A GENETIC ANALYSIS OF AUTOSOMAL RECESSIVE FORMS OF RETINITIS PIGMENTOSA

ELSPETH A. BRUFORD

THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF EDINBURGH

1996

DECLARATION

I declare that

a) I have composed this thesis myself; and

b) the work is my own, except where stated.

Elspeth A. Bruford, B.Sc.(Hons)

ABSTRACT

Retinitis pigmentosa (RP) is a heterogeneous group of genetic disorders characterised by a progressive pigmentary degeneration of the retina. The disease manifests as night-blindness, tunnel vision, and often total loss of sight. RP is a common cause of blindness, affecting around 1 in 4000 of the general population.

RP can be inherited as an autosomal dominant, recessive or X-linked condition, with autosomal recessive RP (arRP) accounting for at least a third of all cases. RP also occurs as a feature of several syndromes in association with non-ocular features. One example is the autosomal recessive condition Bardet-Biedl syndrome (BBS), in which RP is associated with polydactyly, mental retardation, hypogonadism and obesity. To date, five genes have been identified as causing autosomal recessive forms of RP, and linkage to another twelve chromosomal regions has been established, four of these involving BBS. The aim of this project was to identify loci for recessive forms of RP by genetic linkage analysis, using patients with arRP and BBS, and to test for mutations in any resultant candidate genes.

Pedigrees with arRP from south-central Sardinia, an ethnic outlier with a high prevalence of recessive disease, were studied initially. Linkage analysis was carried out using genome-wide microsatellite markers, and genetic heterogeneity was identified among the 11 families studied. Examination of the data revealed potential linkage to D14S80, on chromosome 14q11, in a subset of families. Fine mapping with further markers in this region identified a region of homozygosity in one consanguineous family, suggesting identity-by-descent.

A strong candidate gene for RP, coding for the neural retina specific leucine zipper (NRL), was found to be located within the region of homozygosity. NRL is an evolutionarily conserved protein which is expressed in all layers of the neural retina, and is thought to have a role in the transcriptional regulation of retinal genes, including rhodopsin. The NRL gene was studied in the consanguineous family by direct sequencing and mutation analysis, but no mutations were identified within the coding region.

A total of 29 BBS pedigrees collected worldwide were studied using markers in regions where linkage was established during the course of the study. These loci are located on chromosomes 11q13, 16q21, 3p13-p12 and 15q22.3-q23. The results revealed significant genetic heterogeneity, with most families showing linkage to 11q13 and 15q22.3-q23, and others being consistent with linkage to the 16q chromosomal region. Some of the larger pedigrees could be assigned to specific loci, but many of the smaller families were too small for a definite assignment. The results of multipoint linkage and haplotype analyses in these three linked chromosomal regions, especially in consanguineous pedigrees, have helped to narrow down the region of search for candidate genes, in particular refining the *BBS4* locus from an interval of 4.3 cM to 1.6 cM.

Both of these studies highlight the difficulties of linkage mapping in autosomal recessive disorders, and the power of homozygosity mapping in genetically heterogeneous conditions.

This thesis is dedicated to the memory of my father,

Dr. Alan J. Bruford

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, Alan Wright and Jeff Haywood, for creating a project I was interested in, and for all their help - when I could track them down! I would also like to thank the head of the MRC Cell Genetics section, Veronica van Heyningen, for all the consideration she has shown me during my time at the MRC.

Secondly, I would like to thank everyone I've worked with during the various stages of this project, especially the "girls" in the office, and Kate and Forbes. The following people get a special mention - David Mansfield, for showing me the ropes at the MRC; Stewart Morris, for all his invaluable and foolproof computing advice; Claire Whitton for all those much appreciated "breaths of fresh air"; and Karen Porter - when she wasn't singing - for her enthusiasm and boundless energy. Also a big thank-you to Sandy, Norman and Douglas in the MRC photography department for all their patience and consideration, and to Peter Teague for all those lovely lod scores.

Thirdly, a lot of credit must go to my friends in the real world, and PAMS, for keeping me sane(ish) and laughing. Three people in particular - Karen, Bea, and especially Paul - you were all there for me when I needed you most, I can never thank you enough. And of course Niall, for being my inspiration at the bar, and the pool table - I beat you!!

Finally, this is really for my Mother, without whom I wouldn't be here (in many ways!). Thank-you for everything you've given me, especially the love and support, and for believing in me.

ABBREVIATIONS

a/A	adenine
adRP	autosomal dominant retinitis pigmentosa
ALF	automated laser fluorescent sequencer (Pharmacia)
APS	ammonium persulphate
arRP	autosomal recessive retinitis pigmentosa
ATP	adenosine triphosphate
bp	base pairs
BBS	Bardet-Biedl syndrome
c/C	cytosine
cen	centromere
CEPH	Centre d'Etude Polymorphisme Humain
cGMP	cyclic guanosine monophosphate
cM	centiMorgan
CSNB	congenital stationary night blindness
(d)dNTP	(di)deoxy-nucleotide triphosphate
DMSO	dimethyl sulphoxide
(c)DNA	(complementary) deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediamine tetra-acetic acid
EOG	electroculogram
ER	endoplasmic reticulum
ERG	electroretinogram
g/G	guanine
GDB	Genome Database
GMP, GDP, GTP	guanosine mono/di/tri-phosphate
HA	heteroduplex analysis
HLA	human leucocyte antigen
HPLC	high pressure liquid chromatography
IB	inclusion bodies
IBD	identity-by-descent
IPTG	isopropylthio-β-D-galactoside

kb	kilobase pairs
kDa	kiloDalton
Mb	megabase pairs
MR	mental retardation
NARP	neuropathy, ataxia and retinitis pigmentosa
ORF	open reading frame
p	short arm of chromosome
PCR	polymerase chain reaction
PDE	phosphodiesterase
PPRPE	preserved para-arteriolar retinal pigment epithelial RP
q	long arm of chromosome
RFLP	restriction fragment length polymorphism
(m)RNA	(messenger) ribonucleic acid
ROS	rod outer segment
RP	retinitis pigmentosa
RPA	retinitis punctata albescens
RPE	retinal pigment epithelium
RT-PCR	reverse-transcription PCR
SDS	sodium dodecyl sulphate
SSCP	single-stranded conformation polymorphism
t/T	thymine
TAE	Tris acetate EDTA
TBE	Tris borate EDTA
TBS	Tris buffered saline
TE	Tris EDTA
tel	telomere
TEMED	tetra-methyl ethylenediamine
θ	recombination fraction
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
xlRP	X-linked retinitis pigmentosa
Z	lod score

CONTENTS

CHAPTER TWO - MATERIALS AND METHODS

CHAPTER ONE

INTRODUCTION

Retinitis pigmentosa (RP) is the name given to a set of hereditary progressive pigmentary disorders of the retina. With a prevalence of around 1 in 4000 in most populations (Boughman, Conneally and Nance, 1980) and affecting approximately 1.5 million people worldwide (Kumar-Singh *et al*, 1993), RP is a major cause of blindness. As the retina has a limited response to disease the term retinitis pigmentosa covers a genetically very heterogeneous group of disorders. RP can be inherited as an autosomal dominant, autosomal recessive or X-linked recessive condition and is often categorised on this basis, although in many simplex cases the mode of inheritance cannot be determined (Heckenlively, 1988a). Most cases are thought to be monogenic but there is also a report of digenic RP inheritance, involving mutations in two separate loci (Kajiwara, Berson and Dryja, 1994). Much clinical variability exists both between and within each genetic subgroup, although it is generally accepted that the X-linked and autosomal recessive forms result in a more severe phenotype (Bird and Jay, 1994). The first symptoms of RP are usually a progressive deterioration in night vision (night blindness), followed by constriction of the peripheral visual fields (tunnel vision) due to degeneration of the rod photoreceptors within the eye. The disease frequently results in loss of central vision causing total blindness in the middle decades of life. To date, mutations in ten genes have been shown to cause RP when mutated, but linkage between RP and many other chromosomal locations has been reported. Identification of more causative genes could lead to a better understanding of the disease mechanisms involved, and also allow for genetic counselling for the many families affected by this disorder. This study is concerned with attempting to map and identify genes responsible for autosomal recessive RP in a Sardinian population, and for Bardet-Biedl syndrome, an example of an RP syndrome where other organ systems are also affected (reviewed in Green *et al*, 1989).

1.1 - THE GENETICS OF RP

Retinitis pigmentosa is usually inherited as an autosomal dominant, autosomal recessive or X-linked recessive condition, and is frequently classified by its mode of inheritance. It is often possible to categorise cases of RP on the basis of the family history of the patient, but commonly there is not sufficient information about the family, or insufficient cooperation for this to be done, and the genetic type of RP must remain undetermined. There is also the possibility of maternal inheritance, which occurs when the gene causing RP is present in the

mitochondrial genome, such as in NARP syndrome (neuropathy, ataxia and retinitis pigmentosa).

If there is only one affected individual presenting this can make it difficult for the disease to be classified genetically. Such cases are referred to as "simplex", and many may be due to autosomal recessive inheritance, but it is important not to presume that this is the case without sufficient evidence. The term "sporadic" should not be used unless a mutational or environmental event has been suggested as the cause of the RP in an individual. This can include rubella or cytomegalovirus infection, or exposure to chloroquine (used to treat arthritis) or thioridazine (used in the treatment of psychoses) (Heckenlively, 1988a).

RP is generally accepted as being a monogenic disorder, with mutations in only one gene being responsible in each pedigree. However, Kajiwara, Berson and Dryja (1994) reported digenic inheritance, involving mutations in two gene loci, in three RP families. This poses the possibility that RP could in some cases be an oligogenic trait, with more than one gene influencing the phenotype. This could account for the clinical variability seen within families, but would greatly complicate the search to identify specific causative genes.

1.2 - CLINICAL SYMPTOMS OF RP

Typical retinitis pigmentosa primarily affects the rod photoreceptors, with the cones only being involved as the disease progresses (Deutmann, 1977). Ophthalmoscopic examination reveals progressive pigmentary changes in the retinas of both eyes, with the migration of pigment-laden cells from the retinal pigment epithelium (RPE) to the mid-periphery of the retina where many of the rods are found. This pigment often accumulates around the blood vessels in bone corpuscle-like arrangements as shown in **figure 1.1B**, but can also be seen in irregular clumps or spots, or is even absent in some cases (retinitis pigmentosa sine pigmento). The blood vessels become narrowed but probably only as a secondary feature of the disease (Deutmann, 1977). As the rod cells act as receptors for dim light, one of the first symptoms of RP is night-blindness, followed by a narrowing of the visual fields (tunnel vision) as the rods are also essential for peripheral vision, being more concentrated in the mid-peripheral retina (Krill, 1972). Central vision is usually preserved until the end stages of RP. The loss of rod function can be measured by a scotopic electroretinogram (ERG) which measures electrical activity in the retina in response to flashes of light, following dark-adaptation of the patient (Heckenlively, 1988a). Photopic (cone) responses can also be measured in light-adapted conditions. In a typical RP patient, the rod responses show reduced or absent amplitudes, and may also be temporally delayed. It is sometimes possible to detect ERG changes in presymptomatic cases of RP (Humphries, Kenna and

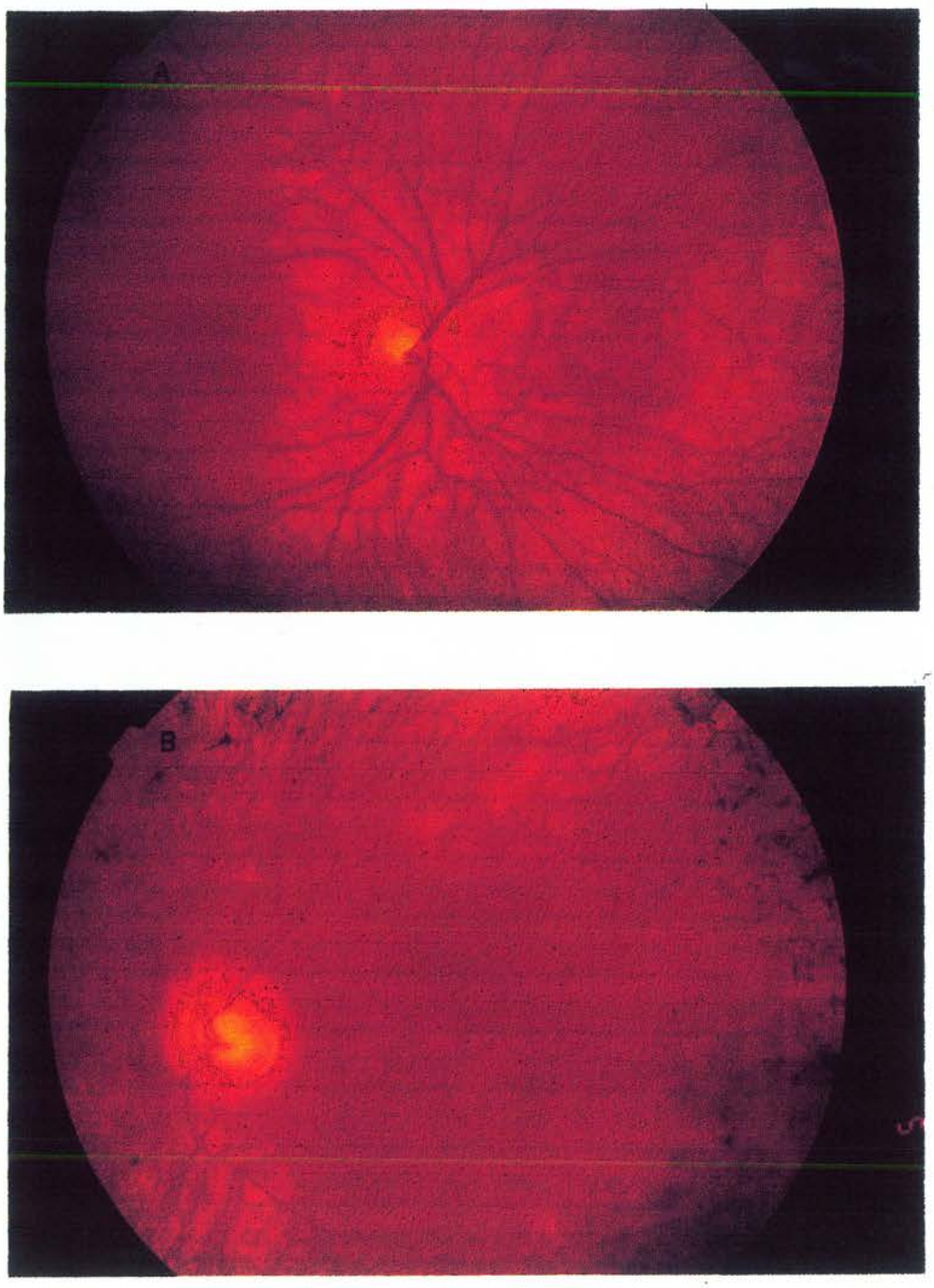


Figure 1.1

A: normal optic disc and retinal vessels

B: RP patient - optic disc pallor and attenuation of retinal vessels, with typical bone spicule pigmentary deposits

Farrar, 1992). Other diagnostic tests that can be used include a visual field test, measurement of visual acuity, a dark adaptation test, fluorescein angiography, and the electro-oculogram (EOG) which measures the standing potential generated by the RPE (Heckenlively, 1988b).

As the course of the disease progresses with continuing loss of the rod cells, more extensive changes take place in the retina, as seen by fundus photography (**fig. 1.1**). The optic nerve develops a waxy pallor, the blood supply to the retina diminishes due to the attenuation of the blood vessels, and the retina visibly thins. The cone cells, which mediate daylight (colour) and central vision, also begin to die, causing further reduction in visual acuity. Eventual loss of central visual acuity is usually due to macular involvement, macular cysts or posterior subcapsular cataracts. These cataracts, characterised by yellow crystalline changes in the lens, are commonly seen in all types of RP with an overall frequency of 41%, although the cause of their formation is unknown (Heckenlively, 1988b).

1.2.1 - Clinical Classification of Genetic Subgroups

The age of onset, severity and progression of RP varies between the genetic sub-types, with X-linked RP and autosomal recessive RP usually being more severe than autosomal dominant RP. However, clinical heterogeneity has also been identified within each of the three genetic classes. Studies of autosomal dominant RP (adRP) have revealed two broad subdivisions of the disease with type I, a "diffuse" form, and type II, a "regional" form (Massof and Finkelstein, 1981; Lyness *et al*, 1985). Type I shows a widespread loss of rod function, with night-blindness in childhood and later cone involvement. Type II shows a regionalised loss of both cone and rod function, with night-blindness onset in adulthood. A further variant that has been identified with autosomal dominant inheritance is sector RP, where the dystrophy is restricted to specific parts of the fundus (usually the lower half) and the progression of the disease is very slow. Sector RP has also been inherited in rare cases as an autosomal recessive condition, and has been observed in X-linked carriers (Deutmann, 1977). It is worth considering that, as with many dominant conditions, the expressivity of adRP varies between generations, and Ernst and Moore (1988) have suggested that up to 30% of obligate carriers show only mild symptoms of the disease.

Excluding cases of recessively inherited RP where the disease appears as part of a syndrome, subtypes of autosomal recessive RP (arRP) have also been identified. One example is preserved para-arteriolar retinal pigment epithelial RP (PPRPE), a severe early-onset form of RP (Heckenlively, 1988c). Kaplan *et al* (1990) have suggested that most cases of arRP can be subdivided into two main groups, precocious onset (average age 7.5

years) with severe progression, and later onset (average age 17 years) with mild progression. A third type, also reported by Grondahl (1987a) and Bonneau *et al* (1992) is the "senile" form of arRP, where onset of symptoms does not occur until the fifth or sixth decade.

A sub-division of X-linked RP (xlRP) has been made based on the clinical symptoms of female carriers. The presence of a tapetal reflex, a golden metallic sheen seen in the fundus, and first observed by Falls and Cotterman (1948), has been noted in the heterozygotes of some families. Linkage studies originally suggested the presence of two separate xlRP loci, *RP2* and *RP3* (a third locus, *RP15*, has now been reported by McGuire *et al*, 1995b), and some studies predict that families with a tapetal reflex are linked to the *RP3* locus, recently identified as the *RPGR* gene (Meindl *et al,* 1996), as opposed to the *RP2* locus (Nussbaum *et al*, 1985, Denton *et al*, 1988, Musarella *et al*, 1989).

1.3 - PREVALENCE OF RP

Retinitis pigmentosa appears in all ethnic groups with comparable gene frequencies. Overall prevalence rates have ranged from 1 in 1500, from a study in Birmingham (UK) (Bundey and Crews, 1984), to 1 in 7000 in Switzerland (Ammann, Klein and Franceschetti, 1965), but generally an average of around 1 in 4000 is reached (**figure 1.2**).

Many studies have been carried out in an attempt to ascertain the relative proportions that each genetic subgroup accounts for in the RP population. Results from various studies differ considerably, and several reasons for this are possible, such as the method of recruitment used by the researchers; the sampling area covered; the method of ascertainment used; and the population under consideration (e.g. level of consanguinity, sex and age distribution). The high degree of variability seen between these figures (**figure 1.3**) shows that they can be regarded only as guidelines to the relative proportions of genetic groups. It is important to discriminate between figures for the number of affecteds as opposed to the number of probands, or families, as this can introduce a bias into the analysis of dominant and simplex cases. The large proportion of arRP in some studies may be attributable to high levels of consanguinity, or inbreeding in the population, as seen in Sardinia.

Overall, it seems apparent that xlRP is generally the least common form of RP, and that simplex cases make up the largest proportion of RP patients. While some of the simplex cases will be due to recessive inheritance in small families, in genetic analyses there still remains an excess of simplex RP beyond probabilistic expectations (Boughman and

Author	Area Studied	Prevalence Rate
Ammann, Klein and Franceschetti (1965)	Switzerland	1 in 7000
Merin and Auerbach (1976)	Israel	1 in 4500
Boughman, Conneally and Nance (1980)	USA and Canada	1 in 3700
Bundey and Crews (1984)	Birmingham, UK	1 in 1500
Bunker *et al* (1984)	Maine, USA	1 in 3544
Hu (1987)	China	1 in 4016
Grondahl (1987b)	Norway	1 in 4440
Fossarello *et al* (1993)	South-Central Sardinia	1 in 3355

Figure 1.2

Prevalence of retinitis pigmentosa

Author	Area	% simplex	% arRP	% adRP	% xlRP
Jay (1982)	England	42.1	15.5	24.4	18.0
Hu (1982)	China	48.3	33.1	11	7.7
Boughman and Fishman (1983)	Illinois	50.0	16.0	21.7	9.0
Grondahl (1986)	Norway	53	30	15	2
Kaplan *et al* (1990)	France	41.9	21.5	20.4	12.9
Haim (1992)	Denmark	43.2	22.6	6.9	10.8
van den Born, Bergen and Bleeker-Wagemaker (1992)	Netherlands	58.1	21.9	13.4	6.6
Fossarello *et al* (1993)	South-Central Sardinia	53.8	21.5	5	1.3

Figure 1.3

Distribution (in percentages) of genetic subgroups in families with retinitis pigmentosa

Fishman, 1983; Jay, 1982). Several reasons have been suggested for the high number of simplex cases:

Some may be phenocopies of RP, or multifactorial in nature

New mutations in dominant RP genes

Reduced penetrance of adRP (mildly affected individuals are not diagnosed)

Simplex females may be xlRP heterozygotes

Simplex males may have xlRP (Pagon, 1988).

It is clear that genetic diagnosis would be greatly aided by the development of genetic tests for commonly mutated genes. This could in turn allow the refinement of clinical correlates for specific subtypes of RP and also allow for a more accurate prognosis of the disease.

1.3.1 - arRP in the Sardinian Population

Sardinia is a recognised ethnic outlier among European populations, second only to the Lapps in genetic isolation (Cavalli-Sforza and Piazza, 1993). The island has been inhabited for at least 10,000 years, probably initially by a heterogeneous population of settlers, but isolation from the mainland, low immigration, the mountainous geography, and malaria have all contributed to genetic drift from the rest of Europe (Workman *et al*, 1975). Consanguinity has been comparatively common, even in the middle of this century, particularly in isolated mountain villages, increasing endogamy within these communities. Genetic disease is common among the Sardinian population, with high rates of insulin-dependent diabetes and haemoglobinopathies. In studies of genetic disease among the Sardinian population, such as β-thalassaemia, relative genetic homogeneity is observed, with one specific mutation (β^{39}) accounting for 95% of β-thalassaemic chromosomes (Rosatelli *et al*, 1992).

The overall prevalence of RP in south-central Sardinia is 1 in 3,355, with the relative proportions of each genetic type listed in **figure 1.3** (Fossarello *et al*, 1993). It is likely that the high proportion of recessive cases of RP is due to the extent of inbreeding and isolation within the small communities in this part of the island. Furthermore, religion and the rural economy have favoured large families, making the mode of inheritance easier to establish. However, this figure is still likely to be an underestimate of the true level of arRP, as many families may have only one affected child. Interestingly, in a study of adRP in Sardinia, a single rhodopsin mutation was found to account for RP in three of a total of eight families (Fossarello *et al*, 1993). A different rhodopsin defect accounted for RP in another of the families, so that this one gene alone was responsible for 50% of adRP, a figure much higher than in other populations, where rhodopsin mutations are thought to cause a quarter of

adRP cases (Dryja, 1992). This result agrees with previous reports of genetic homogeneity in the Sardinian population (Figus *et al*, 1995), and makes it a suitable candidate population in which to study genetic disease.

1.4 - RP SYNDROMES

Pigmentary retinopathy is a feature common to many human disorders, where other organ systems are also affected. These syndromes can be subdivided into four general classes:

i) lipid or mucopolysaccharide disorders

ii) syndromes involving the central nervous system

iii) syndromes with muscular involvement

iv) RP in association with deafness

However, the retinopathy observed in such cases is often not "typical RP", and may be secondary to metabolic disturbances produced by the disease, which can in some cases be treated (Deutmann, 1977). Also, the spectrum of retinal degeneration seen among patients may include true RP as well as other types of pigmentary retinopathy. The two most common RP syndromes, Usher syndrome and Bardet-Biedl syndrome, are discussed, with reference to genetic linkage studies carried out for each condition.

1.4.1 - Usher Syndrome

This is an example of an autosomal recessive syndrome, with classic RP associated with deafness. It is the most common RP syndrome with a prevalence of 6-10% among RP patients, and 3-6% among the deaf population (Heckenlively, 1988d). This highly heterogeneous disorder is usually seen in two equally prevalent forms, type I with profound hearing loss, no speech, vestibular ataxia and RP (usually diagnosed pre-puberty), and type II with milder nonprogressive hearing loss and RP (usually diagnosed post-puberty) (Kimberling, Weston and Moller, 1994). The clinical distinctions between types I and II have been established by the identification of six separate loci, four for type I: *USH1A* at 14q32 (Kaplan *et al*, 1992), *USH1B* at 11q13.5 (Kimberling *et al*, 1992), *USH1C* at 11p15.1-p14 (Keats *et al*, 1994) and *USH1D* at 10q21-q23 (Wayne *et al*, 1996); and two for type II: *USH2A* at 1q41 (Kimberling *et al*, 1995) and *USH2B*, which has not yet been localised (Kimberling, Weston and Moller, 1994). A third form of Usher syndrome, type III, involving RP and progressive hearing loss, is estimated to account for 2% of Usher

patients. Linkage between this condition and 3q21-q25 has been established using Finnish pedigrees (Sankila *et al,* 1995).

The gene for Usher type IB on chromosome 11q was recently identified as the unconventional myosin, myosin VIIA, after null mutations were seen in this gene in patients (Weil *et al*, 1995). This protein is expressed predominantly in the inner cochlea, RPE, lung, kidney and testis (Hasson *et al*, 1995). Studies of the mutated myosin VII gene in mice resulted in congenital deafness, but no retinal degeneration (Gibson *et al*, 1995), and El-Amraoui *et al* (1996) subsequently showed that this protein is expressed only in the RPE, and not the photoreceptors, of adult mice, but is expressed in both cell types in man. This suggests that the RP seen in Usher type IB patients is due to the presence of defective myosin VIIA in human rods and cones.

1.4.2 - Bardet-Biedl Syndrome

Bardet-Biedl syndrome (BBS) is an autosomal recessive condition generally characterised by the presence of the five cardinal features of retinopathy, post-axial polydactyly (**figure 1.4**), hypogenitalism, obesity and mental retardation. The severe tapetoretinal degeneration seen is often described as "atypical" retinitis pigmentosa, with early macular involvement, later pigmentation, and commonly resulting in blindness in the second or third decade (Riise, 1987, Green *et al*, 1989). It has been suggested that renal anomalies should also be added to the diagnostic criteria, as they are a significant cause of death among BBS patients (Harnett *et al*, 1988, Green *et al*, 1989, Odea *et al*, 1996, Riise, 1996), and that syndactyly and brachydactyly are more commonly seen than polydactyly (Green *et al*, 1989). Green *et al* (1989) have further suggested that the apparent mental retardation is often simply due to visual impairment, and that hypogenitalism is much more pronounced in males. It is possible that many female cases of BBS are diagnosed as having non-syndromal RP, due to the lack of obvious hypogenitalism, and the high variability of the other cardinal features. Other symptoms occasionally associated with the disorder include diabetes mellitus, hearing impairments, dental anomalies, liver disease and hypertension (R. Riise, personal communication). However all these symptoms are highly variable, and clinical heterogeneity is seen both within and between families. This can lead to misdiagnosis of BBS with overlapping conditions such as Laurence-Moon syndrome, where spastic paraplegia is present, but obesity and dysmorphic extremities are absent, or other clinically similar disorders such as Biemond syndrome type II, Alstrom syndrome or Weiss syndrome (Schachat and Maumenee, 1982). Obviously, identification of the genetic defect would greatly aid classification in uncertain cases.

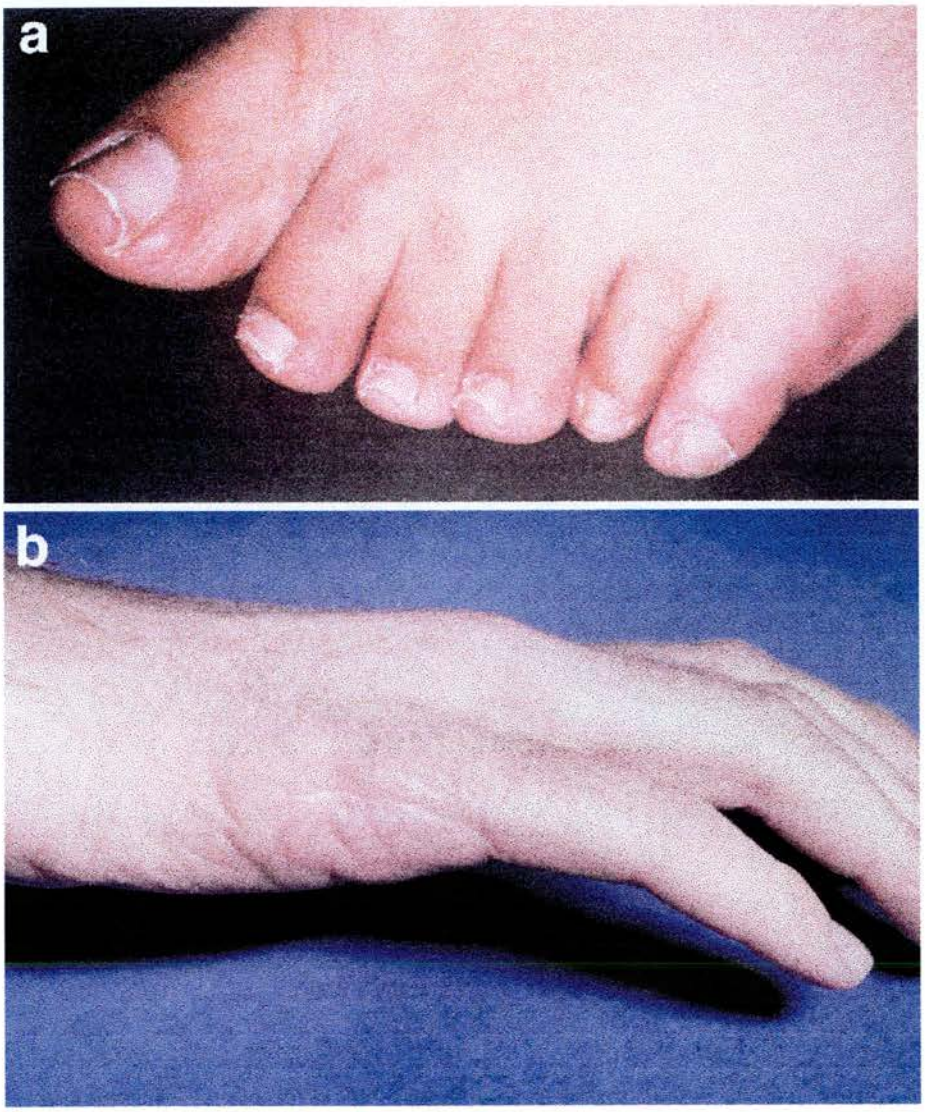


Figure 1.4

Polydactyly in Bardet-Biedl Syndrome patients.

a). Extra post-axial digit on foot.

b) Scar left following surgical removal of extra post-axial digit on hand.

The incidence of BBS differs greatly between populations. Particularly high prevalences are seen among the Bedouin tribes throughout the Middle East (1 in 13,500) (Farag and Teebi, 1989) and in Newfoundland (1 in 17,500) (Green *et al*, 1989) especially when compared to a study in Switzerland which found the incidence to be 1 in 160,000 (Amann, 1970). As BBS is known to be inherited as an autosomal recessive condition, this is probably due to increased levels of consanguinity in these populations (and possible under-diagnosis in other populations).

Linkage has been reported to four distinct BBS loci. The first gene to be localised (*BBS2*) was mapped, using a large inbred pedigree from the Negev region of Israel, to chromosome 16q21 (Kwitek-Black *et al*, 1993). However this locus was excluded in another consanguineous family from a different Bedouin tribe. The second study involved thirty-one multiplex North American families, and found linkage to 11q13 (*BBS1*) in seventeen of these, but no evidence of linkage to chromosome 16 in the remaining unlinked families, indicative of yet another BBS locus (Leppert *et al*, 1994). The presence of this *BBS1* locus has since been confirmed by homozygosity mapping with two large inbred pedigrees from Puerto Rico (Cornier *et al*, 1995). Sheffield *et al* (1994) then mapped the second of the Bedouin pedigrees to a ten centiMorgan (cM) interval on chromosome 3p13-p12 (*BBS3*), and the same group (Carmi *et al*, 1995a) then used a DNA pooling strategy to identify a fourth BBS locus (*BBS4*) at 15q22.3-q23 in a third Bedouin family. This high level of genetic heterogeneity is unexpected for such a comparatively rare disorder, and as yet none of the actual genes have been identified. Carmi *et al* (1995b) have suggested that the genetic heterogeneity in BBS is at least partially responsible for the clinical variation observed. However, their study was carried out using a small number of patients, from three inbred Bedouin kindreds (each linked to *BBS2, BBS3 or BBS4*), and the suggested correlations between phenotype and genotype can only be regarded as observations until a larger population has been studied, and any findings confirmed.

1.5 - MAPPING AND ISOLATION OF RP GENES

1.5.1 - Methods

Two approaches have commonly been used to identify and isolate causative genes for specific diseases. The first of these is the positional cloning ("reverse genetics") approach, which initially uses linkage analysis to localise the disease gene to a small region of the genome. This region can then be examined in more detail by physical mapping, using techniques such as yeast and bacterial artificial chromosome contig mapping and pulsed

field gel electrophoresis. The linkage work is ideally carried out in large pedigrees, as combining data from several small families can be complicated by genetic heterogeneity. Known cytogenetic rearrangements associated with the disease can also help narrow down the region of search, as can deletions, or expanded triplet repeats (Bates and Lehrach, 1994). This approach is guaranteed to be successful providing that sufficient samples are available, and that markers at a suitably high genetic density are studied. Over forty disease genes have been cloned using this technique (Collins, 1995), including Duchenne muscular dystrophy (Monaco *et al,* 1986), cystic fibrosis (Rommens *et al*, 1989, Riordan *et al*, 1989, Kerem *et al*, 1989) and retinoblastoma (Friend *et al*, 1986). However, this method is time-consuming and expensive.

The second method is the candidate gene approach where specific genes of known function are screened for mutations in patients with the disease. As each patient's DNA is examined individually for mutations, genetic heterogeneity is not a complication. Also, if a mutation is found, then deductions can be made concerning the physiology of the disease, as the function of the gene product is already known. However, a negative result could be due to limitations of the mutation screening technique used, or perhaps the sample size was not large enough to include any individuals with defects in that specific gene.

It is becoming increasingly common to use a combination of linkage analysis and candidate gene screening to identify causative genes - the "positional candidate" approach (Collins, 1995). The initial part of the study involves the use of markers to narrow down the region of interest within the human genome, and any known candidate genes from this region are then examined for mutations, or further linkage with the disease. This cuts out the need for large scale physical mapping projects, where linkage may only have narrowed down the region of interest to several megabases of DNA. This type of approach has been used to identify over twenty genes (Collins, 1995), including the RP genes rhodopsin (Dryja *et al*, 1990) and peripherin/RDS (Farrar *et al*, 1991)(see **fig. 1.5**).

1.5.2 - Rhodopsin

The first report of linkage in adRP was from McWilliam *et al* (1989), who were working with a large early-onset and severely affected Irish pedigree. They achieved a two-point lod score of 14.7 between the marker D3S47 and RP (at zero recombination), and this locus on the long arm of chromosome 3 was designated *RP4*. In the region of D3S47 there were four potential candidate genes for RP including two coding for proteins involved in visual transduction, the alpha subunit of transducin (Blatt *et al*, 1988), and rhodopsin (Nathans and Hogness, 1984, Sparkes *et al*, 1986). Dryja *et al* (1990) decided to sequence one of

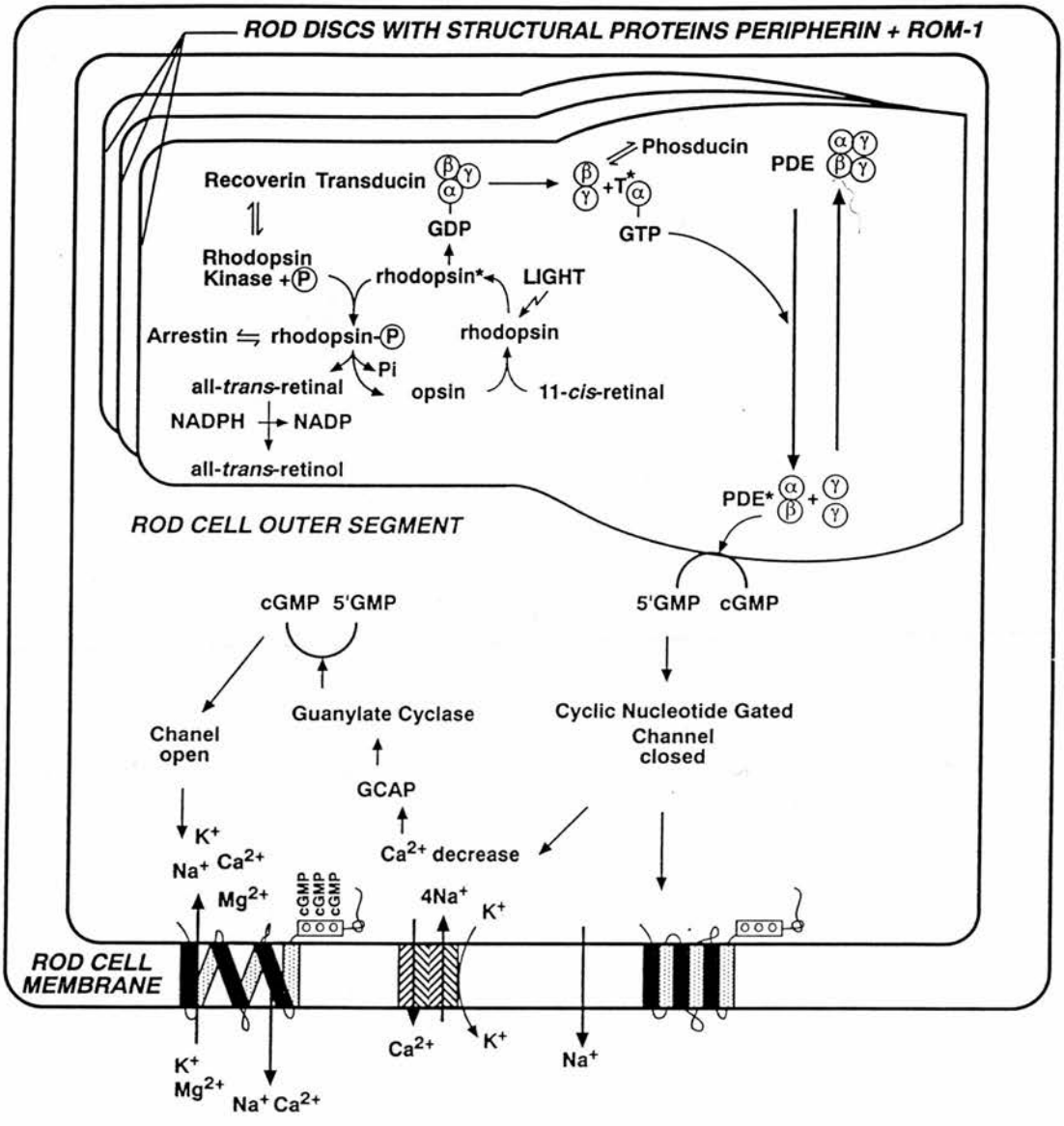


Figure 1.5

Key steps in visual transduction in the rod photoreceptor outer segment (after Wright, 1992 and Humphries, Kenna and Farrar, 1992)

these candidates, rhodopsin, in 20 North American affecteds, and found 5 patients to be heterozygous for a point mutation changing a highly conserved proline to a histidine at residue 23. They then extended their panel of patients to 148, and found this mutation in a total of 17, but in none of their 102 unaffected controls.

Following this original finding, almost a hundred mutations have been identified in the rhodopsin gene (Dryja and Li, 1995) among patients in Europe, North America and Japan, accounting for around a quarter of all cases of adRP. Nearly all cause dominant RP, and are mostly missense mutations, but three recessive alleles have also been identified, two of which are "null" mutations, coding for a non-functional protein (Rosenfeld *et al*, 1992, Kumaramanickavel *et al*, 1994). Interestingly, two missense mutations in the rhodopsin gene have also been identified in patients with autosomal dominant congenital stationary night blindness, a benign disorder diagnosed by night blindness, reduced visual acuity and characteristic ERG changes (Dryja *et al*, 1993, Sieving *et al*, 1995). Recently, Souied *et al* (1996) studied a family with a rhodopsin mutation where three affected members had retinitis punctata albescens (RPA), a form of RP with white dots deep in the retina, and one member had "typical" RP. RPA has previously been reported in a family with a peripherin/*RDS* mutation (Kajiwara *et al,* 1993). Souied *et al* (1996) suggested that the RPA phenotype may be induced by another genetic, or environmental, factor such as specific alleles of the apolipoprotein E gene (apoE), as the lipoprotein coded for by this gene may be involved in the formation of the white deposits seen in RPA. Huq *et al* (1993) previously reported an association between certain haplotypes of this gene (E2 and E4) and the RP phenotype, and the E4 allele of the apoE gene was seen to cosegregate with RPA in this family, suggesting a possible role for this gene in the RPA phenotype.

Rhodopsin is a major component of rod outer segments, and an intrinsic protein of the disc membrane. It is composed of two covalently linked molecules, the apoprotein opsin (coded for by the rhodopsin gene on 3q) and 11-*cis* retinal, a vitamin A derivative. Light causes the 11-*cis* retinal to isomerise to all-*trans* retinal and this photo-excited rhodopsin molecule activates the G-protein transducin, which then triggers the phototransduction pathway (Nathans, 1992). It appears that *in vitro* many of the adRP mutations cause the rhodopsin protein to accumulate in the rough endoplasmic reticulum (ER), as opposed to being transported to the cytoplasmic membrane, or alter the opsin molecule so that it is unable to bind 11-*cis* retinal (Sung *et al*, 1994). Studies on dominant rhodopsin mutations in *Drosophila* have suggested that the mutant protein interferes with the wild-type molecules, perhaps by forming complexes, and prevents them being processed by the ER (Colley *et al*, 1995). Work with transgenic mice has led Sung *et al* (1994) to suggest that the carboxy-terminal region of rhodopsin is required for correct transport of the protein, and that alteration or loss of this region leads to accumulation of the defective protein in the ER,

eventually causing cell death. However, unlike the situation in *Drosophila*, the transport of the wild-type protein is not impaired, suggesting that mammals have a different mechanism for transporting the normal newly synthesised opsin molecules.

1.5.3 - Rod cGMP Phosphodiesterase α and β Subunits

After the activation of rhodopsin in visual transduction, the G-protein transducin is stimulated, releasing its GTP-bound alpha subunit, which in turn stimulates the multimeric protein cGMP phosphodiesterase (PDE). This protein consists of three types of subunits, α, β and γ. The holoenzyme contains one of each of the large catalytic α and β subunits, which are very similar in structure, and both of these bind a small inhibitory γ subunit (Lem *et al*, 1992). On stimulation of the protein in response to light, the two γ subunits dissociate, the active αβ dimer hydrolyses cyclic GMP, hence reducing the intracellular level of cGMP, in turn causing the cGMP-gated cation channels to close (**Fig. 1.5**) (for review see Palczewski, 1994).

After the discovery that the mouse *rd* gene, which causes a severe retinal degeneration, codes for the β subunit of PDE (*PDEB*) (Lem *et al*, 1992) this became a strong candidate gene for human RP. The gene, which comprises 22 exons covering around 43 kb (Weber *et al*, 1991), was mapped to the telomeric region of chromosome 4 (4p16), close to the Huntington disease gene locus (Altherr *et al*, 1992). Riess *et al* (1992a) attempted to identify mutations in the *PDEB* gene and its 5' flanking region, using single-stranded conformation polymorphism (SSCP) analysis in RP families but only identified polymorphic variants (Riess *et al*, 1992b).

McLaughlin *et al* (1993) then used the same technique in a partial screen (7 exons) of the gene in 99 unrelated arRP patients, and identified four mutations segregating with the disease, two of which were nonsense mutations. In the same year, Suber *et al* (1993) also identified a causative nonsense mutation in the *PDEB* gene in Irish setter dogs with a recessive rod/cone dysplasia. The next mutation identified in the *PDEB* gene was found to cause autosomal dominant congenital stationary night blindness (CSNB) (Gal *et al*, 1994), as seen with some rhodopsin mutations (Dryja *et al*, 1993). It was suggested that the missense mutation (His258Asp), occurring in a highly conserved residue, in this form of CSNB may inhibit the binding of the γ subunit to the β subunit, such that about half of the PDE enzyme is permanently in a partially active state. In a further screen of 92 arRP patients, McLaughlin *et al* (1995) identified a total of seven mutations, and this has led to the suggestion that mutations in *PDEB* account for approximately 5% of arRP mutations in the North American population. Bayes *et al* (1995a) have also identified a tandem

duplication of 71 base-pairs in exon 1 of the *PDEB* gene in a Spanish consanguineous arRP family. As mutations in this gene only accounted for 1 out of 19 pedigrees studied, this suggests that a similar proportion of arRP could be due to *PDEB* mutations in European populations.

Because of the sequence and functional similarities between the β subunit and the α subunit Huang *et al* (1995) have also examined the α subunit gene, on chromosome 5q31.2-34, using SSCP analysis. They used DNA from 340 unrelated RP patients (173 arRP patients) and screened 19 of the 22 exons. A total of three mutations were identified in two recessive RP patients, two of which were nonsense mutations, and the other of which altered a highly conserved amino acid.

