth
CNTF	Ciliary neurotrophic factor
DNA	Deoxyribonucleic acid
dCTP	Deoxycytosine triphosphate
dNTP	Deoxynucleotide triphosphate
dUTP	Deoxyuridine triphosphate
EDTA	Ethylenediaminetetraacetic acid
HMSN	Hereditary motor and sensory neuropathies
HNPP	Hereditary neuropathy with pressure palsies
IBD	Identical by descent
LOD	Logarithm of the odds
MCV	Motor nerve conduction velocity
MPZ	Myelin protein zero
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PMP2	Peripheral myelin protein 2
PMP22	Peripheral myelin protein 22
TBE	Tris-Borate-EDTA
TEMED	Tetramethylethylenediamine

APPENDIX B

Full pedigree Bulgarian Gypsies (a)

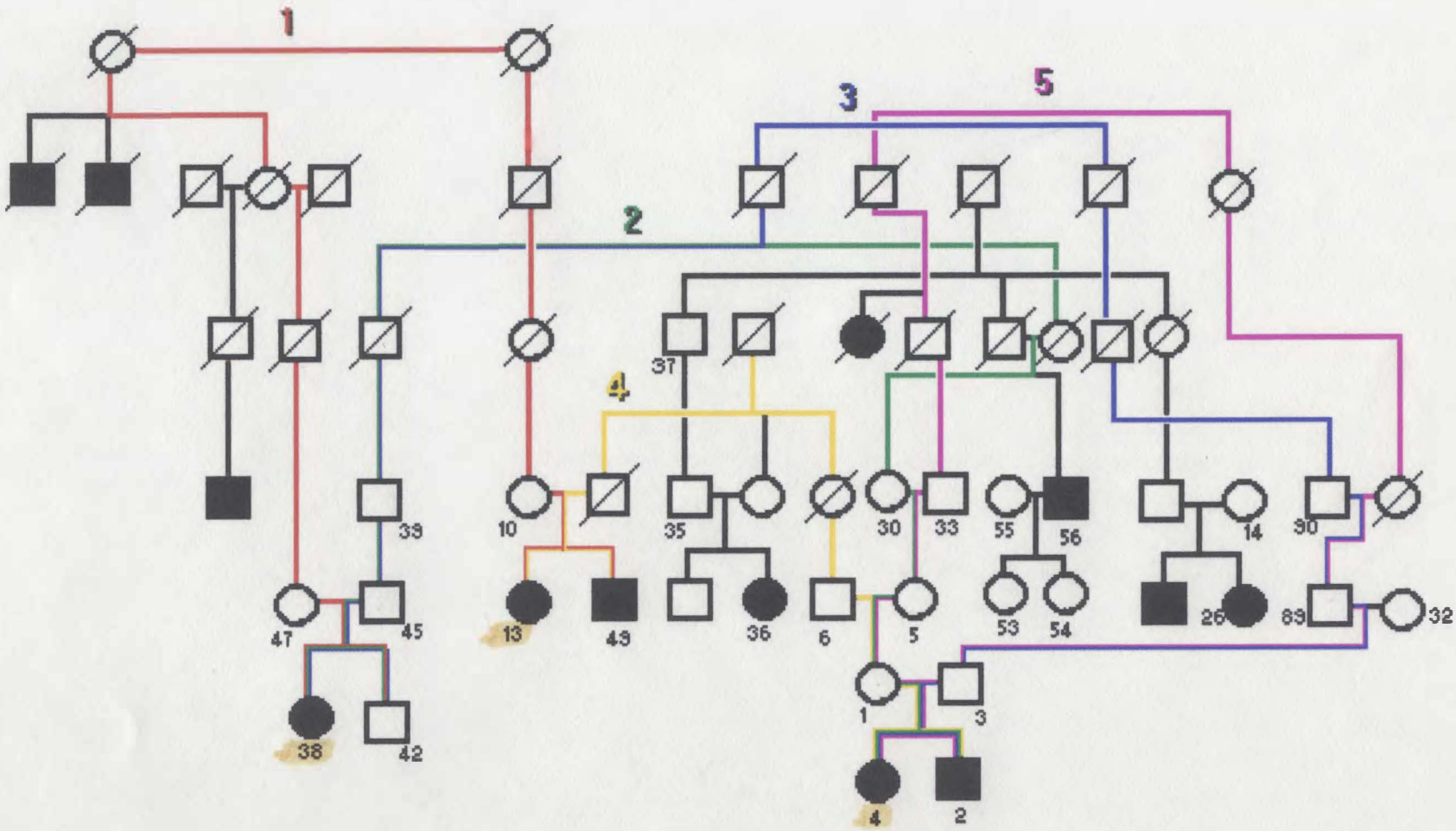


Full pedigree Bulgarian Gypsies (b)



APPENDIX C

Subset of pedigree : Segment Sharing



GYPSY C M T FAMILY: D N A CONCENTRATIONS

Family	Ped#	Link#	Concentration ng/μl	Dilution 10μlDNA + ... H_2
1	99	2	197	187
1	96	4	294	284
1	20	5	647	637
1	91	7	579	569
1	51	8	122	110
1	50	9	86	76
1	52	10	76	66
2	6	1	158	148
2	5	2	194	184
2	89	3	496	486
2	32	4	578	568
2	1	5	204	194
2	3	6	321	311
2	4	7	16	6
2	2	8	303	293
3	10	4	438	428
3	57	5	373	363
3	35	6	489	479
3	12	7	755	745
3	13	8	436	426
3	11	9	182	172
3	7	10	260	250
3	49	11	145	135
3	34	12	456	446
3	36	13	280	270
4	45	1	369	359
4	47	2	438	428
4	38	3	235	225
4	42	4	336	326
5	43	2	367	357
5	46	3	302	292
5	44	4	412	402

Solutions

TRIS-BORATE (TBE)

TO MAKE 1 LITRE OF X10 :

110g Tris Base
55g Boric Acid