1.5.4 - Rod cGMP-Gated Cation Channel α Subunit

The hydrolysis of cyclic GMP to 5′GMP by active phosphodiesterase causes the closure of cGMP-gated cation channels that are located in the plasma membrane of the rod outer segment. This results in hyperpolarisation of the rod photoreceptor cell, inhibiting the release of neurotransmitter at the cell synapse (for review see Yau, 1994). The rod cGMP-gated cation channel protein is made up of two similar subunits, α and β, although a γ subunit of unknown function has also been identified (Chen *et al*, 1994). The α subunit gene has been mapped to human chromosome 4p14-q13, and the sequence of its ten exons determined (Pittler *et al*, 1992). The SSCP technique was used to screen the entire coding sequence and exon-intron boundaries of the α subunit cGMP-gated channel (*CNCG1*) in 173 autosomal recessive and 94 dominant RP patients (Dryja *et al*, 1995). In addition to polymorphic changes, they identified five mutations in four arRP pedigrees. Three of these mutations were null alleles (two nonsense mutations and a large deletion), while the other two were a missense mutation (Ser316Phe) and a frameshift that should truncate the protein by 32 amino acids.

In one of the arRP families in which a paternally-derived mutation was identified, Dryja *et al* (1995) were unable to find the other mutated allele of the *CNCG1* gene, and indeed saw from polymorphic analysis that both two affected sibs inherited different maternal alleles. They suggest that the other recessive mutation could be in another gene coding for a protein that interacts with the product of *CNCG1* in some way. This type of digenic inheritance has previously been proposed for RP with mutations in the peripherin/*RDS* gene and the *ROM-1* gene being found in three families (Kajiwara, Berson and Dryja, 1994). Another possibility is that the heterozygous nonsense mutation in the *CNCG1* gene in this family does not cause RP. Interestingly, one of the panel of patients with adRP also carried

the Ser316Phe missense mutation in heterozygous form. This individual belonged to a family where a known rhodopsin mutation had previously been found, and there appeared to be no clear phenotypic difference between individuals in the family who only had the rhodopsin mutation or both the rhodopsin and the *CNCG1* mutation. Also, the unaffected mother carried only the *CNCG1* mutation. It seems likely that the rhodopsin defect is the cause of RP in this family and that the *CNCG1* mutation is a variant with no obvious phenotypic effect (Dryja *et al*, 1995).

The *CNCG1* gene product has previously been expressed *in vitro* in human embryonic kidney (HEK) cells (Dhallan *et al*, 1992) to measure its channel activity, and Dryja *et al* (1995) decided to express the missense Ser316Phe allele and the Arg654(1-bp del) allele in cultured HEK cells. They detected very low levels of mutant channels in the plasma membrane, although any found were functional, and found that most of the mutant protein failed to be incorporated into the plasma membrane. Thus, it seems that a lack of cGMP-gated channels correlates with the photoreceptor degeneration for these two recessive mutations. In both of these mutations, and in the null alleles, the cells would be expected to behave as if exposed to constant light, as there would be few or no channels open, as happens when cGMP levels fall due to hydrolysis by activated PDE. However, the cause of the rod photoreceptor degeneration remains unknown.

1.5.5 - Peripherin/*RDS*

The peripherin/*RDS* gene encodes a photoreceptor specific glycoprotein found in the outer membrane of rods and cones in the rim region of outer segment discs (Connell and Molday, 1990). The human gene was first identified by homology to the mouse *rds* (retinal degeneration slow) gene, which, when homozygously mutated, causes abnormal development and then a slow degeneration of the photoreceptors. The gene maps to human chromosome 6p12, consists of 3 exons, and encodes a 346 amino acid protein with over 90% homology to the predicted mouse protein (Travis *et al*, 1991). Comparisons of these sequences with bovine and rat mRNAs have shown 85% identity overall, with the conserved features of four putative membrane spanning regions, a potential glycosylation site, and 12 cysteine residues (Travis *et al*, 1991). The native protein has a molecular weight of 39 kD, and forms a disulphide-linked homodimer in the disc membrane, non-covalently bonding with a disulphide-linked rom-1 dimer (Goldberg, Moritz and Molday, 1995). It has been suggested that this complex has a role in anchoring the discs to the cytoskeleton of the photoreceptor cell, through homotropic or heterotropic protein interactions (Moritz and Molday, 1996).

Linkage to 6p12 was reported for a large Irish kindred by Farrar *et al* (1991) who went on to identify a three basepair deletion of a highly conserved cysteine residue in the peripherin/*RDS* gene using SSCP. Affected members of this family showed late-onset adRP, although cone and rod ERG responses were greatly reduced in young affected members. These features are similar to the phenotype seen in heterozygous *rds* mice, where outer segments are present but disorganised and shortened, and degeneration of photoreceptors is much slower than in homozygotes (Hawkins, Jansen and Sanyal, 1985). A further three mutations were identified by Kajiwara *et al* (1991) in another three adRP families, after studying 139 unrelated adRP patients. Each mutation affected a highly conserved residue, either removing or substituting a proline. However, in 1994 Kajiwara, Berson and Dryja reported that one of these mutations affecting a leucine at position 185 (changing it to a proline) was also present in another two families, and in asymptomatic carriers of RP. Each of these three families transmitted the disease to less than 50% of their offspring. The probands were born to unaffected parents, and the mutation was present in unaffected members. This led Kajiwara *et al* to the conclusion that an additional gene defect at another locus could be responsible (as opposed to reduced penetrance of the peripherin/*RDS* gene), and discovery of mutations in the *ROM1* gene (discussed below) in these families seemed to confirm this.

While most mutations in the peripherin/*RDS* gene have been seen to cause dominant RP, other dominant retinal dystrophies have also been associated. These include RP with bull's eye maculopathy (Kikawa *et al,* 1994), macular dystrophy (Wells *et al*, 1993), adult vitelliform macular dystrophy (Wells *et al*, 1993), central areolar choroidal dystrophy (Reig *et al*, 1995), pattern macular dystrophy (Nichols *et al*, 1993, Weleber *et al*, 1993), fundus flavimaculatus (Weleber *et al*, 1993) and retinitis punctata albescens (Kajiwara *et al*, 1993). Different phenotypes, including macular and peripheral dystrophies, have even been reported in a single family with a three base-pair deletion (Weleber *et al*, 1993). This suggests that other genetic and/or environmental factors could influence the clinical course of the disease. One possibility is the involvement of proteins that interact with peripherin in the rods and cones of the retina. This is exemplified by the proposed digenic inheritance involving peripherin/*RDS* and *ROM-1*, as it is has been shown that peripherin non-covalently interacts with rom-1, a photoreceptor-specific protein (Bascom *et al*, 1992a, Goldberg, Moritz and Molday, 1995). It further suggests that other genes implicated in RP could also be responsible for macular disorders.

1.5.6 - ROM-1

The *ROM-1* gene was first identified by differential hybridisation to select human retina-specific cDNAs (Bascom *et al*, 1992a). The isolated cDNA encoded a 351 amino-acid polypeptide, which was found to be 35% identical to human peripherin (55% identity at nucleotide level) with four regions of extended homology. Affinity-purified antibodies originally localised the gene product to the outer segments of rod photoreceptors, and this led to the designation "retinal outer segment membrane protein", rom-1 (Bascom *et al*, 1992a). Recently, bovine rom-1 (or a closely related homologue) was also identified in cone outer segment disc membranes (Moritz and Molday, 1996). Because of the similarity to peripherin, Bascom *et al* (1992a) also studied rom-1 and peripherin with a view to identifying an association between the two proteins. Their results led them to suggest that disulphide-linked dimers of both rom-1 and peripherin bind non-covalently to each other, and that other proteins may associate with this complex. Further, they have hypothesised a major role for this complex in the morphogenesis of the disc rim.

Bascom *et al* (1992b) then mapped the rom-1 gene to 11q13, a region to which four retinopathies have been linked (Bascom *et al*, 1993) including Usher syndrome type IB (Kimberling, Weston and Moller, 1994) and Best's disease (BMD), a juvenile vitelliform macular degeneration (Stone *et al*, 1992). Because of the mutations already identified in the peripherin gene (Farrar *et al*, 1991, Kajiwara *et al*, 1991), the similarities between the two proteins, and the disorders linked to the region of the gene, Bascom *et al* (1993) screened the gene for variants in 180 patients with inherited retinal degenerations. Six polymorphisms and three rare variants were identified, but none were associated with the disease.

Nichols *et al* (1994) also screened the rom-1 gene for mutations in 11 unrelated Best's disease patients, but found no alterations in the coding sequence, and Graff *et al* (1994) effectively excluded the gene in a multigeneration Swedish BMD family linked to 11q13, because of a recombinational event between an intragenic *ROM-1* polymorphism and the disease phenotype.

ROM-1 mutations have however been identified in three families with RP in association with peripherin mutations (Kajiwara, Berson and Dryja, 1994). This digenic inheritance pattern required that both a mutation in *ROM-1* and in the peripherin gene was present in affecteds, with either mutation alone not causing RP. Two distinct one base pair insertions were found in exon 1 of *ROM-1*, both resulting in a prematurely terminated protein ending at codon 131. As the protein is severely truncated, it is likely that these mutant alleles do not code for a functional gene product (i.e. are null alleles). Subsequently, Sakuma *et al*, (1995) found one of these *ROM-1* mutations in a simplex RP patient, but could not identify

any mutation in peripherin/*RDS*. On examination, the mutation was also present in other younger members of the family, only one of whom showed mild disease symptoms, suggesting phenotypic heterogeneity. Possible explanations for this are that another unidentified locus influences the phenotype, or that the mutation is not causing RP in this case (Sakuma *et al*, 1995). More recently, Bascom *et al* (1995) identified four potentially pathogenic *ROM-1* alleles among patients with adRP, without detectable peripherin defects, but the affected families were too small to establish co-segregation of the disease with the alleles.

1.5.7 - Myosin VIIA

Recessive mutations, predicted to code for non-functional proteins, have been identified in the gene coding for myosin VIIA in patients with Usher syndrome type I (USH1)(Weil *et al*, 1995). This is a disease which involves profound congenital deafness, vestibular dysfunction and RP. Linkage to USH1 has been localised to four chromosomal regions; 14q32 (Kaplan *et al*, 1992), 11p15 (Smith *et al*, 1992), 11q13.5 (Kimberling *et al*, 1992) and 10q21-q23 (Wayne *et al*, 1996). In humans, the region 11q13.5 (*USH1B*) is syntenic to part of mouse chromosome 7, where the mutation responsible for the *shaker-1* (*sh-1*) mouse is located. This mouse exhibits deafness and vestibular dysfunction, but no retinal degeneration. Gibson *et al* (1995) identified three mutations (two missense and one splice acceptor) in a gene encoding an unconventional myosin, type VII, in the *shaker-1* mouse. This led Weil *et al* (1995) to consider the human homologue of this gene as a candidate for USH1B, especially since some patients have microtubule defects in photoreceptor, nasal ciliar, and sperm cells. They identified the human myosin type VIIA gene from a yeast artificial chromosome covering the *USH1B* locus, and screened the gene for mutations in nineteen families that showed linkage to this region. Five mutations were found: two nonsense mutations, two missense mutations altering highly conserved residues, and one deletion of two amino acids, one of which is highly conserved. It was suggested that all of these mutations encode null alleles.

Myosin VIIA is an unconventional myosin, a family of proteins related to striated-muscle myosin in head sequence, but divergent in tail sequence. All myosins move along actin filaments by hydrolysing ATP, acting as motor molecules. Antibodies to the tail portion of myosin VIIA have been produced to study its expression pattern (Hasson *et al*, 1995) and have shown that it is expressed in the cochlear hair cells, RPE, testis, lung and kidney. The presence of the protein in the RPE and cochlea confirms its role in Usher syndrome, and abnormalities in sperm and bronchial function have been noted in some patients (Hasson *et*

al, 1995). No kidney abnormalities have been reported, but this is possibly due to other unconventional myosins having overlapping functions in kidney cells.

Unexpectedly, *sh-1* mice with defects in the homologous gene do not have a retinal degeneration. Possible explanations for this discrepancy included: the specific mouse mutations observed, since each of the mutants still expressed myosin VIIA, the timescale of the retinal degeneration, and that the corresponding human mutations could cause a nonsyndromic deafness, as the gene for this condition (*DNFB2*) also maps to 11q13 (Guilford *et al*, 1994). However, a study of the expression of myosin VIIA in both mouse and man revealed that the protein is expressed in the RPE only in mouse, but in the RPE and photoreceptors of humans (El-Amraoui *et al,* 1996).

1.5.8 - RPGR

The most recently identified RP gene is responsible for the commonest form of X-linked RP, previously designated *RP3*. *RPGR* (RP GTPase regulator) was identified by positional cloning, using information from microdeletions in RP3 patients, and has been shown to be mutated in seven affected individuals (Meindl *et al*, 1996). At present, the function of the protein is unknown, but it appears to be ubiquitously expressed. The N-terminal half of the predicted protein shows homology to RCC1 (regulator of chromatin condensation), which regulates a GTPase, while the C-terminal half is highly charged and contains a putative isoprenylation site for membrane anchorage. Meindl *et al* have suggested that the RPGR protein may interact with a small GTPase, anchoring it to a membrane or vesicle, and hence acting as a regulator of a specific type of membrane transport. This gene may also be responsible for potentially allelic retinal dystrophies mapping to Xp21.

1.5.9 - Chromosomal regions linked to RP

In addition to the genes discussed above, mutations have recently been identified in the microsomal triglyceride transfer protein gene in abetalipoproteinemia (Bassen-Kornzweig syndrome), a rare recessively inherited disorder, characterised by fat malabsorption, cerebellar ataxia and retinal degeneration (Narcisi *et al*, 1995). Mitochondrial mutations in the gene coding for subunit 6 of ATPase have also been observed in NARP syndrome (neuropathy, ataxia and retinitis pigmentosa) (Holt *et al*, 1990). Due to the high ATP-dependence of the retina, pigmentary retinopathy is often a symptom of mitochondrial DNA

mutations affecting the genes involved in oxidative phosphorylation (Brown, Lott and Wallace, 1994).

Apart from the ten genes previously mentioned, where mutations have been identified, many other chromosomal regions are linked to RP (see **figure 1.6**). However, this still leaves many pedigrees that do not show linkage to any of these reported loci, such as the postulated RP8 locus involved in a family of Irish origin (Kumar-Singh *et al*, 1993).

1.6 - CANDIDATE GENES FOR RETINITIS PIGMENTOSA

From the results of linkage studies it is apparent that many RP genes are still to be identified. After localisation, the next stage is to search for candidate genes in the region of interest. Many genes fall into this category, including all the genes coding for proteins involved in the phototransduction cascade (see **figure 1.5**), all retina-specific proteins, other unconventional myosins, genes implicated by vertebrate and invertebrate models (see **figure 1.7** for mouse models), and those involved in clinically related disorders. This covers numerous possible loci, many of which have been mapped in the human genome.

1.6.1 - Genes Involved in Phototransduction

Rhodopsin is a protein consisting of opsin (coded for by the rhodopsin gene) covalently bound to 11-*cis*-retinal, which isomerizes to all-*trans*-retinal following photoactivation. Retinol (vitamin A) and its derivatives are water-insoluble, and must be transported around the body by carrier proteins. Seven of these specific binding proteins have been characterised and mapped: serum retinol-binding protein (*RBP4*) to 10q23-q24, cellular retinol-binding protein (*RBP1*) to 3q21-q22 (Rocchi *et al*, 1989), cellular retinoic acid-binding protein (*RBP5*) to 15q22-qter (Guerts van Kessel *et al*, 1991), cellular retinoic acid-binding protein II (*CRABPII*) to 1q21 (Elder *et al*, 1992), cellular retinal-binding protein specific to visual tissue (*CRALBP*) to 15q26 (Sparkes *et al*, 1992), interphotoreceptor retinol-binding protein (*RBP3*) to chromosome 10q11.2 (Liou *et al*, 1987, GDB), and cellular retinol-binding protein type II (*RBP2*) to the short arm of chromosome 3, close to *RBP1* (Demmer *et al*, 1987). Deficiency of vitamin A is known to result in clinical features similar to RP, as seen in abetalipoproteinemia, a defect in fat absorption, due to mutations in the microsomal triglyceride transfer protein (MTP) (Narcisi *et al*, 1995). Ringens *et al* (1990) and Cotran *et al* (1990) investigated RBP3 and CRALBP in patients

Gene Locus	Chromosomal Location	Phenotype	Reference
RP1	8p11-q21	adRP	Blanton *et al*, 1991
RP2	Xp11.3-p11.2	xlRP	Coleman *et al*, 1990
RPGR (RP3)	Xp21.1	xlRP	Meindl *et al*, 1996
Rhodopsin *(RP4)*	3q21-q24	adRP arRP CSNB	Dryja *et al*, 1990 Rosenfeld *et al*, 1992 Dryja *et al*, 1993
Peripherin/*RDS* *(RP7)*	6p12	adRP Digenic RP Macular Dystrophies	Farrar *et al*, 1991 Kajiwara *et al*, 1994 Dryja and Li, 1995
RP9	7p15.1-p13	adRP	Inglehearn *et al*, 1994
RP10	7q31-q35	adRP	McGuire *et al*, 1995a
RP11	19q13.4	adRP	Al-Maghtheh *et al*, 1994
RP12	1q31-q32.1	PPRPE	van Soest *et al*, 1994
RP13	17p13.1	adRP	Kojis *et al*, 1996
RP14	6p21.3	arRP	Shugart *et al*, 1995
RP15	Xp22.13-p22.11	X-Linked Cone-Rod Dystrophy	McGuire *et al*, 1995b
RP17	17q22-q23	adRP	Bardien *et al*, 1995
RP18	1p13-q23	adRP	Xu *et al*, 1996
PDEB	4p16.3	arRP CSNB	McLaughlin *et al*, 1995 Gal *et al*, 1994
PDEA	5q31.2-q34	arRP	Huang *et al*, 1995
CNCG1	4p14-q13	arRP	Dryja *et al*, 1995
ROM-1	11q13	Digenic RP, (adRP)	Kajiwara *et al*, 1994 Sakuma *et al*, 1995 Bascom *et al*, 1995
USH1A	14q32	USH1	Kaplan *et al*, 1992
Myosin VIIA	11q13.5	USH1	Weil *et al*, 1995
USH1C	11p15.2-p14	USH1	Keats *et al*, 1994
USH1D	10q21-q23	USH1	Wayne *et al*, 1996
USH2A	1q41	USH2	Kimberling *et al*, 1995
USH3	3q21-q25	USH3	Sankila *et al*, 1995
BBS1	11q13	BBS	Leppert *et al*, 1994
BBS2	16q21	BBS	Kwitek-Black *et al*, 1993
BBS3	3p13-p12	BBS	Sheffield *et al*, 1994
BBS4	15q22.3-q23	BBS	Carmi *et al*, 1995a

Figure 1.6

Loci Mapped for RP, Usher Syndrome and Bardet-Biedl Syndrome

Mouse Mutant	Chromosome/ Gene	Phenotype
retinal degeneration *rd-1*	*PDEB* (chromosome 5)	Onset 7-8 days, complete photoreceptor loss by day 20, more widespread changes later.
retinal degeneration *rd-3*	chromosome 1	Normal photoreceptor development, degeneration begins by week three, slower cone degeneration.
retinal degeneration *rd-4*	chromosome 4	Homozygous lethal, reduced heterozygote ERG by three weeks.
retinal degeneration slow *rds*	*RDS*/peripherin (chromosome 17)	Failure of normal rod outer segment formation, slow degeneration of photoreceptors over 7-12 months.
purkinje cell degeneration *pcd*	chromosome 13	Ataxia, male sterility, rapid degeneration of cerebellar Purkinje cells, slower photoreceptor degeneration.
cribiform degeneration *cri*	chromosome 4	Ataxia, anaemia, normally lethal by three months, severe degeneration of brain stem, spinal cord and inner nuclear layer of retina.
nervous *nr*	chromosome 8	Ataxia, hyperactivity, loss of cerebellar Purkinje cells weeks 3-7, earlier loss of photoreceptor outer segments.
vitiligo mi^{vit}	chromosome 6	Progressive depigmentation due to melanocyte abnormality, slow degeneration of photoreceptors, abnormal RPE cells.
retinal degeneration *rd-5* (Heckenlively *et al*, 1995)	*tub* (Noben-Trauth *et al*, 1996) (chromosome 7)	Failure of normal photoreceptor outer segment formation, progressive degeneration with extinguished ERG at 6 months. Degeneration and loss of cells of organ of Corti, virtually deaf by 5-6 months.
motor-neuron degeneration *mnd* (Messer *et al*, 1993)	chromosome 8	Limb dysfunction onset 6 months. Paralysis by 9 months. Persistence and progressive shortening of photoreceptor outer segments from day 15, most photoreceptors lost by 9 months.

Figure 1.7

Summary of Mouse Mutants with Retinal Degeneration (after Wright, 1994)

with autosomal dominant, recessive, and simplex RP, and Ushers type I, using RFLPs at each locus, and Southern blotting. No gene rearrangements or deletions were found, and no linkage disequilibrium was observed within the sets of patients, or in the pedigrees studied, suggesting these genes are not a major cause of RP.

Transducin, a guanine nucleotide-binding protein, is a heterotrimer composed of an α, β and γ subunit. The α subunit exchanges GDP for GTP following interaction with activated rhodopsin, and then dissociates from the β and γ subunits to activate cGMP phosphodiesterase. This step amplifies the visual signal, as one activated rhodopsin molecule can in turn activate up to 500 transducin molecules. The α subunit varies between rods and cones, with the rod specific gene (*GNAT1)* located on chromosome 3p21 (OMIM, 1996). Recently, Dryja *et al* (1996) identified a missense mutation in a highly conserved residue (gly38asp) in the *GNAT1* gene, in affected individuals with the Nougaret form of congenital stationary night blindness. It is possible that, as with rhodopsin and *PDEB*, this gene could cause both RP and CSNB. The cone specific α subunit (*GNAT2*) has been mapped to chromosomal region 1p13 (Morris and Fong, 1993), and is a candidate for *RP18*, although Ringens *et al* (1990) could not identify any gene rearrangements, or suggestion of linkage to this gene, in a set of 526 patients with autosomal dominant, recessive or simplex RP. The α, β and γ subunits are cell specific in the rods and cones (Peng *et al*, 1992), but rod transducin is also found in vertebrate taste cells (Ruiz-Avila *et al*, 1995).

The γ subunit of cGMP PDE is a candidate for RP as causative mutations have been identified in the α and β subunits. The gene, located on chromosome 17q21.1, has been studied by Hahn, Berson and Dryja (1994) in 704 patients with RP and related hereditary retinal diseases. None of the sequence variants correlated with any phenotype, so it seems unlikely that this gene is a common cause of RP, although it is still a candidate for the *RP17* locus which maps to 17q22-q24 (Bardien *et al*, 1995). Studies of transgenic mice homozygous for a null mutation in the *PDEG* gene have shown that they display biochemical changes similar to those seen in mice homozygous for mutations in the *PDEB* gene, suggesting that mutations in the *PDEG* gene could result in an RP phenotype in humans (Tsang *et al*, 1996). Also, this work has shown that the gamma subunit is required for normal function and stability of cGMP phosphodiesterase.

The cyclic nucleotide gated channel is composed of a β subunit, in addition to the α subunit. The β subunit is coded for by a complex locus on chromosome 16q21, previously thought to consist of two separate loci coding for a β subunit, and a glutamic-acid rich γ subunit, coded for by *GAR1* (Ardell *et al*, 1995). Recent studies have shown that the *GAR1* locus in fact encodes the N-terminal portion of the β subunit. This portion, and the C-terminal two-thirds of the protein can be transcribed separately, or as a polycistronic message, separated from each other by six exons which are highly evolutionarily divergent

(Ardell *et al*, 1996). Highly homologous channels are also found in kidney cells, and Ardell *et al* (1995) have suggested that mutations in these genes could cause kidney defects as well as retinopathy, making it an ideal candidate gene for the *BBS2* locus.

1.6.2 - Genes Involved In Recovery Phase

All the genes involved in re-opening of the cGMP channels and dark adaptation following phototransduction activation are potential retinopathy genes. The removal of cations from the cell during phototransduction results in activation of recoverin (cancer-associated retinopathy protein), which is thought to bind to rhodopsin kinase, and hence inhibit rhodopsin phosphorylation (Chen *et al*, 1995). The three exons of the recoverin gene on chromosome 17p13 were screened by SSCP for mutations in 42 Spanish arRP families, but none were detected (Bayes *et al*, 1995b). This gene also seemed to be a good candidate for RP13, which maps to 17p13, but analysis in the large South African family linked to this locus failed to find any mutations (Greenberg *et al*, 1994). However, a particulate (membrane) human photoreceptor guanylate cyclase (retGC), which catalyses the formation of cGMP, has also been assigned to chromosome 17p13 (Oliviera *et al*, 1994), so that this gene (*GUC2D*) is worth investigating. Indeed, Perrault *et al* (1996) have recently identified mutations in this gene in patients with Leber's congenital amaurosis, the earliest and most severe form of inherited retinopathy. Guanylate cyclase activating proteins stimulate synthesis of cGMP in photoreceptors by activating retGC. Two genes coding for GCAPs (GCAP1 and GCAP2), are located on chromosome 6p21.1, within two centiMorgans of the peripherin/*RDS* gene (Subbaraya *et al*, 1994). They have identical gene structures and chromosomal localisations, and 50% sequence similarity, suggesting they arose relatively recently from gene duplication (Surguchev *et al*, 1996). GCAP1 is specific to rod outer segments (ROS), and GCAP2 is found in the retina, but not the ROS (Gorczyca *et al*, 1995) - both are good candidates for RP. A gene for guanylate kinase (*GUK1*), the protein that catalyses the conversion of GMP to GTP, has been localised to 1q32-q41, making it a candidate for RP12, and Usher syndrome type II (Fitzgibbon *et al*, 1996).

Rhodopsin kinase (RK) helps inactivate rhodopsin by phosphorylation of the carboxy terminal region of the protein. The *RK* gene has been assigned to chromosomal band 13q34 (Khani *et al,* 1996), the same region as Stargardt's dominant macular dystrophy has been mapped to, suggesting *RK* as a possible candidate for this disease. Following this phosphorylation, arrestin (S-antigen) can then bind to rhodopsin to block transducin activation. Mutations in the arrestin gene on chromosomal region 2q37 have been identified in patients with Oguchi disease, a rare form of CSNB, with very slow dark

adaptation and fundus discolouration (Maw *et al*, 1995). Fuchs *et al* (1995) found a homozygous one base-pair deletion in arrestin in five unrelated Japanese patients, which was thought to result in a nonfunctional protein.

1.6.3 - Retina-specific Proteins

While many retina-specific proteins are known to exist, a large number have not yet been characterised, or assigned a chromosomal location. A few that have been studied in more detail, and seem to be good candidates for retinal degenerations, are mentioned here.

Phosducin is a protein present in the photoreceptors and the embryologically related pineal gland. It is thought to have a role in phototransduction, by binding to the γ and β subunits of transducin following activation of the α subunit by rhodopsin. The human phosducin gene has been mapped to chromosome 1q25-q32.1, and is highly homologous to the mouse gene (Abe, Kikuchi and Shinohara, 1994). Two forms of RP have been linked to this region, Usher type II and RP12, although the phosducin gene has recently been excluded in RP12 (van Soest *et al*, 1996).

A 61 kiloDalton protein which is abundant and specific to the retinal pigment epithelium has been identified, although its function is unknown (Nicoletti *et al*, 1995). It is conserved in vertebrates, has been mapped to chromosome 1p31 and appears to be developmentally regulated. Therefore this gene (*RPE65*) is a candidate for disease involving the RPE.

1.6.3.1 - *NRL*

Using subtractive hybridisation, Swaroop *et al* (1991) isolated several retina-specific cDNA clones. One of these has been shown to code for the neural retina leucine zipper gene (*NRL*) (Swaroop *et al*, 1992), encoding an evolutionarily-conserved protein which is expressed in all neural layers of the retina with strong expression in photoreceptor cells. The gene has been localised to chromosomal region 14q11.1-q11.2 (Yang-Feng and Swaroop, 1992) and consists of three exons, the first of which is untranslated (Farjo *et al*, 1993). It contains a basic motif and a leucine zipper domain, both of which show strong homology to the v-*maf* oncogene product. This has led to Nrl being designated a member of the bZIP family of proteins, which includes the fos and jun proto-oncogene transcription factors. Members of this family can form hetero- or homo-dimers, and bind to AP-1 like elements, which are thought to have a role in transcriptional regulation (Kerppola and Curran, 1994a). The Nrl homo- and hetero-dimers can bind to the consensus sequence

$TGCN_{6-8}GCA$, and such a sequence is seen in the upstream promoter region of the rhodopsin gene. Rehemtulla *et al* (1996) have shown that Nrl homodimers can mediate transactivation of the rhodopsin promoter, and that a naturally occuring truncated form of the protein can inhibit this transactivation in a "dominant negative" manner.

Studies of the expression of murine Nrl in the developing nervous system of mice have shown that it is expressed exclusively in post-mitotic central nervous system neurons in 12.5 day embryos, and is down-regulated postnatally (Liu *et al*, 1996). High levels persist in the photoreceptors and other retinal neurons. This suggests that Nrl plays a role in maturation and/or establishment of the differentiated phenotype in neuronal cells, and may maintain the differentiated state in the adult retina by regulating cell-specific gene expression. It seems likely that Nrl has further target genes, in addition to rhodopsin, and is a strong candidate for human retinopathies.

1.7 - METHODS OF MUTATION DETECTION

Once a candidate gene for a retinopathy has been identified, the next step is to try and find mutations in that gene in patients. Methods of mutation detection in genes in which no previous defects have been reported involve scanning hundreds of basepairs of DNA, whilst trying to avoid the laborious process of sequencing the entire gene. This is particularily important if the gene is large and contains many exons, or if there are many patients to be screened. Nearly all commonly used techniques involve an initial polymerase chain reaction (PCR) step to amplify exponentially a specific segment of the target DNA. This step can also incorporate a label, such as biotin, fluorescein or radioactivity, into the DNA product so that it can be tracked following polyacrylamide gel electrophoresis. Screening methods can be broadly divided into three categories; those involving chemical modification, electrophoretic mobility or enzymatic cleavage (Prosser, 1993), each of which will be briefly discussed.

The most commonly used chemical modification method is chemical cleavage of mismatch (CCM), also known as HOT. CCM (Cotton, Rodrigues and Campbell, 1988) involves amplification of normal (wild-type) and test alleles, followed by annealing to form heteroduplexes. Mismatched cytosine residues in the heteroduplex are then modified with hydroxylamine, mismatched thymine residues with osmium tetroxide, and the resulting product treated with piperidine which cleaves at the modified mismatched sites. The products are then analysed on a denaturing polyacrylamide gel, and determination of their length allows localisation of the mutation. This method can in theory detect 100% of mismatches, provided both alleles are labelled, and can be used on fragments of over one

kilobase (Roberts, Bobrow and Bentley, 1992), but the chemicals involved are toxic, high quality reagents are necessary to avoid background contamination, and the procedure is time-consuming. Another much less commonly used technique involves addition of carbodiimide (CDI), a large molecule which binds to double-stranded DNA at sites of mismatched G and T residues. This bound adduct can then be detected by antibodies to CDI, or by blocking primer extension. This technique has been used to detect mutations in large fragments of up to 7.2kb, but it is not widely used (Cotton, 1993).

Techniques based on electrophoretic mobility include single-stranded DNA conformation polymorphism (SSCP), heteroduplex analysis (HA), and denaturing gradient gel electrophoresis (DGGE). SSCP (Orita *et al*, 1989) is a simple and straightforward method, involving PCR, denaturation to single-stranded products, followed by analysis on a non-denaturing gel. Under non-denaturing conditions, single-stranded DNA folds in a manner specific to its nucleotide sequence, which affects its mobility in a gel matrix. Fragments with alterations in the nucleotide sequence frequently have different electrophoretic mobilities when compared to normal fragments, which can be detected as a band shift in a non-denaturing polyacrylamide gel. The smaller the fragments under test, the more apparent is the change in mobility, and fragments of less than 200 basepairs are optimal (Prosser, 1993). Also, gel conditions significantly affect the rate of detection, with variations in gel and buffer constituents, and running temperature being important (Glavac and Dean, 1993). Optimisation of these factors is necessary for efficient detection rates.

Heteroduplex analysis is very similar to SSCP, but uses double-stranded nucleic acids, either formed during PCR or by post-PCR annealing of normal and test products. This technique was first described (Keen *et al*, 1991) using a special type of polyacrylamide gel, Hydrolink, manufactured by AT Biochem who have subsequently produced a more specific product (MDE gel) for this method of mutation detection. Heteroduplex analysis also works most effectively on fragments of a few hundred base pairs, but separation of duplexes can be achieved on shorter gels, making non-isotopic detection methods an option (Glavac and Dean, 1995).

Denaturing gradient gel electrophoresis (DGGE) (Myers *et al*, 1985) separates DNA fragments in polyacrylamide gels containing an increasing gradient of denaturant, such as formamide, urea or temperature. The fragment initially migrates according to its size, until it reaches a point in the gel where it starts to denature, and is retarded in the gel. The melting point of the fragment can be predicted, and optimised, using computer simulation programmes (Fodde and Losekoot, 1994). Often GC-rich clamps of DNA are added to the PCR primers, to increase the sensitivity of detection by preventing complete denaturation of the fragment (Sheffield *et al*, 1989). This method works best on double stranded DNA of

less than 500bp in length and, after optimisation of conditions, detects up to 100% of mutations. Again, this method does not require the use of radioactivity.

The first version of enzymatic cleavage of mismatched hybrids involved the use of RNase A (Myers, Larin and Maniatis, 1985). Heteroduplexes are formed using wild-type and test RNA (or RNA and DNA), preferably both labelled, and subjected to RNase A digestion, which should cleave at single base mismatches. The fragments are then run on denaturing polyacrylamide gels, allowing for localisation of the site of mismatch. However, the cleavage is often inefficient, especially at purine bases. This has led to the use of other enzymes, such as T4 endonuclease VII (Babon, Youil and Cotton, 1995), which cleaves at single base mismatches and insertions and deletions. As this technique allows for sensitive detection and localisation of mutations in large fragments of DNA, other bacterial enzymes are now being investigated as possible alternatives, the most popular being MutS, which is discussed below.

1.7.1 - MutS

In *E. coli*, several pathways exist to correct mismatches introduced during replication. The best characterised of these is the methyl-directed mismatch repair system (for review, see Claverys and Lacks, 1986, and Modrich, 1987). This pathway uses the products of the *mutS, mutH, mutL* and *mutU* genes, and is dependent on the methylation of adenine residues in d(GATC) sequences (carried out by DNA adenine methyltransferase). The *mutS* gene encodes a 97 kD protein which recognises and binds to double-stranded, hemimethylated regions of DNA containing a mismatch. It is then joined by MutH protein, a d(GATC) specific endonuclease, which cuts the unmethylated (newly synthesised) strand of DNA, 5' to the d(GATC) sequence. Excision by MutH proceeds from the nick to the mismatch (1-2 kb), with the help of MutU (DNA helicase II), which is present to unwind the duplex DNA. Other proteins, such as single strand DNA-binding protein, DNA polymerase III holoenzyme and DNA ligase are known to be involved in the reaction (Au, Welsh and Modrich, 1992). The role of MutL is yet to be revealed, but it is known to interact with the MutS-mismatch complex, possibly forming an interface with MutH. All four *mut* genes are essential for efficient mismatch repair, as mutations in any of these genes result in a mutator phenotype (Au, Welsh and Modrich, 1992).

This methyl-dependent repair pathway recognises and repairs single base mismatches, and insertions or deletions of up to four bases (Parker and Marinus, 1992). However, different base mismatches are recognised with different affinities, and the sequence context is also thought to affect repair efficiency (Au, Welsh and Modrich, 1992). *In vivo*, G-T and A-C

mismatches are poorly recognised (Modrich, 1987), and base heterologies of over four nucleotides, and C-C mismatches, are not corrected by this pathway (Smith and Modrich, 1996).

MutS is known to be involved in two other, less efficient, repair pathways, including the very short patch (VSP) system, where MutL is also required (Modrich, 1987). Homologues of the *E. coli* MutS protein have been identified in *S. typhimurium* (MutS), *S. pneumoniae* (HexA), *S. cerevisiae* (MSH1, 2 and 3), *S. pombe* (Swi4) (Fleck, Michael and Heim, 1992), mouse (Rep3) and man (hMSH2) (Reenan and Kolodner, 1992). In man the hMSH2 gene is known to have a role in Lynch syndrome (hereditary non-polyposis colorectal cancer) when mutated (Leach *et al*, 1993). The most highly conserved region of these proteins lies at the C-terminus and contains a putative helix-turn-helix DNA binding motif, and a nucleotide binding site, for which ATPase activity has been reported (Fleck, Michael and Heim, 1992). The ability to hydrolyse ATP is characteristic of enzymes that travel along DNA (Haber and Walker, 1991). The bacterial enzymes also share a region of N-terminal homology (Haber *et al*, 1988).

MutS has obvious applications in mutation detection systems, because of its ability to recognise and bind to single-base mismatches. Related proteins, such as MutT and MutY, can only recognise a small number of mismatches, and hence are of limited use in scanning for unknown mutations (Au *et al*, 1989; Lu and Hsu, 1992). Several groups have already experimented with MutS as a mutation detection system. The first of these involved identifying mismatches in denatured and re-annealed PCR-amplified sections of the *CFTR* gene (Lishanski, Ostrander and Rine, 1994). PCR products of up to 500 bp were incubated with MutS, and binding detected by a band shift assay. All heterozygote mutations tested were detected, due to each mutant allele producing two different heteroduplex mismatches, by reannealing with the normal allele. MutS has also been used in an exonuclease protection assay, again detecting mutations in the *CFTR* gene (Ellis *et al*, 1994). Following binding of MutS to heteroduplex DNA, the mixture is exposed to T7 polymerase digestion, digesting homoduplexes to completion. The remaining MutS-bound fluorescently labelled PCR fragments are then detected by electrophoresis on an automated DNA sequencer. The position of the mutation can be estimated by sizing the fragments detected.

Immobilisation of MutS has been shown to enhance its ability to detect mismatches, and reduce background binding (Wagner, Debbie and Radman, 1995). The protein was bound to nitrocellulose, and washed with buffer containing biotinylated 30-mers, with and without mismatches. Eleven of the twelve possible mismatches were detected by chemiluminescence, even in the presence of 1000-fold excess of normal DNA, the exception being C-C mispairs. However, as noted by Lishanski, Ostrander and Rine (1994),

each heterozygote will result in two mismatches, in this case C-C, and G-G, which is easily detected. This mutation detection system could also be adapted for homozygote detection, by adding wild-type DNA to test DNA before denaturation. Feng and Winkler (1995) have since published a protocol for production and purification of His_6-MutS, His_6-MutH and His_6-MutL fusion proteins. The binding properties of the MutS fusion protein were indistinguishable from the native protein, providing an excellent source of protein for testing and use in routine mutation detection assays. It may also help in the development of genomic mismatch scanning, an approach to cloning genes using MutS, by identifying regions of idenity-by-descent among closely related, affected individuals (Nelson *et al*, 1993). As this system has only been tested in *S. cerevisiae* it remains to be seen if it can be applied to the human genome, which is much larger, and richer in repetitive DNA sequences.

1.8 - AIMS

The aims of this project were:

1) To produce a MutS fusion protein, with a view to using this in mutation screening candidate genes for autosomal recessive forms of RP.

2) During the course of this project, linkage to four loci was identified for Bardet-Biedl syndrome, a syndromic form of RP. Thus, another aim of this study was to identify linkage to these loci in a set of 29 BBS pedigrees collected world-wide. In particular, we hoped to confirm genetic heterogeneity for this disorder, and if possible, further refine the location of any of the BBS loci and perhaps localise a further BBS gene.

3) To identify linkage within a set of eleven Sardinian autosomal recessive RP families. As the Sardinian population is a genetic outlier within Europe, with comparative homogeneity for other genetic diseases, it was predicted that there would be reduced non-allelic heterogeneity for arRP, hence increasing the likelihood of detecting linkage to causative loci. Any candidate genes in regions where linkage was identified would then be screened for mutations.

CHAPTER TWO

MATERIALS AND METHODS

2.1 - CHEMICALS AND SOLUTIONS

Chemicals were purchased from BDH Chemicals Ltd. or from Sigma Chemical Co. Ltd., unless otherwise stated. Media components were purchased from Difco Labs. Sterilising was carried out by autoclaving at 15 lb/sq. in. for 20 minutes, where appropriate.

2.2 - BACTERIAL CULTURE

2.2.1. - Media and additives

Unless otherwise stated, bacteria were grown on Luria-Broth (L-Broth) or Agar, supplemented with appropriate antibiotics and chemicals.

L-Broth/Agar

1% w/v Bacto®-tryptone, 0.5% w/v Bacto-yeast extract, 0.5% w/v sodium chloride. 1.5% w/v agar was added if required.

Kanamycin

Kanamycin was added to broth or agar at a final concentration of 50 μg/ml to select for bacteria carrying resistance to this antibiotic. A stock of kanamycin at 10 mg/ml in water was filter sterilised and stored at -20°C.

X-gal (5-Bromo-4-chloro-3-Indolyl-β-D-galactoside)

X-gal is a substrate for the enzyme β-galactosidase, and was added to broth or agar at a final concentration of 0.005% to select for recombinant plasmids using blue/white colour selection. A stock solution of 10 mg/ml in DMSO was prepared in a polypropylene tube, wrapped in aluminium foil, and kept at -20°C.

IPTG (Isopropylthio-β-D-galactoside)

IPTG derepresses the Lac operon, and was added to broth or agar at a final concentration of 0.25mM to select for recombinant plasmids using blue/white colour selection. A stock solution of 250mM in water was filter sterilised, and stored at -20°C.

2.2.2. - Bacterial strains

***E. coli* NM522**	*sup*E *thi* Δ(*lac-pro*AB) Δ*hsd*5 (r^-, m^-)
(Gough and Murray, 1983)	F′[*pro*AB *lac*I^q ZΔM15]
***E. coli* JM109**	*end*A1 *rec*A1 *gyr*A96 *thi hsd*R17 (r_k^-, m_k^+)
(Yanisch-Perron, Vieira	*rel*A1 *sup*E44 Δ(*lac-pro*AB)
and Messing, 1985)	F′[*tra*D36 *pro*AB^+ *lac*I^q ZΔM15]
***E. coli* XL1-Blue**	*sup*E44 *hsd*R17 *rec*A1 *end*A1 *gyr*A46 *thi*-1
(Bullock, Fernandez	*rel*A1 *lac*$^-$ F′[*pro*AB^+ *lac*I^q ZΔM15 Tn*10*(*tet*^{r})]
and Short, 1987)	
***E. coli* pop 2136**	*sup*E44 *hsd*R17 *mcr*A^+ *mcr*B^+ $r_k^-m_k^+$
(Kusters, Jager and	*end*A1 *thi*-1 *aro*B *mal*-1
van der Zeijst, 1989)	
***E. coli* DH5α™**	(ϕ80d *lac*ZΔM15) *rec*A1 *end*A1 *gyr*A96 *thi*-1 *sup*E44
(Gibco, BRL)	*hsd*R17 (r_k^-, m_k^+) *rel*A1 *deoR* Δ(*lac*ZYA-*arg*F) U169

Key: **Δ(*lac-pro*AB)** deletion removes the *lac* operon and two genes required for proline biosynthesis. These genes are usually carried on the F′ episome.

***hsd* (r^-m^-)** restriction minus - transformed DNA is not cleaved by endogenous restriction endonucleases; modification minus - DNA can be used to transform r+ strains.

***hsd* (r^-m^+)** restriction minus (as above); modification positive - DNA is unmethylated by *hsd* methylases.

***end*A1** mutation increases plasmid yield, especially from mini-preps.

***lac*I^q** mutation causes overproduction of the *lac* repressor, and thus minimises levels of *lac* expression in uninduced wild-type cells.

***rec*A1** strains are recombination deficient, which decreases their ability to rearrange cloned sequences.

***lacZ*ΔM15** deletion removes the amino-terminal α-peptide of β-galactosidase. Vectors with *lac* selection carry a gene that encodes this peptide and rescues the cells by α-complementation.

thi-1 strains cannot make thiamine, and hence won't grow on minimal media.

***sup*E** mutation suppresses amber (UAG) mutations.

***gyr*A96** mutation confers resistance to nalidixic acid.

***rel*A1** strains can synthesise RNA in the absence of protein synthesis.

Tn10 - this transposon confers resisitance to tetracycline

***mcr*A/B** mutations block restriction of methylated DNA at specific sites.

***deo*R** strains can uptake large plasmids.

2.3 - PREPARATION OF DNA

2.3.1 - Preparation of high molecular weight bacterial DNA

200 mls of an overnight culture of NM522 was harvested at 4,700 × g for 7 mins in a Beckman JA-14 rotor, and resuspended in 25 mls ice-cold buffer A (10mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0, 0.1M NaCl). This process was repeated twice, and the cells finally resuspended in 25 mls buffer A. 5 mls of 10 mg/ml lysozyme was added, and left for 10 mins at 37°C. 30 mls of buffer A with 2% sarcosyl and 20 µg/ml DNase-free RNase was added and incubated at 42°C for one hour. An equal volume of Tris-buffered phenol was added, the mixture was vortexed and spun for 5 mins at 2,400 × g. The top aqueous layer of volume V was retained and the phenol treatment repeated a further two times. 1/10 V 3M sodium acetate, pH 5.2, and 2.5 V ice-cold ethanol were added, and left for 30 mins at -70°C. This was then spun at 6,500 × g for 15 mins, the ethanol removed from the DNA pellet, and the DNA vacuum dessicated. The pellet was resuspended in 250 µl TE buffer (10mM Tris-HCl, pH 8.0, 1mM EDTA, pH 8.0), transferred to a new tube and a further 250 µl TE buffer added. DNA concentration was estimated in a spectrophotometer using the following equation :

$A_{260} \times 50 \times$ dilution factor = Double-stranded DNA concentration (µg/ml)

$A_{260} \div A_{280}$ gives an estimate of the protein or phenol contamination in the sample. A pure preparation of DNA or RNA should have a ratio of 1.8 - 2.0. If the ratio was lower than this, the sample was phenol/chloroform extracted again until this ratio was reached.
The DNA was diluted to 1 µg/µl in TE buffer, homogenised in a glass-glass homogeniser for ten rounds, and run out on a 0.8% agarose gel to check shearing was sufficient for use in PCR, aiming for < 100 kb fragments.

2.3.2 - Small scale preparation of plasmid DNA

This method was modified from the alkaline-lysis method of Sambrook, Fritsch and Maniatis (1989). 1 ml overnight culture grown in selective broth was harvested at 13,400 × g (Scotlab MicroCentaur) for 3 mins in an Eppendorf® tube, and the pellet resuspended in 180 µl of solution 1 (25mM Tris-HCl, pH 8.0, 10mM EDTA, pH 8.0, 50mM glucose) with 10 mg/ml lysozyme added. This was then incubated on ice for 2 mins, 400 µl freshly prepared solution 2 (0.2M sodium hydroxide, 1% SDS) was added, and the mixture vortexed. The tube was left on ice for 5 mins, 300 µl 3M sodium acetate, pH 5.2, was added, vortexed, and incubated on ice for 10 mins. Following centrifugation at 13,400 × g for 10 mins, the supernatant was decanted into a new tube, 500 µl isopropanol was added, mixed and the sample respun for 5 mins at 13,400 × g. The pellet was resuspended in 200 µl TE, 100 µl Tris-buffered phenol added, and vortexed. 100 µl chloroform:iso-amyl alcohol (24:1, v/v) was added and vortexed, and then spun for 5 mins at 13,400 × g. The top aqueous layer of volume V was extracted to a new tube and 1/10 V 3M sodium acetate, pH 5.2, and 2.5 V ice-cold ethanol added. The tube was incubated at -70°C for 30 mins, centrifuged for 10 mins at 13,400 × g, drained and the pellet of DNA dried in a vacuum dessicator. The DNA was finally resuspended in 20 µl TE.

2.3.3. - Medium scale preparation of plasmid DNA

The protocol used was essentially the same as **2.3.2**, but on a larger scale. 50 mls overnight culture was harvested at 5,000 × g for 10 mins in a Beckman JA-20 rotor, resuspended in 2 mls solution 1 with lysozyme, and left on ice for 10 mins. 4 mls freshly made solution 2

was added, mixed and left on ice for 10 mins. 3 mls ice-cold 3M sodium acetate, pH 5.2, was added, and the mixture left on ice for a further 10 mins. After centrifugation for 15 mins at 5,000 × g, the supernatant was decanted to a clean JA-20 tube. Half a volume of Tris-buffered phenol was added, the solution vortexed, half a volume of chloroform: IAA (24:1) added, vortexed, and respun for 10 mins at 5,000 × g. The top layer was removed to another clean tube, two volumes of ice-cold ethanol and 1/10 volume 3M sodium acetate, pH 5.2, added, and left at -20°C for 30 mins. After centrifugation for 10 mins at 5,000 × g, the DNA pellet was resuspended in 200 µl TE buffer, 20 µl 1mg/ml DNase-free RNase added, and incubated at 37°C for 30 mins. 20µl 3M sodium acetate, pH 5.2, and 500 µl cold ethanol were added, and incubated at room temperature for 10 mins. The DNA was pelleted by spinning for 5 mins at 13,400 × g in a microfuge, vacuum dessicated, and finally resuspended in 500 µl sterile TE.

DNA concentration was estimated using a spectrophotometer (**2.3.1**).

2.3.4 - Extraction of DNA from blood and cultured cells

Blood samples were frozen in sodium EDTA tubes until ready for DNA extraction, when they were defrosted on ice. DNA was extracted using the Nucleon™ II kit for large volumes (Scotlab), following the Nucleon II "protocol a" for DNA extraction as detailed in the booklet. Cultured lymphoblastoid cells were harvested at 550g for 5 mins, and DNA extracted using "protocol b" (for 1×10^7 to 3×10^7 cells) as detailed in the booklet. This kit uses a silica matrix to bind contaminating proteins, and form a partially solid interphase between the lower organic phase and upper nucleic acid phase, following chloroform extraction. Once extracted, the DNA concentration was determined, stocks in TE buffer were stored at -20°C, and working dilutions (in sterile water) kept at 4°C

2.3.5 - Extraction of PCR products from agarose gels

PCR products (100 bp - 1 kb) were extracted from ≤ 2% low-melting point agarose (Ultrapure™ LMP agarose, BRL) gels in 1 × TAE (40mM Tris-acetate, 1mM EDTA, pH 8.0) containing 0.5 µg/ml ethidium bromide (from a 10 mg/ml stock). DNA bands were

excised after brief UV transillumination and treated with a QIAquick™ gel extraction kit (Qiagen), which adsorbs nucleic acids to a silica-gel membrane. The protocol was followed as detailed in the QIAquick handbook.

2.3.6 - Preparation of cDNA

A solid-phase cDNA template was constructed using Dynabeads® Oligo$(dT)_{25}$ (Dynal A.S.) as detailed in the technical handbook (second edition). Briefly, 10mls of cultured lymphoblastoid cells were harvested, washed in PBS (phosphate-buffered saline), re-harvested, and frozen at -20°C. When required, the cells were thawed on ice, and resuspended by pipetting in 1ml lysis/binding buffer in an Eppendorf tube. Poly A^+ mRNA was isolated from the cell lysate using 60 µl of Dynabeads, thoroughly washed to remove salts and detergent, and then solid-phase cDNA synthesised using SuperScript™ II reverse transcriptase (Gibco, BRL).

2.4 - POLYMERASE CHAIN REACTION (PCR)

2.4.1 - Preparation and extraction of PCR primers

Most primers were synthesised within the MRC Human Genetics Unit by A. Gallagher, on an ABI PCRMate™ by phosphoramidite linkage. Fluorescein (FITC) labelling was carried out during primer synthesis by addition of FluorPrime™ (Pharmacia AB). Following deprotection, each primer in ammonium hydroxide solution was passed through a NAP®-10 column (Pharmacia), and eluted in 10mM Na_2HPO_4 buffer, pH 6.8. Some primers were provided HPLC purified, making this step unnecessary. The absorbance of the primers at 260 nm was measured using a spectrophotometer and the concentration calculated using the following formula:

Concentration (M) = A_{260}/ml ÷ [(nA × 16,000) + (nG × 12,000) + (nC × 7,000) + (nT × 9,600)] , where n = number of bases (A, G, C, T) in the oligonucleotide sequence.

Other primers were obtained from Oswel DNA Services, Edinburgh. They were supplied HPLC purified and at pre-determined concentrations.

The melting temperature of each oligonucleotide (in °C) was calculated using the equation [4(G+C) + 2(A+T)] °C (Itakura, Rossi and Wallace, 1984). The initial test annealing temperature for the PCR reaction was then taken as four degrees lower than the melting temperature. Primer pairs were designed to have as similar melting temperatures as possible, and some sequences were checked for self-annealing and dimer formation using the Oligo program, designed by Rychlik and Rhoads (1989).

2.4.2 - PCR of *E. coli mutS* gene

This reaction was carried out in a 50 µl volume in a microcentrifuge tube, containing 1 × GeneAmp® Perkin Elmer (Roche) PCR Buffer II (0.01M Tris-HCl, pH 8.3, 0.05M KCl), 1mM $MgCl_2$, 200µM dNTPs (Advanced Biotechnologies), 150 pmoles each primer, 1.5 units AmpliTaq® DNA polymerase (Perkin Elmer, Roche) and 50 ng sheared high molecular weight *E. coli* DNA (see section **2.3.1**). The reaction was overlaid with mineral oil and placed in a thermal cycler using the following programme:-
(95°C, 3 mins) × 1; (95°C, 2 mins, 65°C, 2.5 mins, 72°C, 3 mins) × 25; (72°C, 5 mins) × 1.
The primers used had the following sequences, based on the putative sequence of the *E. coli mutS* gene, as described by Schlensog and Bock (1991). They were also designed to introduce an *Eco*RI site at the 5' end of the PCR product, and a *Bam*HI site at the 3' end, for forced ligation into the pAX11 vector (Zueco and Boyd, 1992) (see **fig. 3.1**).
Forward 5' TGG GAA TTC AGT GCA ATA GAA AAT TTC GAC 3'
Reverse 5' TCT GGA TCC TTA CAC CAG GCT CTT CAA GCG 3'

2.4.3 - PCR of microsatellite markers

PCR reactions were carried out in 25 µl reaction volumes (or 10µl volumes with scaled down recipe) in 96-well microtitre plates (Thermowell™, Costar) containing 50 ng template DNA, 10 pmoles each primer (one of which is FITC labelled), 1 × PCR Buffer II (Perkin Elmer, Roche), 1mM $MgCl_2$, 200µM dNTPs and 1.5 units AmpliTaq DNA polymerase (Perkin Elmer, Roche). The reactions were overlaid with mineral oil, and placed in a Hybaid Omnigene thermal cycler with the following program:-

(94°C, 2 mins) × 1; (94°C, 1 min; 55°C, 2 mins; 72°C, 2.5 mins) × 24; (94°C, 1 min; 55°C, 2 mins; 72°C, 10 mins) × 1. The number of cycles was sometimes extended to 30 if too little PCR product was produced. If reactions still failed, attempts were made to optimise the magnesium concentration, annealing temperature and concentration of the primers.

2.4.4 - PCR amplification of *NRL* gene from genomic DNA

All PCR reactions were carried out as described below. For SSCP and heteroduplex analysis 1% (0.5 µl in 50 µl reaction volume) radioactively labelled [α-^{32}P]dCTP (10 µCi/µl)(Amersham) was added to the reaction - radioactive PCR reactions were carried out by K. Porter. If necessary PCR products were digested with appropriate restriction enzymes to give fragments of a suitable size for analysis.

Exon 1

Primer sequences: (expected product size = 254 bp)

Forward 5'ACAGATGACCTCAGAGAGCTGG 3'

Reverse 5'GAAAACAGACAACAGAGAGCAGGTG 3'

Reactions were carried out in a 50 µl reaction volume in microcentrifuge tubes placed in a Perkin Elmer Cetus DNA Thermal Cycler. The reaction contained 50-75 ng template DNA, 20 pmoles each primer, 1 × PCR Buffer II (Perkin Elmer, Roche), 1mM $MgCl_2$, 200µM dNTPs and 1.5 units AmpliTaq DNA polymerase (Perkin Elmer, Roche). The following PCR program was used:

(94°C, 2 mins) × 1; (94°C, 1 min; 64°C, 2 mins; 72°C, 2.5 mins) × 40; (94°C, 1 min; 64°C, 2 mins; 72°C, 10 mins) × 1.

Exon 2A

Primer sequences: (expected product size = 271 bp)

Forward 5'GAAGAGGGACTTGGTGAAGAGG 3'

Reverse 5'CATGCCTGGTTCACTGAAGGTG 3'

The reaction was carried out as for Exon 1.

Exon 2B

Primer sequences: (expected product size = 337 bp)

Forward 5'CTCACTGGGCTCCACACCTTAC 3'

Reverse 5'CTTCCTCTCTTGGGCAGTCCTC 3'

The reaction was carried out as for Exon 1, except an annealing temperature of 66°C for 2 mins was used (instead of 64°C).

This PCR product was digested with *Pvu*II (at 37°C), to give fragment sizes of 204 bp and 133bp, for SSCP analysis.

Exon 3

Primer sequences: (expected product size = 573 bp)

Forward 5'ACCGAAACAGACTGCGTGGAAG 3'

Reverse 5'CTGGTAACGATGCAGAGAACCG 3'

The reaction was carried out as for Exon 2B, with the following alterations:

10% DMSO (final concentration) was added to the reaction mixture and the magnesium concentration was increased to 1.5mM.

This PCR product was digested with *Taq*I (at 65°C), to give fragment sizes of 186bp, 254bp and 133bp, for SSCP analysis.

2.4.5 - PCR amplification of *NRL* cDNA

Primers were designed to amplify the exon boundaries of the *NRL* gene from a solid-phase cDNA template (see **fig. 6.4**). The primer sequences and cycling conditions are detailed below. For both reactions, nested primers were designed to eliminate multiple PCR products seen after the first round of PCR.

Exon 1/exon 2

First round: Forward 5'GAAGCTGTGCCTGTCTGGCAC 3' (Primer 1)

Reverse 5'GTAGCCAGCCAGTACAGCTCC 3' (Primer 3)

The reaction contained 2.5 µl of cDNA Dynabeads (13 µg cDNA) in 1 × GeneAmp PCR buffer II, 200µM dNTPs, 1.5mM $MgCl_2$, 25 pmoles each primer, 5% DMSO, 5% glycerol and 1 unit of AmpliTaq, in a 25 µl reaction volume. The PCR cycling conditions were as follows:

(94°C, 2.5 mins; 64°C, 1 min; 72°C, 1.5 min) × 1; (94°C, 1 min; 64°C, 1 mins; 72°C, 1.5 mins) × 35; (64°C, 1 min; 72°C, 5 mins) × 1.

Second round: Forward 5'CTGTCTGGCTCTGGCACTGA 3' (Primer 5'0)
Reverse 5'ACTGAGCTGTAAGGTGTGGAG 3' (Primer 2)

This reaction was carried out as above, except that 0.5 µl of the first round of PCR was used (in a 25 µl reaction volume), and the annealing temperature of the PCR reaction was reduced to 60°C.

Exon2/exon3

First round: Forward 5'CTGGCCATGGAATATGTCAATG 3' (Primer 5)
Reverse 5'GCCTTGTAGAGATCGCGCTC 3' (Primer 6)

This PCR reaction was carried out as for the second round of exon1/exon2, but with 2.5 µl cDNA Dynabeads used as the template in a 25 µl reaction.

Second round: Forward 5'TCACTGGGCTCCACACCTTAC 3' (Primer 4)
Reverse 5'GTTTAGCTCCCGCACAGACATC 3' (Primer 7)

This PCR reaction was carried out as for the second round of exon1/exon2, but using an annealing temperature of 63°C.

2.5 - ELECTROPHORESIS

2.5.1 - Agarose gel electrophoresis

DNA fragments were visualised following electrophoresis on 0.8%-2.0% agarose gels in 1 × TAE (40mM Tris-acetate, 1mM EDTA, pH 8.0), containing 0.5 µg/ml ethidium bromide (from a 10 mg/ml stock). DNA was loaded in 1 × buffer (type II, Sambrook, Fritsch and Maniatis, 1989) from a 6 × stock containing 0.25% bromophenol blue, 0.25% xylene cyanol FF and 15% Ficoll® (Type 400, Pharmacia) in water. The gels were run in 1 × TAE buffer. Following electrophoresis DNA was visualised by UV transillumination.

2.5.2 - Polyacrylamide gel electrophoresis

Small SDS (sodium dodecyl sulphate) polyacrylamide gels were prepared using the discontinuous buffer system, first described by Laemmli (1970), to analyse the protein fractions collected following induction of transformed bacterial cells. 8% separating gels, with 5% stacking gels, were made using the following recipe:

Separating gel	Stacking gel	
2.65 mls	1.2 mls	acrylamide:N,N'-bis-acrylamide (30:8 v/v) solution
2.5 mls	2 mls	gel buffer (stacking gel: 0.5M Tris-HCl, pH 6.8, 5mM EDTA, 0.4% SDS; separating gel: 1.5mM Tris-HCl, pH 8.8, 5mM EDTA, 0.4% SDS)
1.65 mls	1.3 mls	3% polyacrylamide
3.2 mls	3.4 mls	deionised water
10 µls	10 µls	TEMED
90 µls	125 µls	10% ammonium persulphate

Butan-2-ol was layered over the top of the separating gel to ensure a smooth interface between the stacking and separating gels, and was thoroughly removed before the stacking gel was added.

75 µl sample was mixed with 25 µl 4 × sample buffer (0.2M Tris-HCl, pH 6.8, 8% SDS, 0.4% bromophenol blue, 40% glycerol) and 5 µl 1M DTT, and heated for 3 mins at 100°C. The gels were run in electrophoresis buffer (25mM Tris, 250mM glycine, 0.1% SDS, pH 8.3) at 20 mA for around 90 mins. Following electrophoresis, gels were fixed in 10% acetic acid, 20% methanol for 30 minutes and then stained with Coomassie Brilliant Blue (50% methanol, 7% acetic acid, 0.25% w/v Coomassie Brilliant Blue R®) for 5 mins (shaking). Gels were destained overnight in 10% methanol, 7% acetic acid with pieces of sponge added to absorb the dye as it leached from the gel. Alternatively, gels were not stained in Coomassie Blue, but were used for electrophoretic transfer of proteins onto nitrocellulose.

2.5.3 - Preparation and running of ALF gels

Electrophoresis of fluorescein labelled microsatellite alleles was carried out using an Automated Laser Fluorescence (ALF™) DNA sequencer (Pharmacia). Glass plates were

washed thoroughly and wiped with isopropanol before use. Gels were prepared with 6% Hydrolink® (Long Ranger™, AT Biochem Ltd.), a type of high resolution polyacrylamide, in 0.6 × TBE (0.06M Tris base, 0.05M boric acid, 0.6mM EDTA) containing 7M urea (Ultrapure, Gibco, BRL) and polymerised with 0.0875% TEMED and 0.035% ammonium persulphate (APS). Gels were run in the same buffer at constant power (55W) and constant temperature (50°C), with a 1.75 second sampling interval. Samples (usually 0.5 µl each PCR product) were added to 3-6 µl loading buffer (95% formamide, 0.8% dextran blue and 25mM EDTA, pH 9.0), with appropriate size standards amplified from M13 vector DNA using primers designed by Dr. D. Mansfield (MRC Human Genetics Unit, Edinburgh). This mixture was then denatured at 95°C for 3 minutes, and cooled on ice prior to loading the total mixture on the gel.

2.5.4 - Preparation and running of sequencing gels

Sequencing reaction products were run out on denaturing 6% polyacrylamide (acrylamide:bisacrylamide 40:2, NBL Gene Sciences Ltd.) gels containing 7M urea in 1 × TBE. Gel plates were cleaned thoroughly with detergent, wiped with ethanol before use and sealed together on three sides with tape. The smaller gel plate was siliconised with dimethyldichlorosilane solution to ensure that the gel would adhere to the larger plate. Gels were polymerised using 0.15% TEMED and 0.035% APS, and run in 1 × TBE at 60 W constant power for at least 30 mins prior to loading. After electrophoresis at 60 W for 90-180 mins, gels were fixed in 5% acetic acid, 15% methanol for 20 mins before drying. Once dried onto filter paper, gels were placed next to BioMax™ (Kodak) autoradiographic film.

2.5.5 - Preparation and running of heteroduplex analysis gels

Heteroduplex analysis gels were prepared using Mutation Detection Enhancement (MDE™) gel solution (AT Biochem) as detailed in the protocol. This gel is thought to improve the resolution of conformationally altered DNA molecules, when compared to conventional polyacrylamide. Gel plates were prepared as for sequencing gels. 70 mls of 1 × gel

solution was polymerised with 0.04% TEMED and 0.04% APS, and run in 0.6 × TBE at a constant voltage of 20 V/cm for an appropriate length of time. 5 µl of each PCR product was heated at 95°C for 3 minutes, slowly cooled to 37°C, and added to 1 µl gel loading buffer (AT Biochem) ready for loading. After electrophoresis, gels were dried onto filter paper and placed next to X-Omat (Kodak) autoradiographic film.

2.5.6 - Preparation and running of SSCP gels

SSCP analysis gels were also prepared using non-denaturing MDE gel solution (AT Biochem) as detailed in the protocol. 60 mls of 0.5 × gel solution (without urea) was polymerised with 0.04% TEMED and 0.04% APS, and run in 0.6 × TBE at constant power of 7 W for 14 hours at room temperature. 2 µl of each PCR was added to 10 µl stop solution (95% formamide, 10mM NaOH, 0.25% bromophenol blue, 0.25% xylene cyanol), heated to 94°C for 2 minutes, and then placed on ice prior to loading. After electrophoresis, gels were dried onto filter paper and placed next to BioMax autoradiographic film.

2.6 - MANIPULATIONS OF DNA

2.6.1 - Restriction enzyme digestion

Restriction endonuclease enzymes were obtained from Boehringer Mannheim, and came with an appropriate buffer supplied. Digestion reactions were carried out in 20 µl volumes in the presence of 1 × buffer concentration at the optimum temperature for at least two hours. One unit of restriction endonuclease activity is defined as the amount of enzyme required to fully digest 1 µg substrate DNA in one hour at the appropriate temperature under optimal conditions. Hence, a suitable amount of enzyme was added for the quantity of DNA in the reaction. Following each digestion, DNA was cleaned up using the GeneClean® kit (Bio101)(as detailed in protocol) to remove salts and small DNA fragments. This method involves binding DNA to a silica matrix in the presence of sodium iodide, washing the DNA, and eluting in TE or water.

2.6.2 - Ligation reactions

Ligations of the vector and insert DNA were carried out using T4 DNA ligase (1 unit/µl)(Boehringer Mannheim) in a 25 µl reaction volume. The enzyme was supplied with a 10 × buffer (20mM Tris-HCl, 1mM EDTA, 5mM DTT, 60mM KCL, 50% glycerol, pH 7.5) which was used at 1 × concentration. Usually 1 µg plasmid was used with a 1-3 times molar ratio of insert DNA, and ligation catalysed by 1 Weiss unit of ligase, which was added last to the reaction mixture on ice. Reactions were carried out at 16°C overnight.

2.6.3 - DNA sequencing of plasmids

This protocol was carried out using the Sequenase® T7 DNA polymerase kit, Version 2.0, from USB (Amersham), based on the Sanger chain-termination method (Sanger, Niklen and Coulson, 1977). 5-10 µg plasmid DNA was first digested with RNase A (final conc. 0.1mg/ml) at room temperature for 15 mins, and then ethanol precipitated.

This DNA was resupended in 8 µl of deionised water, and 1 µl sequencing primer (10 ng/µl) and 1 µl 1M NaOH added. This mixture was left at 68°C for 10 mins to denature the template in presence of the primer. The annealing reaction was carried out with the addition of 4 µl TDMN (0.3M TES, 0.05M DTT, 80mM $MgCl_2$, 0.2M NaCl, 1% chloroform, pH 1.6) at room temperature for 10 mins. The reaction mixture was then stored on ice until ready to use. To the reaction mix, 4 µl labelling mix was added, containing 1 µl 0.1M DTT (from kit), 0.4 µl 5 × Sequenase labelling mix, 2.1 µl LoTE (3mM Tris, pH 7.5, 0.2mM EDTA) and 0.5 µl [α-^{35}S]dATP (12.5 µCi/µl)(Amersham). Finally, 2 µl Sequenase enzyme, diluted 1:5 with enzyme dilution buffer (from kit) was added, and the extension reaction incubated for 5 mins at room temperature. 4 µl of this reaction was added to 2.5 µl of each dideoxynucleotide termination mix (from kit, prewarmed at 37°C for 5 mins) and incubated at 37°C for 10 mins before addition of 5 µl stop buffer (from kit).

Reactions were stored at -20°C, and before electrophoresis were heated at 95°C for two mins, and then placed on ice. 5 µl of each termination reaction was loaded per well.

2.6.4 - DNA sequencing of *NRL* gene

This protocol was carried out using the Sequenase kit described in **2.6.3**. PCR products of the *NRL* gene were sequenced using the following method. 100 ng of PCR DNA in 6 µl volume was added to 1 µl primer (0.5 pmoles), 2 µl reaction buffer (from kit) and 1 µl DMSO, and the mixture kept on ice. This annealing reaction was left at room temperature for 1 minute, placed at 100°C for three minutes, and then plunged into a bath of dry ice and methanol. 3.5 µl sequencing mix, containing 1 µl DTT (from kit), 0.5 µl [α^{35}S]dATP and 2 µl 1:9 labelling mix:deionised water, was added to the annealing reaction and pipetted up and down to thaw. 1.5 µl Sequenase enzyme, diluted 1:5 in dilution buffer (stored on ice), was then added, and the labelling reaction incubated at room temperature for 45 seconds. 3 µl of this mixture was then added to 2 µl of each dideoxynucleotide termination mix (prewarmed at 42°C for 1 min), mixed, centrifuged, and incubated at 42°C for 5 mins. After this termination reaction was completed, 4 µl stop buffer was added, and the reactions were frozen at -20°C until ready to use. Samples were heated to 100°C for two minutes before loading 2.5 µl to each well.

If problems were encountered with premature "stops" in the sequence (bands in all four lanes), 1 µl formamide (or less, depending on the G/C content of the PCR product) (Zhang, Hu and Deisseroth, 1991) was added to the annealing reaction, which was then denatured for 10 mins at 100°C before being placed on dry ice.

Another sequencing protocol for difficult-to-sequence templates involved using the IsoTherm™ DNA sequencing kit (Cambio) using a modified thermostable T7 DNA polymerase, so that reaction steps can be carried out at 65°C. The reaction was carried out as detailed in the protocol, except that the annealing mix was left at room temperature for 1 min, and then denatured for 3 mins (10 mins if using formamide) at 100°C. Briefly, the template, buffer and primer were left at RT for 1 min, heated at 100°C, and then plunged into a dry-ice/methanol bath. The IsoTherm DNA polymerase and [α^{35}S]dATP were added, and the mixture divided between each of the four ddNTP tubes and incubated at 65°C for 2 mins. Chase mix, containing 0.5mM dNTPs, was added to each tube, incubated for a further 2 mins at 65°C, and the reaction stopped with formamide loading buffer. Samples were heated to 70°C for 5 mins before loading 3µl, to avoid degredation of the sequencing reaction products.

2.6.5 - Preparation and sequencing of single-stranded DNA

This protocol was kindly provided by Dr. S. Kerr (MRC Human Genetics Unit, Edinburgh). For single-stranded DNA sequencing, PCR reactions were carried out using a primer labelled with biotin at the 5' end, at 0.1μM final concentration, and an unlabelled primer (final conc. 0.5μM) in a 50 μl reaction. The biotin-labelled strand of the product was then captured using 20 μl Dynabeads M-280 Streptavidin (Dynal AS) prewashed in 100 μl buffer A (250mM Tris, pH 8.0, 0.1% Tween® 20)). The beads were resuspended in 40 μl deionised water, and the same volume of PCR and 6M lithium chloride added. This mixture was shaken on a Vibrax mixer for 30 mins, and the beads were then pulled onto the Dynal Magnetic Particle Concentrator (MPC®). The beads were washed in 100 μl buffer A, washed again in 100 μl deionised water, and resuspended in 100 μl 0.1M NaOH (freshly made) and left for 4 mins. They were then washed in 100 μl 0.1M NaOH, 100 μl buffer A, 100 μl TE (pH 7.5) and finally washed in 100 μl deionised water. The beads could be stored in this way for up to a week until they were ready to be sequenced. They were then resuspended in 7 μl deionised water, 1 μl sequencing primer (at least 0.8μM) added, 2 μl Sequenase reaction buffer added, and the sequencing reaction carried out as detailed in the Sequenase manual, or using the method described in **2.6.4**.

2.7 - TRANSFORMATIONS

2.7.1 - Preparation of competent cells

Competent *E. coli* cells were prepared using the calcium chloride method, as described in Sambrook, Fritsch and Maniatis (1989). Cells were harvested using a Beckman JA-20 rotor at 2,000 × g for 5 mins, and 50 mls of cells were finally resuspended in 1 ml ice-cold $CaCl_2$ and left on ice for at least one hour, or up to a maximum of overnight.

2.7.2 - Transformation of cells

For each ligation reaction, 150 µl of competent cells were used. The cells and the ligated plasmid were left on ice for 60 mins, heat shocked at 42°C for two minutes, 1ml L-Broth added and incubated at 37°C (shaking) for 60 mins. The cells were then diluted in L-Broth and plated on L-agar containing Kanamycin, IPTG and X-gal, to select for colonies containing recombinant plasmids.

2.8 - PROTEIN EXPRESSION

2.8.1 - Induction of fusion protein production

A culture of *E. coli* harbouring a recombinant plasmid under the control of the *lac Z* promoter was grown in selective medium to an optical density of 0.6 - 0.8 at 600 nm. Expression was induced by the addition of IPTG to a final concentration of 0.25mM, and the culture was left growing at 37°C (unless stated) for a further four hours, with samples extracted at various time-points.

2.8.2 - Fractionation of Bacterial Culture

Cells were harvested by centrifugation (7 mins at 3,800 × g in a Beckman JA-20 rotor), and resuspended in 1/10 culture volume of 10mM HEPES, pH 7.2, with 1 mg/ml lysozyme. The cells were left on ice for 20 mins, and then lysed by sonication with a micro-tip sonicator (5 bursts of 5 seconds at full power, with 50 second cooling intervals on ice). The lysate was clarified by centrifugation for 5 mins at 700 × g, to yield the pellet P3. The resultant supernatant was respun for 20 mins at 31,300 × g, yielding the soluble S20 fraction and the insoluble P20 fraction. Both pellets were resuspended to the same volume as the S20 fraction, in 10mM HEPES, pH 7.2. They were then analysed by electrophoresis on an SDS polyacrylamide gel.

2.8.3 - Western blotting

After SDS polyacrylamide gel electrophoresis, gels were placed next to a sheet of nitrocellulose (Hybond™-ECL™, Amersham)(cut to size of gel) sandwiched between two sets of four sheets of Whatman 3MM paper, and several porous pads. This whole assembly was placed in a tank of transfer buffer (20mM $NaHPO_4$, 20% methanol, 0.02% SDS), and run at 0.4 A overnight, or 1 A for 2 hours. The nitrocellulose was then rinsed in water to remove excess SDS, and stained in Ponceau S (0.2 % w/v Ponceau S, 3% trichloroacetic acid) for 5 mins to allow visualisation of transferred proteins, and destained in water and 1 × TBS (20mM Tris-HCl, pH 7.4, 0.15M NaCl). The filter was then washed in blocking buffer (0.05% bovine serum albumin, 0.5% Tween 20, 1 × TBS, with 0.05% sodium azide if blocking overnight) for 60 mins, or overnight. The blocking buffer was then replaced with 15 mls of antibody solution (1.5 µl rabbit anti-bovine IgG conjugated to horseradish-peroxidase (Sigma Immunochemicals), 0.05% BSA in 1 × TBS) and left shaking for 30 mins. It was then washed twice in 50 mls 1 × TBS, twice in 1 × TBS with 0.1% Tween 20, and finally once in 1 × TBS.

2.8.4 - Detection of recombinant proteins

After Western blotting and adding anti-horseradish peroxidase labelled antibody, the filter was treated with Enhanced Chemiluminescence (ECL, Amersham) reagents as detailed in the manufacturer's instructions, wrapped in cling-film, and exposed to autoradiography film, initially for 30 secs, and then for longer periods if necessary.

2.8.5 - Purification of recombinant proteins using IgG-Sepharose

IgG-Sepharose (Pharmacia) was stored as a 10% slurry in 1 × TBS at 4°C, and centrifuged at 13,400 × g for 1 min. in a microcentrifuge prior to use, to remove the TBS. A soluble fraction (S20, see **2.8.2**) of volume v was prepared from a bacterial culture, and 1/9 v 10 × TBS added, to give a final concentration of 1 × TBS, in a total volume of V. This was incubated at 4°C overnight, on a rotating platform with 1/10 V packed IgG-Sepharose. This

mixture was centrifuged at 13,400 × g for 1 min., and the supernatant retained - this fraction should contain any proteins present in the soluble fraction of the culture that have not bound to the IgG-Sepharose. The IgG-Sepharose was then washed twice, with V (1 × TBS), and then twice in V (5mM ammonium acetate, pH 5.0). 0.5V ammonium acetate, pH3.4 was added to the gel and incubated on a rotating table for 5 mins. This slurry was then spun at 13,400 × g for 1 minute, and the supernatant retained. This fraction should contain the eluted proteins. The IgG-Sepharose was then washed twice in V (1M Tris-HCl, pH 8.0), and finally in V (1 × TBS), and resuspended in V (1 × TBS). The protein from each fraction was precipitated by adding a final concentration of 10% Trichloroacetic Acid (TCA), and centrifuging for 10 minutes at 13,400 × g in a microcentrifuge. The pellets were then resuspended in 1 × loading buffer (25% 4 × loading buffer, see section **2.5.2**, 2.5% 1M DTT, 1.25% 2M Tris (unbuffered)), and run on a polyacryalmide gel.

2.9 - LINKAGE ANALYSIS

2.9.1 - arRP families

Eleven nuclear arRP families from South-Central Sardinia, with ancestors of Sardinian origin, were sampled by venopuncture after obtaining ethical approval and their informed consent, and lymphoblastoid cell lines established at the European Collection of Animal Cell Cultures at Porton Down, UK. Clinical examinations, including electrophysiological and visual field testing, fundus examination and fluorescein angiography (see Fossarello *et* al, 1993) were carried out at the University Eye Clinic of Cagliari. DNA was extracted as detailed in **2.3.4**, or extracted previously by M.Deans, as described by Wright *et al* (1987).

2.9.2 - BBS families

29 families with Bardet-Biedl syndrome, from the UK, Denmark, Norway, Switzerland, Sweden, Egypt and Brazil were used in this study. The criteria for inclusion of families were the presence of at least two affected members meeting the diagnostic criteria for three out of five of the cardinal features (see **1.4.2**). Sibs of affected individuals were required to meet two instead of three of the cardinal signs. Subjects, including parents and unaffected

siblings where available, were sampled by veno-puncture after obtaining ethical approval and their informed consent. Lymphoblastoid cell lines for some of the patients were established at the European Collection of Animal Cell Cultures at Porton Down, UK. DNA was extracted as for arRP families (**2.9.1**) or sent by collaborating groups.

2.9.3 - Markers used in linkage analysis

All microsatellite marker sequences used in linkage analyses were obtained from Genethon (Gyapay *et al*, 1994; Dib *et al*, 1996) or from the on-line Genome Database (GDB).

2.9.4 - ALP

Software for peak analysis and genotype filtering of ALF data was written in C and developed on Sun Sparc and IBM PC compatible computers (Mansfield *et al*, 1994). These programs have been unified to run under Microsoft Windows on IBM PC compatibles and are freely available on request.

2.9.5 - Programs used in linkage analysis

Two-point linkage analyses were carried out using MLINK from the LINKAGE program package, version 5.2 (Lathrop *et al*, 1984). Allele frequencies for microsatellite markers were taken from the Genome Database (GDB) or from parents of the families if significant differences were noted. Allele frequencies at the arRP disease locus were set at 0.992 (normal) and 0.008 (affected), and at 0.9999 (normal) and 0.0001(affected) for BBS. Multipoint lod scores were derived from markers using the LINKAGE program by means of serial three-point analyses on adjacent loci, moving along one locus at a time. A linear genetic map over equally spaced points derived from these lod scores was used for heterogeneity analysis with a Fortran program applying the formulae for a group of families with a proportion linked and a proportion unlinked (Terwilliger and Ott, 1994; Teague *et al*, 1994). In the case of BBS, the 3 chromosomes were combined as a single continuous map, and a maximum likelihood Fortran search program applied an extension of Ott's formulae to 3 linked loci (as in HOMOG4), in order to estimate the proportions of families linked to each locus.

CHAPTER THREE

PRODUCTION OF A MutS FUSION PROTEIN

3.1 - INTRODUCTION

Bacterial DNA repair enzymes are increasingly being recognised as potential tools for mutation detection. In this project, the *Escherichia coli* enzyme MutS was chosen for study. As described in section **1.7.1**, this protein works together with MutH, MutL and MutU (DNA Helicase II) to identify, and excise, short tracts of mismatched bases in newly synthesised strands of DNA. As MutS is the component that recognises, and binds to, the site of mismatch, it is a good candidate for detecting gene mutations and polymorphisms in the human genome. Proposed uses of this protein have been to enrich a DNA mixture (from two individuals) for regions of identity-by-descent, or to detect mutations in short stretches of heteroduplex DNA (see section **1.7.1**).

The production of a MutS fusion protein was undertaken as an approach to identify RP genes. Staphylococcal protein A (SPA) was chosen as the affinity handle for the fusion protein. This polypeptide can bind tightly to the F_c portion of immunoglobulin G (IgG), to aid in purification by affinity chromatography. It is thought to have a minimal steric effect on the protein, is stable in *E. coli*, and may inhibit proteolysis of the fusion product (Nilsson and Abrahmsen, 1990). It can also act as an adjuvant for production of specific antibodies. The aim of this study was to create an active fusion protein, bind it to an IgG-Sepharose column, and use this to affinity purify fragments of DNA containing a region of mismatched heteroduplex.

3.2 - PCR OF MutS GENE

Polymerase chain reaction primers were designed using the DNA sequence submitted by Schlensog and Bock (1991), that putatively coded for the *E. coli mutS* gene. This sequence was first reported for the *fdv* gene in *E. coli*, though the function of this gene was unknown. However, its position in the *E. coli* chromosome, and 90% sequence similarity with the *S. typhimurium* MutS protein, led Schlensog and Bock to suggest that this gene actually codes for *E. coli* MutS. Reenan and Kolodner (1992) have also cited the *fdv* sequence as the *mutS E. coli* gene in their report describing yeast MutS homologues. Hence, the sequence (Genbank accession no. M8642) was used to design forward and reverse primers which

added an *Eco*RI site to the 5' end of the gene and a *Bam*HI site to the 3' end, for forced cloning into pAX11 (Zueco and Boyd, 1992). The primers' sequences are shown below:

*Eco*RI
MS1F 5' TGG G↓AA TTC AGT GCA ATA GAA AAT TTC GAC 3'
FIRST CODON

*Bam*HI
MS1R 5' TCT G↓GA TCC TTA CAC CAG GCT CTT CAA GCG 3'
3' AAT GTG GTC CGA GAA GTT CGC 5'
STOP CODON

pAX11 is a SPA fusion vector, constructed from a pK19 backbone, designed by Dr. A. Boyd and Dr. J. Zueco in the Dept. of Biochemistry, University of Edinburgh, who kindly donated some of the vector for use in this project. A plasmid map of pAX11 is shown in **figure 3.1**, with the multiple cloning site highlighted. This vector encodes fusion proteins from the start codon (ATG), with the IgG-binding domains of protein A (spa, also sometimes known as 'ZZ', Hammerberg *et* al, 1990) at the N-terminus. Expression is controlled by the P_{lac} promoter. When the DNA fragment of choice is ligated in-frame into the multiple cloning site (MCS) this inactivates the LacZα polypeptide, yielding white recombinant colonies following transformation into an appropriate *E. coli* strain with the LacZM15 deletion.

Figure 3.2 (lane 4) shows the results of PCR using the primers designed, which yielded a single band from sheared *E. coli* genomic DNA. This product ("ms1"), predicted to be 2,576 bp, lies between the 2.7 kb band and the 1.9 kb band of the DNA size marker, as expected. As the distance migrated by a DNA fragment in an agarose gel is inversely proportional to the $\log_{10}$ of the number of base pairs in the fragment (Helling *et al*, 1974), the size of the PCR product was calculated using a graphical method to be approximately 2,500 bp. ms1 was digested with restriction enzymes *Eco*RI and *Bam*HI, as shown in lane 5. This digest also confirmed that ms1 did not contain any internal *Eco*RI or *Bam*HI sites, as expected. Simultaneously, pAX11 (lane 2) was digested with the same restriction enzymes, yielding a fully digested product (lane 3) of the expected size (3.13 kb).

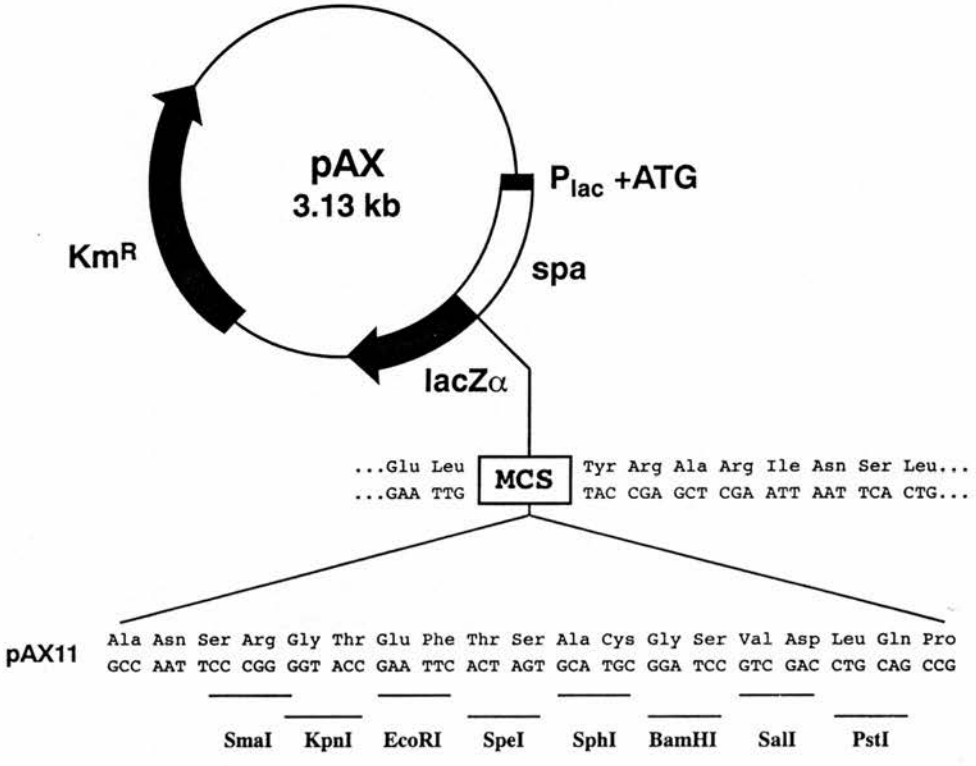


Figure 3.1 - Plasmid map of pAX11 (from Zueco and Boyd, 1992).

Thin black lines and blackened arrows represent pK19 sequences. The open box represents the *spa* gene. This plasmid confers resistance to kanamycin in transformed *E. coli*, and yields blue or white colonies in strain NM522, in medium containing IPTG and X-Gal.

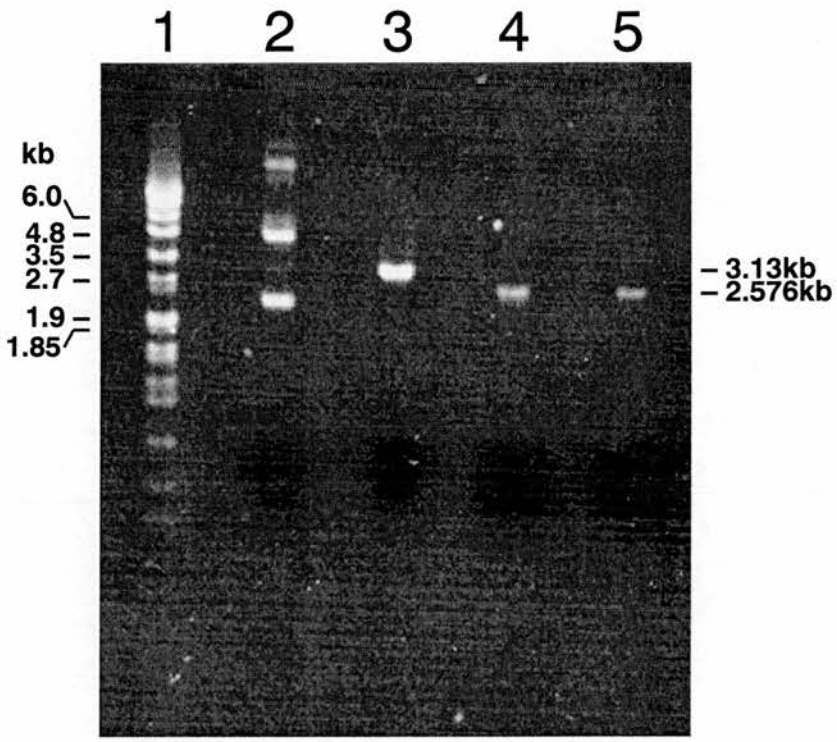


Figure 3.2 - PCR of *E. coli mutS* gene

Photograph of ethidium bromide stained agarose gel containing the following:

Lane 1: DNA marker (band sizes as marked)

Lane 2: pAX11 plasmid

Lane 3: pAX11 plasmid following digestion with *Eco*RI and *Bam*HI

Lane 4: PCR product ms1

Lane 5: ms1 following digestion with *Eco*RI and *Bam*HI

3.3 - CONSTRUCTION OF pEB1

Digested ms1 and pAX11 were purified (to remove buffer, salts, DNA fragments) and used in a ligation reaction, at an approximately 1:1 molar ration of vector:insert. This reaction was used to transform freshly prepared competent NM522 *E. coli* cells, with a heat shock at 42°C for 2 minutes to increase transformation efficiency. The cells were then plated at various dilutions onto L-agar plates containing kanamycin, to select for the presence of the plasmid, and IPTG and X-gal to select for recombinants.

After overnight incubation, a variety of white and blue colonies were visible on the plates. The negative control, of competent untransformed NM522 cells, had not grown, and the positive control, of cells transformed with pAX12 (Zeuco and Boyd, 1992), yielded only blue colonies. A selection of the white and white/blue colonies were picked for further study, and inoculated into L-broth with kanamycin. Mini-preps of DNA were prepared using these cultures, and digested with *Eco*RI and *Bam*HI to check the size of the insert fragment, as shown in **figure 3.3**. One of the plasmids containing an insert of the correct size, and hence assumed to be ms1, was then chosen for use in protein production, and designated pEB1.

3.4 -SEQUENCING OF pEB1 INSERT

Before progressing further, it was important to check that the insert in the vector had the correct DNA sequence expected for the putative MutS gene. This was carried out in two ways. Dr. D. Porteous (MRC Human Genetics Unit) had kindly donated a sample of the original MutS plasmid (pMS312) first reported by Su and Modrich in 1986. This was used as a substrate for PCR using the primers (MS1F and MS1R) described in section **3.1**, and yielded a single band of the same size as ms1. However, it is possible that this could have been due to contamination of the plasmid prep with genomic *E. coli* DNA. pMS312 was also digested using a set of restriction enzymes, to compare with the digest pattern found on digestion of ms1. It was hoped that this would further confirm that the *E. coli fdv* gene was the MutS gene, whilst avoiding sequencing the pMS312 plasmid. However, all attempts at digestion of pMS312 yielded no visible bands on an agarose gel, perhaps due to low yield of DNA from plasmid preps.