40 ml 0.5M EDTA (pH8.0)
Make up to 1 litre with distilled H_2O.

(modified from Sambrook *et al.*, 1989, B.23)

DEIONISED FORMAMIDE

Add 5 g/litre mixed bed resin beads and stir on a magnetic stirrer for 1 hour.

Filter twice through Whatman No. 1 paper.

Store small aliquots under nitrogen at -70°C.

(Sambrook *et al.*, 1989, p. 1.102)

FORMAMIDE LOADING BUFFER

80% Formamide (deionised)
10mM EDTA (pH 8.0)
1mg/ml Xylene Cyanol FF
1mg/ml Bromophenol Blue

(Sambrook *et al.*, 1989, p. 7.76)

TRIS (1M)

TO MAKE 1 LITRE:

Dissolve 121.1g of Tris base in 800ml of H_2O.

Adjust the pH to the desired value by adding concentrated HCl.

For pH 7.4 add approximately 70 ml HCl
pH 7.6 " " 60 ml HCl
pH 8.0 " " 42 ml HCl.

(Sambrook *et al.*, 1989, B. 13)

EDTA (0.5 M, pH 8.0)

TO MAKE 1 LITRE:

Add 186.1 g of disodium ethylenediaminetetraacetate.$2H_2O$ to 800 ml of H_2O.

Stir vigorously on a magnetic stirrer. Adjust the pH to 8.0 with NaOH (~20 g of NaOH pellets). Sterilize by autoclaving.

(Sambrook *et al.*, 1989, B. 11)

25% AMMONIUM PERSULPHATE

TO MAKE 10 ML:

To 2.5 g of ammonium persulphate, add H_2O to 10 ml. Store at 4°C. Make up fresh solution weekly.

(modified from Sambrook *et al.*, 1989, B. 10)

LIQUID ACRYLAMIDE (for radioactive labelling)

TO MAKE 1 LITRE:

Dissolve 420.5 g urea in H_2O and make up to 700 mls.
Add 150 ml "Liqui-mix-19" for a 6% gel or 100 ml "Liqui-mix-19" for a 4% gel.