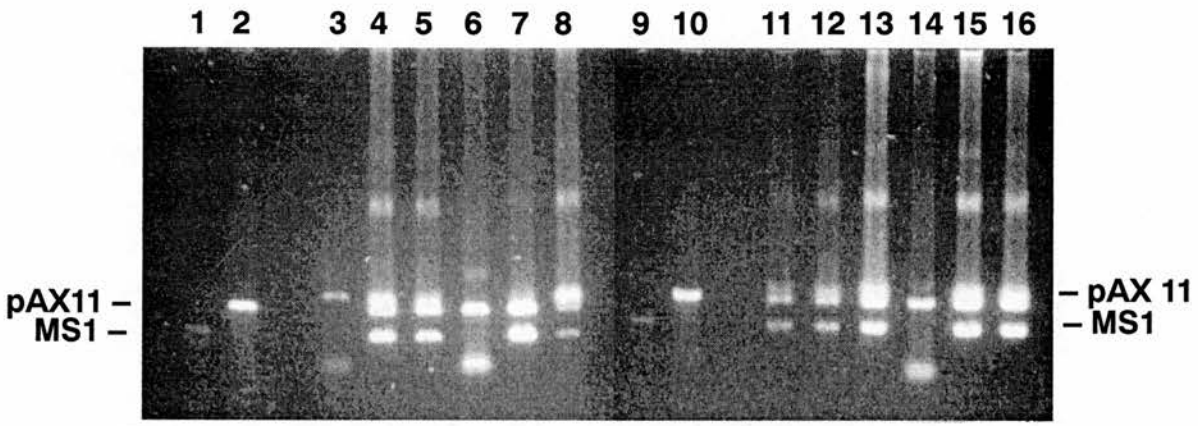


Figure 3.3 - Restriction digestion of recombinant plasmids

Photograph of ethidium bromide stained agarose gel.

Lanes 1 and 9 contain PCR product ms1

Lanes 2 and 10 contain linearised pAX11 (following digestion with *Eco*RI and *Bam*HI)

Lanes 3-8 and 11-16 contain DNA from a variety of colonies, picked following transformation of NM522 with ms1 ligated into pAX11. This DNA has been digested with *Eco*RI and *Bam*HI, to yield the inserted fragment and the linearised plasmid.

The second method involved direct sequencing of pEB1, which was carried out following preparation of sufficient quantities of plasmid DNA. The M13 -40 and +40 primers, which are supplied with the Sequenase DNA sequencing kit, were suitable for use with this plasmid. Several different protocols for dideoxy termination DNA sequencing were used, with varying results. Around 400 basepairs of DNA sequence was obtained from each end of the plasmid insert, and confirmed that it had the expected *mutS* sequence.

3.5 -PROTEIN INDUCTION

Plasmid pEB1 was transformed into competent *E. coli* cells. Cells containing the plasmid were grown up in L-Broth with kanamycin, until they had reached an optical density at 600nm (OD_{600}) of ~0.6, corresponding to a cell concentration of around 5×10^8/ml. At this point IPTG was added to the culture medium to induce production of the fusion protein for up to four hours. Cells were extracted, harvested, and incubated with lysozyme, followed by sonication to disrupt the cells. The resultant suspension was subjected to a low speed clearing spin at 700 × g (3,000 rpm in a Beckman JA-20), the supernatant decanted, and respun at 31,300 × g (20,000 rpm in a Beckman JA-20). This yielded three fractions: P3, the pellet containing cell debris from the clearing spin; P20, the pellet from the second spin, containing insoluble intracellular proteins; and S20, the supernatant from the second spin, containing soluble proteins.

The P20 and S20 fractions were analysed simultaneously on two polyacrylamide gels. One gel was stained with Coomassie blue dye, to reveal all proteins present. The other gel was Western blotted onto nitrocellulose, which was then probed with an IgG antibody conjugated to horseradish peroxidase (HRP), to detect proteins tagged with protein A. The enzymatic activity of this antibody was then converted to light, using the enhanced chemiluminescence (ECL) kit, and detected by brief exposure to autoradiographic film. The results of these experiments are detailed below.

3.6 - INDUCTION OF NM522(pEB1)

The initial induction experiment involved inducing NM522 cells, transformed with pEB1, at 37°C for four hours. The cells were harvested, processed as described above, and

analysed on an SDS polyacrylamide gel and Western blot, both shown in **fig. 3.4**. The Coomassie blue stained gel (a) does not reveal any obvious bands of the expected size of 113 kDa (97 kDa MutS and 16 kDa spa tag) in any of the fractions. However, the Western blot (b) reveals the presence of a major protein A-tagged product, lying close to the 116 kDa marker, in the P3 and P20 fractions. No corresponding band was visible in the S20 fraction, even with prolonged exposure of the blot to autoradiographic film. The experiment was repeated to confirm this result. This suggests that following four hours induction at 37°C, all the fusion protein has formed into insoluble inclusion bodies, discussed below.

3.7 - INCLUSION BODIES

Overproduction of a recombinant protein in *E. coli* often leads to the formation of intracellular insoluble protein aggregates, known as inclusion bodies (IBs). IB formation has been observed in the expression of fusion proteins, heterologous genes, and also high level expression of endogenous genes, usually when under the control of a strong promoter (Geisow, 1991). This phenomenon can be advantageous or disadvantageous depending upon the protein and its proposed use. IBs can protect proteins from degradation by proteolysis, probably due to their density preventing protease access (Enfors, 1992). This density, and the purity of the protein in the IB, can also lead to straightforward purification procedures. However, as the protein in the IB is biologically inactive and insoluble, problems can be encountered in converting it to an active, soluble protein. The fusion protein in this project was to be used in DNA-binding experiments, and hence had to be non-denatured, and in its native conformation. Often the methods for denaturation and refolding can be prohibitively time-consuming.

It is thought that formation of IBs is related to protein folding and assembly, with folding intermediates being involved in aggregation (for review, see Hockney, 1994). Indeed, mutations that alter elements of secondary structure have been shown to influence IB formation (Chrunyk *et al*, 1993), and the introduction of a translational pause, promoting correct intracellular protein folding, helps decrease aggregation (Hockney, 1994). However, protein folding is also influenced by environmental factors, such as cellular temperature, pH, and presence of other proteins, especially chaperones. These factors have all been used as approaches to prevent IB formation (Hockney, 1994). In effect, most

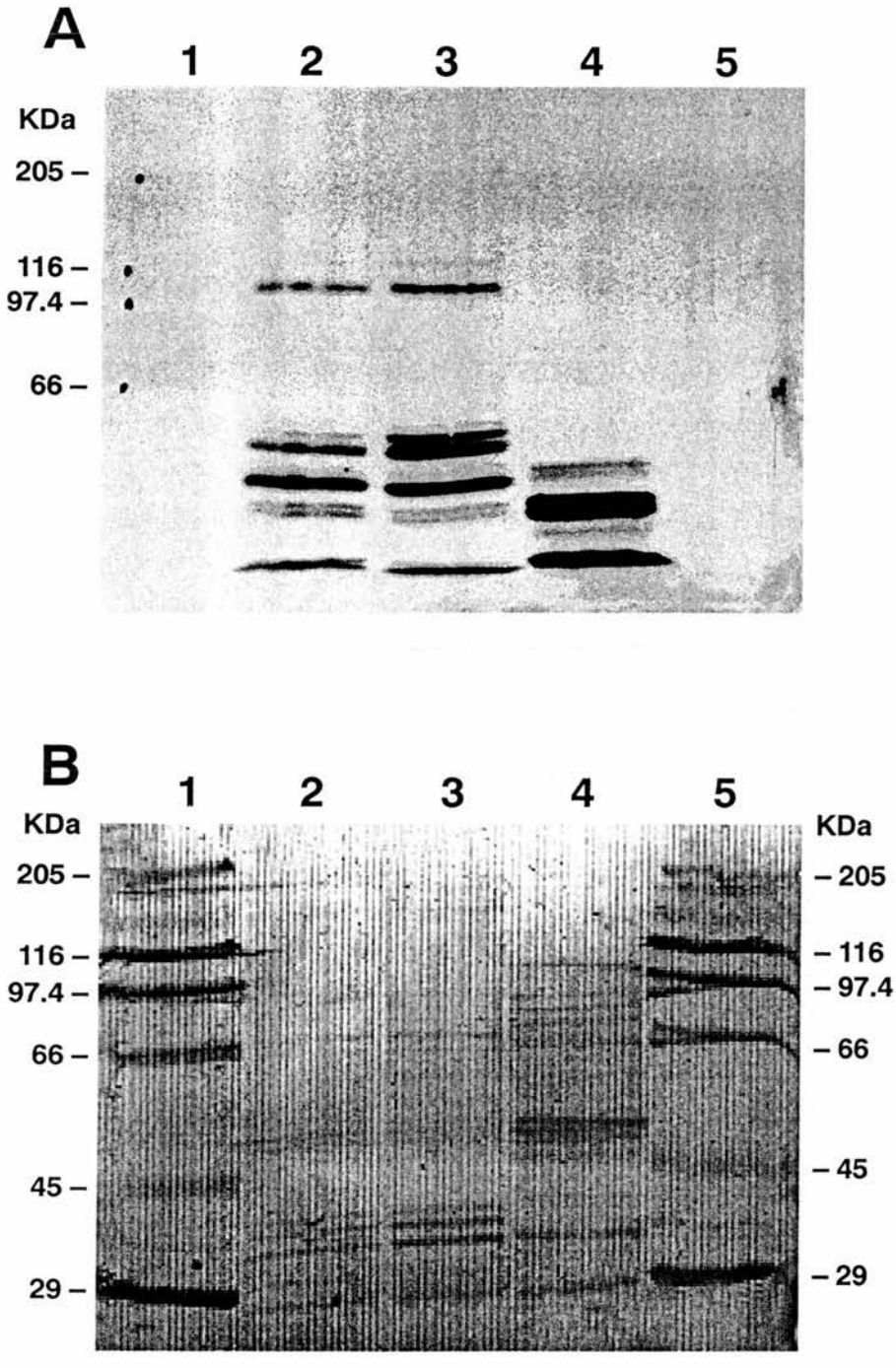


Fig 3.4 - Induction of NM522(pEB1) for 4 hours at 37°C

(a) Autoradiograph of Western blot of gel identical to (b)

(b) Coomassie blue stained 8% polyacrylamide gel. Lanes 1 and 5 contain high molecular weight markers. Lanes 2, 3 and 4 contain fractions P3, P20 and S20 respectively.

methods work by reducing the rate at which the protein is introduced into the cytoplasm, or by altering the environment into which the protein is released.

3.8 - TIME-COURSE INDUCTION

The first approach to produce a soluble product was to carry out a time-course induction. As inclusion body formation is associated with high levels of expression, the protein may be produced in a soluble form soon after induction, when only low levels are present in the cytoplasm. To investigate this possibility, induction was carried out as before, but samples were extracted from the culture at 1, 2, 3 and 4 hour time-points. The gels run with the samples P20 and S20 from this experiment are shown in **figures 3.5** (a) and (b). The Coomassie blue stained gel shown in (a) reveals faint bands of the correct size in the supernatant fractions, with much greater amounts visible in the P20 fractions. The Western blot (b) shows that this protein in the S20 appears to represent a protein A tagged protein, with the largest soluble amounts present in the 2 and 3 hour time-point fractions.

Attempts were made to purify this protein from these two fractions using IgG-Sepharose. However, the resultant material eluted from the IgG-Sepharose appeared to contain more protein bands than the original S20, as shown in the Western blot in **figure 3.6** (lane 1). Some of the additional protein bands probably represent degradation products of the fusion protein, or fragments of IgG that have become detached from the Sepharose gel matrix. Lane 3 of **fig. 3.6** also shows the unbound material from the IgG-Sepharose column, showing that the 113 kDa band observed in **figure 3.5** (b) did not run through the column without being bound.

The time-course experiment was repeated to confirm that there was fusion protein present in the soluble fraction following induction at 37°C for 1-4 hours. Only one gel was run, with half the amount of each sample loaded, and then Western blotted as before, to detect protein A tagged proteins. This Western blot is shown in **figure 3.7**, and shows that no product of the expected size is visible in the soluble fractions. This suggests that the protein observed in **figure 3.5** (b) was due to overloading of the insoluble fraction in adjacent lanes.

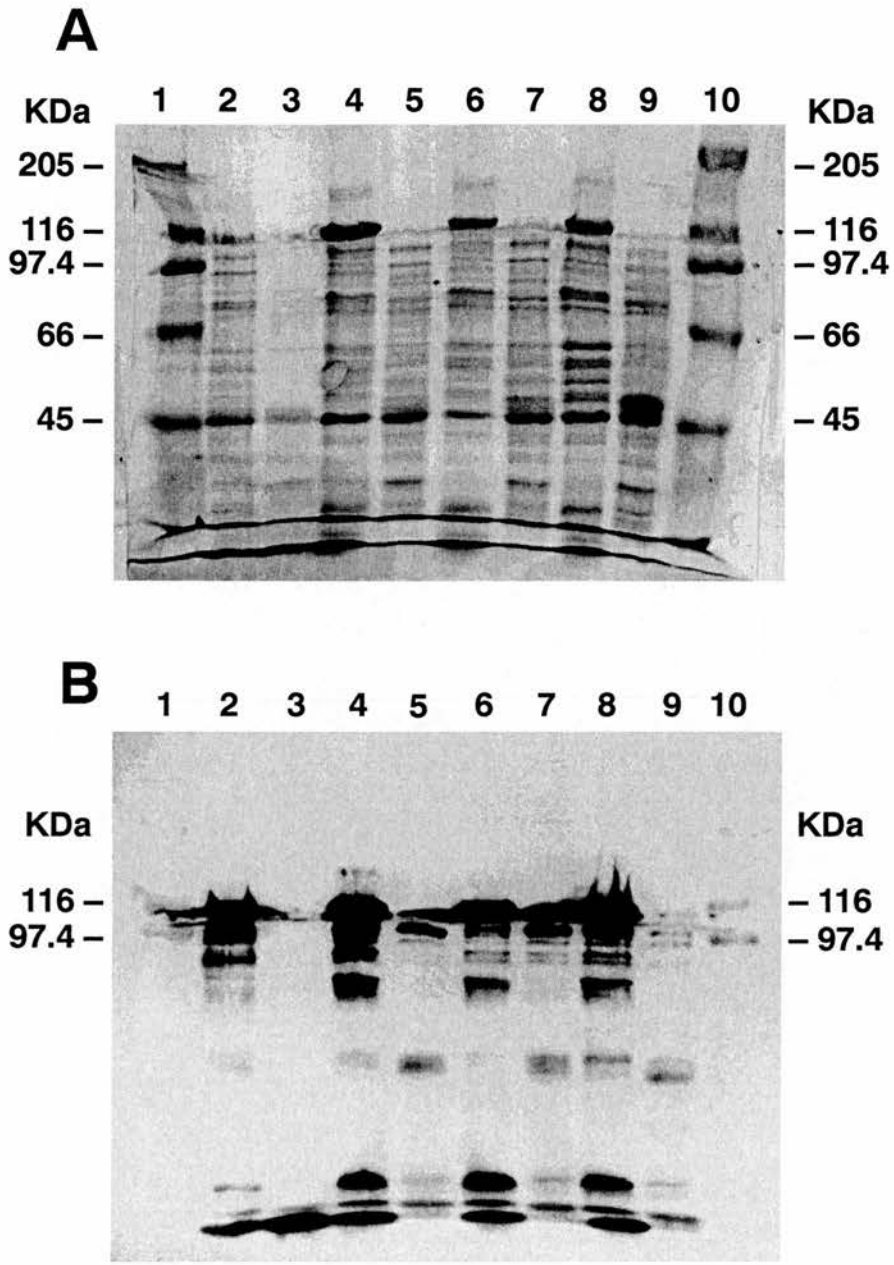


Fig 3.5 - Time-course induction of NM522(pEB1) at 37°C

(a) Coomassie blue stained 8% polyacrylamide gel

Lanes 1 and 10 contain high molecular weight protein markers. Lanes 2, 4, 6, and 8 contain the P20 fractions from NM522(pEB1) cells induced for 1, 2, 3 and 4 hours respectively. Lanes 3, 5, 7, and 9 contain the S20 fraction from the same cells, induced for 1, 2, 3 and 4 hours respectively

(b) Autoradiograph of Western Blot of gel identical to (a)

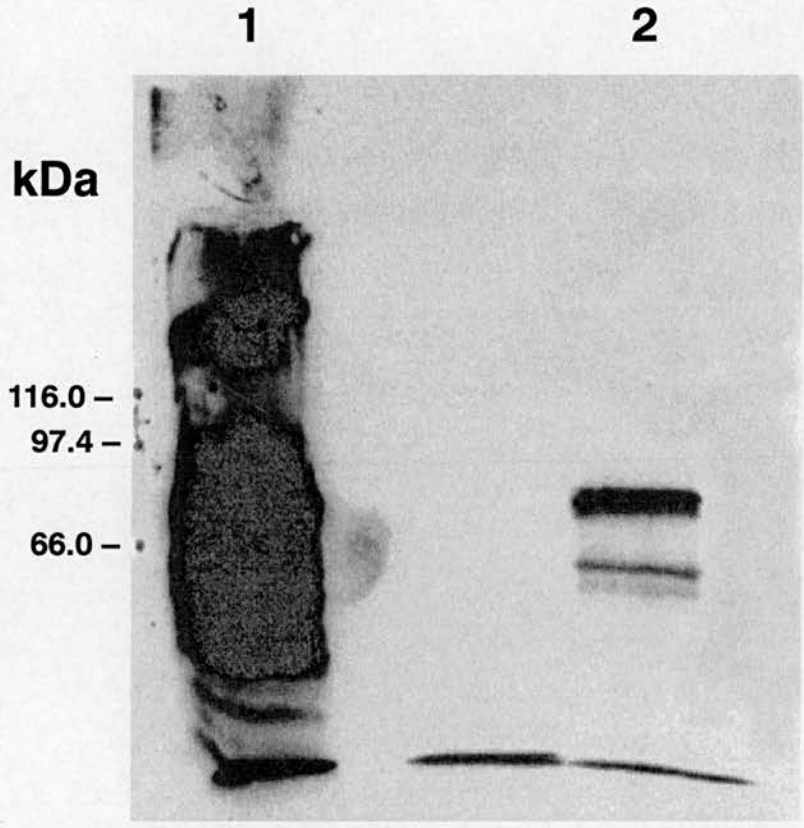


Fig. 3.6 - IgG-Sepharose affinity chromatography

Autoradiograph of Western blot of gel containing

lane 1 : S20 fractions from 2 and 3 hours induction, following purification with IgG-Sepharose, and TCA precipitation

lane 2 : material from S20 fractions not bound by IgG-Sepharose column (TCA precipitated)

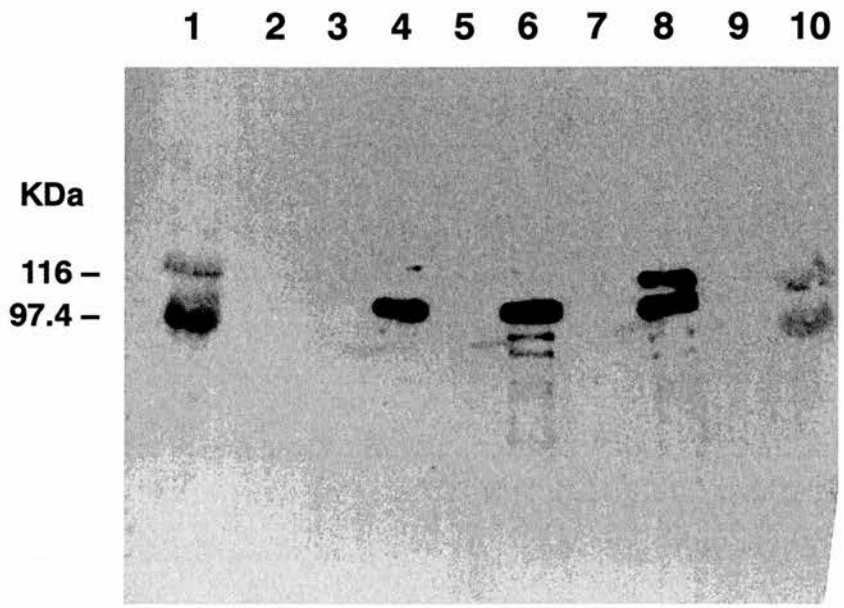


Fig. 3.7 - Time-course induction of NM522(pEB1)

Repeat of Western blot shown in fig. 3.5 (b), with half the amount of protein loaded

Lanes 1 and 10 contain high molecular weight protein markers. Lanes 2, 4, 6 and 8 contain the P20 fractions from the NM522(pEB1) cells induced for 1, 2, 3 and 4 hours respectively. Lanes 3, 5, 7 and 9 contain the S20 fraction from the same cells induced for 1, 2, 3 and 4 hours respectively.

3.9 - INDUCTION OF NM522(pEB1) AT 25°C

The next approach to produce the fusion protein in a soluble form involved inducing protein production at a lower incubation temperature of 25°C. This approach has been successfully used for several proteins, including interleukin-1β, where Chrunyk *et al* (1993) showed that the percentage of recombinant protein produced in inclusion bodies decreased from 50-60% at 42°C to around 20% at 32°C. Similar results have been found for production of human interferons, where fermentation temperatures of 23-30°C resulted in 30-90% of the protein being produced in a soluble form (Schein and Noteborn, 1988).

Induction of cells at a fermentation temperature of 25°C was carried out as described previously, with culture samples extracted at 1 - 4 hour time points. The Coomassie blue stained polyacrylamide gel (a) and corresponding Western blot (b) are shown in **fig 3.8**. The gel (a) shows bands just below the 116 kDa marker in each of the insoluble fractions, with most protein visible at the 3 and 4 hour time-points. This corresponds to the bands seen in the Western blot (b) following probing with an IgG antibody. While this clearly shows that the protein is not produced in a soluble form, it also shows that reduced temperature greatly reduces degradation of the fusion protein, when compared to **fig 3.5** (b), showing the results of the induction at 37°C.

3.10 - USE OF DIFFERENT *E. COLI* STRAINS

Following the results of reduced temperature incubation, the next approach chosen was to transform the pEB1 plasmid into two different compatible *E. coli* strains. The strains chosen, XL1-Blue and JM109, are both commonly used for production of recombinant proteins. The use of alternative *E. coli* strains has been successfully employed in the production of β-lactamase, polio virus 3C protease, and HIV GAG-9, where it was shown that protein solubility varied in 11 *E. coli* strains tested (Kenealy *et al,* 1987). It is not known which genetic differences between strains cause this variation.

The induction experiments were carried out at 37°C, as for NM522 cells, except that samples were extracted at 30 minutes, 1 hour, 2 hours and four hours. The Coomassie blue stained gel and Western blot are shown in **figures 3.9** (XL1-Blue) and **3.10** (JM-109). In each gel, only the soluble fraction was loaded, to avoid contamination by overloading of the

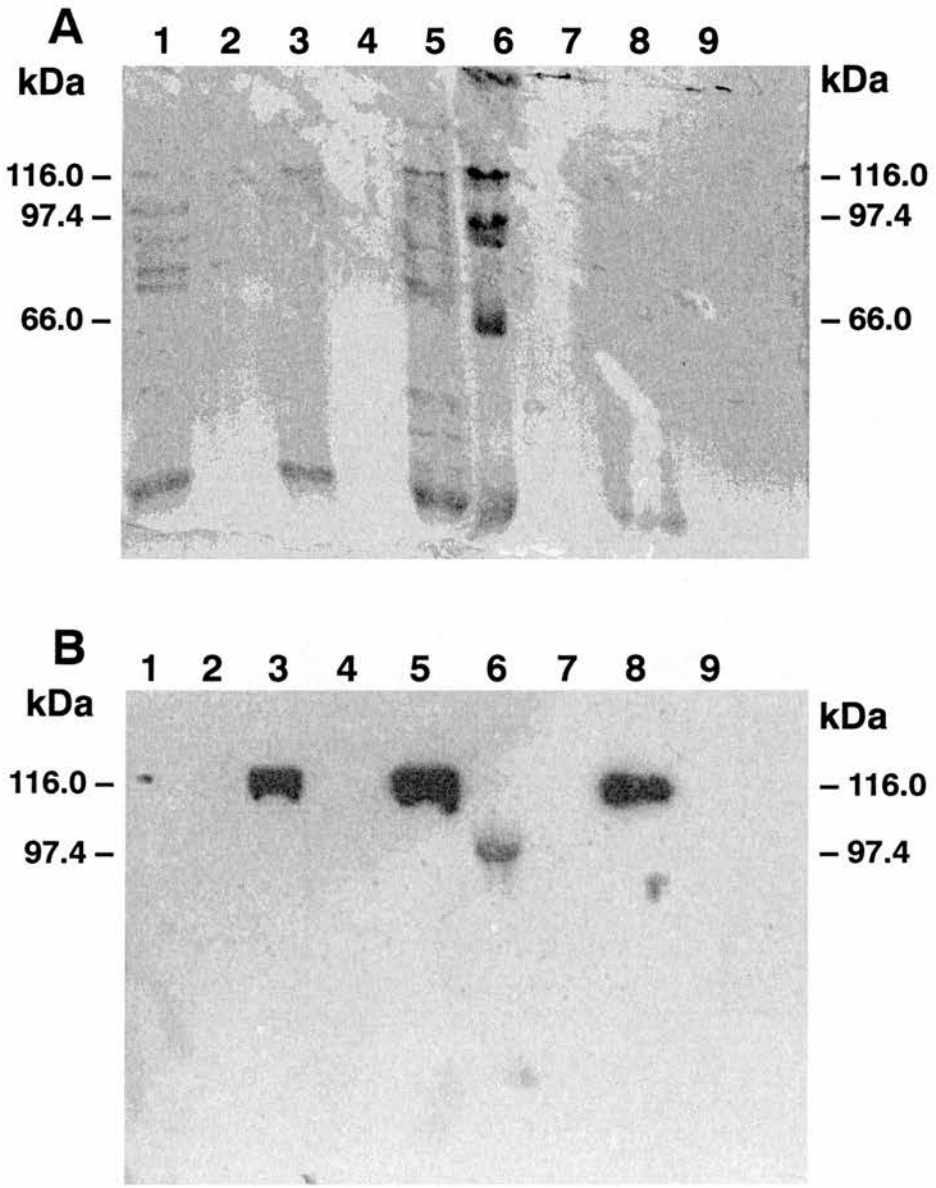


Fig. 3.8 - Time-course induction of NM522(pEB1) at 25°C

(a) Coomassie blue stained 8% polyacrylamide gel. Lane 6 contains a high molecluar weight protein marker. Lanes 1, 3, 5, and 8 contain the P20 fraction from NM522 (pEB1) cells induced for 1, 2, 3, and 4 hours respectively, at 25°C. Lanes 2, 4, 7, and 9 contain the S20 fractions from the same cells, induced for 1, 2, 3, and 4 hours respectively, at 25°C

(b) Autoradiograph of Western blot of gel identical to (a)

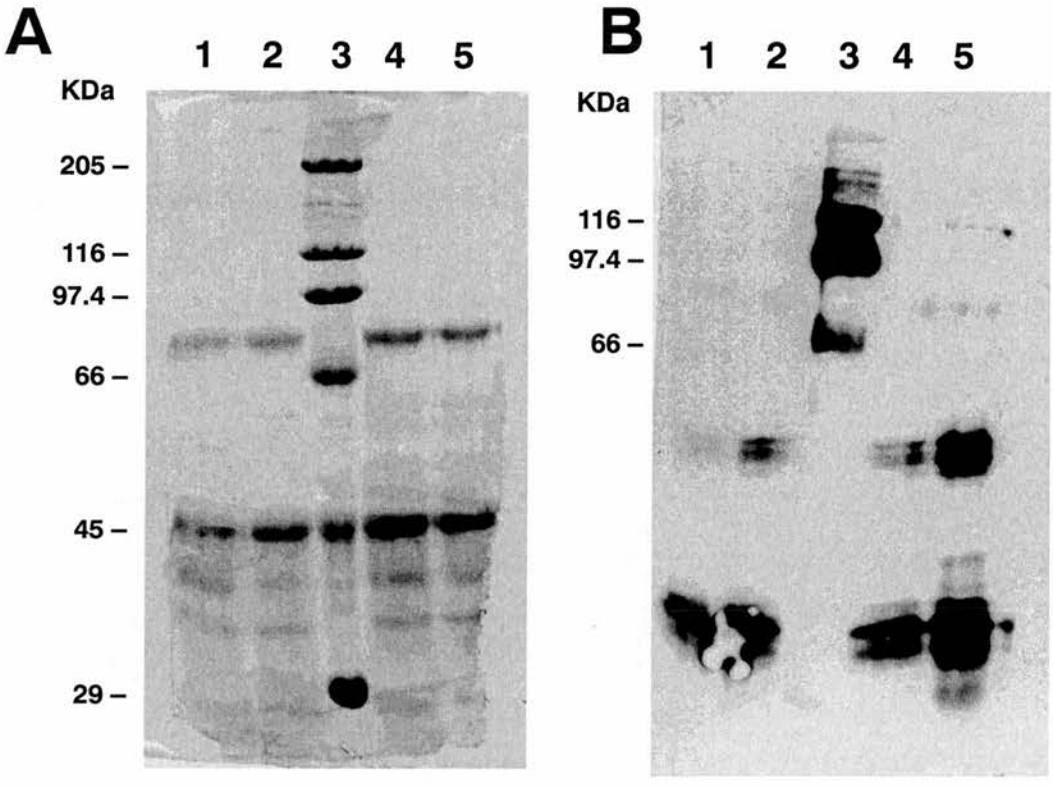


Figure 3.9 - Time-course induction of XL1-Blue(pEB1)

(a) Coomassie blue stained 8% polyacryalmide gel. Lanes 1, 2, 4 and 5 contain the 30 min. and 1, 2 and 4 hour S20 samples respectively. Lane 3 contains a high molecular weight protein marker.

(b) Autoradiograph of Western blot of gel identical to (a).

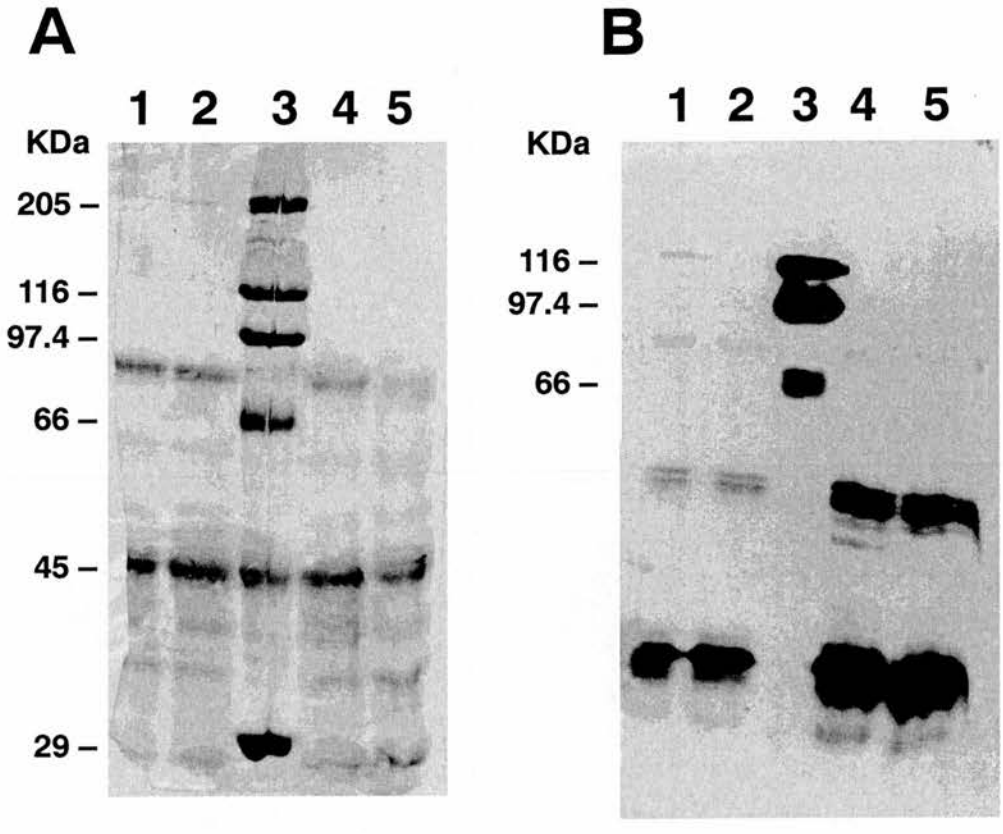


Figure 3.10 - Time-course induction of JM109 (pEB1)

(a) Coomassie blue stained 8% polyacrylamide gel. Lanes 1, 2, 4 and 5 contain the 30 min. and 1, 2 and 4 hour S20 samples respectively. Lane 3 contains a high molecular weight protein marker.

(b) Autoradiograph of Western blot of gel identical to (a).

insoluble fractions. Both Coomassie blue stained gels reveal no obvious protein of the correct size. However, the Western blot for the transformed XL1-Blue cells shows a faint band near the 116 kDa marker, in the 4 hour time-point sample (**fig. 3.9**), and the corresponding blot for the transformed JM-109 cells reveals a faint band at the same position following thirty minutes induction (**fig. 3.10**). The induction experiment was repeated in the XL1-Blue cells for 3-6 hours, and in the JM-109 cells for 5-40 minutes, to check for higher levels of soluble protein. However, in each case no increase was seen in the levels of soluble fusion protein produced. The induction experiments were also repeated at 25°C, but this resulted in a decrease in overall protein production, with only degradation products visible following overnight exposure of the Western blot.

3.11 - SOLUBILISATION OF INCLUSION BODIES

When all efforts aimed at production of a soluble fusion protein fail, attempts can be made to solubilise the protein from the IBs. Several protocols are available for extracting protein in a denatured form from IBs, and subsequently renaturing it, to restore normal activity. The most common approach is to use strong denaturants, such as 5-8M guanidine hydrochloride or 6-8M urea. Other alternatives include SDS and acetonitrile/propanol, to disrupt hydrophobic interactions, or pH > 9.0, to disrupt ionic interactions. A number of variables can also influence solubilisation, in particular the incubation temperature, concentration of protein and solubilising agent, and pH, with each protein requiring different conditions.

Following solubilisation, the next step is to refold the protein by diluting or dialysing out the solubilising agent. Again, factors such as the rate of dilution, and pH, and concentration of components must be considered. In this study, attempts were made to solubilise the protein from inclusion bodies using urea and guanidine hydrochloride, followed by overnight dialysis. However, in both cases the protein precipitated out of solution, and although further conditions could have been tried, it was decided that a less time-consuming approach would be to subclone the pEB1 *mutS* gene into a different type of fusion vector, in an attempt to produce soluble fusion protein. This was not carried out however, due to time constraints.

3.12 - ALTERNATIVE APPROACHES

All previous attempts at MutS fusion protein induction had produced very low levels of soluble protein. In some cases, fusion proteins remain insoluble with certain fusion tags, but are more soluble when expressed from different constructs. A variety of commonly used fusion systems are available for use, including the 10 amino-acid (aa) Streptavidin tag, the 40 kDa Maltose Binding Protein, the 25 kDa glutathione-S-transferase, and the 6 aa Histidine tag (for review see Nygren, Stahl and Uhlen, 1994). Each system has its own specific method for affinity purification. The system planned for further use in this study with MutS was the His tag, consisting of a string of six histidine residues, attached to the C or N terminus of the protein. This polyhistidine tail has a high affinity for metal ions, and is usually bound by a resin containing nickel ions. The strength of the affinity for this resin means that a tagged protein can still be efficiently purified from solutions containing strong denaturants such as 6M guanidine hydrochloride or 8M urea, allowing for purification from inclusion bodies if necessary. Bearing in mind the experience of working with MutS fused to SPA, this feature made the system particularly suitable. Once bound and washed, the protein is then eluted from the column by competition with imidazole, or by reducing the pH slightly (Hoffmann and Roeder, 1991).

At the end of 1995, a report was published detailing a purification protocol for producing a His_6-MutS fusion protein (Feng and Winkler, 1995). Unfortunately, time-constraints had prevented work from progressing on the production of a MutS fusion protein prior to this publication. Feng and Winkler found that most of the expressed fusion protein remained in the soluble fraction, and discarded any inclusion bodies that were formed. This level of solubility allowed purification under non-denaturing conditions, yielding purified protein accounting for 15%-30% of the total soluble protein in cell extracts. The His_6 tag was subsequently removed by thrombin cleavage, without cleaving the MutS protein internally. This protocol produced MutS with mismatch-binding activity, in a form suitable for use in mutation and polymorphism detection experiments.

3.13 - DISCUSSION

This chapter describes attempts to produce an active MutS-protein A fusion protein in *E. coli*. PCR primers have been designed that successfully amplify the *E. coli* MutS gene, which could then be cloned into fusion vectors. In this project the pAX11 fusion vector

was originally chosen. This vector tags the cloned fragment with 16 kDa IgG-binding domains of Staphylococcal protein A, enabling straightforward detection and purification of recombinant proteins. Attempts were made to produce such a fusion protein in an active, soluble form, for use in DNA-binding. Early results indicated that the desired protein was present in the insoluble fraction of the cell extract, suggesting it was being directed to inclusion bodies upon synthesis within the cytoplasm. Despite efforts to prevent IB formation, by altering the levels of expression, temperature of incubation, and host strain used for synthesis, we were unable to produce suitable quantities of the fusion protein in a soluble form. Attempts were made to solubilise the protein in the IBs with denaturing agents, with the final aim of purification by affinity chromatography (with IgG-Sepharose), but these too proved unsuccessful. FPLC was also employed to purify trace quantities in the soluble cell extract, but this failed to yield perceptable amounts of protein.

The next stage of this work was to involve subcloning the MutS gene into a His_6 fusion vector, but no progress was made with protein expression before Feng and Winkler (1995) published a protocol detailing purification of such a fusion protein. This will enable production of active MutS for use in mutation detection, as described by Wagner, Debbie and Radman (1995), who used MutS immobilised to nitrocellulose. The His_6 fusion protein could be used in a similar way, but immobilised to nickel Nitrilo-Tri-Acetic acid (NTA) resin (Hoffmann and Roeder, 1991).

In summary, a MutS fusion protein presents an ideal tool for mutation detection. Preliminary results from groups working with MutS (see section **1.7.1**) have shown that it reliably detects and binds to most mismatches *in vitro*. The production of a MutS fusion protein, for easy purification, and handling, widens the possible applications of this protein in the field of human genetics, and provides an ideal tool for quick and routine detection of single base mismatches.

CHAPTER FOUR

LINKAGE MAPPING IN BARDET-BIEDL SYNDROME

4.1 -INTRODUCTION

Bardet-Biedl syndrome (BBS) is a genetically heterogeneous disorder with four loci, designated *BBS1-4*, situated on chromosomes 11q13, 16q21, 3p13-p12 and 15q22.3-q23 respectively (Leppert *et al*, 1994, Kwitek-Black *et al*, 1993, Sheffield *et al*, 1994, Carmi *et al*, 1995a) (see section **1.4.2**). All of these loci have been localised solely by linkage analysis, and no patients with BBS have been reported with cytogenetic abnormalities in these regions. None of the corresponding BBS genes have been identified as yet.

The aim of this part of the study was to identify one or more BBS loci by genetic linkage analysis in a set of BBS families collected world-wide. By the end of the project there was a total of 29 families provided by collaborating physicians in U.K., Norway, Denmark, Sweden, Switzerland, Brazil, Egypt and Italy. This represents one of the largest resources of BBS patients currently available. At the start of the study no linkage had been reported for BBS, and Dr. David Mansfield (formerly of MRC Human Genetics Unit) had analysed ten polymorphic DNA markers on chromosome 18, nine markers on chromosome 2 and six markers on chromosome 17. These specific chromosomes were analysed because of a report of an attenuated form of BBS in a patient with a mosaic deletion of chromosome 18 (Dr. A. Wright, personal communication), and a 46, XX t(2p-;17p+) translocation in a patient with BBS (Fannemel *et al*, 1993). However, no significant linkage was identified within the set of families to any of these chromosomes.

The next stage of the project was to involve using microsatellite markers distributed at 20 cM intervals on every chromosome, in a genome-wide scan for a BBS gene. This was interrupted at the end of 1993 when Kwitek-Black *et al* reported linkage between BBS and markers on chromosome 16q21 in a large inbred Bedouin kindred from the Negev region of Israel. Three microsatellite markers (D16S408, D16S400 and D16S421) were reported as being linked to this locus, later designated *BBS2*, and so were analysed in all the available families.

The next report of linkage came several months later, and led the search to chromosome 11q13, where Leppert *et al* (1994) had found significant linkage with two markers, in a subset of 17 out of 31 North American families. These families were similar in size and ethnic origin to the families used in the present study, and it was decided to analyse six markers on chromosome 11q13. A further marker was later analysed, when results

suggested that the locus could be several cM centromeric to the location originally suggested by Leppert *et al.*

Later that year another report of linkage for BBS came from the same group who had reported linkage for 16q, again using a large Bedouin family from the same region of Israel (Sheffield *et al,* 1994). They found linkage between BBS and markers on the short arm of chromosome 3. Again, our set of families was analysed using the three most closely linked dinucleotide repeat markers, D3S1254, D3S1251 and D3S1271. However, these markers were relatively uninformative in the families studied, but suggested no significant linkage.

The final report of a BBS locus came in 1995, again from the Sheffield group, using a third inbred Bedouin kindred (Carmi *et al,* 1995a). They identified linkage on the long arm of chromosome 15, using a DNA pooling method to find the common chromosomal region inherited by the affected members of the family. The four most tightly linked microsatellite markers in this region (D15S125, D15S131, D15S204 and D15S114) were analysed, and suggested linkage in several of our families.

This chapter describes the results of these analyses using markers linked to three of the four BBS loci, and attempts to assign, or exclude, specific BBS loci in each family.

4.1.1 - Families

Twenty-nine families diagnosed with Bardet-Biedl syndrome were used in this study. Most are of European descent, and the ethnic origin of the family is indicated in the results. The criteria for inclusion of families were the presence of at least two affected members of whom at least one met the diagnostic criteria for three of the cardinal signs (retinal dystrophy, polydactyly, obesity, hypogenitalism and mental retardation). Sibs of affected individuals were required to meet two instead of three of the cardinal signs. Many associated features were also observed in the families, including brachydactyly, renal anomalies, dental anomalies and liver disease. Brief clinical details for each family, where available, are listed in the results (**4.2.1 - 4.2.28**). Parental consanguinity was present in six of the pedigrees: BB7, BB13, BB18, BB20, BB25, BB31. The degree of consanguinity is detailed in the pedigree figures or results section, where known. All available affected and unaffected individuals were sampled by venopuncture after obtaining ethical approval and informed consent.

4.1.2 - Microsatellite Markers

Sequences for microsatellite markers were obtained from Genethon or the on-line Genome Database (GDB). Markers were synthesised with one primer of each pair fluorescently labelled for use in an automated laser fluorescence DNA sequencer (ALF, Pharmacia Biotech). **Figure 4.1** shows a typical pattern obtained with a dinucleotide repeat microsatellite marker in a nuclear family. Each colour represents a separate lane in the gel, loaded with the PCR reaction of one individual. "Stutter" or "shadow" peaks (as seen in **fig. 4.1**) are usually observed when running dinucleotide repeat markers, possibly due to polymerase slippage during the PCR reaction (Mansfield *et al*, 1994). These stutter peaks are often used to aid in sizing the larger peaks, which correspond to the true alleles.

If accurate sizing of the peaks was required, the ALP program (Mansfield *et* al, 1994) was used to size alleles, with internal size standards run in each lane, flanking the expected product (see **fig 4.1**). One lane in each gel would also be devoted to a "gold standard" sample, using DNA from CEPH individual 134702 as the template for the microsatellite marker. Genethon markers are published with the allotype of this individual, which has been accurately determined by sequence analysis. Following sizing, the ALP program output was checked manually, and then converted to data in a form suitable for input into a linkage program.

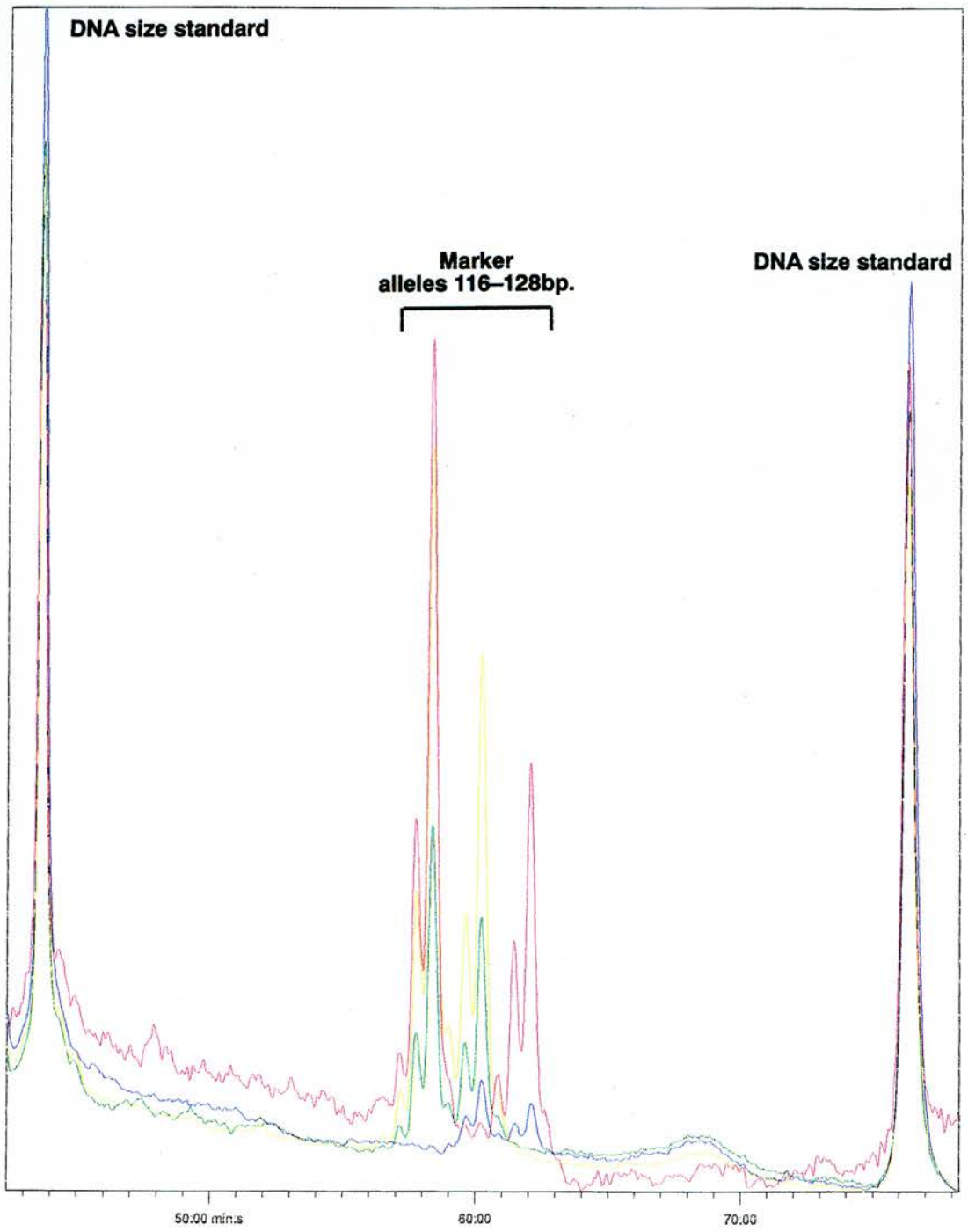
4.1.3 - Linkage Mapping

Linkage analysis was carried out by multipoint analysis using the LINKAGE program package (Lathrop *et al*, 1984), by P. Teague (MRC Human Genetics Unit). LINKMAP was used to obtain multipoint lod scores by a series of three-point analyses on adjacent loci moving along the chromosome one locus at a time. A linear genetic map for each chromosome studied was derived from these lod scores, and used for the analysis. Markers and genetic distances were used as detailed below (section **4.2**).

Heterogeneity analysis was carried out by P. Teague, using either the HOMOG program package (Terwilliger and Ott, 1994), or a Fortran program which located the maximum lod score using the formulae employed by HOMOG (Teague *et al*, 1994).

Figure 4.1

Example of microsatellite alleles in a nuclear family, run with internal DNA size standards on the ALF. Father = green lane (alleles 116 bp, 122 bp); mother = blue lane (122, 128); child 1 = yellow lane (116, 122); child 2 = red lane (116, 128)



4.2 - RESULTS

The results for each family for the loci on chromosomes 11, 15 and 16 have been displayed in the form of haplotypes in **figures 4.5 - 4.32**. These haplotypes have been deduced so as to minimise the number of crossovers between markers. Seven markers on chromosome 11 were analysed, and these are shown in **figure 4.2** with their relative distances apart in cM, as calculated from Genethon data (Dib *et al*, 1996) and James *et al* (1994). Four markers on chromosome 15 were analysed, and they are shown in **figure 4.3**. The three markers studied on chromosome 16 are shown in **figure 4.4**, with genetic distances taken from Genethon (Dib *et* al, 1996) and the radiation hybrid maps generated by the Whitehead Institute at MIT (http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map). Further markers relevant to the subsequent mapping of these loci are also included in the figures.
Positive multipoint lod scores (> 0.10) are listed in the results for each family (**4.2.1 - 4.2.28**), with the maximum value and chromosomal location in cM. The conditional probability of linkage to each locus is also included, with a probability < 0.10 suggestive of exclusion of that locus, and > 0.50 suggestive of linkage. The position of the markers studied along each chromosome is listed below, as used in the multipoint linkage analyses:
Chromosome 11
D11S480 (61cM from 11pter) - D11S913 (69cM) - D11S987 (69.1cM) - D11S916 (80cM) - D11S527 (85cM) - D11S906 (85.1cM) - D11S901 (92cM)
Chromosome 15
D15S125 (89cM from 15cen) - D15S204 (96cM) - D15S131 (96.1cM) - D15S114 (98.6cM)
Chromosome 16
D16S408 (74cM from 16pter) - D16S400 (81cM) - D16S421 (85cM)

The recombination distances between these markers differ in some cases from those detailed in **figures 4.2-4.4**, as these figures use the most recent data, and the multipoint analyses were carried out before this information became available.

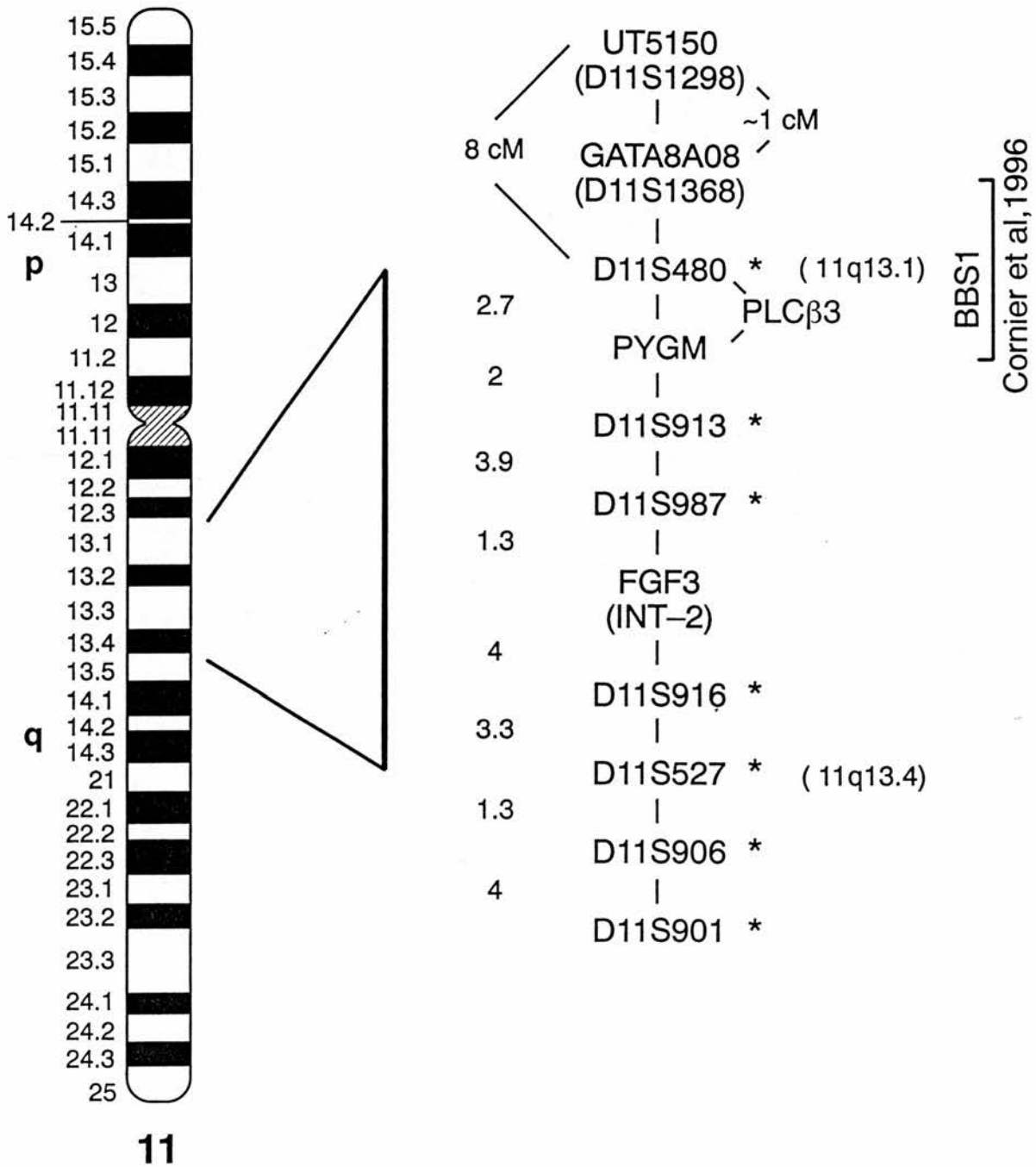


Figure 4.2 - Location of linked genetic markers in chromosomal region 11q13

Recombination distances are shown in centiMorgans. Markers used in this study are marked with an asterisk.

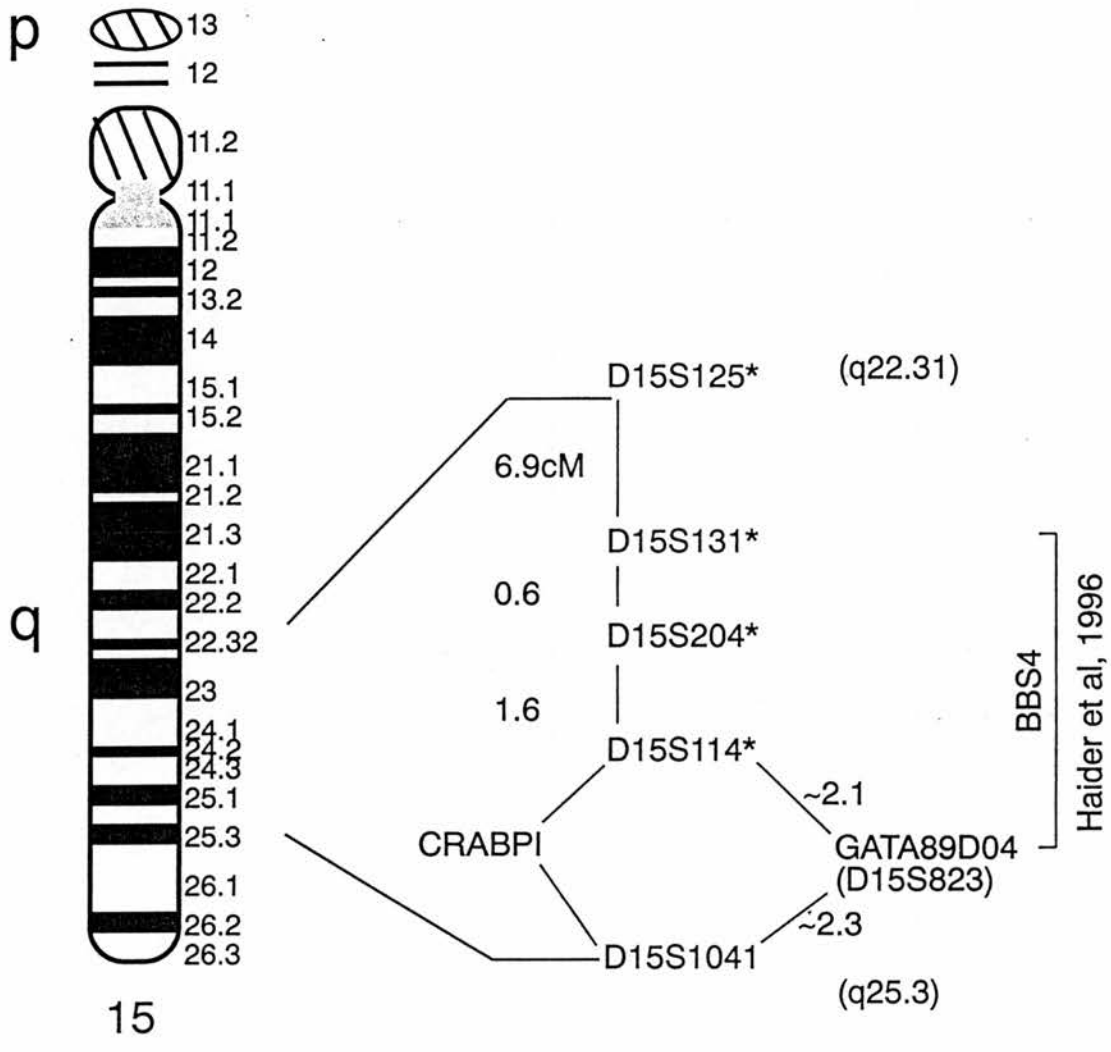


Figure 4.3 -Location of linked genetic markers in chromosomal region 15q22.31-q25.3
Recombination distances are shown in centiMorgans. Markers used in this study are marked with an asterisk.

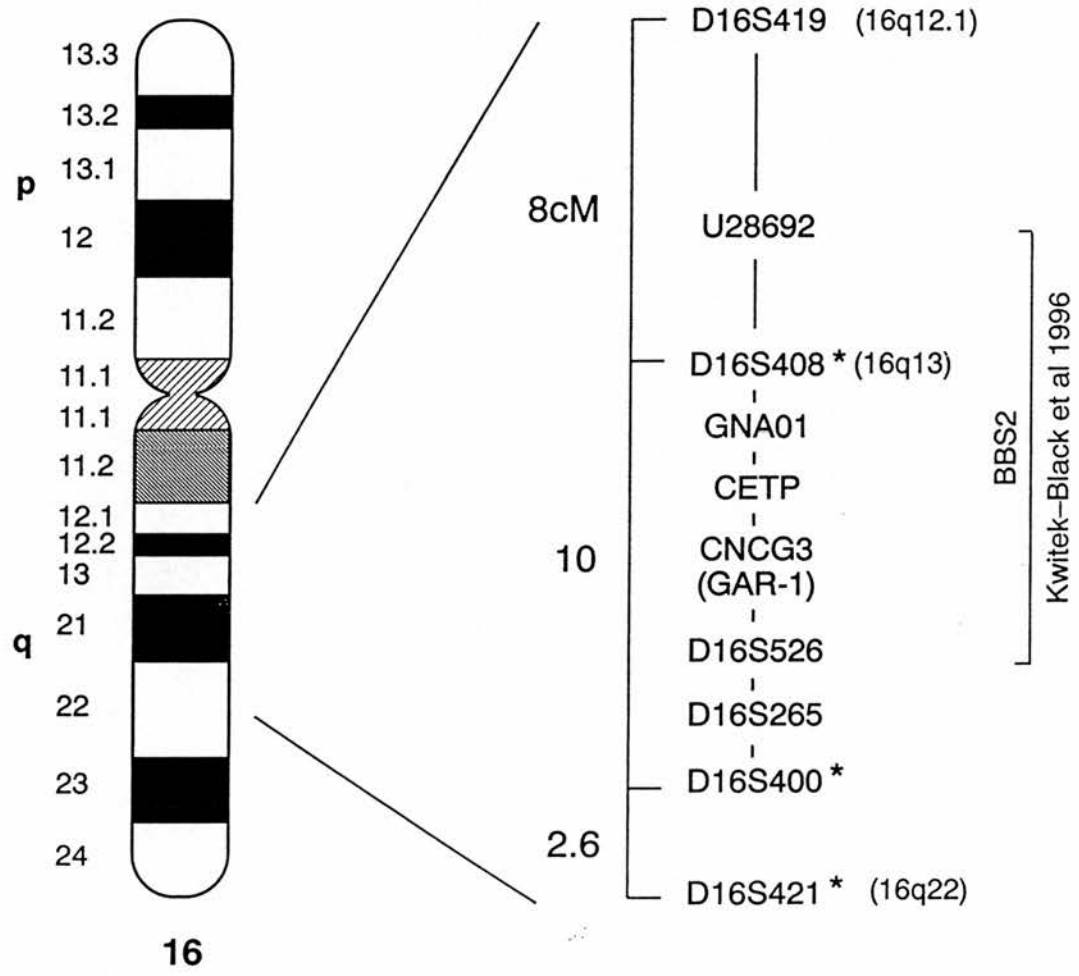


Figure 4.4 - Location of linked genetic markers in chromosomal region 16q12.1-q22

Recombination distances are shown in centiMorgans. Markers used in this study are marked with an asterisk.

4.2.1 - Family 1 - British (see **fig. 4.5**)

Clinical features: RP, obesity, polydactyly, mental retardation (MR), very small kidneys

Chromosome	11	15	16
max. multipoint lod score	0.53		0.28
chromosomal location, cM	63-69		72-76
conditional probability of linkage	0.75	0.04	0.13

4.2.2 - Family 2 - British (see **fig. 4.6**)

Clinical features: RP, polydactyly, diabetes mellitus, hypogonadism

Chromosome	11	15	16
max. multipoint lod score		0.28	0.28
chromosomal location, cM		98-100	72
conditional probability of linkage	0.03	0.52	0.27

4.2.3 - Family 4 - British (see **fig. 4.7**)

Clinical features: RP, borderline MR, hypogenitalism, polydactyly, syn/brachydactyly, obesity

Chromosome	11	15	16
max. multipoint lod score		0.04	0.40
chromosomal location, cM		120-130	72-76
conditional probability of linkage	0.03	0.28	0.50

4.2.4 - Family 5 - British (see **fig. 4.8**)

Clinical features: RP, polydactyly of both hands, mild syn/brachydactyly, renal anomalies, obesity, mild MR, genito-urinary abnormalities

Chromosome	11	15	16
max. multipoint lod score	0.73	0.71	0.62
chromosomal location, cM	69	88	76-80
conditional probability of linkage	0.51	0.34	0.12

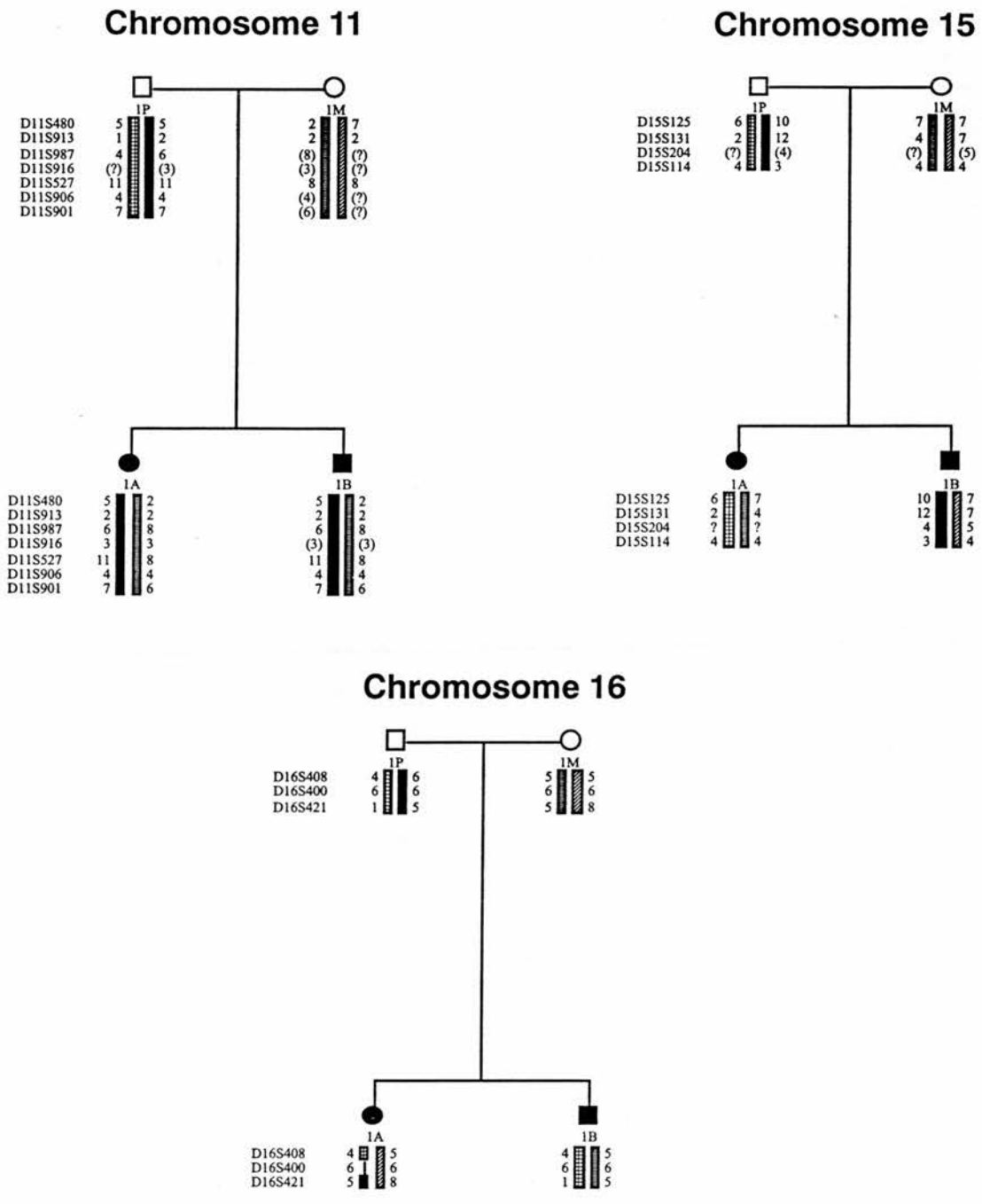


Figure 4.5

Haplotype analysis for family BB1, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

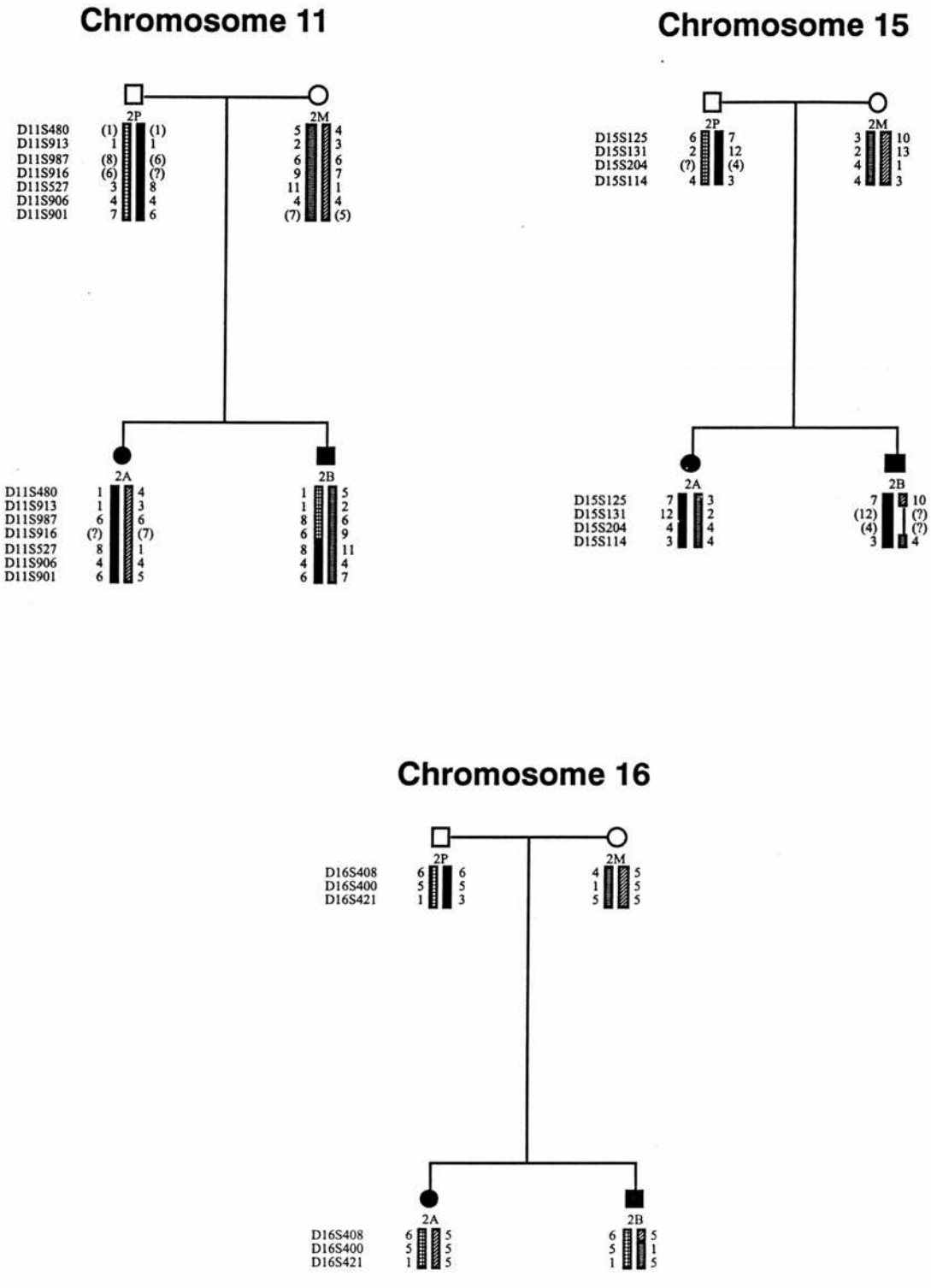


Figure 4.6

Haplotype analysis for family BB2, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

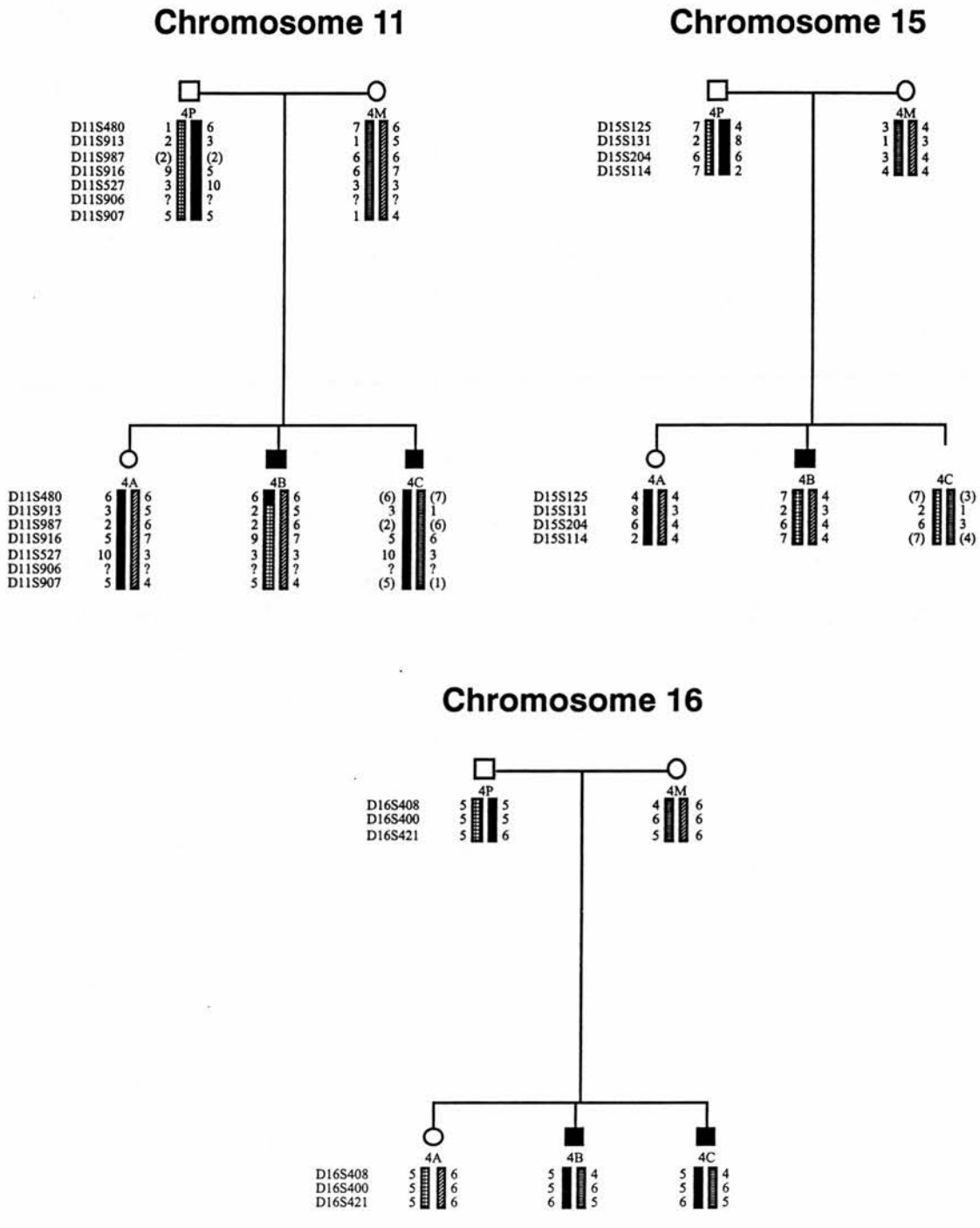


Figure 4.7

Haplotype analysis for family BB4, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

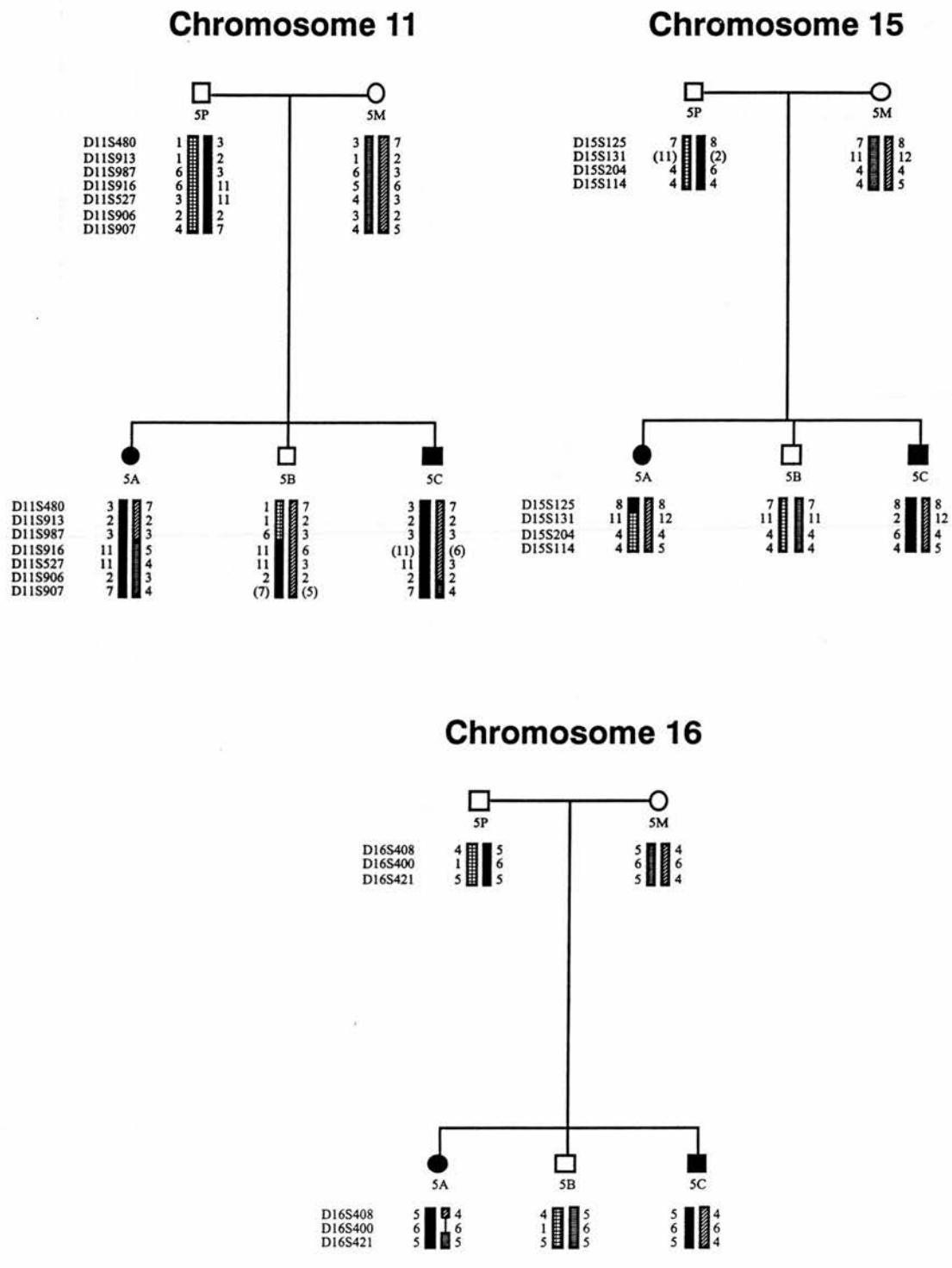


Figure 4.8

Haplotype analysis for family BB5, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

4.2.5 - Family 6 - British (see fig. **4.9**)

Clinical features: cone-rod dystrophy, moderate MR, obesity, polydactyly, brachydactyly

Chromosome	11	15	16
max. multipoint lod score	1.80		
chromosomal location, cM	72, 84		
conditional probability of linkage	0.97	0.00	0.01

4.2.6 - Family 7 - Asian - Consanguineous (Parents are first cousins) (see fig. **4.10**)

Clinical features: polydactyly of hands and feet, obesity, RP, mild MR

Chromosome	11	15	16
max. multipoint lod score		1.70	0.37
chromosomal location, cM		90-92	88
conditional probability of linkage	0.00	0.98	0.01

4.2.7 - Family 8 - Brazilian (see fig. **4.11**)

Clinical features: RP, MR, polydactyly, obesity

Chromosome	11	15	16
max. multipoint lod score	0.71		
chromosomal location, cM	81		
conditional probability of linkage	0.55	0.06	0.09

4.2.8 - Family 9 - British (see fig. **4.12**)

Clinical features: RP, obesity, menstrual problems

Chromosome	11	15	16
max. multipoint lod score	0.42	0.10	0.72
chromosomal location, cM	60	50	80
conditional probability of linkage	0.38	0.14	0.40

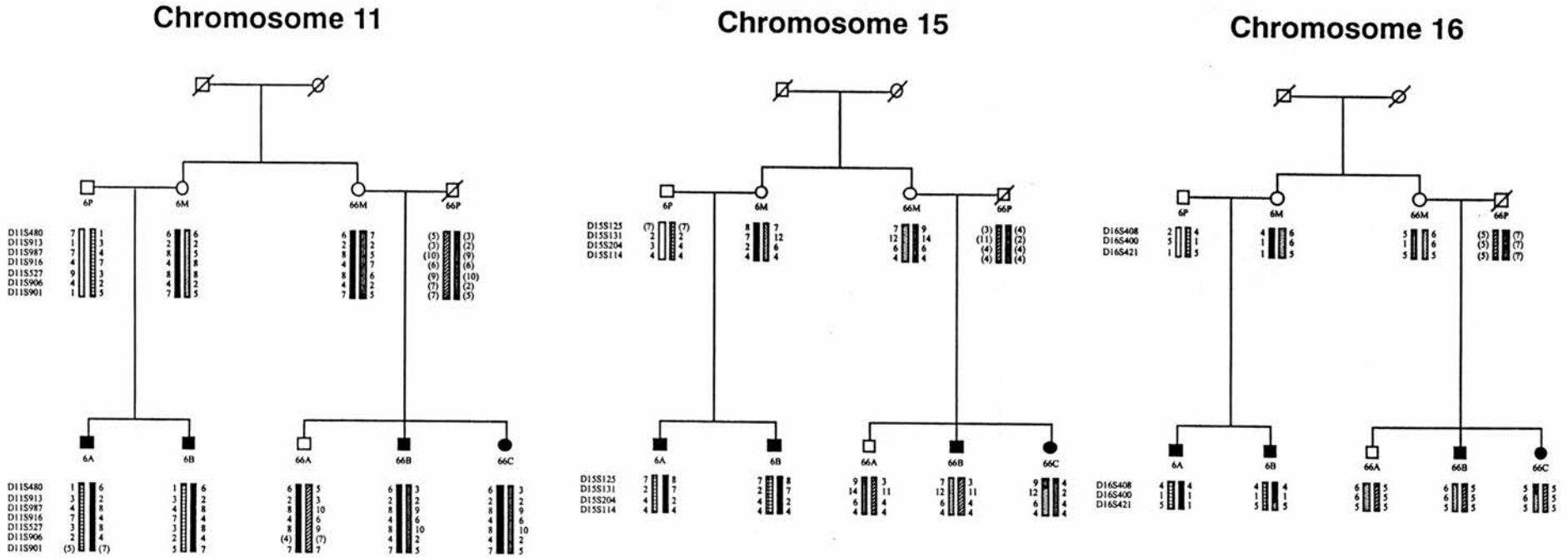


Figure 4.9

Haplotype analysis for family BB6, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

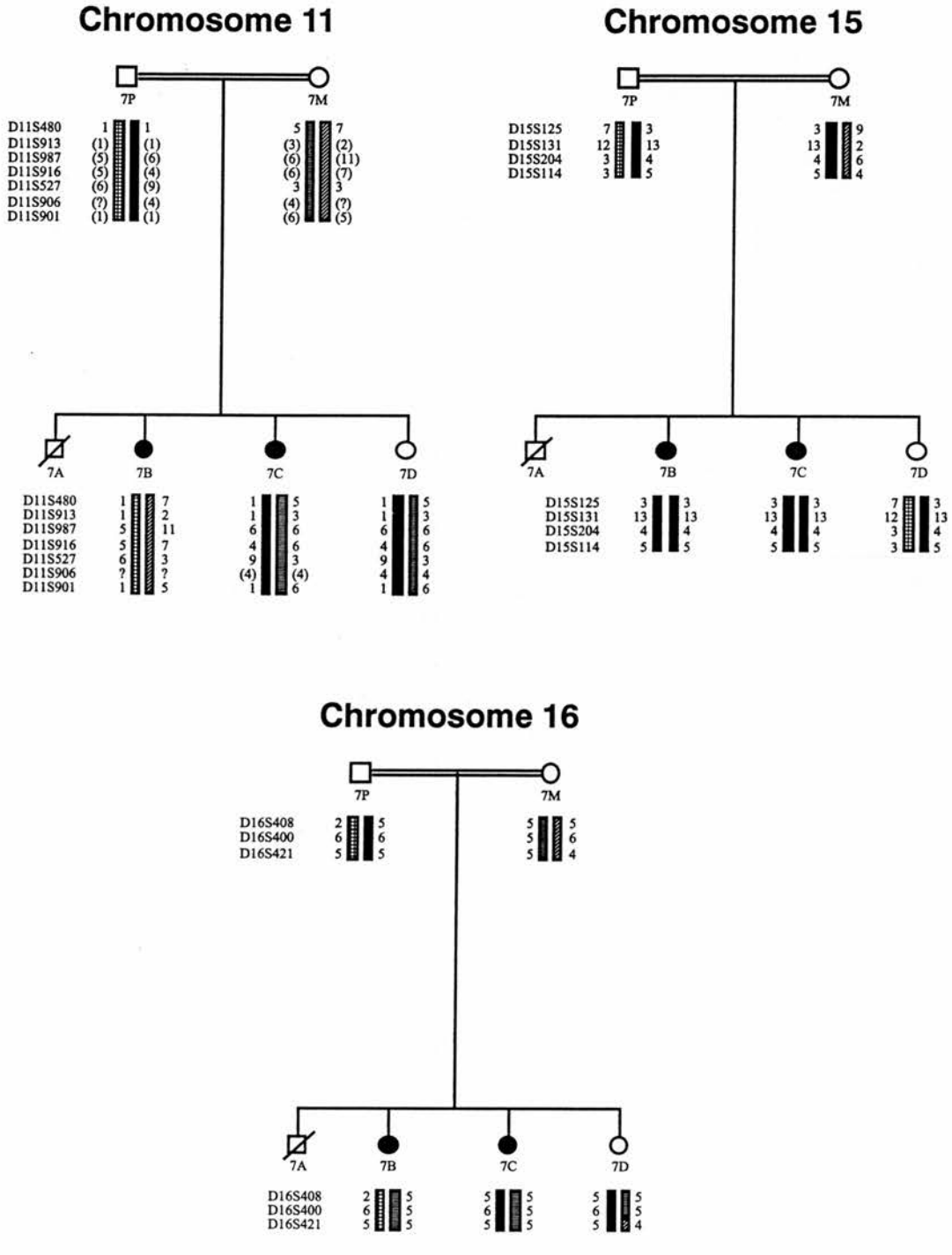


Figure 4.10

Haplotype analysis for family BB7, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

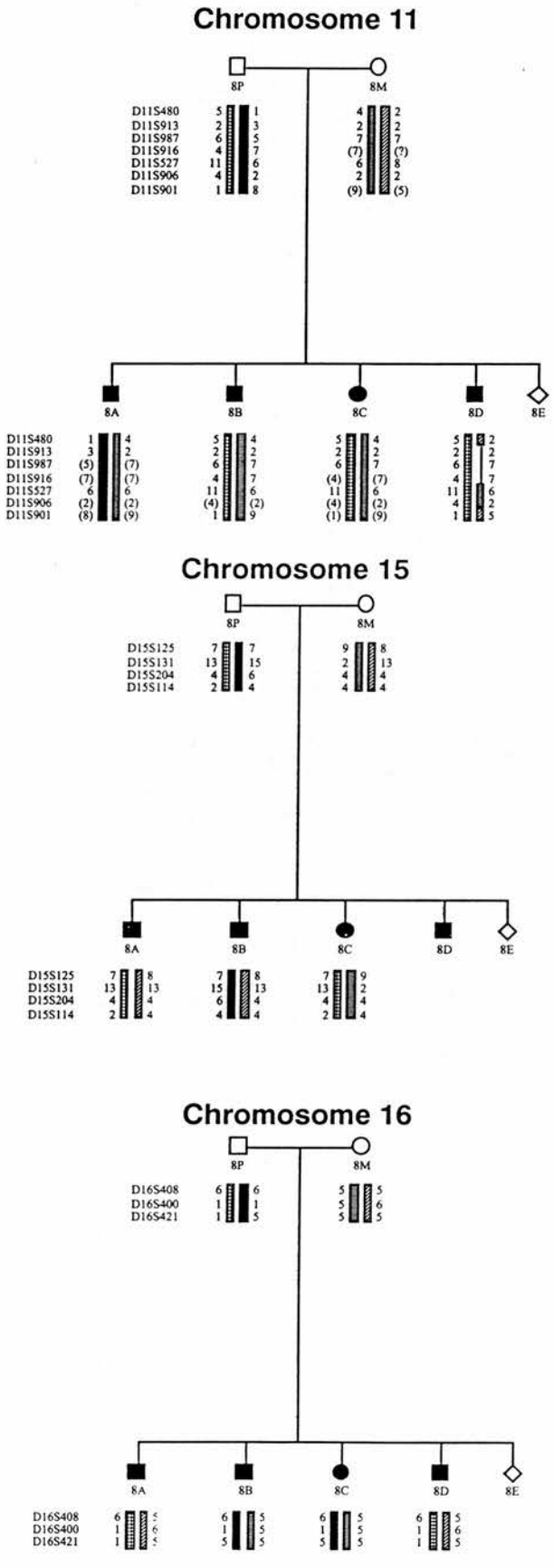


Figure 4.11

Haplotype analysis for family BB8, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

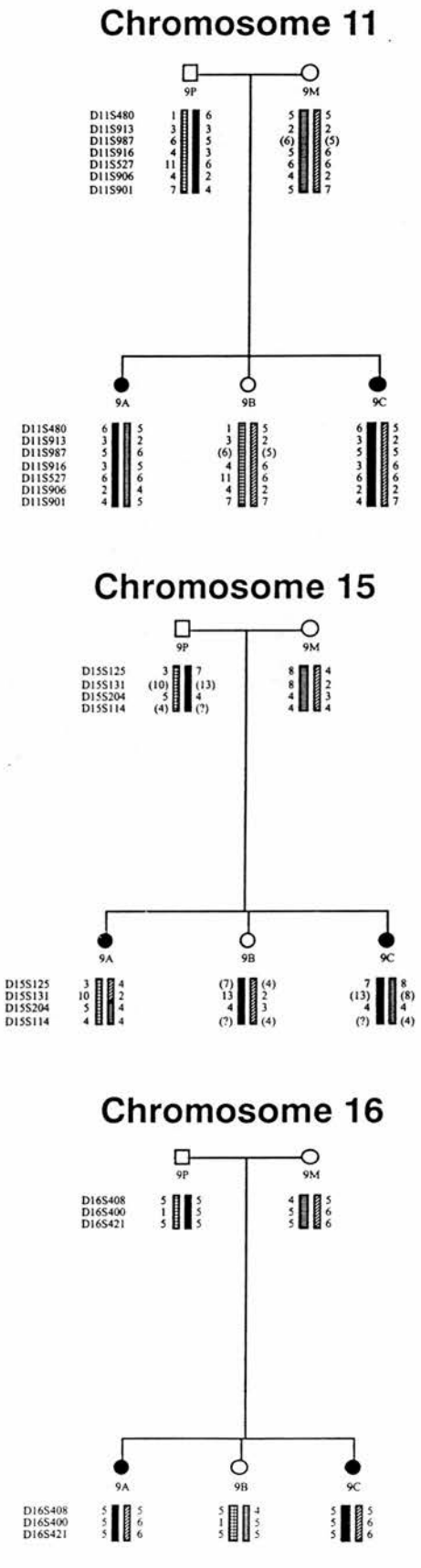


Figure 4.12 Haplotype analysis for family BB9, for loci *BBS1, BBS2* and *BBS4.* Inferred alleles are in brackets, question marks represent unknown results.