Deionise for 30 minutes by adding 5 g/litre mixed bed resin beads and stirring on a magnetic stirrer.

Filter once using Whatman No. 1 paper. Make up volume to 900 ml. Add 100 ml of 10X TBE.

Degas (under vacuum) for 30 minutes. Wrap bottle in alfoil and store at 4°C.

(modified from Sambrook *et al.*, 1989, p. 13.47)

5.1[20] AMPLIFICATION BUFFER

(This buffer is used in PCR reactions. 5 = 5X buffer, 1 = 1mM magnesium chloride, 20 = 20 *u*M of each dNTP in the final reaction solution)

TO MAKE 1 ML:

10X Amplification buffer (Biotech)	500 *u*L
H_2O	475 *u*L
5 mM dNTP mix	20 *u*L
1 M $MgCl_2$	5 *u*L

5.1.5[250] AMPLIFICATION BUFFER

(This buffer is used in PCR reactions. 5 = 5X buffer, 1.5 = 1.5 mM magnesium chloride, 250 = 250 *u*M of each dNTP in the final reaction solution)

10X Amplification buffer (Biotech)	500. uL
H_2O	448.5 *u*L
5 mM dNTP mix	50.0 *u*L
1 M $MgCl_2$	1.5 *u*L

GEL FIXATIVE

10% Methanol
10% Glacial Acetic Acid

TO MAKE 2.5 LITRES:

250 ml Methanol
250 ml Glacial Acetic Acid
2 Litres H_2O

APPENDIX E

Capillary system (PCR modification)

The standard 0.65 mL reaction tubes were replaced by a capillary system which used a programmable pipettor (Corbett Research). This system was optimised in the laboratory (by D.Chandler) in order to increase efficiency and reduce costs. This was achieved firstly by a reduction of reagents from 20 *u*L/reaction to 5 *u*L/reaction. Secondly, a reduction in thermal cycling time from an average of ~7 minutes/cycle to ~3 minutes/cycle was achieved. Finally, a 36-well block replaced the standard 16-well block in the thermal cycler (MJ Research). These last two modifications greatly increased throughput.

PCR with radioactive labelling

Master mix X1

These quantities are for one reaction volume of **5** ***u*L**.

Add the following 5 reagents to a 0.65 mL microcentrifuge tube in order:

H_2O	1.6 *u*L
5.1^{20} amplification buffer	1.0 *u*L (see "Solutions" Appendix D)
Primer [1.6 *u*M]	0.25 *u*L

Tth plus or Taq enzyme	0.05 *u*L	(0.3 units/reaction)
^{32}P dCTP	0.1 *u*L	(specific activity 3000Ci/mM)
	3.0 *u*L	
Target DNA	2.0 *u*L	

Master mix X60

This is a 60 times mix used for 45-50 samples.

60 *u*l 5.1^{20} Amplification Buffer

15 *u*l Primer (1.6 *u*M)

3 *u*l Tth Plus Enzyme (0.3 units/reaction)

4 *u*l ^{32}P (specific activity 3000 Ci/mM)

98 *u*l H_2O

180 *u*l total volume

PCR Method:

* Pipette - 2 *u*l of the target DNA into capillary (Corbett capillary pipettor)
* Pipette - 3 *u*l of master mix
* Cap capillary to prevent leakage
* Place into a thermocycler (MJ Research) programmed as follows:

Step 1	94^0 C for 5 minutes
Step 2	94°C for 20 seconds
Step 3	55°C for 40 seconds
Step 4	72^{0}C for 1 minute

Repeat steps 2 to 4, 29 times

Post PCR:

* Expel amplified sample into labelled 0.65 mL microcentrifuge tube containing 8 *ul* of formamide loading buffer (see "Solutions" Appendix D).

  While gel is warming pre-run:

* Place tube into a heating block for 5 minutes denaturing at 94°C.

Preparation of glass plates (used with Hoefer Poker Face sequencer for radioactively labelled products)

* Wash two glass plates (37cm x 44cm and 37cm x 41cm) in hot water and liquid pyroneg (Diversey) detergent.
* Rinse with hot water
* Spray with 70% ethanol
* Wipe dry with tissues
* Silanise the smaller back plate with "Sigmacote" (Sigma Chemicals)
* Allow to dry
* Place the two plates together with 0.35mm nylon spacers between them at the long edges.

* Place three bulldog clamps at each side leaving the bottom third of the plates free.
* Tape bottom one third with masking tape
* Place a small wedge centrally at the top between the two plates to assist with the pouring of the gel.

Pouring a polyacrylamide gel (used with Hoefer Poker Face sequencer for radioactively labelled products)

* Pour 70 mLs of liquid acrylamide[1] into a 250 mL beaker
* Add 170 *u*L tetramethylethylenediamine (Temed)
* Add 170 uL 25% ammonium persulphate
* Mix carefully and draw up without delay into a 50mL syringe
* Inject between the glass plates held at a slight incline
* Place the plates horizontally, remove the wedge
* Insert the well-forming spacer
* Leave to set at room temperature for at least one hour before use.

[1](19:1 acrylamide:bisacrylamide, Gradipore Liqui-Mix, see "Solutions" Appendix D)

Assembling the electrophoresis apparatus (Radioactive labelling)

The SE1600 Poker Face II sequencer (Hoefer Scientific Instruments) was used with the 3000Xi electrophoresis power supply (Bio-Rad).

* Add enough 1X TBE buffer (see"Solutions" Appendix D) to cover electrode in lower buffer chamber.
* Remove masking tape from the glass plates
* Place aluminium and perspex backing plate against the smaller glass plate (this functions to provide even heat distribution across the gel).
* Clamp backing plate into place with the two side clamps provided.
* Place this unit into a vertical position in the lower buffer tank and secure.
* Fill upper buffer tank with 1X TBE buffer
* Remove well-forming spacer and flush out well with 1X TBE (this removes excess urea).
* Place top safety lid and connectors and bottom connectors into position
* Plug power leads into the power supply.

Polyacrylamide gel electrophoresis (used with Hoefer Poker Face sequencer for radioactively labelled products)

* Pre-warm gel for 20 minutes (connect to power supply and set the voltage at 1400V)
* Place tubes containing amplified products and formamide loading buffer into a heating block at 90°C for 5 minutes to denature the double-stranded DNA.
* Place tubes into ice to prevent evaporation and/or re-annealing of the complementary strands.
* Load 1 to 2 *u*L of each sample was loaded (1 *u*L for small wells of the 66-well comb, 2 *u*L for the 44-well comb).
* Replace top safety cover
* Run gel for 1 hour 30 minutes to 3 hours 30 minutes at 1400 volts depending on the size range of the alleles and percentage acrylamide.

4%	>200bp	2.5 - 3 hours
6%	>200bp	3 - 3.5 hours
6%	150->200bp	2.5 - 3 hours
8%	<100bp	1.5 - 2 hours

Note: For tetranucleotide repeat markers less time will be needed compared to dinucleotide repeats for separation of alleles.

Post-run (procedures for radioactive labelling)

* Disconnect power supply
* Remove upper safety lid and connectors
* Empty upper buffer tank
* Remove side clamps
* Prize apart the two glass plates using a wedge and remove the plate without the gel
* Fix gel with 10% methanol 10% glacial acetic acid for 15 minutes
* Drain excess moisture and place a piece of blotting paper slightly larger than the gel on top of the gel.
* Peel off paper with gel attached and cover gel side with plastic wrap.
* Place gel on a slab gel dryer (Savant SGD4050) at 80°C for 1 hour under vacuum.
* Place dry gel in an x-ray cassette (Amersham Hypercassette™) with a sheet of x-ray film and leave at room temperature overnight.
* Develop x-ray film (Curix benchtop automatic processor, Agfa).