4.2.9 - Family 10 - Swiss (see **fig. 4.13**)

Clinical features: RP, MR, obesity, polydactyly

Chromosome	11	15	16
max. multipoint lod score	0.57	0.29	0.53
chromosomal location, cM	87	98-100	80
conditional probability of linkage	0.23	0.35	0.32

4.2.10 - Family 11 -Swiss (see **fig. 4.14**)

Chromosome	11	15	16
max. multipoint lod score	0.60	0.60	0.57
chromosomal location, cM	63-69	90	72-76
conditional probability of linkage	0.34	0.49	0.14

4.2.11 - Family 12 - Norwegian (see **fig. 4.15**)

Clinical features - RP, obesity, polydactyly, brachydactyly, dental anomalies, liver disease, hypogenitalism

Chromosome	11	15	16
max. multipoint lod score	0.60	0.58	0.30
chromosomal location, cM	63-69	98	76-84
conditional probability of linkage	0.54	0.26	0.14

4.2.12 - Family 13 -Norwegian - Consanguineous (parents are second cousins)(see **fig. 4.16**)

Clinical features - RP, polydactyly, obesity

Chromosome	11	15	16
max. multipoint lod score	1.30	0.46	0.44
chromosomal location, cM	69	100	76
conditional probability of linkage	0.72	0.14	0.10

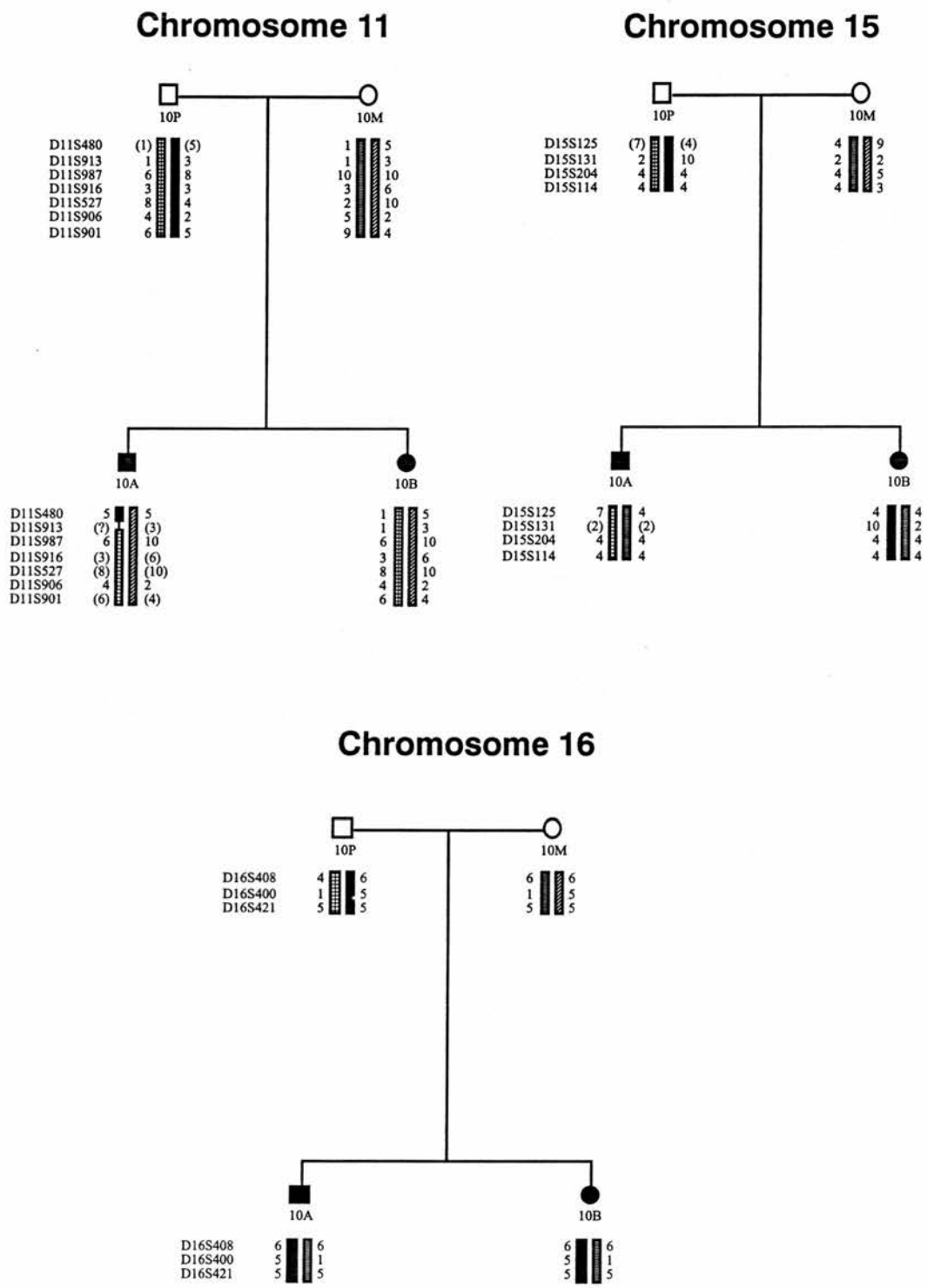


Figure 4.13

Haplotype analysis for family BB10, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

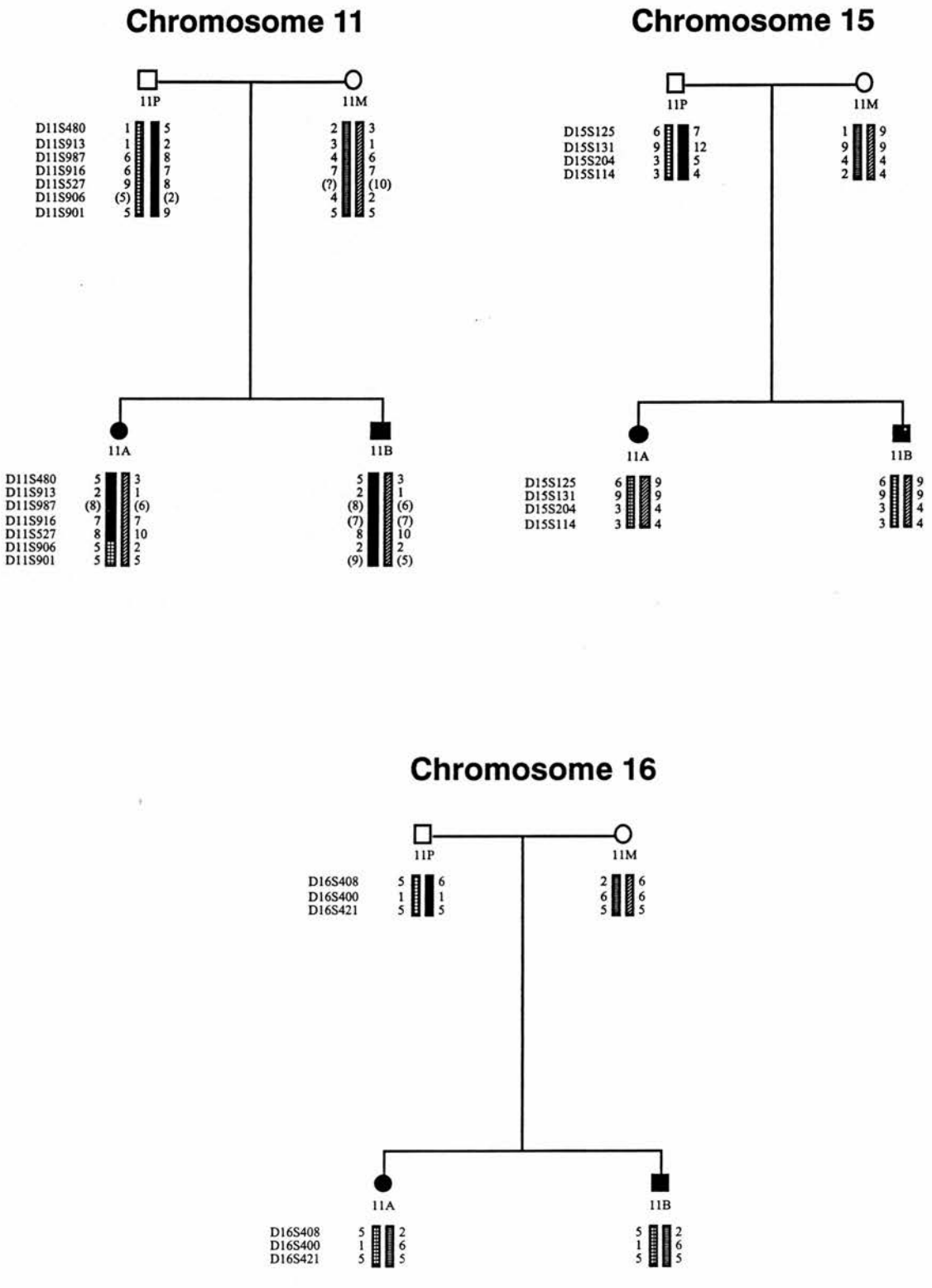


Figure 4.14

Haplotype analysis for family BB11, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

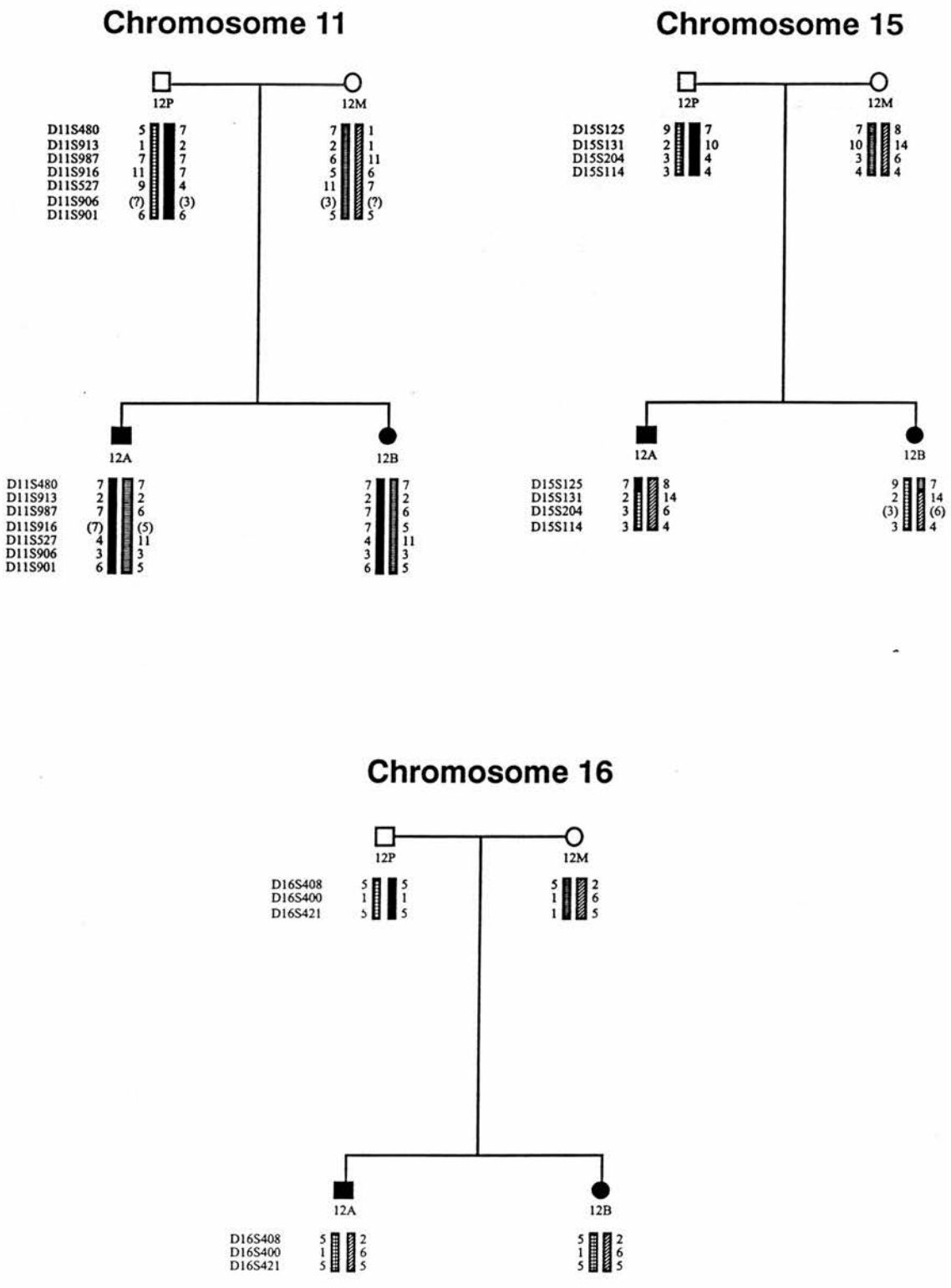


Figure 4.15

Haplotype analysis for family BB12, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

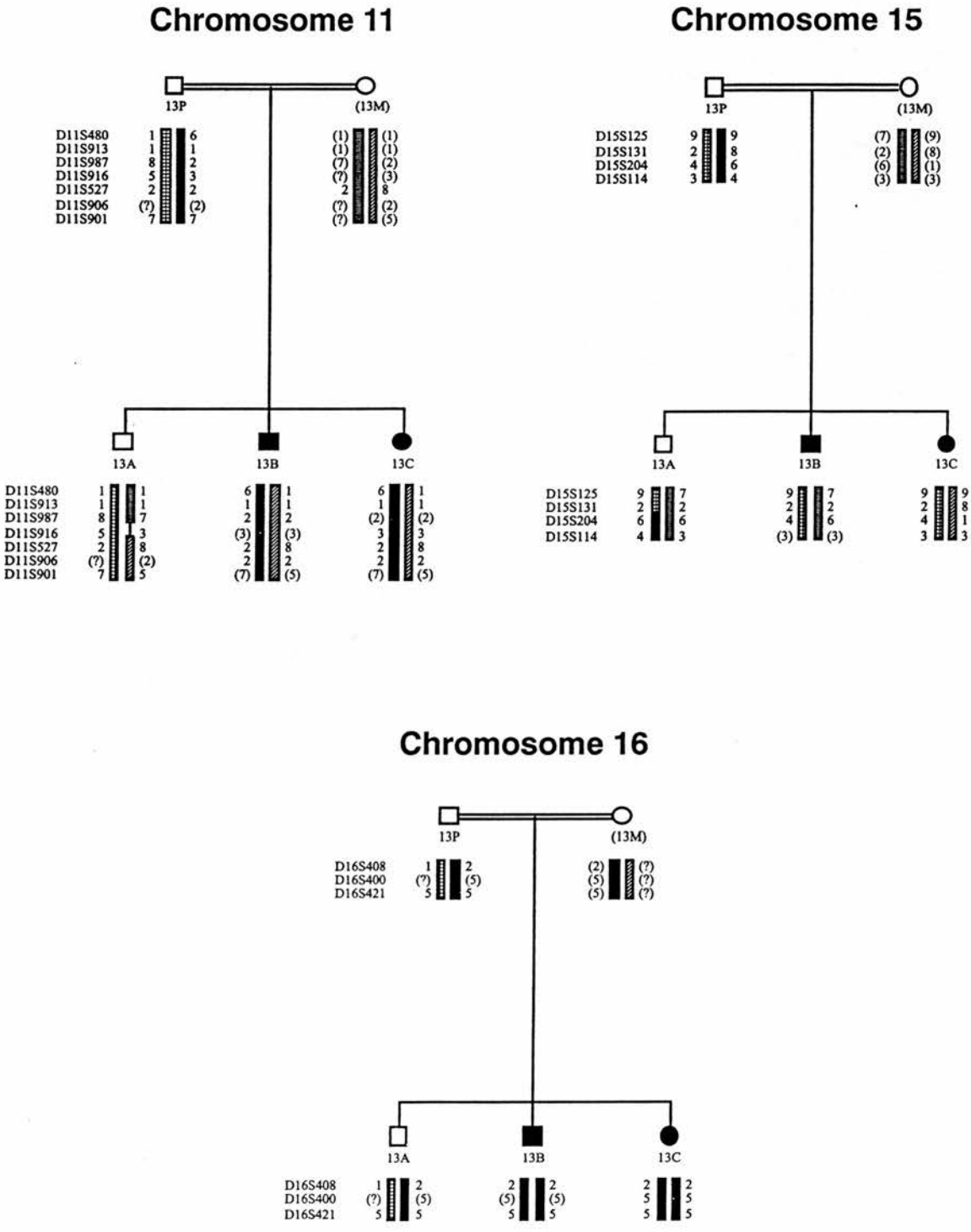


Figure 4.16

Haplotype analysis for family BB13, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

4.2.13 - Family 14 - Norwegian (see fig. **4.17**)

Clinical features - RP, obesity, polydactyly, brachydactyly, dental anomalies, hypogenitalism

Chromosome	11	15	16
max. multipoint lod score	0.71	0.72	
chromosomal location, cM	60-63	88-90	
conditional probability of linkage	0.41	0.54	0.01

4.2.14 - Family 15 - British (see **fig. 4.18**)

Chromosome	11	15	16
max. multipoint lod score	0.59	0.60	
chromosomal location, cM	81	96-98	
conditional probability of linkage	0.51	0.41	0.03

4.2.15 - Family 17 - British (see **fig. 4.19**)

Clinical features: RP, polydactyly, obesity, mild MR, renal dysplasia, hypertension, syndactyly

Chromosome	11	15	16
max. multipoint lod score	1.20		
chromosomal location, cM	69		
conditional probability of linkage	0.87	0.01	0.04

4.2.16 - Family 18 - Norwegian - Consanguineous (see **fig. 4.20**)

Clinical features - RP, obesity, polydactyly, brachydactyly, hypogonadism, dental anomalies, hypertension

Chromosome	11	15	16
max. multipoint lod score	1.50	2.61	0.30
chromosomal location, cM	69	94	88
conditional probability of linkage	0.04	0.95	0.00

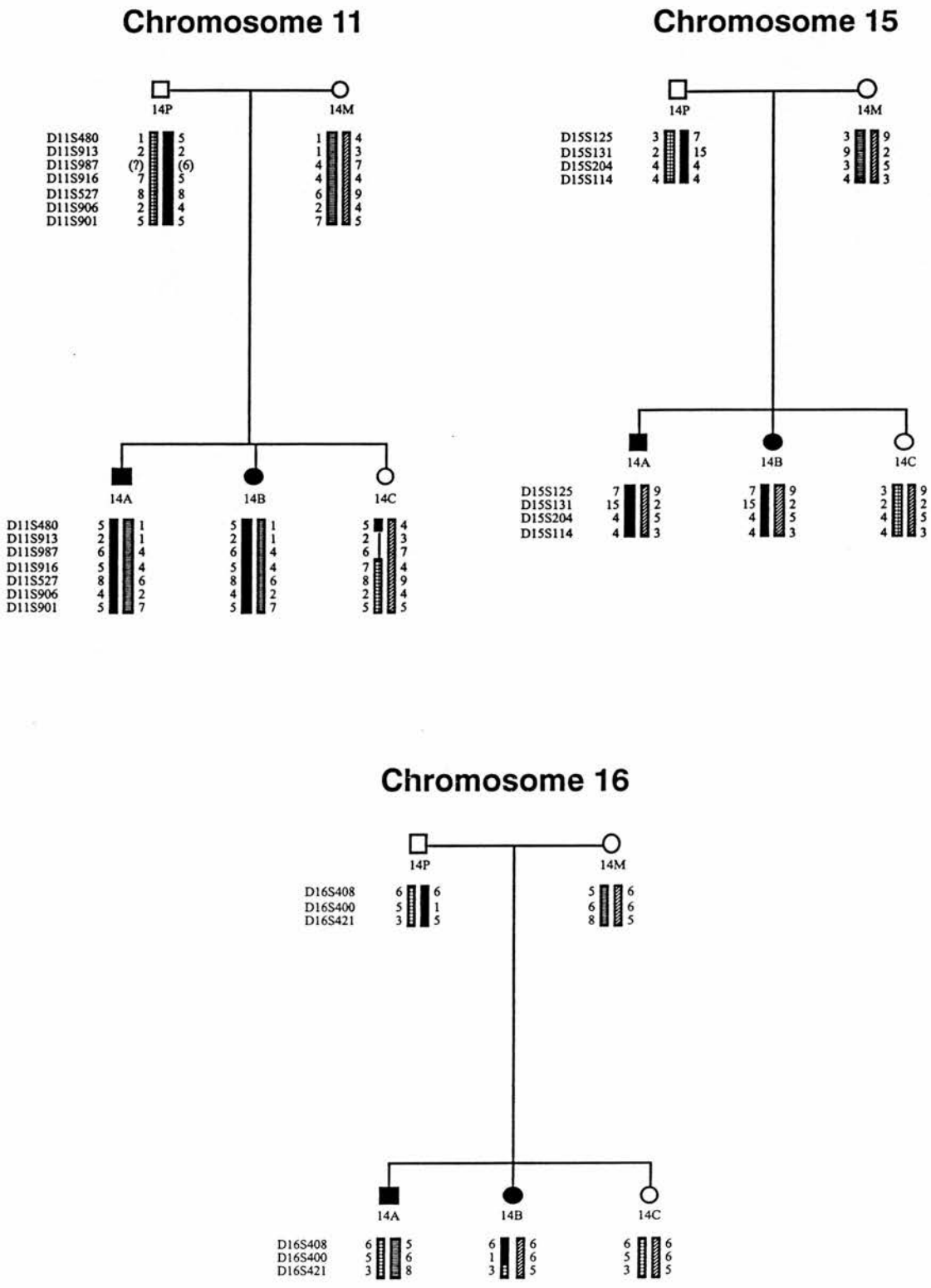


Figure 4.17

Haplotype analysis for family BB14, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

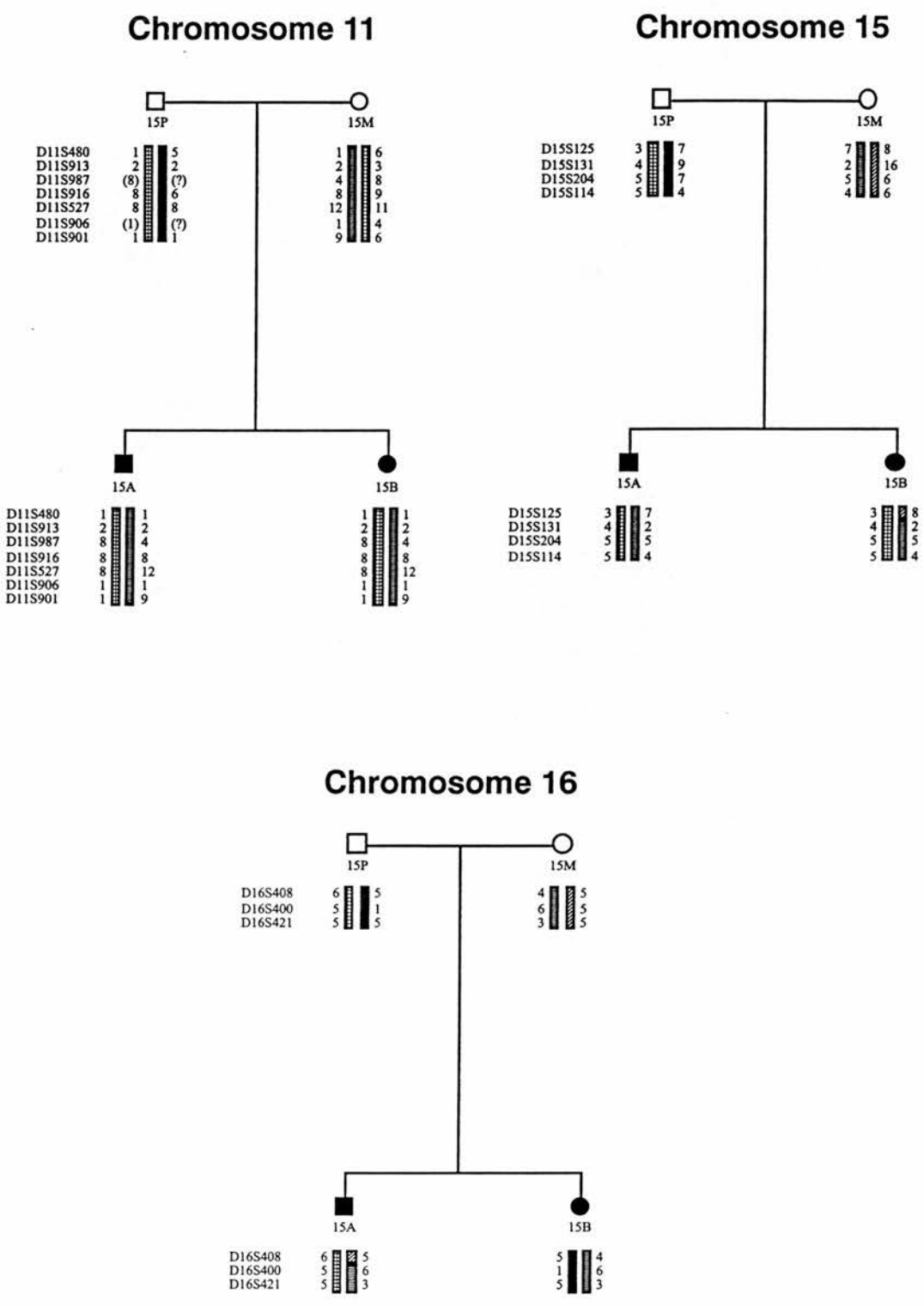


Figure 4.18

Haplotype analysis for family BB15, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

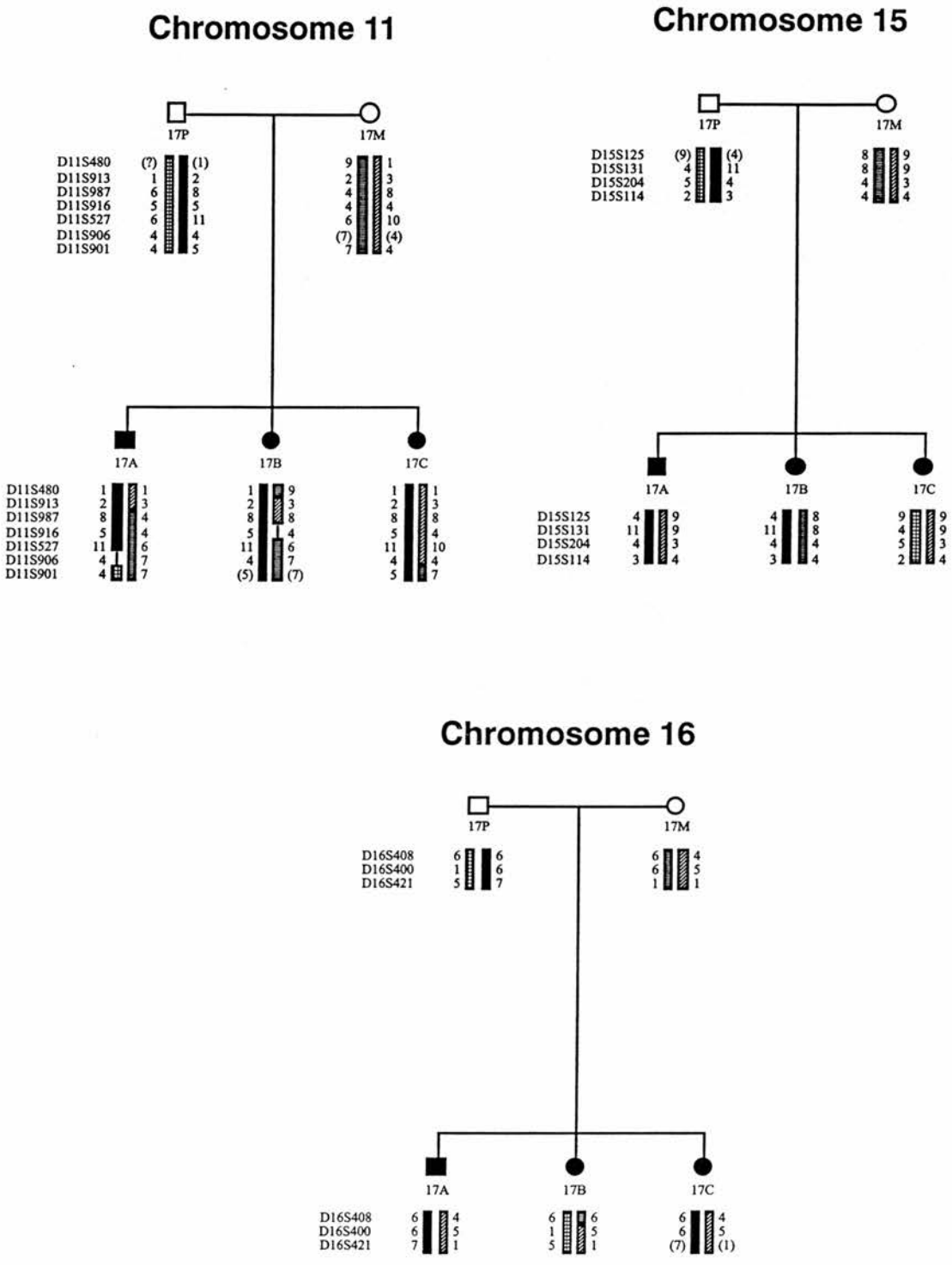


Figure 4.19

Haplotype analysis for family BB17, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

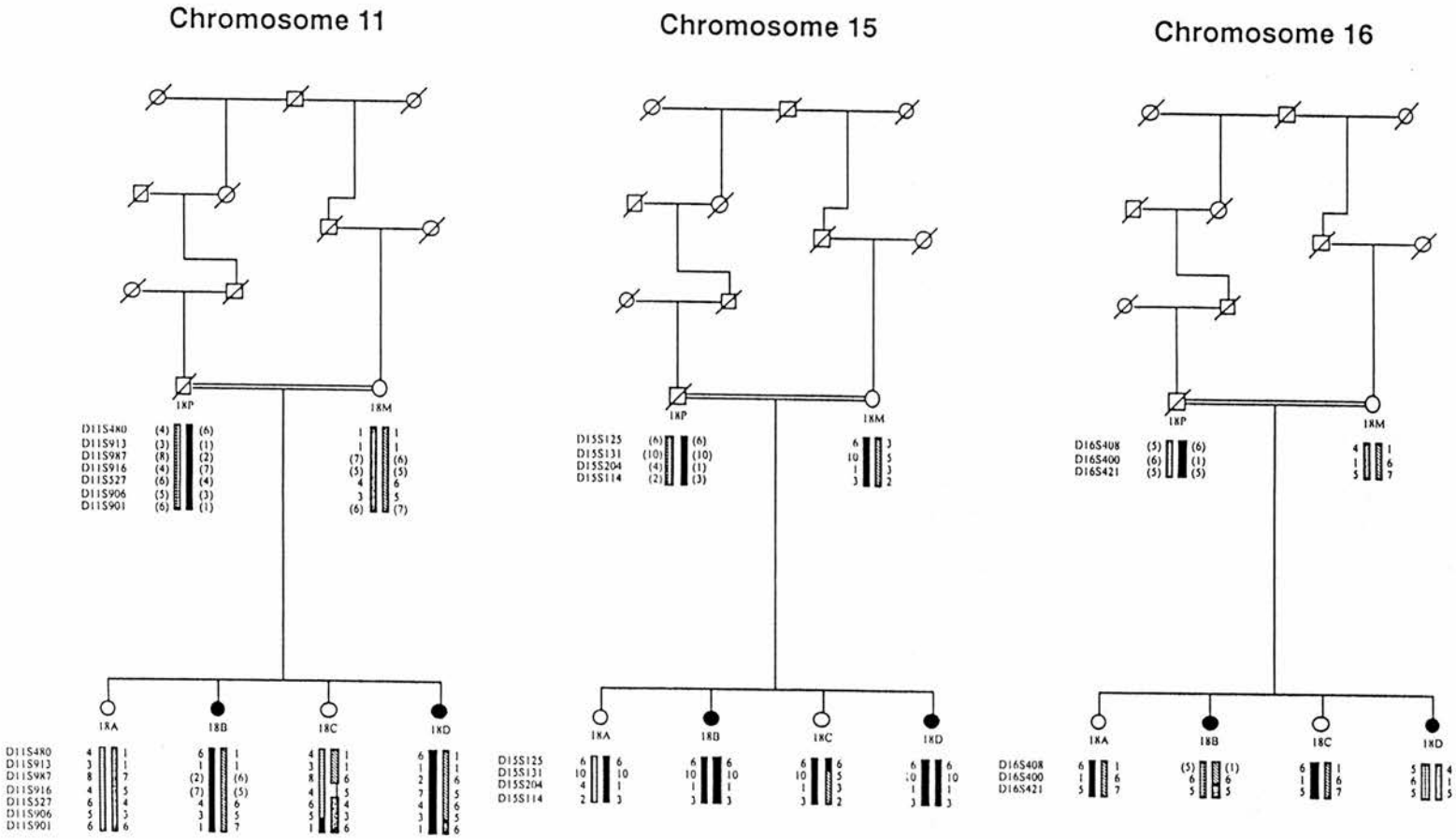


Figure 4.20

Haplotype analysis for famile BB18, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

4.2.17 - Family 19 - British (see **fig. 4.21**)

Clinical features: RP, obesity, polydactyly, hypogonadism, hypertension

Chromosome	11	15	16
max. multipoint lod score	0.69	0.69	
chromosomal location, cM	81	98	
conditional probability of linkage	0.29	0.64	0.03

4.2.18 - Family 20 - Egyptian - Consanguineous (see **fig 4.22**)

Clinical features - RP, polydactyly (right hand), brachydactyly, syndactyly, mild obesity, short stature

Chromosome	11	15	16
max. multipoint lod score	0.58		
chromosomal location, cM	84		
conditional probability of linkage	0.78	0.10	0.02

4.2.19 - Family 21 - British (see **fig 4.23**)

Clinical features: RP, polydactyly, mild MR, renal anomalies

Chromosome	11	15	16
max. multipoint lod score	0.60		
chromosomal location, cM	63-69		
conditional probability of linkage	0.85	0.02	0.05

4.2.20 - Family 22 - Swedish (see **fig 4.24**)

Clinical features - RP, obesity, polydactyly, brachydactyly, hypogonadism, dental anomalies

Chromosome	11	15	16
max. multipoint lod score			0.39
chromosomal location, cM			84
conditional probability of linkage	0.03	0.16	0.55

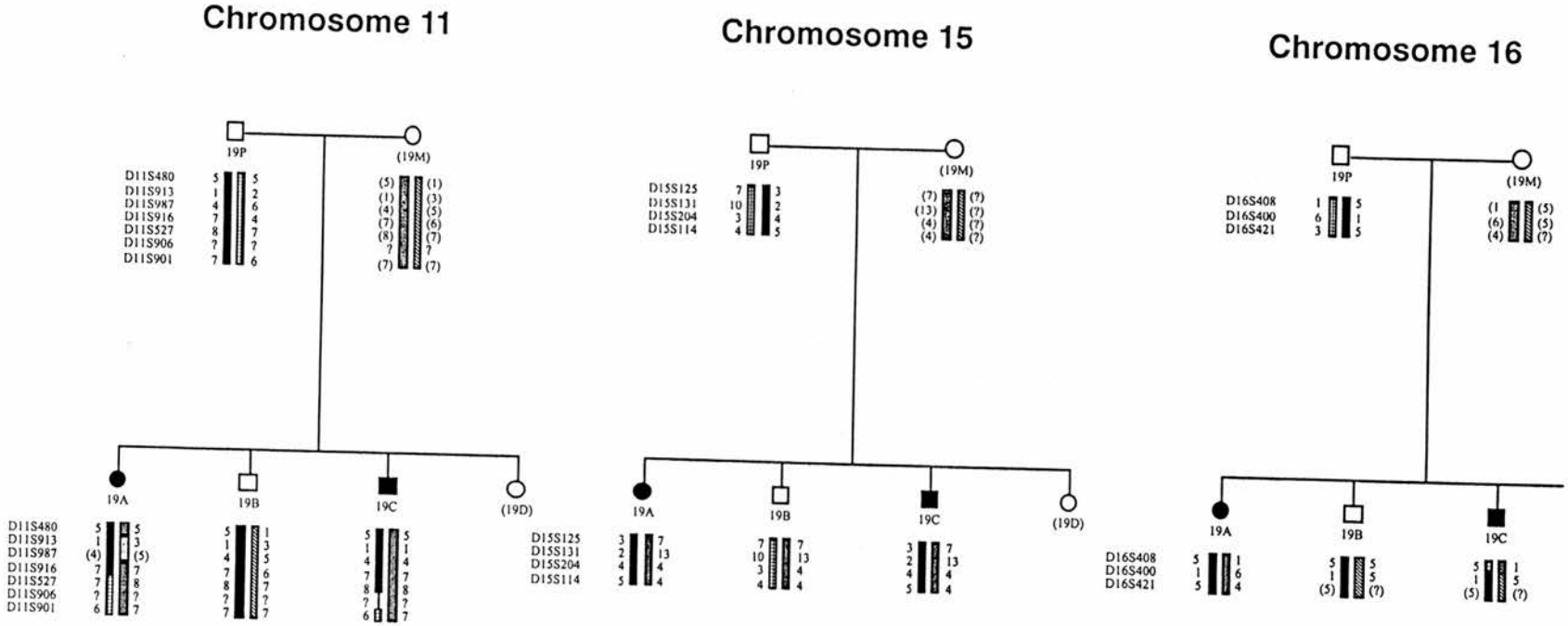


Figure 4.21

Haplotype analysis for famile BB19, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

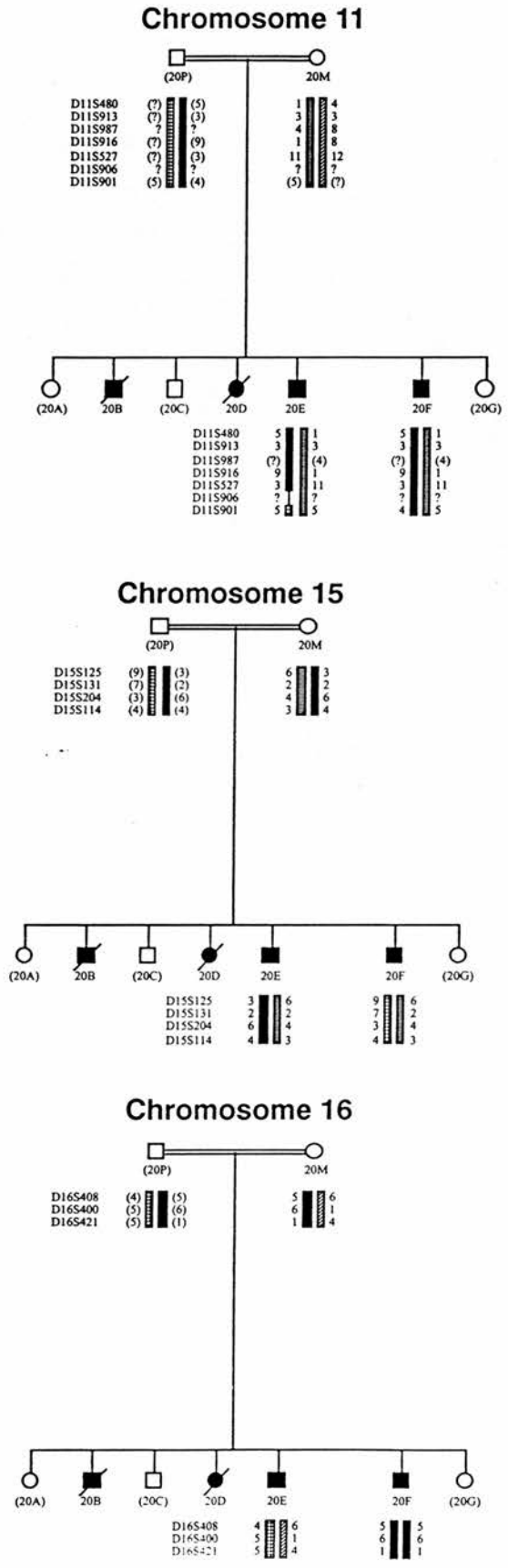


Figure 4.22

Haplotype analysis for family BB20, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

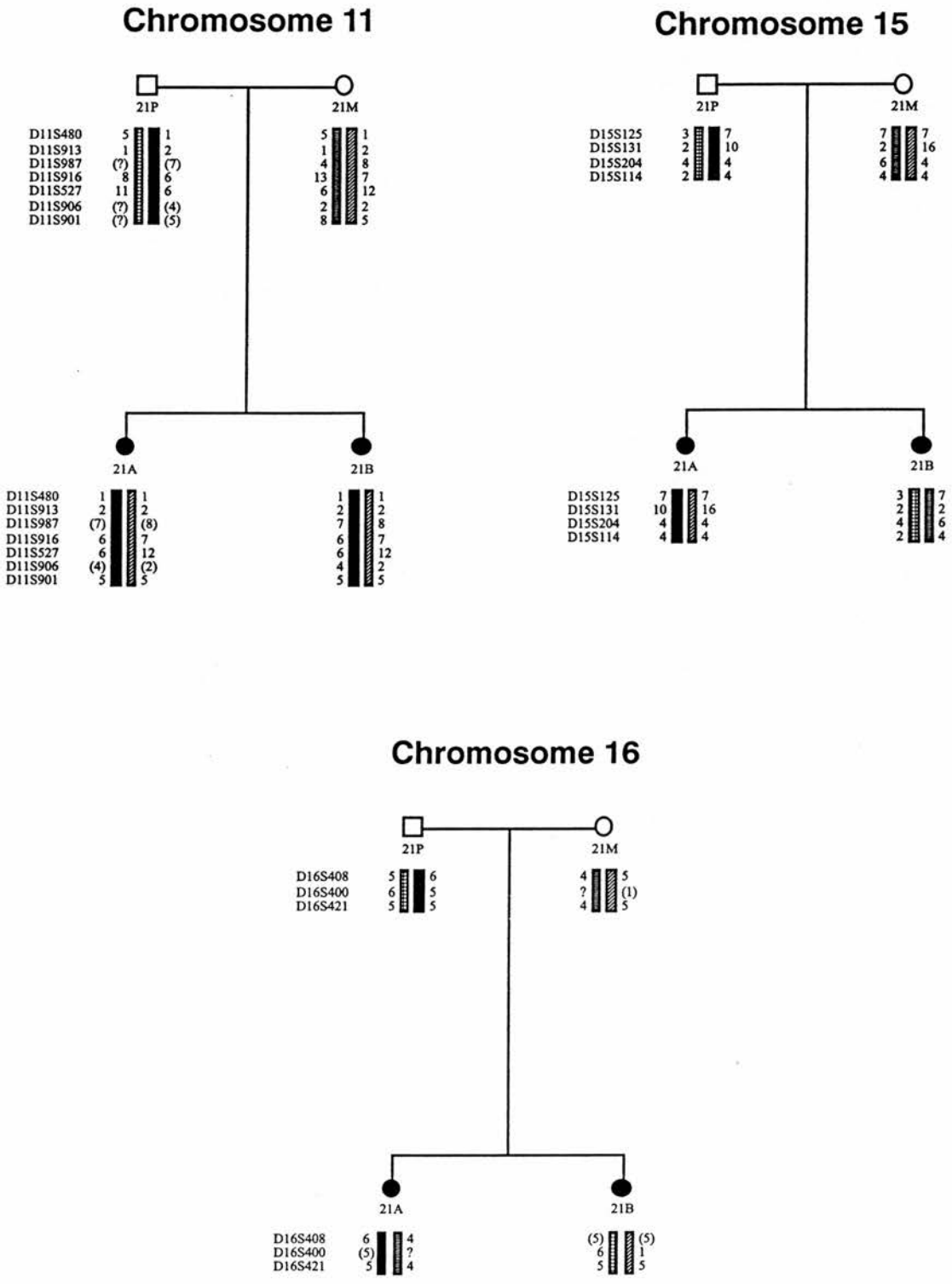


Figure 4.23

Haplotype analysis for family BB21, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

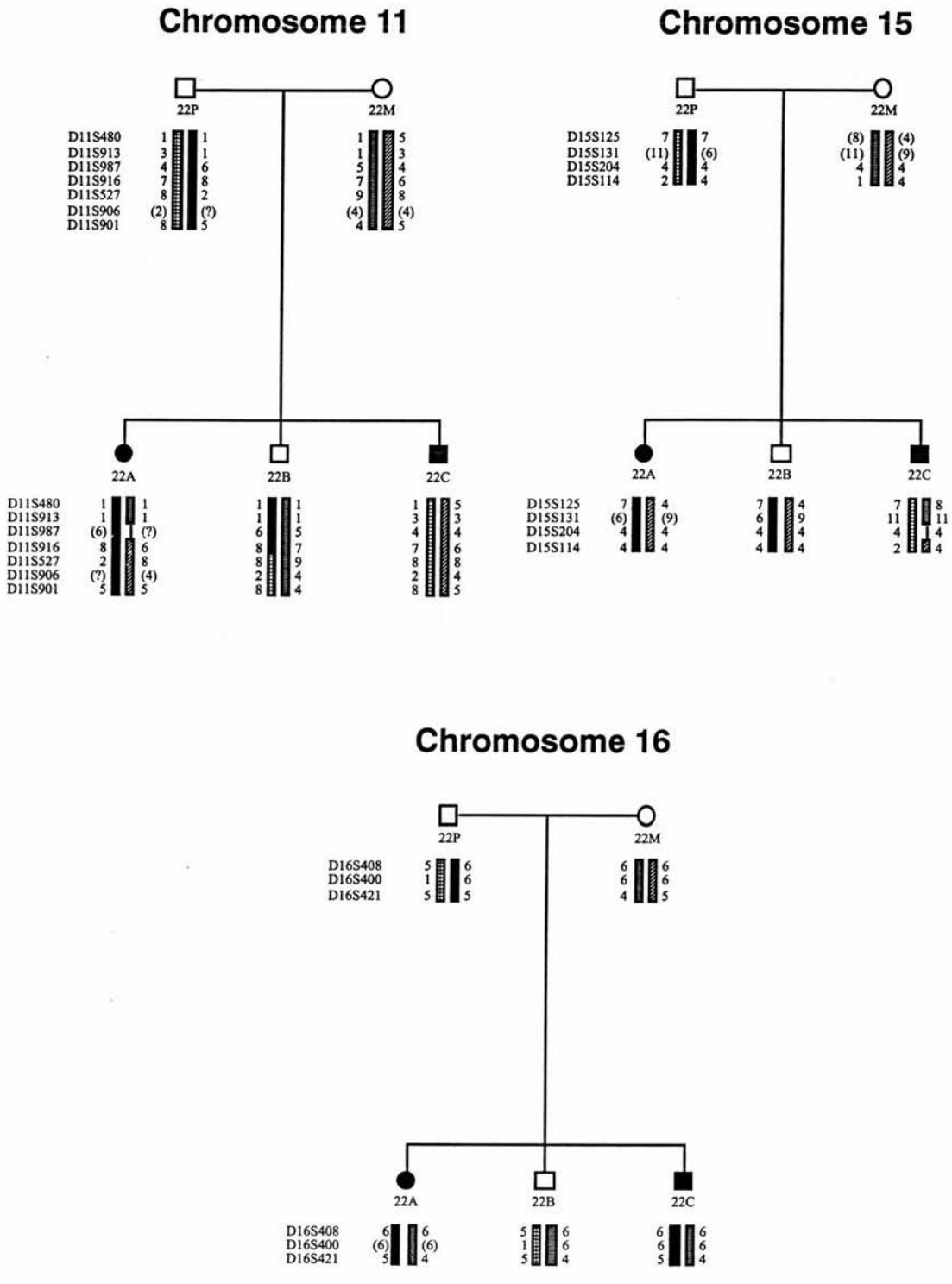


Figure 4.24

Haplotype analysis for family BB22, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

4.2.21 - Family 23 - Danish (see **fig. 4.25**)

Clinical features - RP, obesity, polydactyly, clinodactyly, brachdactyly, hypogonadism, mild MR, gout, hypodontia, liver and kidney disease, high blood pressure, diabetes mellitus

Chromosome	11	15	16
max. multipoint lod score	0.83	0.23	
chromosomal location, cM	60	78-80	
conditional probability of linkage	0.62	0.24	0.07

4.2.22 - Family 24 - Danish (see **fig. 4.26**)

Clinical features - RP, polydactyly (fingers and toes), brachydactyly, mild MR, dental anomalies, high blood pressure, obesity, irregular menstruation.

Chromosome	11	15	16
max. multipoint lod score	1.31		1.30
chromosomal location, cM	81-84		76
conditional probability of linkage	0.61	0.02	0.34

4.2.23 - Family 25 -Danish - Consanguineous (see **fig. 4.27** and †**section 4.5**)

Clinical features - RP, mild MR, polydactyly, brachydactyly, clinodactyly, hypogonadism, hypodontia, mild diabetes mellitus, obesity,

Chromosome	11	15	16
max. multipoint lod score	0.17	1.56†	0.29
chromosomal location, cM	42-45	88	84
conditional probability of linkage	0.01	0.93†	0.04

4.2.24 - Family 27 - Danish (see **fig. 4.28**)

Clinical features - RP, obesity, polydactyly, brachydactyly, MR, high blood pressure, reduced hearing, hypogonadism, dental anomalies

Chromosome	11	15	16
max. multipoint lod score	0.61	0.43	
chromosomal location, cM	72	94	
conditional probability of linkage	0.48	0.42	0.03

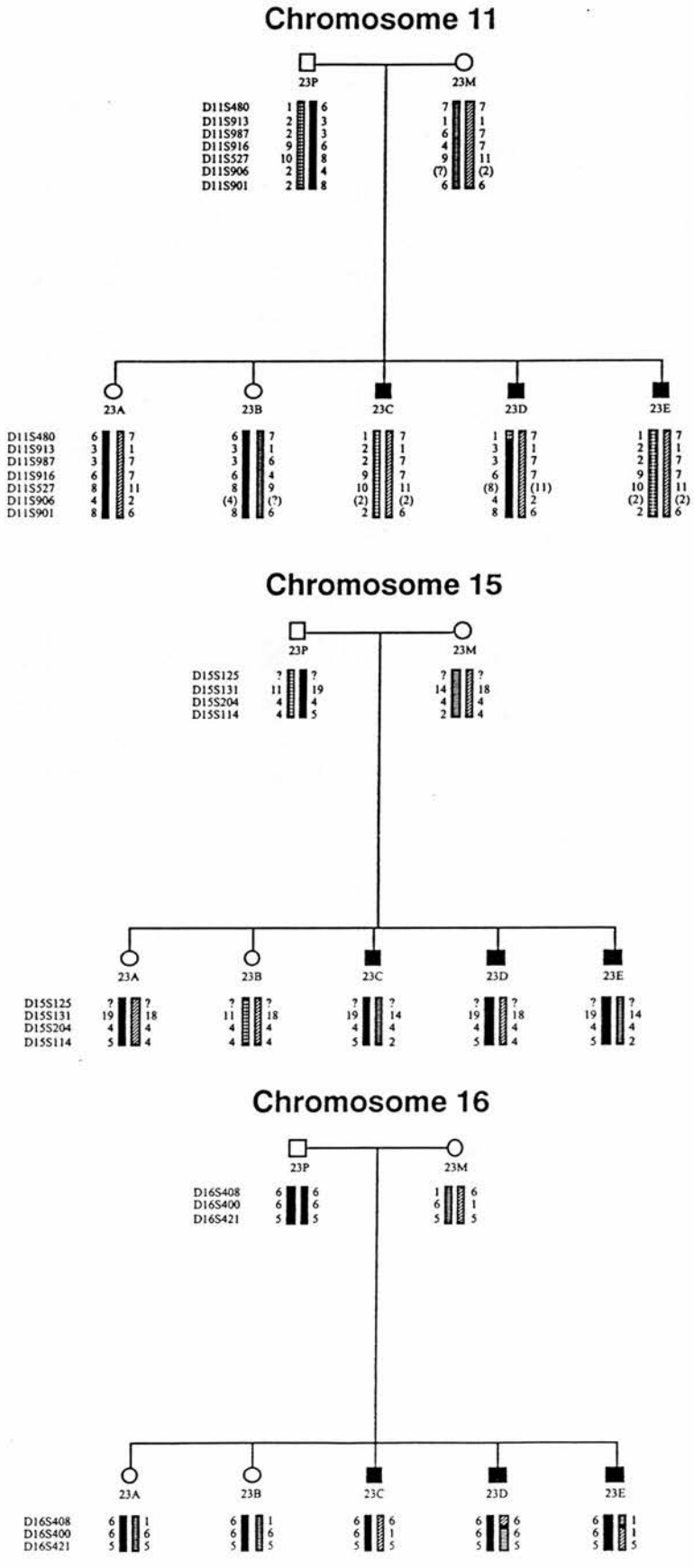


Figure 4.25

Haplotype analysis for family BB23, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

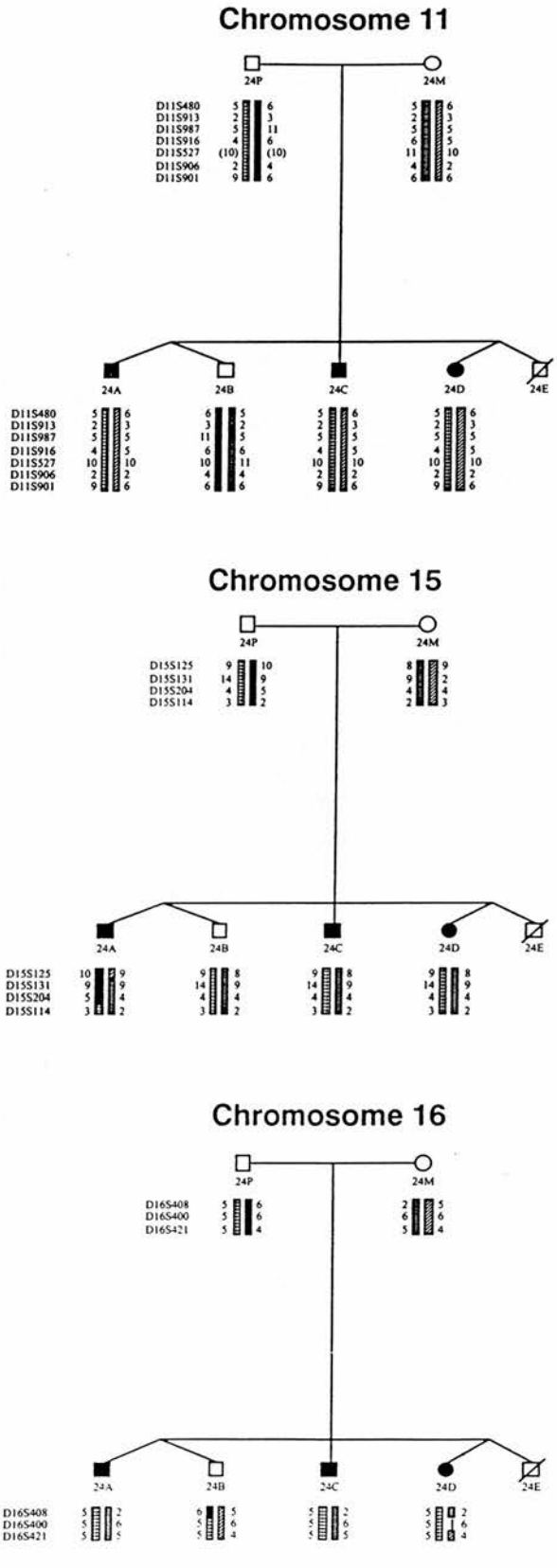


Figure 4.26

Haplotype analysis for family BB24, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

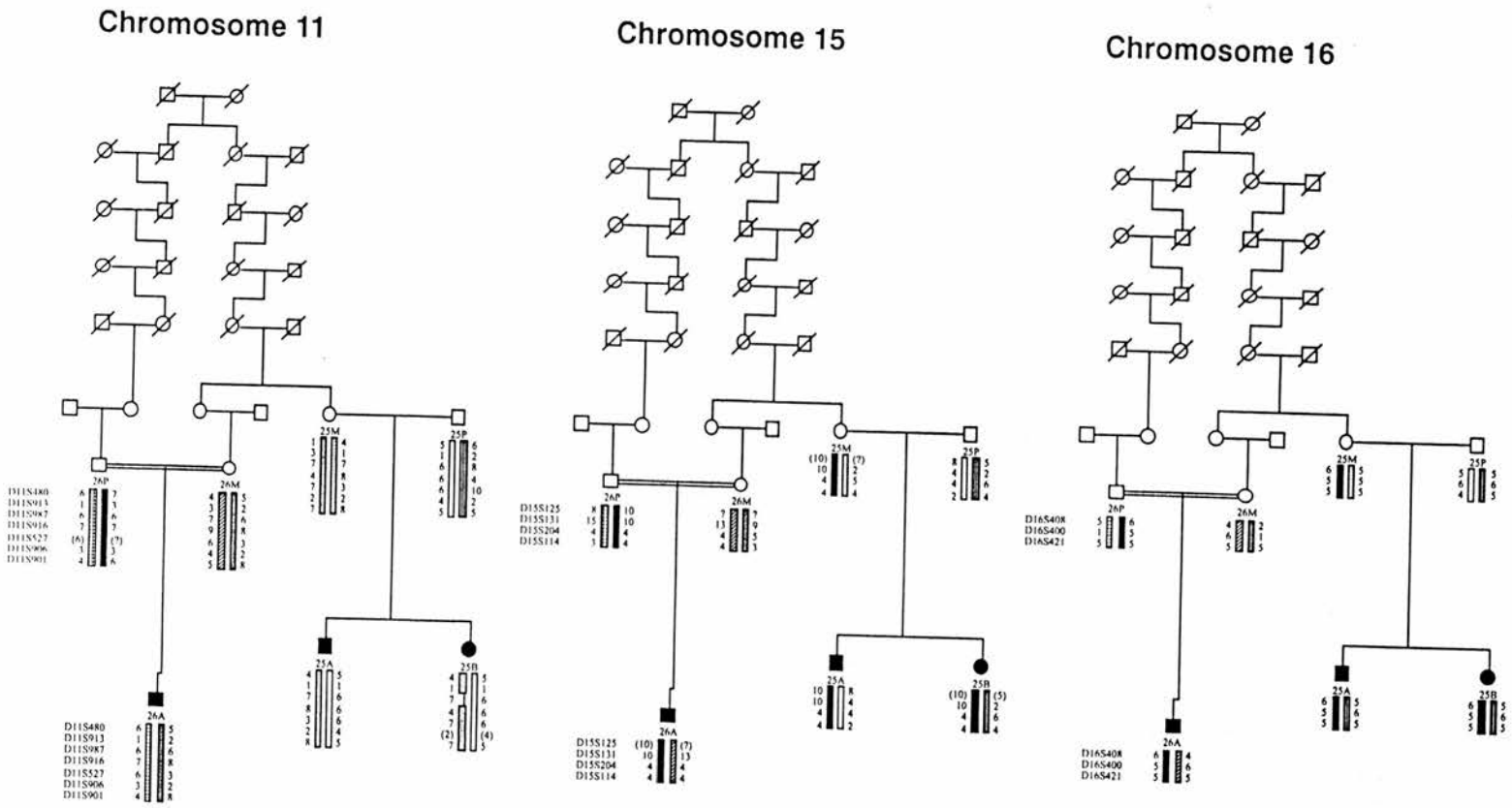


Figure 4.27

Haplotype analysis for famile BB25, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

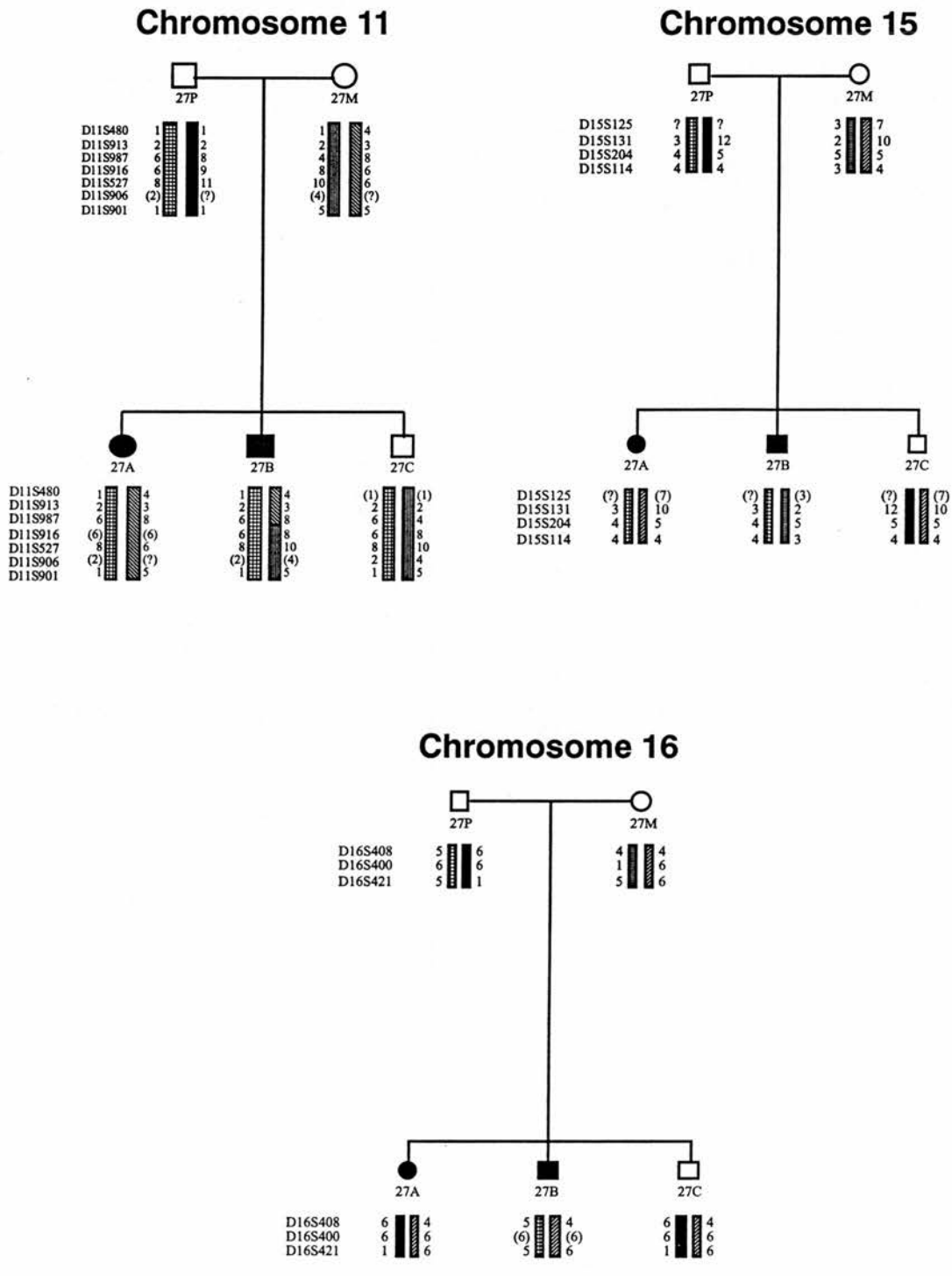


Figure 4.28

Haplotype analysis for family BB27, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

4.2.25 - Family 28 -Norwegian (see **fig. 4.29**)

Clinical features - RP, polydactyly, obesity, mild MR, hypogenitalism

Chromosome	11	15	16
max. multipoint lod score			0.60
chromosomal location, cM			84
conditional probability of linkage	0.08	0.25	0.48

4.2.26 - Family 29- Italian (see **fig. 4.30**)

Chromosome	11	15	16
max. multipoint lod score		1.19	
chromosomal location, cM		90	
conditional probability of linkage	0.01	0.97	0.01

4.2.27 - Family 30 - Sweden (see **fig. 4.31**)

Clinical features - RP, obesity, hypogenitalism, polydactyly, brachydactyly, renal cyst

Chromosome	11	15	16
max. multipoint lod score		0.72	
chromosomal location, cM		90	
conditional probability of linkage	0.05	0.83	0.07

4.2.28 - Family 31 - Norwegian - Consanguineous (see **fig. 4.32**)

Clinical features: RP, hypodontia, polydactyly, brachydactyly, moderate obesity, hypogonadism

Chromosome	11	15	16
max. multipoint lod score	1.82	3.00	
chromosomal location, cM	69	94	
conditional probability of linkage	0.06	0.94	0.00

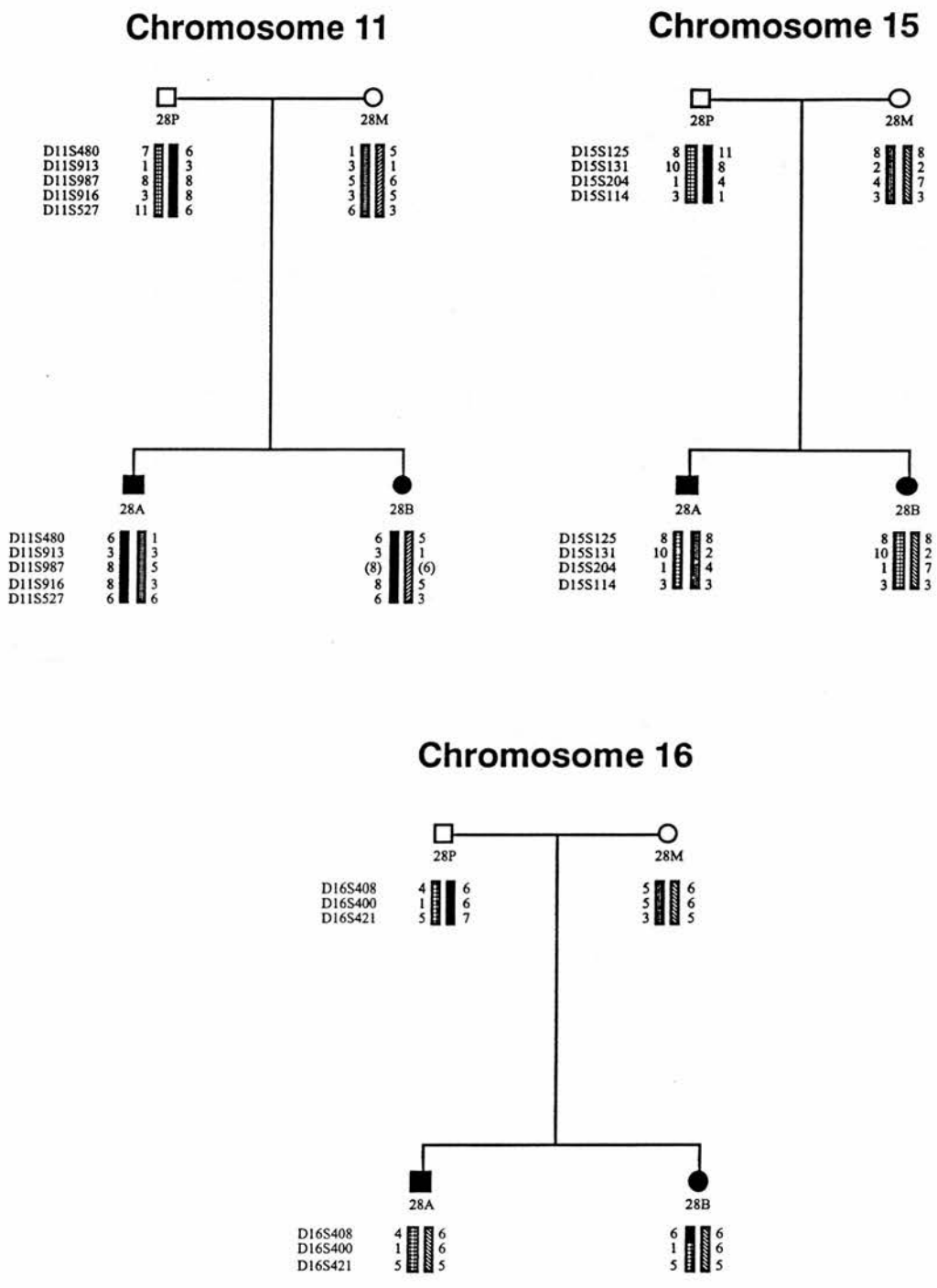


Figure 4.29

Haplotype analysis for family BB28, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

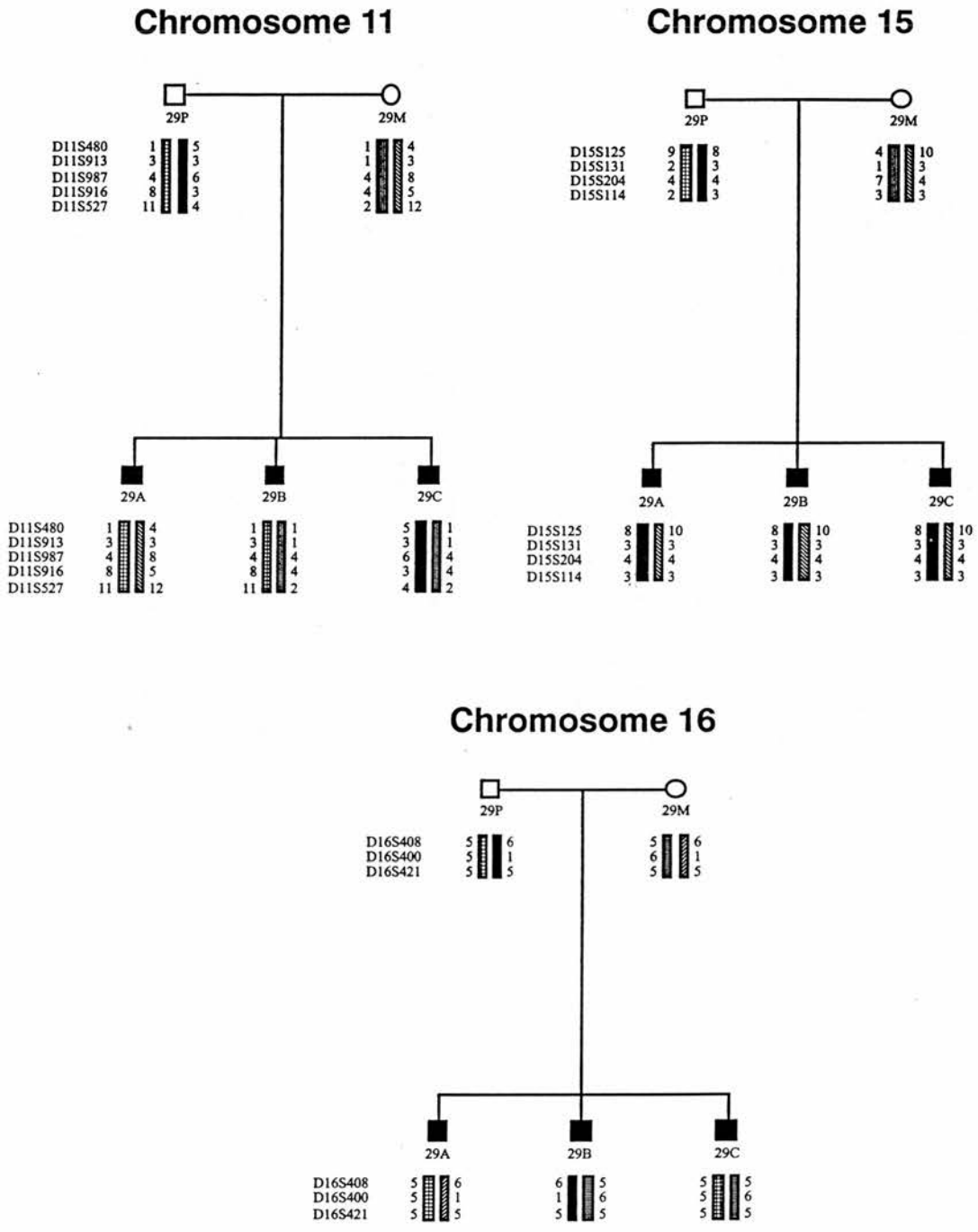


Figure 4.30

Haplotype analysis for family BB29, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

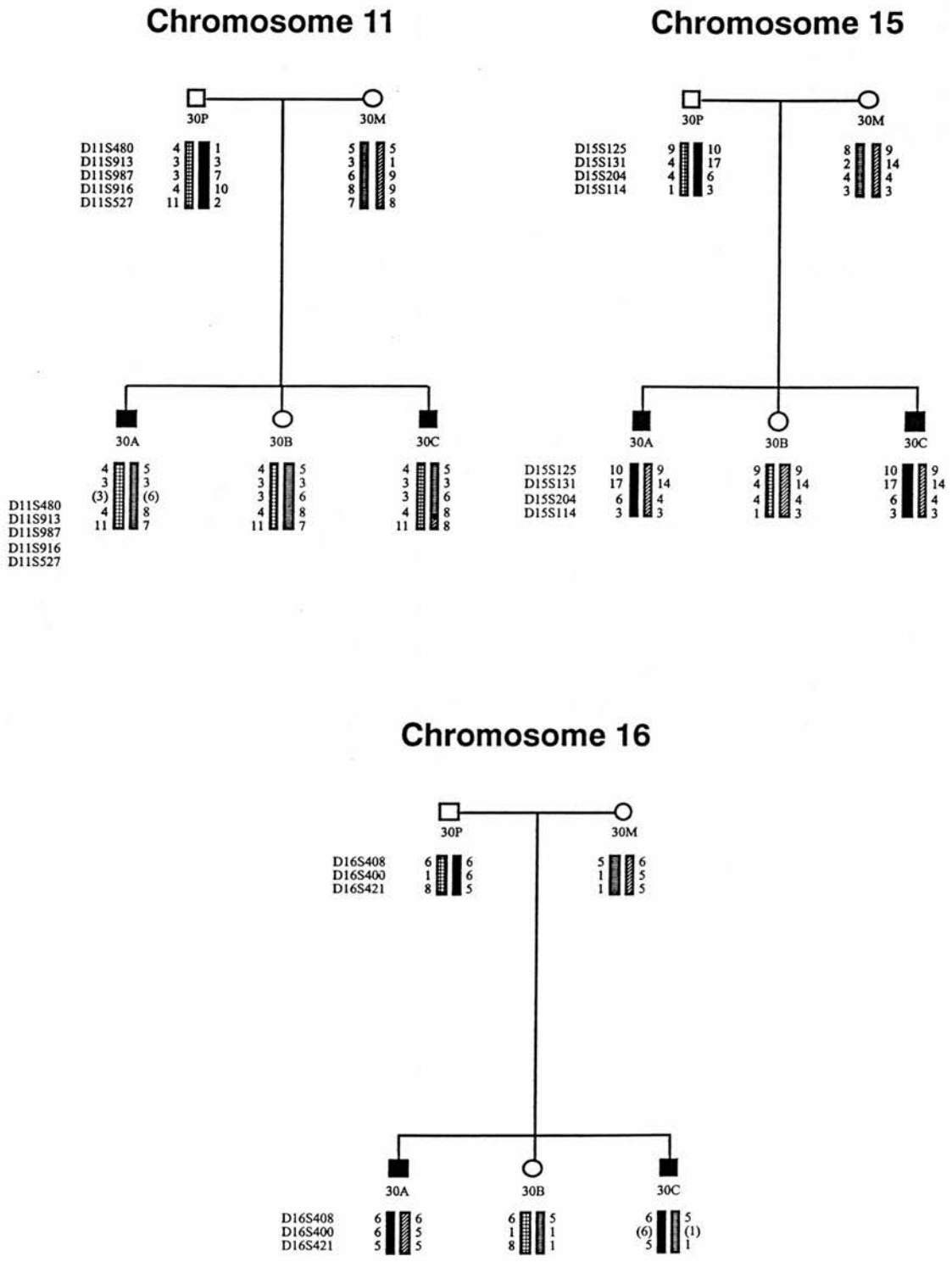


Figure 4.31

Haplotype analysis for family BB30, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results

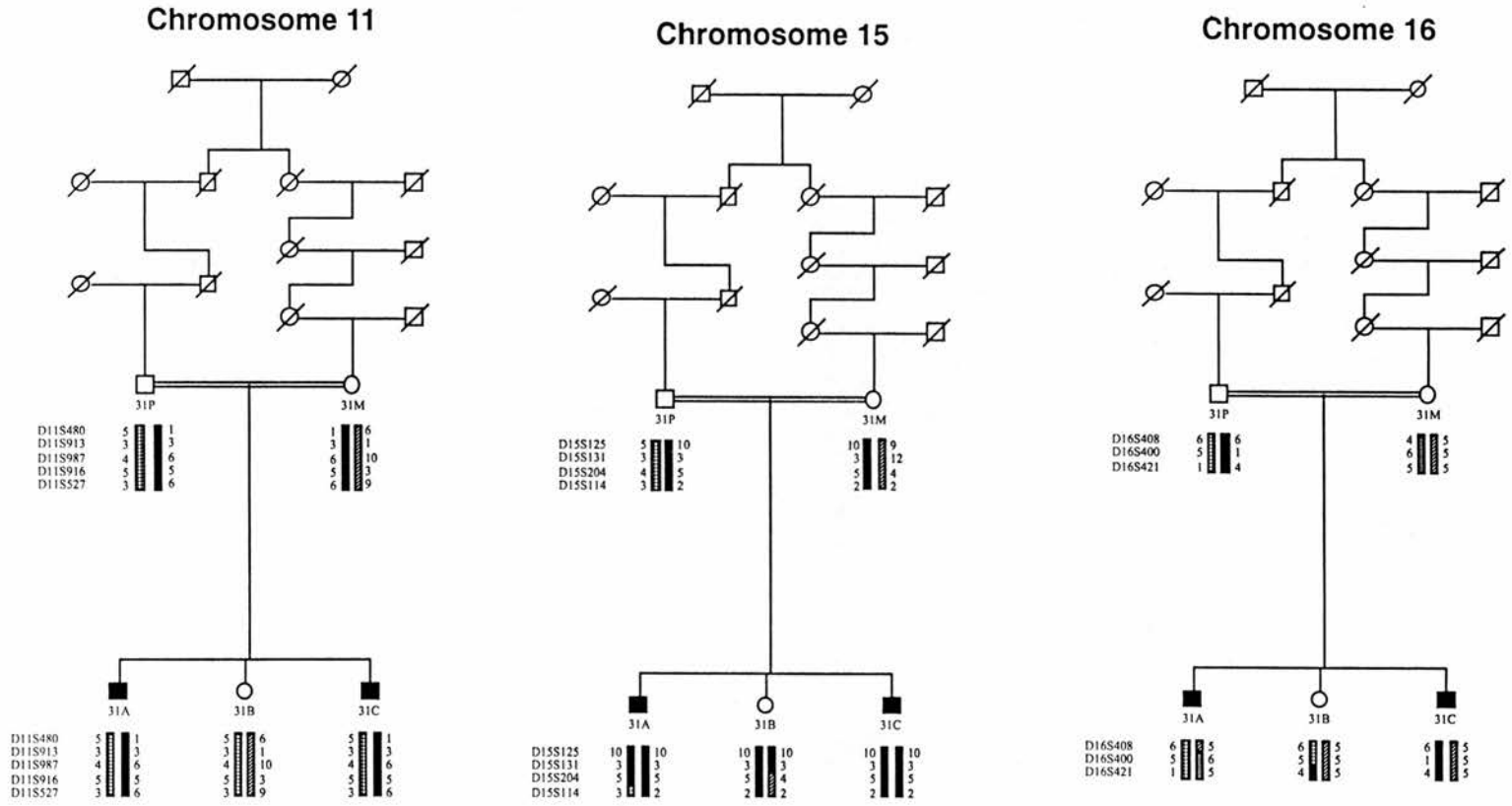


Figure 4.32

Haplotype analysis for famile BB31, for loci *BBS1, BBS2* and *BBS4.*

Inferred alleles are in brackets, question marks represent unknown results.

4.3 - DISCUSSION

4.3.1 - Linkage To Chromosome 11

In the original report of linkage to chromosome 11q, Leppert *et al* identified heterogeneity within their set of 31 North American families. They achieved a lod score of 4.31 for all 31 families, with the gene for human muscle glycogen phosphorylase, *PYGM*, at a recombination fraction (θ) of 0.15, but on closer examination of the results they found that this positive lod score was due to 17 of the families, with the remaining 14 having negative lod scores. Heterogeneity analysis gave a significant result, (χ^2 = 5.62, 1 d.f., p = 0.0089), and after removing the unlinked families from the analysis they obtained a lod score of 12.39 with *PYGM* (θ = 0.001), and a lod score of 6.03 with the microsatellite marker D11S913 at θ = 0.05. Although recombinants had been observed between these two markers, they could not determine their relative order, and chose to estimate the BBS locus relative to *PYGM*. Their analysis placed the locus in an interval 1 cM proximal and 2 cM distal to *PYGM*, between the markers UT5150 and *INT-2*, the gene coding for fibroblast growth factor type 3 (Dickson and Peters, 1987). As shown in **fig. 4.2** it has since been established that D11S913 lies approximately 2 cM distal to *PYGM*.

The original report of linkage to *BBS1* was based on a pooled set of small positive lod scores, with individual family lod scores ranging from 0.12 - 1.79, the average score being 0.73. This result could have occurred by chance, but linkage to this region has since been confirmed by Cornier *et al* (1995) who have identified a region of homozygosity on 11q in two large inbred Puerto Rican families, both of which share a common haplotype, suggestive of a founder effect for BBS in this population. This enabled them to place the *BBS1* locus between the markers GATA8A08 and *PYGM* (see **fig. 4.2**), an interval of around 10 cM.

In this study, statistically significant evidence of linkage has been obtained for a BBS locus at 11q13, between D11S480 (which lies in the region of homozygosity identified by Cornier *et al*) and D11S913. A maximum multipoint lod score of 6.26 was obtained at a position 5 cM distal to D11S480 and 3 cM proximal to D11S913 (see **fig. 4.33**), with a 90% confidence interval (Z_{max} - 1) of 5 cM, and 95% confidence interval (Z_{max} - 2) of 8.5 cM (Conneally *et al*, 1985). Considering the updated genetic positions of these markers (**fig 4.2**) we can place the *BBS1* locus in the interval between D11S480 and D11S913, close to

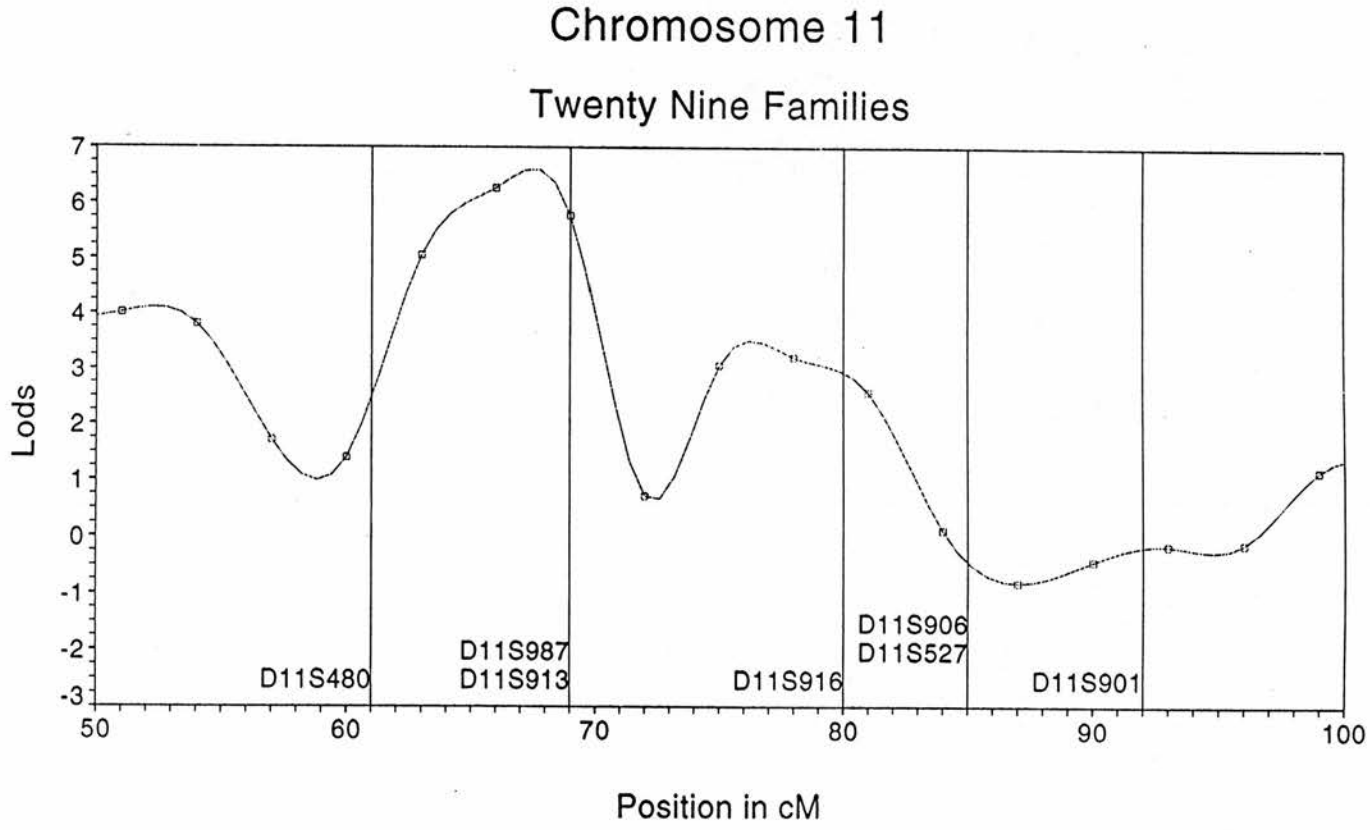


Figure 4.33 - Multipoint analysis of *BBS1*

The genetic distance of the markers used in the analysis from 11pter are represented by vertical lines.

the *PYGM* locus. This result partially overlaps the interval reported by Cornier *et al* (1995), but extends several cM distally.

Heterogeneity analysis also provided statistically significant evidence for genetic heterogeneity, with a lod score of 6.26 assuming two loci (one linked and one unlinked) (χ^2 = 28.84, P = 0.0000), 28 times greater than the lod score of 4.82 obtained assuming a single locus (χ^2 = 6.64, P = 0.005). When only the 11q13 locus is considered, 56% of families appear linked, as opposed to 36% of families when all four loci (*BBS1, BBS2, BBS4* and unlinked) are considered simultaneously. When combining haplotype analysis and the estimated conditional probabilities of linkage for each family, eight families show strong evidence of linkage to *BBS1* (BB1, BB5, BB6, BB12, BB15, BB21, BB23, BB24). BB13 and BB20, both of which are consanguineous (although the relationship between the parents is not known for BB20), also appear linked to *BBS1*, but do not show homozygosity for the markers studied in 11q13. Homozygosity would be expected for BB13, where the parents are closely related (second cousins) (see also results for chromosome 16, **fig. 4.16**), although markers closer to the locus than those studied here may show identity-by-descent. Based on results in other consanguineous families, where homozygosity has been observed for markers flanking a BBS locus (see **figs. 4.10** and **4.20**), and the close proximity of several of the markers studied to the region in which the *BBS1* gene is located, this result suggests that these families are not linked to this locus. Another three families (BB11, BB14 and BB27) may also be linked to chromosome 11, but further work is required to confirm this, and exclude them from linkage to other loci..

4.3.2 - Linkage To Chromosome 15

Carmi *et al* (1995a) used a form of DNA pooling to identify the *BBS4* locus, using DNA from a large inbred Bedouin kindred. All affected members were pooled, with unaffecteds and parents of affecteds comprising another two pools of controls. These DNA pools were used as templates for PCR reactions with microsatellite markers distributed throughout the genome. One marker on chromosome 15q showed a shift towards a particular allele in the affected sample, as compared to the multiple alleles for the control groups, and genotyping of individual DNA samples was then undertaken for this and flanking markers. This identified four linked markers as D15S125, D15S131, D15S204 and D15S114, at chromosome 15q22.3-q23, as shown in **fig. 4.3**. Multiple recombinants with D15S125, and

a single recombinant with D15S114, designated these as flanking markers. Maximum lod scores of 4.09 with D15S131 and 4.66 with D15S204 were achieved at zero recombination fraction, though the relative order of these two markers was undetermined. Genethon has since established that D15S131 is 0.6 cM proximal to D15S204 (>1000:1 odds) (Dib *et al*, 1996), placing the *BBS4* locus in a 8.5 cM interval. Most recently, the same group (Haider *et al,* 1996) have redefined the limits of the *BBS4* locus, using two inbred kindreds. They now place the locus in a 4.3 cM interval, flanked by D15S131 (6.9 cM distal to D15S125) and GATA89D04 (D15S823) (see **fig. 4.3**). This expands the previous limits of the gene distally by 2.8 cM, disregarding the single unaffected recombinant which placed the gene proximal to D15S114, as reported by Carmi *et al* (1995a) in the original family studied.

In this study, a maximum multipoint lod score of 6.10 was obtained at a location 5 cM distal to D15S125, and 2 cM proximal to D15S131/D15S204 (see **fig 4.34**). The 90% confidence interval spans 6 cM around the maximum lod score location, with the 95% confidence limit extending to 7.5 cM (Conneally *et al*, 1985). Heterogeneity testing showed statistically significant evidence for heterogeneity at odds of 180:1, with a lod score of 6.10 under the hypothesis of two loci ($\chi^2 = 28.07$, P = 0.0000), as compared to a score of 3.84 on the assumption of one locus ($\chi^2 = 10.39$, P = 0.0006). When only the *BBS4* locus is considered, 35% of families appear linked, compared to 32% when all four loci are considered. Seven families are consistent with linkage to *BBS4* when haplotypes and conditional probabilities are considered (BB7, BB14, BB18, BB19, BB29, BB30, BB31). Haplotype analysis alone suggests a further three families (BB2, BB11, BB28) may be linked to this locus.

Of these linked families, three are consanguineous (BB7, BB18 and BB31), and show homozygosity for D15S125, D15S131 and D15S204, which is consistent with identity-by-descent. Haplotype analysis in family BB18 (see **fig. 4.20**) confirms the observation of Carmi *et al* (1995a) that D15S125 is a flanking marker, as a crossover in unaffected member 18C places the locus distal to this marker. Haplotype analysis of pedigree BB31 (see **fig. 4.32**) also highlights two informative crossovers, one in affected male 31A, placing the locus proximal to D15S114, and another in unaffected female 31B, placing *BBS4* distal to D15S131, agreeing with the results of Haider *et al* (1996). This places the locus *BBS4* in the interval between D15S114 and D15S131, a genetic distance of 1.6cM (Dib *et al*, 1996), narrowing the region reported by Haider *et al* by approximately 2 cM.

This size of interval (corresponding to a few megabases) would be suitable for a positional cloning project to identify the *BBS4* gene. Before this could be considered, the results in

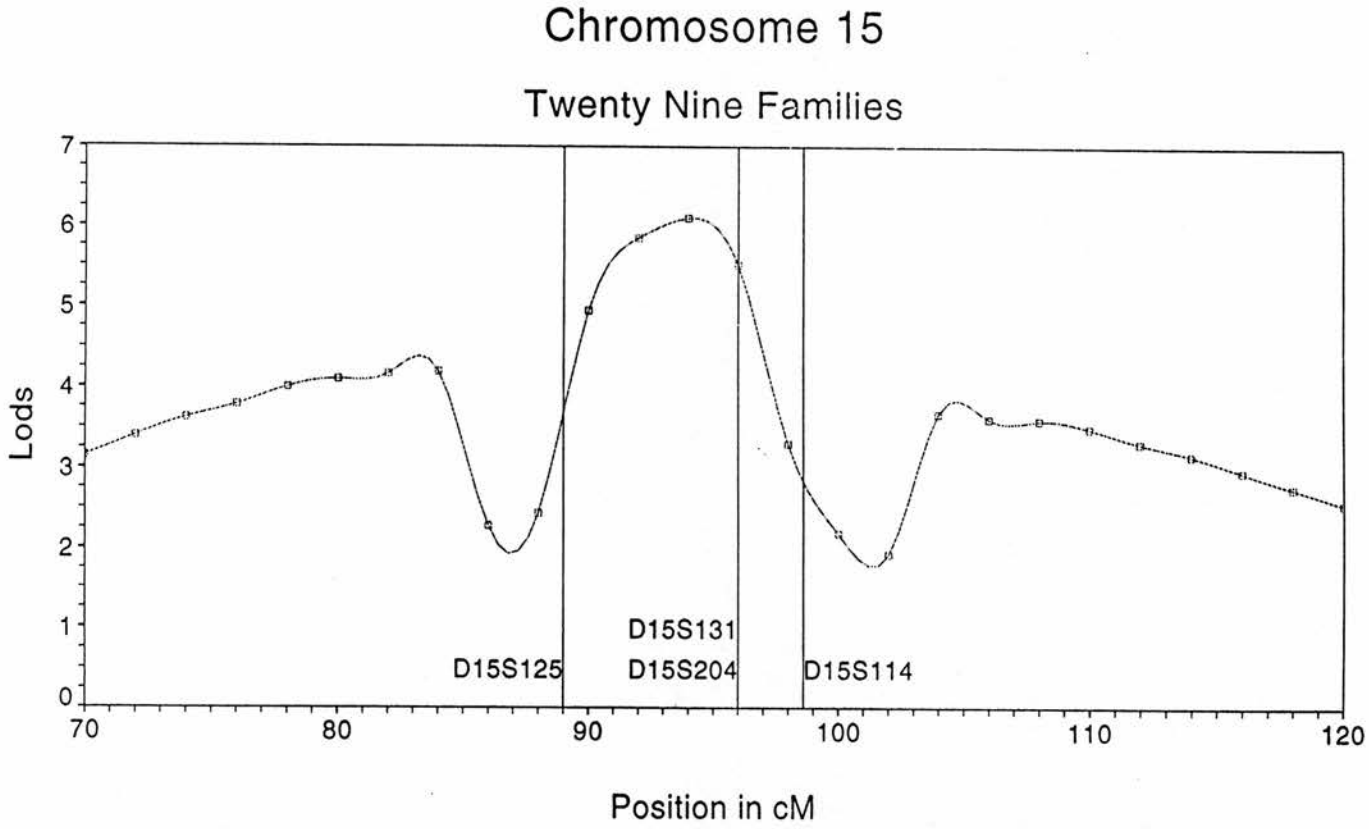


Figure 4.34 - Multipoint analysis of *BBS2*

The genetic distances of the markers used in the analysis from 15cen are represented by vertical lines

this family would have to be repeated, to confirm the data. In BB31, microsatellite marker D15S114 is uninformative in the mother, requiring that the crossover in individual 31B is confirmed by typing a marker distal to D15S114 that is informative in both parents. The diagnosis of Bardet-Biedl syndrome in this family would also have to be checked, to confirm that 31B is unaffected, and that 31A and 31C are both affected with BBS. If all the results indicate that the BBS locus must lie in this interval, the region could then be narrowed further, if possible, using Genethon markers lying between D15S131 and D15S114. Fourteen such new markers have been reported in this interval by Dib *et* al (1996), but the genetic order has not been resolved with odds >1000:1 in this region. Candidate genes localised to 15q22.3-q23 could be mapped using radiation hybrids, to confirm their location within, or outside, the critical region, which could in turn allow for mutation analysis of these genes in this family, and others families linked to the *BBS4* locus.