Preparation of polyacrylamide (fluorescent labelling)

* Place 15 gms urea (molecular biology grade) into a 250 ml Erlenmeyer flask
* Add 4.5 mls of Liqui-gel (40%, 19:1 acrylamide:bis)
* Add 10.75 mls of Milli-Q water
* Add 3 mls of 10X TBE (see "Solutions" Appendix D).
* Cover with parafilm and dissolve under warm running water
* Filter the solution through a 0.22mm pore size filter (Nalgene vacuum filter unit).
* De-gas under vacuum for 10 minutes.

Pouring polyacrylamide gel (fluorescent labelling)

* Make up 250 *ul* of 10% ammonium persulphate (APS) (0.025 gms in 250 *ul* Milli-Q water)
* Pour 20 mLs of freshly prepared acrylamide into a 100 ml beaker
* Add 250 *ul* of APS
* Add 25 *ul* of TEMED
* Draw up solution without delay into a 50mL syringe
* Inject solution between the horizontally placed glass plates
* When the solution reaches the top of the plates remove the wedge
* Insert a well-forming spacer

* Leave gel to set at room temperature for at least one hour before use.

PCR with fluorescent dUTP labelling

* Dilute [f]dUTPs (separate tubes) as follows:

40*u*M for Tamra : Add 1ul of 400uM Tamra to 9ul of Tris buffer (pH 9.5)

10 *u*M for R6G : Add 1ul of 100uM to 9 ul Tris buffer (pH 9.5).

10 *u*M for R110 : Add 1 ul of 100 uM to 9 ul Tris buffer (pH 9.5).

Master mix:

These quantities are for one reaction volume of 5 *u*L.

Add the following 5 reagents to a 0.65 mL microcentrifuge tube in order:

H_2O	1.4 *u*L
5.1.5[250] amplification buffer	1.0 *u*L (refer "Solutions" Appendix D)
Primer [0.5 *u*M]	0.25 *u*L
Tth plus or Taq enzyme	0.1 *u*L (0.3 units/reaction)
[F]Dutp	0.25 *u*L
	3.0 *u*L
Target DNA	2.0 *u*L

PCR Method:

* Pipette - 2 *ul* of the target DNA into capillary (Corbett capillary pipettor)

* Pipette - 3 *ul* of master mix

* Cap capillary to prevent leakage
* Place into a thermocycler (MJ Research) programmed as follows:

Step 1	94°C for 5 minutes
Step 2	94°C for 20 seconds
Step 3	55°C for 40 seconds
Step 4	72°C for 1 minute

Repeat steps 2 to 4, 29 times

Post PCR:

* Expel amplified sample into labelled 0.65 mL microcentrifuge tube containing 10 *ul* of water.
* Add 1 *ul* of each sample to be pooled (run in the same lane) to a new 0.65 mL tube with 5 *ul* loading buffer (1 part Dextran blue, 5 parts distilled and deionised molecular biology grade formamide).
* Add 0.5 *ul* sizing standard dye (ROX 500) to each tube.
* Place tubes into a heating block for 5 minutes denaturing at 94°C.

The gels are run on an ABI Prism™ 373 DNA sequencer with GeneScan™ software (Perkin-Elmer).

APPENDIX F

MATERIALS

List of chemicals

Chemical	Supplier
Deoxynucleotide triphosphates (dNTPs) 100mM (Concentration : 40 *u*Mole)	Biotech International Ltd.
Tth Plus DNA Polymerase (5.5 units/*ul*)	Biotech International Ltd.
Magnesium Chloride ($MgCl_2$) (1.00M)	Sigma Chemical Co.
Glacial Acetic Acid (CH_3COOH)	BDH
Methanol (CH_3OH)	BDH
Liqui-gel (Liqui-mix-19 : Acrylamide/Bis 19:1)	Gradipore Ltd.
Boric Acid (H_3BO_3)	Sigma Chemical Co.
Tetramethylethylenediamine (TEMED) ($C_6H_{16}N_2$)	Sigma Chemical Co.
Ammonium Persulphate $(NH_4)_2S_2O_8$	Sigma Chemical Co.
Analytical Grade Mixed Bed Resin	Bio-rad
Chlorinated organopolysiloxane in Heptane (Sigmacote)	Sigma Chemical Co.
Formamide (CH_3NO)	Sigma Chemical Co.
Xylene Cyanole FF ($C_{25}H_{27}N_2O_6S_2Na$)	Sigma Chemical Co.
Bromephenol blue (Tetrabromophenolsulfo-nephthalein)($C_{19}H_9Br_4O_5SNa$)	Sigma Chemical Co.

Tris[Hydroxymethyl]aminomethane ($C_4H_{11}NO_3$) (Trizma Base)	Sigma Chemical Co.
Sodium Hydroxide (NaOH)	Ajax Chemicals
Urea (CH_4N_2O)	Sigma Chemical Co.

APPENDIX G

```
--------------------------------------------------------------
--------------

           Common ancestors:

           Ancestor  Frequency
            --------------
            Total         0

           Paths:

            Length   Frequency
            --------------
            Total         0

           Inbreeding coefficient for    66 is:
0.00000000

--------------------------------------------------------------
--------------

           Common ancestors:

           Ancestor  Frequency
            --------------
            Total         0

           Paths:

            Length   Frequency
            --------------
            Total         0

           Inbreeding coefficient for    67 is:
0.00000000

--------------------------------------------------------------
--------------

           Common ancestors:

           Ancestor  Frequency
            --------------
            Total         0

           Paths:

            Length   Frequency
            --------------
            Total         0

           Inbreeding coefficient for    51 is:
0.00000000

--------------------------------------------------------------
--------------
```

```
            Common ancestors:

            Ancestor  Frequency
             ----------
             Total         0

            Paths:

             Length    Frequency
             ----------
             Total         0

            Inbreeding coefficient for    33 is:
0.00000000

-------------------------------------------------------------
--------------

            Common ancestors:

            Ancestor  Frequency
             ----------
             Total         0

            Paths:

             Length    Frequency
             ----------
             Total         0

            Kinship coefficient for    31 and    33 is:
0.00000000

-------------------------------------------------------------
--------------

            Common ancestors:

            Ancestor  Frequency
             ----------
             Total         0

            Paths:

             Length    Frequency
             ----------
             Total         0

            Inbreeding coefficient for    31 is:
0.00000000

-------------------------------------------------------------
--------------

            Common ancestors:

            Ancestor  Frequency
             ----------
```

```
            Total        0

           Paths:

            Length   Frequency
            --------------
            Total        0

           Inbreeding coefficient for    14 is:
0.00000000


           Inbreeding coefficient for     9 is:
0.00000000

-------------------------------------------------------------
---------------

           Common ancestors:

           Ancestor  Frequency
            --------------
            Total        0

           Paths:

            Length   Frequency
            --------------
            Total        0

           Inbreeding coefficient for    52 is:
0.00000000


           Inbreeding coefficient for    53 is:
0.00000000

-------------------------------------------------------------
---------------

           Common ancestors:

           Ancestor  Frequency
            --------------
            Total        0

           Paths:

            Length   Frequency
            --------------
            Total        0

           Inbreeding coefficient for    54 is:
0.00000000


           Inbreeding coefficient for    55 is:
0.00000000
```

```
          Inbreeding coefficient for    57 is:
0.00000000

-------------------------------------------------------------
--------------

          Common ancestors:

          Ancestor  Frequency
           --------------
           Total         0

          Paths:

           Length    Frequency
           --------------
           Total         0

          Inbreeding coefficient for    72 is:
0.00000000

-------------------------------------------------------------
--------------

          Common ancestors:

          Ancestor  Frequency
           --------------
           Total         0

          Paths:

           Length    Frequency
           --------------
           Total         0

          Inbreeding coefficient for    73 is:
0.00000000

-------------------------------------------------------------
--------------

          Common ancestors:

          Ancestor  Frequency
           --------------
           Total         0

          Paths:

           Length    Frequency
           --------------
           Total         0

          Inbreeding coefficient for    82 is:
0.00000000
```

```
--------------------------------------------------------------
--------------

           Common ancestors:

           Ancestor  Frequency
            --------------
            Total         0

           Paths:

            Length   Frequency
            --------------
            Total         0

           Inbreeding coefficient for    83 is:
0.00000000

--------------------------------------------------------------
--------------

           Common ancestors:

           Ancestor  Frequency
            --------------
            Total         0

           Paths:

            Length   Frequency
            --------------
            Total         0

           Inbreeding coefficient for    58 is:
0.00000000

--------------------------------------------------------------
--------------

           Common ancestors:

           Ancestor  Frequency
            --------------
            Total         0

           Paths:

            Length   Frequency
            --------------
            Total         0

           Inbreeding coefficient for    59 is:
0.00000000

--------------------------------------------------------------
--------------
```

```
            Common ancestors:

            Ancestor  Frequency
             --------------
             Total         0

            Paths:

             Length    Frequency
             --------------
             Total         0

            Inbreeding coefficient for    74 is:
0.00000000

-------------------------------------------------------------
--------------

            Common ancestors:

            Ancestor  Frequency
             --------------
             Total         0

            Paths:

             Length    Frequency
             --------------
             Total         0

            Inbreeding coefficient for    75 is:
0.00000000


            Inbreeding coefficient for    60 is:
0.00000000

-------------------------------------------------------------
--------------

            Common ancestors:

            Ancestor  Frequency
             --------------
             Total         0

            Paths:

             Length    Frequency
             --------------
             Total         0

            Inbreeding coefficient for    61 is:
0.00000000

-------------------------------------------------------------
--------------
```

```
            Common ancestors:

            Ancestor  Frequency
             ----------
             Total         0

            Paths:

             Length    Frequency
             ----------
             Total         0

            Inbreeding coefficient for    76 is:
0.00000000

------------------------------------------------------------
--------------

            Common ancestors:

            Ancestor  Frequency
             ----------
             Total         0

            Paths:

             Length    Frequency
             ----------
             Total         0

            Inbreeding coefficient for    62 is:
0.00000000

------------------------------------------------------------
--------------

            Common ancestors:

            Ancestor  Frequency
             ----------
             Total         0

            Paths:

             Length    Frequency
             ----------
             Total         0

            Inbreeding coefficient for    78 is:
0.00000000

------------------------------------------------------------
--------------

            Common ancestors:

            Ancestor  Frequency
             ----------
```

```
            Total        0

           Paths:

            Length   Frequency
            ----------
            Total        0

           Inbreeding coefficient for    64 is:
0.00000000

-------------------------------------------------------
--------------

           Common ancestors:

           Ancestor  Frequency
            ----------
            Total        0

           Paths:

            Length   Frequency
            ----------
            Total        0

           Inbreeding coefficient for    65 is:
0.00000000

-------------------------------------------------------
--------------

           Common ancestors:

           Ancestor  Frequency
            ----------
            Total        0

           Paths:

            Length   Frequency
            ----------
            Total        0

           Inbreeding coefficient for    46 is:
0.00000000


           Inbreeding coefficient for    47 is:
0.00000000


           Inbreeding coefficient for    43 is:
0.00000000

-------------------------------------------------------
--------------
```

```
            Common ancestors:

            Ancestor  Frequency
             ----------
             Total         0

            Paths:

             Length    Frequency
             ----------
             Total         0

            Inbreeding coefficient for    44 is:
0.00000000

---------------------------------------------------------------
---------------

            Common ancestors:

            Ancestor  Frequency
             ----------
             Total         0

            Paths:

             Length    Frequency
             ----------
             Total         0

            Inbreeding coefficient for    42 is:
0.00000000


            Inbreeding coefficient for    37 is:
0.00000000
```