4.3.3 - Linkage To Chromosome 16

Linkage was originally identified in another Bedouin kindred using five markers on the long arm of chromosome 16 (Kwitek-Black *et al*, 1993). Three of the markers studied by Kwitek-Black *et al* gave significant lod scores at zero recombination fraction - D16S408 (Z_{max} = 4.22), D16S265 (Z_{max} = 3.14) and D16S186 (Z_{max} = 3.50). Recombinational events within the pedigrees placed the locus between D16S419 (9 cM centromeric to D16S408) and D16S265 (see **fig 4.4**), defining them as flanking markers. The markers used in this study included D16S408, and D16S400 and D16S421, two further Genethon microsatellite markers distal to D16S408. Subsequent work by Kwitek-Black *et al* (1996) has narrowed the region containing the *BBS2* locus to 2.5 megabases, lying between U28692 and D16S526 (see **fig. 4.4**). D16S408 lies within this interval.

This study achieved a maximum lod score of 1.09, under the hypothesis of genetic heterogeneity (χ^2 = 5.02, P = 0.041), a result of borderline significance. This lod score is achieved at a location 2 cM distal to D16S408, and 5 cM proximal to D16S400 (see **fig. 4.35**). When only the chromosome 16 locus is considered, 27% of families appear linked, as compared to 24% when all loci are considered simultaneously. Conditional probabilities and haplotype analysis indicate that three families are linked to *BBS2* (BB4, BB9, BB10), although only one of these has a conditional probability of linkage to *BBS2* greater than 0.5,

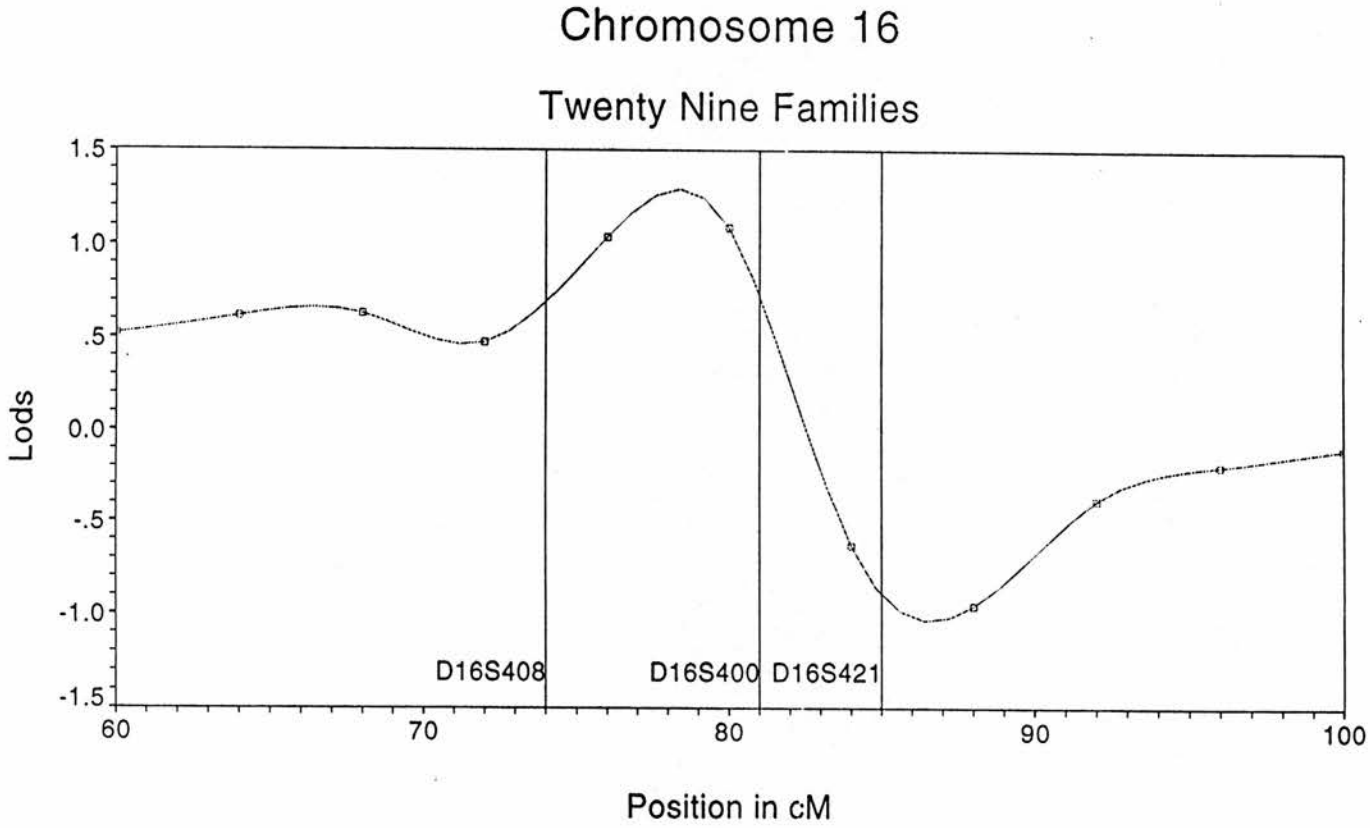


Figure 4.35 - Multipoint analysis of *BBS4*

The genetic distances of the markers used in the analysis from 16pter are represented by vertical lines

which is taken as the criteria for linkage. Several other pedigrees may be linked to this locus (BB1, BB2, BB5, BB11, BB12, BB21, BB22), but more markers in this region will need to be studied to confirm any tentative linkage. However, BB13, a consanguineous family, shows homozygosity for all three markers studied in the affected members. The probability of this haplotype occuring by chance, assuming no linkage disequilibrium, is 0.00434^2 (1.884×10^{-5}). This homozygosity suggests linkage (lod score of 0.46), despite achieving a higher lod score of 1.30 for chromosome 11, where homozygosity is not observed. This is probably due to the number, and informativeness, of markers studied on each chromosome. BB25, another consanguineous pedigree, may be also be linked to *BBS2*, though there is no evidence of identity-by-descent for the markers. Lod scores suggested that BB25 was linked to chromosome 15, but this is clearly not possible when the haplotypes are examined, and is due to inaccurate allele frequencies biasing the lod score.

4.3.4 - Families unlinked to *BBS1, BBS2* and *BBS4*

Following linkage and haplotype analysis for *BBS1, BBS2* and *BBS4* in all 29 families, at least three families (BB8, BB17 and BB28) show no signs of linkage to any of these loci (**figs. 4.11, 4.19 and 4.28**). Initial linkage analysis with markers close to *BBS3* was also carried out during this study, but showed no sign of linkage for any of the families. While more work with further, more informative, markers in this region is required for total exclusion of this locus in these three families, this result also suggests the possibility that there is a fifth BBS locus. This suggestion has also been raised by two other groups who have identified genetic heterogeneity among BBS families. Sunden *et al* (1995) used 23 Canadian BBS families in linkage analysis studies, and found that none of the four previously reported loci was a major disease-causing locus in their set of families. They were also able to exclude linkage to any of the four loci in several families, suggesting the presence of a fifth locus. A similar result has been reported by Bonneau *et al* (1996), who have studied 16 families from France, North Africa and Portugal. Three showed linkage to *BBS3*, two to *BBS2* and one to *BBS4,* but linkage to all four loci could be excluded in seven families, again suggesting a fifth locus that causes BBS.

4.3.5 - Clinical Correlations

As BBS is known to be genetically and clinically heterogeneous, this raises the possibility of identifiable clinical variation due to mutations in different loci. Carmi *et al* (1995b) suggested specific correlations using clinical data from the three inbred Bedouin kindreds whch have been used to identify the *BBS2, BBS3* and *BBS4* loci. They suggested that the chromosome 3 locus is associated with polydactyly of all limbs, the chromosome 15 locus is associated with polydactyly of the hands only, and with early-onset obesity, whereas the chromosome 16 locus results in the "lean" form of BBS.

In an attempt to compare Carmi's data with the findings of this study, the main problem encountered is trying to classify definitively the families studied as being linked to one specific locus. The most reliable way to attempt this is to exclude a family if recombination has occurred with markers that are thought to be flanking the locus, or in the case of consanguineous families, to identify regions of homozygosity. On this basis, 8 families appear to be linked to chromosome 11, and 7 to chromosome 15. There are no obvious clinical differences between these families, or the other families in the study, suggesting that the results of Carmi *et al* are probably more due to the inbred genetic backgrounds of the families, further influenced by the comparatively small sample size studied. Another possibility is allelic heterogeneity, where the specific mutations in the families studied by Carmi *et al* result in slightly different phenotypes to those observed in the patients in this study, who may have different mutations in the same genes.

This study emphasises the importance of recognising additional clinical features in the diagnosis of BBS. Symptoms identified among the patients included renal anomalies, dental anomalies, short stature, diabetes mellitus, deafness, hepatic fibrosis and hypertension. Odea *et al* (1996) recently noted the importance of renal disease in the prognosis of BBS. In a study of 21 families, 96% showed renal anomalies, and 75% of deaths occuring during the study were due to renal failure. Other symptoms they identified included diabetes mellitus in 25% of patients, and hypertension in 66%, probably due to the renal disease. Another study, Riise (1996), found renal disease to be the cause of death in 50% of BBS patients. Hepatic fibrosis has previously been reported in several cases of BBS, and has been diagnosed in two of the families in this study (Pagon *et al*, 1982; Nakamura *et al*, 1990). Elbedour *et al* (1994) also noted the frequency of cardiac abnormalities in BBS, suggesting that there is cardiac involvement in 50% of patients. Hypogenitalism is thought to be more common among males, but Stoler, Herrin and Holmes

(1995) have proposed that genital abnormalities among females are often missed in childhood. This agrees with reports among patients of irregular menstruation (Green *et al*, 1989; R. Riise, personal communication), suggesting that such abnormalities should be sought systematically. Overall, this study serves to emphasise the need for a multidisciplinary approach to BBS to identify and treat the complications associated with this disorder.

A report of obesity and increased height in carriers of BBS has suggested that the BBS gene can influence growth in heterozygotes (Croft *et al*, 1995). This study was conducted in a small sample size of 34 parents of BBS patients, and found four of the 15 fathers to be severely overweight. This represents three times the expected number in a sample of this size of American whites aged 35-74, and hence Croft *et al* have proposed that being a BBS heterozygote carries a three-fold risk of being obese. Assuming a carrier rate of 1% for BBS in the general population, Croft *et al* have suggested that 2.9% of all severely overweight men are BBS carriers. They also found that BBS heterozygotes were taller than comparable controls, and that BBS homozygotes were as tall, or taller, than comparably aged women and men. This finding contradicts reports of short stature in BBS patients from Schachat and Maumenee (1982), Green *et al* (1989) and this study (R. Riise, personal communication). It appears that more research is required, with larger sample sizes, to confirm or deny the findings of Croft *et al* . However, the study has emphasised the importance of research into obesity in Western society, and hence the importance of identifying the genes responsible for obesity syndromes such as BBS, Prader-Willi syndrome and Alstrom syndrome. Preliminary research has suggested abnormal serum leptin levels in BBS patients, but more work is required to find out if this is a general obesity finding, or specific to BBS (R. Riise, personal communication).

4.4 - CANDIDATE GENES

The high level of non-allelic heterogeneity identified in BBS is unexpected for a rare disorder with a comparatively specific diagnosis. Sheffield *et al* (1994) have suggested the involvement of a ligand-receptor complex, protein sub-units, or proteins involved in a common pathway, as there appear to be at least four separate genes causing a very similar phenotype. The involvement of a wide range of organ systems suggests that the genes involved may have widespread effects, a situation similar to that for Grieg

cephalopolysyndactyly syndrome. This syndrome, the symptoms of which include polysyndactyly, macrocephaly and facial anomalies, is caused by mutations in the *GLI3* gene, which codes for a zinc finger protein (Vortkamp, Gessler and Grezschik, 1991). Studies in mice have confirmed that this gene is expressed in the regions affected in Grieg syndrome (Hui and Joyner, 1993).

Several candidate genes have been suggested for BBS, due to their genetic location, expression pattern and phenotypic effects when mutated. One of these is the gene which codes for the β-subunit of the rod cGMP-gated cation channel (see **section 1.6.1**). Mutations have been identified in the α-subunit of the CNCG protein in one form of arRP, making the *CNCG2* gene, coding for the β subunit of this protein, a candidate for the retinal dystrophy seen in BBS (Ardell *et al*, 1996). Ardell *et al* (1995) have suggested that this gene may also be involved in the renal disease seen in BBS, as highly homologous channels have been found in the kidney. The *CNCG2* gene maps to a complex locus on chromosome 16q21, in the same region as linkage has been reported for BBS, and Kwitek-Black *et al* (1996) recently mapped it within the region (covering 2.5 megabases) in which the *BBS2* gene must lie.

The gene coding for phospholipase C β-3 (PLCB3) is found on chromosome 11q13, very close to *PYGM* (Sinke and Geurts van Kessel, 1995), where the *BBS1* gene has been mapped. This enzyme plays an important role in signal transduction, and is activated in many cells as a result of stimulation by ligands such as growth factors, hormones and guanine nucleotide binding proteins. Correspondingly, the gene is widely expressed in tissues including the retina, brain and kidney, and comprises 31 exons, ranging from 36 basepairs to 571 basepairs in length (Mazurak *et al*, 1995). Another gene mapping to chromosome 11q13 is the gene for the fibroblast growth factor 3 (FGF3, also known as the oncogene *INT-2*)(Peters *et al*, 1983). This was originally considered as a candidate for BBS due to its expression pattern and genomic location. FGF3 is generally not expressed in adult tissues, but is produced at precise stages of embryonic development, and fetal development, where it is found in the retina, cerebellum, sensory regions of the inner ear, and the teeth (Wilkinson, Bhatt and McMahon, 1989). A related protein of the fibroblast growth factor family, FGF4, which lies close to FGF3 in 11q13 , is known to be involved in limb development, and several genes coding for receptors for FGFs result in human skeletal disorders when mutated (for review of FGF family and receptors, see Muenke and Schell, 1995). However, both these genes have been excluded as candidates for BBS by recent fine

mapping which places them outside the region where the BBS gene lies (Cornier *et al*, 1995).

A further candidate gene on chromosome 11q is the human homologue of the *Drosophila* gene *rdgB*. The *rdgB* gene is thought to code for a phosphatidylinositol transfer protein, which causes light and age dependent degeneration of the rhabdomeres and axons, by affecting calcium ion mobilisation (Vihtelic, Hyde and Otousa, 1991). Two human homologues of this gene have been identifed, one at 17p13.3, the human phosphatidylinositol transfer protein (Fitzgibbon *et al*, 1994) and another at 11q13.5, Drosophila-related expressed sequence 9 (*DRES9*) (Banfi *et al*, 1996). The *DRES9* gene was identified by searching an expressed sequence tag (EST) database for sequences with homology to *Drosophila* genes, and was later linked to marker D11S913 by radiation mapping. In *Drosophila*, this gene is expressed in the photoreceptors, antennae, and sensory processing centres of the brain (Pak, 1995), making it a plausible BBS candidate in man. Another gene coding for a protein involved in photoreceptor structure is the gene for *Rom-1* on 11q13, several cM proximal to *PYGM*. This is a candidate for the retinal dystrophy observed in BBS, but it seems unlikely that this gene alone could also be causing the many phenotypic manifestations of BBS, as Bascom *et al* (1989) have shown that this gene is not expressed in the brain, kidney, liver, muscle, lymphoblasts or fibroblasts, and appears to be photoreceptor specific.

Other genes involved in limb development, such as the homeobox (*Hox)* genes, are candidates for BBS. For example, *HoxD13* causes synpolydactyly in man when mutated (Muragaki *et al*, 1996), and *HoxD6* overexpression results in an extra digit in the chick wing bud (Morgan *et al*, 1992). Also, the *Hox4* gene family are involved in hand and foot morphogenesis, and are co-ordinately expressed in the CNS and elsewhere in the body mesenchyme (Tabin, 1992). These developmentally important genes are regulated in part by retinoic acid (RA) and its derivatives (for review see Conlon, 1995), which are also known to have a role in retinal function. Excess and deficiency of RA can cause a syndrome with features such as craniofacial and limb abnormalities, heart and central nervous system defects (Lammer *et al*, 1985). RA signal transduction is mediated in part by cytoplasmic RA-binding proteins, CRABP-I and CRABP-II. Transgenic mice which are homozygous null mutants for CRABP-II are viable and fertile, but have an extra postaxial digit, as seen in BBS (Fawcett *et al*, 1995). CRABP-I is a closely related peptide, the gene for which maps to chromosome 15q22, several cM distal to D15S114, in the region where

the *BBS4* locus could lie (Haider *et al*, 1996). All these genes could be considered as candidates for BBS.

Obesity has recently been shown to be influenced by several genes, including the genes coding for leptin and its receptor(s) (Lonnqvist *et al*, 1995, Chen *et al*, 1996). As obesity is a symptom of BBS, it would be interesting to see if this syndrome is linked to any of these genes. A study conducted by Reed *et al* (1995) suggested that genes that are commonly involved in non-syndromic human obesity are not present in the regions where linkage to BBS genes has been identified. As previously mentioned, work is ongoing to investigate serum leptin levels, to find out if expression is altered in BBS patients.

Recently, the *tub* gene has been identified in the mouse mutant *tubby*. The human homologue of this gene is a strong candidate for BBS because of the similarities between the *tubby* phenotype and human BBS. This mouse mutant was a spontaneous autosomal recessive mutation, first reported by Coleman and Eicher in 1990. It was initially characterised by slow, progressive obesity, with mild hyperinsulinaemia and hypoglycaemia in the absence of hyperphagia. Coleman and Eicher (1990) mapped the gene to mouse chromosome 7, several cM from the *Hbb* (haemoglobin β chain complex) locus. This region in the mouse is syntenic to chromosome 11p15 in humans, although some of the genes in this region map to 11q13 (where *BBS1* maps to), suggesting a break in the syntenic relationship (Chung *et al*, 1996). The genetic map surrounding *tub* was recently refined by Chung *et al* (1996) and allowed exclusion of several candidate genes in the region. Heckenlively *et al* (1995) identified a new retinal degeneration mouse mutant *rd5*, which showed early progressive hearing loss and progressive retinal degeneration, as seen in Usher syndrome type I. When they mapped *rd5* to chromosome 7 they could find no animals with the tubby phenotype without a retinal degeneration, although animals were observed with retinal degeneration but no obesity. It seems likely that the *rd5* and *tubby* genes are either very closely linked, or that the phenotypes are pleiotropic effects of a single gene.

Most recently Noben-Trauth *et al* (1996) identified two mutations in a candidate gene for *tubby*. This gene is expressed predominatly in the brain, eye and testis, and the C-terminus shows 62% identity at the amino-acid level to a putative mouse testis-specific phosphodiesterase (p4-6), while the amino-terminal half is novel. Kleyn *et al* (1996) also identified a mutation in the *tubby* candidate gene, and analysed the human homologue of *tub*, which is 89% identical to the mouse gene. The encoded protein is predicted to be 94% identical to the mouse protein, and Kleyn *et al* also found homology between these two

proteins and a human retinal transcript of unknown function. Ohlemiller *et al* (1996) have examined the cochlear and retinal degeneration in *tubby* mice, and concluded that while the retinal phenotype is consistent with that seen in Usher type I, the progression and severity of the hearing loss, without vestibular defects, suggests that clinically it is more similar to Usher type II or III. If *tubby* and *rd5* are caused by one gene, then BBS, and similar syndromes such as Alstrom syndrome (which consists of retinal degeneration, sensorineural deafness, obesity and diabetes) are suitable candidates, although hearing impairment is an inconsistent feature of BBS.

4.5 - LINKAGE AND HOMOZYGOSITY MAPPING

The linkage analysis carried out in this study illustrates some of the difficulties encountered when working with a genetically heterogeneous disease. This is further complicated by the disorder being both rare and autosomal recessive, so that few large multi-generational pedigrees are available. With this kind of condition, possibly the simplest way to attempt gene mapping is to study large inbred families, using the technique of homozygosity mapping, first suggested by Lander and Botstein (1987). Homozygosity mapping relies on the parents of an affected individual both inheriting the disease gene from a common ancestor. Thus, any markers closely linked to the disease gene will be identical in the affected individual, as the same DNA flanking the gene has been passed down to each parent, and then to the affected child.

This information is taken into account when calculating lod scores, and greatly increases the power of consanguineous families in linkage mapping. For example, families BB14 and BB31 both appear linked to the *BBS4* locus on chromosome 15, and both contain two affected sibs, and one unaffected child. Family BB14, which is non-consanguineous, achieves a lod score of 0.72 for this locus, whereas BB31 alone reaches a significant lod score of 3.00, as the parents are second cousins once-removed. However, problems are encountered in lod score calculations, as inbred pedigrees take longer, and more memory, to compute, especially if several markers with multiple alleles are being studied, as in this project. Furthermore, if the marker and disease gene frequencies are inaccurate, this can lead to errors in the lod scores (see Terwilliger and Ott, 1994). This occurred with one of the inbred pedigrees studied, BB25, for chromosome 15. This family has two branches with affected children, one in which both parents are related, and the other with one

unrelated parent. As all affected children have inherited one common haplotype with comparatively rare alleles, this has increased the lod score, despite the two affected children in one branch each inheriting different chromosomes from the unrelated parent (see **fig.4.27**). Indeed, this family looks more likely to be linked to chromosome 16, despite a lack of identity-by-descent. This result exemplifies the difficulty in selecting a suitable population from which to estimate marker allele frequencies, especially when using pedigrees from several different countries. In this study, allele frequencies were taken from GDB (if published) or all assumed to be equal. Ideally, a group of unrelated individuals from similar genetic populations would have been available for typing with the markers, to gain an accurate allele frequency estimate.

Several human disease genes have been identified using homozygosity mapping, including the genes for alkaptonuria (Pollak *et al*, 1993), spinal muscular atrophy (Gilliam *et al*, 1990) and Hirschsprung disease (Puffenberger *et al*, 1994). Usually the pedigrees are from distinct populations, such as Japan where 1 in 25 marriages involve first cousins, compared to 1 in 200 in the U.K. (Farrall, 1993). Other commonly studied communities include the Mennonites, the Mormons of Salt Lake City, the Ashkenazi Jews and Bedouin tribes, such as the three kindreds used in mapping BBS loci.

In the set of families studied in this project, six were consanguineous, although the degree of relationship was not known for one of them (BB20). Nevertheless, this information still meant that identity-by-descent was expected for markers flanking the locus, if the family was linked to that particular locus. Because of homozygosity mapping, assignments to specific loci were possible in four of the families. It also seems possible that several of the other families studied are consanguineous, as the parents share a common haplotype for at least one of the chromosomes studied (see **figs 4.8** (BB5); **4.13** (BB10); **4.30** (BB29)). This identity-by-state for chromosome markers suggests that each parent has inherited this piece of DNA from a common ancestor, though they may be unaware of being distantly related. This is to be expected with such a rare disorder.

4.6 - SUMMARY

This study has attempted to assign 29 families with Bardet-Biedl syndrome to specific, previously reported, BBS loci. On the assumption of genetic heterogeneity, 36 % are linked to *BBS1* (11q13), with a lod score of 6.26. From haplotype anlysis, eight families are

consistent with this result. For the locus on chromosome 16q21, *BBS2*, a lod score of 1.09 was reached (assuming genetic heterogeneity), and only three families appear linked to this region, though one consanguineous family shows homozygosity for all three markers studied. For *BBS4,* on chromosome 15, at lod score of 6.10 was achieved, assuming genetic heterogeneity, and seven families appear linked from haplotype ananlysis. Furthermore, haplotype analysis in consanguineous pedigree BB31 highlighted two informative crossovers, enabling us to narrow the *BBS4* region from 4.3 cM (Haider *et al* , 1996) to approximately 2.2 cM. Several families do not show linkage to any of the regions studied, suggesting the possibility of a fifth BBS locus.

This work has highlighted problems encoutered in mapping a rare autosomal recessive condition, especially when non-allelic genetic heterogeneity is present. The most reliable gene assignments have been possible in consanguineous families, emphasising the power of large inbred pedigrees in this type of gene mapping. However, care must be taken in relying on linkage data alone, as small errors in gene frequency can dramatically alter the final result obtained. Also, the number, and informativeness, of markers studied can alter the lod score obtained for a specific locus. For reliability, this data should be considered in parallel with haplotype analysis for each family, to confirm linkage results. Furthermore, gene assignment is difficult in small families, where a lack of recombination does not necessarily equate to linkage. This problem can be reduced where possible by using highly informative markers, so that information can be obtained from all relevant meioses. This study has also suggested the presence of a fifth BBS locus, and further linkage mapping or candidate gene analysis could be carried out in the families that appear unlinked to any of the previously reported loci.

Future work in attempting to identify causative genes for BBS could utilise the consanguineous families that appear linked to one specific locus, by typing multiple closely linked, highly informative, microsatellite markers in the regions of homozygosity. In this way, the region of homozygosity may be narrowed, hopefully to several megabases, a distance suitable for physical mapping efforts, or testing of candidate genes within the region. In particular, the results of haplotype analysis in family BB31 have suggested that further mapping work is required in this family to reduce the interval of homozygosity in which the *BBS4* locus may lie.

Finally, while genetic heterogeneity complicates the possibilty of prenatal diagnosis, this study emphasises the need for a multidisciplinary examination of suspected BBS patients. A thorough examination could allow early, and possibly preventative, treatment of associated conditions to increase the quality of life, and prognosis, for patients.

CHAPTER FIVE

LINKAGE MAPPING IN AUTOSOMAL RECESSIVE RETINITIS PIGMENTOSA

5.1 - INTRODUCTION

The major problem in mapping genes for autosomal recessive retinitis pigmentosa is the expected nonallelic genetic heterogeneity. To date, four specific genes, rhodopsin, the alpha- and beta-subunits of cGMP phosphodiesterase, and the alpha subunit of the cyclic nucleotide gated channel, have been shown to cause arRP when mutated. These four genes are thought to account for less than 10% of arRP cases (Dryja and Li, 1995). In addition to these loci, linkage has also been reported to 6p21.3 (RP14) in a large kindred from the Dominican Republic (Knowles *et al*, 1994), and to 1q31-q32.1 (RP12) in an extended inbred Dutch pedigree, and more recently also in a large Pakistani pedigree (Leutelt *et al*, 1995), with preserved para-ateriolar retinal pigment epithelium (PPRPE), a clinically distinguishable subtype of RP (van Soest *et al*, 1994). For the RP14 locus, the peripherin/*RDS* gene has been excluded, but the guanylate cyclase activating protein genes (*GUCA1* and 2) which lie in the same region are currently being investigated as candidates by Shugart *et al* (1995). Another gene coding for a protein which is abundantly and exclusively expressed in the retina has also been mapped to the 400kb interval in which RP14 must lie (Shugart *et al*, 1995), and is presently being screened in families with arRP (A. Wright and M. North, personal communication). With at least six separate loci causing arRP, and a high probability that many more genes will be involved, it is important to carry out linkage studies for arRP loci in a genetically homogeneous population.

As with most autosomal recessive conditions, few extensive arRP pedigrees are available. To reduce the complication of genetic heterogeneity without using extended families, a genetically homogeneous and inbred population can be used for study. In this project the south-central Sardinian population was selected, as it is known to be a genetic outlier, with a comparatively high level of inbreeding. Furthermore, economic, religious and social factors have encouraged establishment of large families in Sardinia. Finally, a high proportion of the RP observed in Sardinia is thought to be recessively inherited, providing a large selection of families to study.

5.2 - RETINITIS PIGMENTOSA IN SARDINIA

Sardinia is the second largest island in the Mediterranean, lying 150 miles off the North-West coast of Italy, 10 miles south of Corsica. Much of the terrain is mountainous, with a low population density, although there are several fertile alluvial plains in coastal areas. Most of the population of around 1.5 million people is concentrated in the capital towns of the four provinces - Cagliari, Sassari, Nuoro and Oristano (Bernardinelli *et al*, 1994). The origin of the Sardinian population is thought to be multicentric, with evidence of colonisation since the pre-Neolithic period (over 9000 years ago). Archealogical findings indicate a "nuragic" culture, in common with other coastal and island areas of the Mediterranean, around 6000 years later. Colonisation of the island introduced the Phoenicians, Carthagians and Romans to Sardinia (Bernardinelli *et al*, 1994). The inhabitants also suffered invasions by the Vandals, Byzantines, Saracens, and later domination by the Pisans, Genoese, and Spanish, until finally becoming a region of Italy in 1861 (Cappello *et al*, 1996). However, the local inhabitants are thought to have taken refuge in the central moutainous regions of the island, also avoiding outbreaks of malaria in the lowlands (Workman *et al*, 1976).

Genetic analysis of polymorphisms, such as immunoglobulin allotypes, confirm this history of the island's inhabitants, showing a strong association with Italian, North African and Middle Eastern populations (Piazza *et al*, 1976; Menozzi *et al*, 1978). The comparative isolation of the island from the rest of Europe, also seen through linguistic studies, combined with a small initial population size, and genetic drift, has resulted in the modern population being one of the most genetically deviant in Europe, second only to the Lapps (see **fig. 5.1**) (Cavalli-Sforza and Piazza, 1993). Low migration has also meant that the population has remained relatively stable, with low rates of immigration due to a poor economy. Within the island, the geography has resulted in isolation of communities, with increased levels of endogamy, resulting in genetic heterogeneity between small villages. Consanguinity rates have remained quite high, particularly in the highlands, where 10% of marriages were consanguineous in the middle of this century (Workman *et al*, 1976).

Many genetic disorders are prevalent in Sardinia due to this unique genetic background. Particularly common are glucose-6-phosphate dehydrogenase deficiency and thalassaemia (both of which are thought to confer increased resistance to malaria in the heterozygote), multiple sclerosis, insulin-dependent diabetes and Wilson's disease (Bernardinelli *et al*, 1994). Some of these conditions have been shown to be genetically homogeneous in the

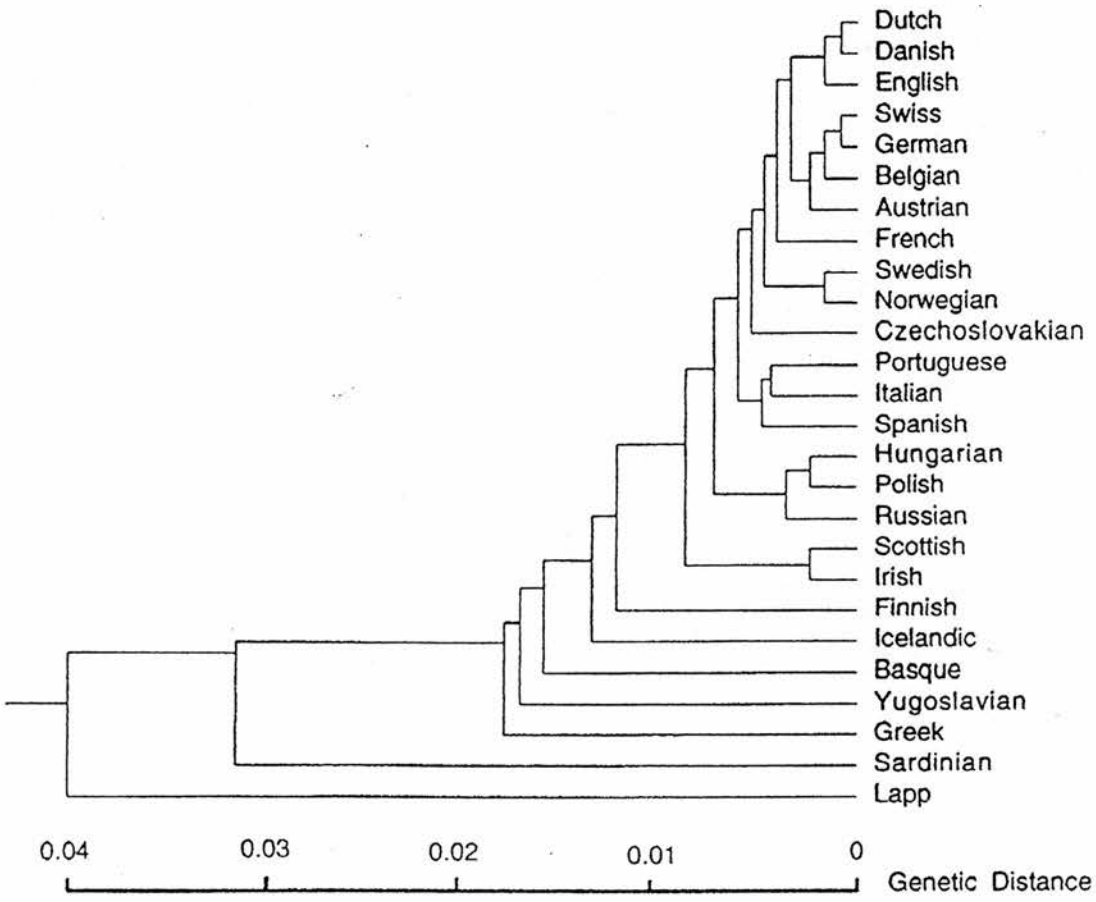


Figure 5.1 - Sardinia is a genetic outlier in Europe

Genetic tree of 26 European populations. F_{ST} distances are based on an average of 88 genes. From Cavalli-Sforza, Menozzi and Piazza (1994).

Sardinian population, probably due to a founder effect. This genetic homogeneity makes the population ideal for the study of autosomal recessive diseases, as do the levels of consanguinity, and the establishment of large families.

Eleven arRP families from south-central Sardinia (shown in **fig. 5.2** and **fig. 5.3**), comprising 28 affected and 44 unaffected members, were identified and sampled by Dr. M. Fossarello (Dept. of Clinical Ophthalmology, University of Cagliari, Sardinia). Clinical characterisation included complete ophthalmological examination with electrophysiological and visual field testing, and fluorescein angiography (Fossarello *et al*, 1993). The prevalence of arRP in south-central Sardinia has been reported to be 21.5% of all RP families, but this is likely to be an underestimate, as many recessive cases will have only one affected, and have been designated as simplex. However, when compared with 5% of families having autosomal dominant RP, and 1.3% having X-linked recessive RP, arRP is clearly the most common type of genetic RP within this population. Genetic analyses of eight families with adRP from this region have shown that 50% have rhodopsin mutations (Fossarello *et al*, 1993), a much higher proportion than in most populations, where rhodopsin accounts for around a quarter of all adRP (Dryja, 1992). Indeed, three of the four families in which rhodopsin defects were identifed shared exactly the same mutation. This can be taken as further evidence of genetic homogeneity within this population.

5.3 - EXCLUSION OF CANDIDATE LOCI

Before this project began, previous work carried out by Dr. D. Mansfield involved exclusion of candidate loci for arRP using markers in the regions surrounding candidate genes. A dinucleotide microsatellite repeat in the first intron of the rhodopsin gene was analysed, and allowed for the exclusion of 3 out of 9 of the Sardinian families. The β-PDE gene in chromosomal region 4p16 could not be excluded due to a positive lod score with the neighbouring marker HOX7, but a mononucleotide repeat in the 3' untranslated region of the peripherin/*RDS* gene allowed the exclusion of this gene in 5 of the 9 families. A further thirty markers were analysed on chromosomes 2, 7, 11, 16, 18 and 21, all resulting in negative lod scores, with the exception of one marker on chromosome 11. These results meant that large exclusions could be made, on the basis of $Z < -2$, with the assumption of genetic homogeneity.

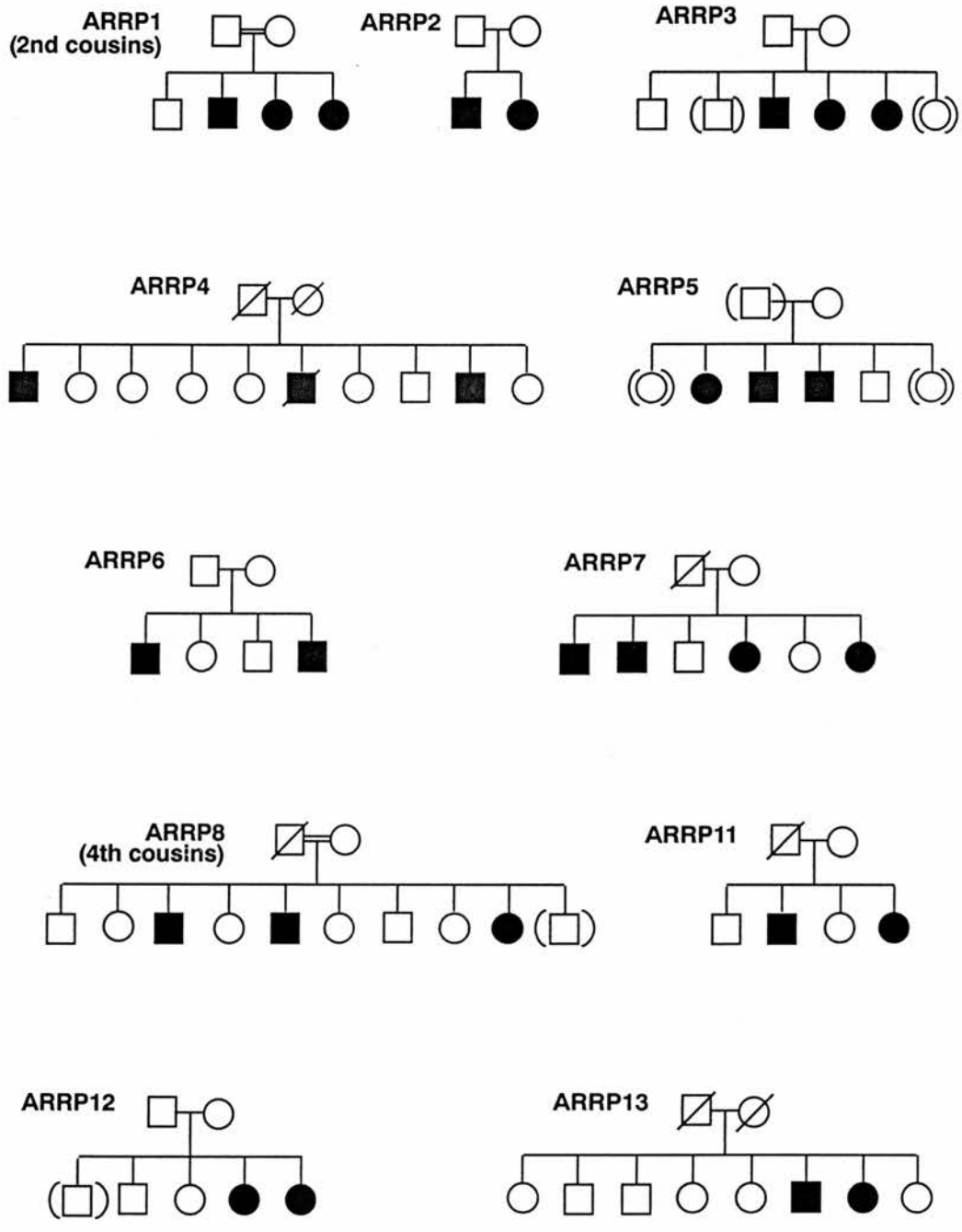


Figure 5.2 - Pedigrees of the eleven Sardinian families used in this study.
Individuals in brackets were not included in this study. For the two consanguineous families the relationship between the parents is indicated.

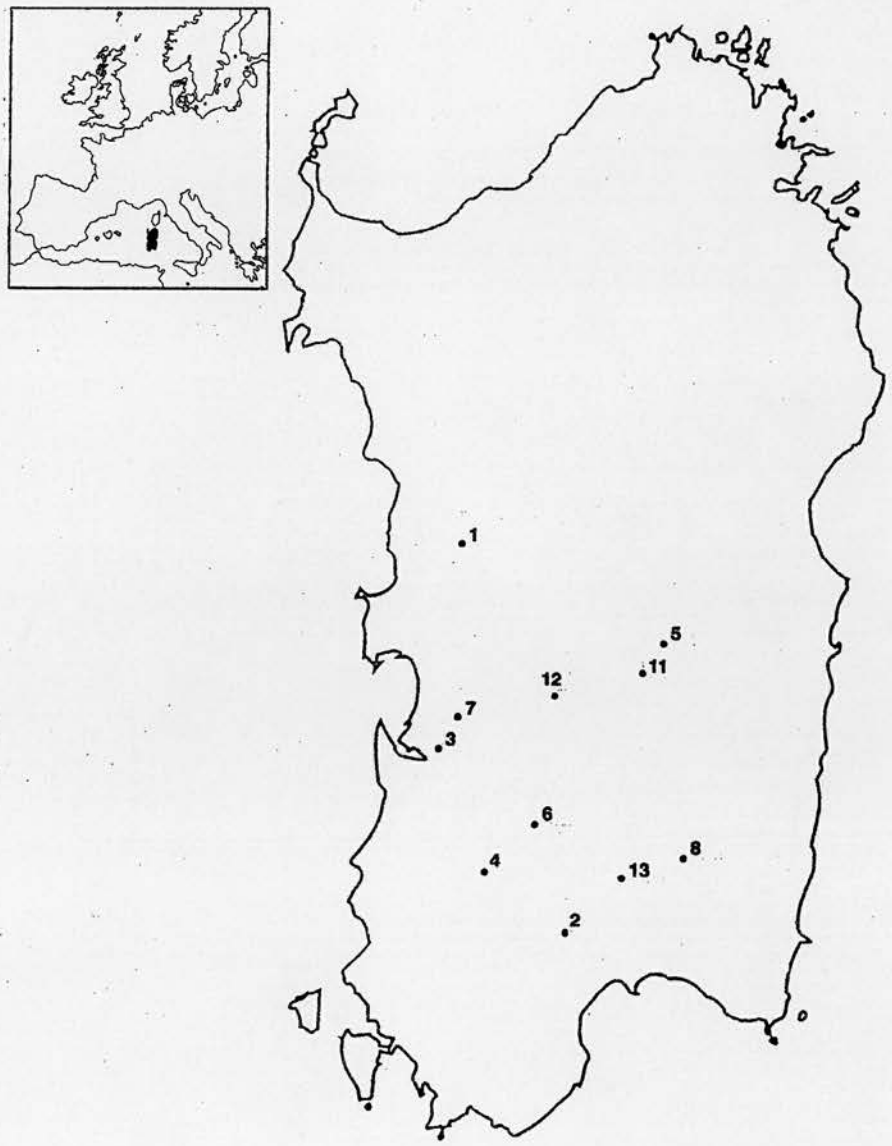


Figure 5.3 - Map of Sardinia, showing the relative geographic locations of the families used in this study.

Each number represents the pedigree number, see **fig. 5.2**.

5.4 - GENOME SCANNING FOR ARRP GENES

In an attempt to identify linkage among the set of 11 Sardinian families studied, 195 microsatellite markers were typed in collaboration with Dr. D. Mansfield. Most of the highly informative markers used were taken from the Genethon catalogue, with an average spacing of 20 cM. No significant lod scores ($Z > 3$) were obtained with the complete set of families. However, considerable exclusions (as calculated by P. Teague) were possible using $Z < -2$, excluding a total of 2,284 cM, representing 62 % of the genome (assuming a sex-averaged length of 3,699 cM) (Dib *et al*, 1996), on the assumption of genetic homogeneity. The genome exclusions are detailed in **figure 5.4**, showing the region excluded from the results of each marker. Any lod scores > 0.50 have been highlighted in bold. The approximate location of any candidate genes has been included, relative to the markers studied. Regions excluded on the basis of genetic homogeneity included the genes coding for β-PDE, rhodopsin, CNCG1, α-PDE, γ-PDE, phosducin, arrestin, peripherin/*RDS* and the whole short arm of chromosome six, as linkage has been reported for arRP to 6p21 in a large inbred kindred from the Dominican Republic (Knowles *et al*, 1994). **Figure 5.5** also shows these exclusions as a representation of each chromosome, highlighting the areas not yet excluded. Although further analysis of regions not yet excluded could yield positive lod scores, it seems highly likely that there is genetic heterogeneity among the families in this study.

This suggestion of genetic heterogeneity led us to examine the individual family lod scores for each marker, with particular attention paid to markers close to candidate loci. The loci considered were phosducin (*PDC*), RP12 (van Soest *et al*, 1996), arrestin (*SAg*), rhodopsin (*RHO*), the alpha, beta and gamma subunits of cGMP phosphodiesterase (*PDEA, PDEB* and *PDEG*), the alpha and beta subunits of the cyclic nucleotide gated channel (*CNCG1* and *CNCG2*), RP14 (Shugart *et al*, 1995), peripherin (*RDS*), recoverin (*RCV1*), guanylate cyclase (*GUC2D*), interphotoreceptor retinol-binding protein (*RBP3*), *ROM-1* , rhodopsin kinase (*RK*) and the alpha and gamma subunits of rod transducin (*GNAT1* and *GNGT1*). Genetic locations for each locus were obtained from the Human Transcript Map (http://www.ncbi.nlm.nih.gov/SCIENCE96/) or the Genetic Location Database (LDB) (http://cedar.genetics.soton.ac.uk/). For all but one of the loci at least one of the families showed signs of no recombination with the closest marker studied (see **fig. 5.6**). These families deserve further investigation, primarily with more markers flanking the candidate

MARKER NAME	CHROMO-SOMAL POSITION	LOD SCORES AT RECOMBINATION FRACTIONS						REGION OF EXCLUSION
		0.01	0.05	0.10	0.20	0.30	0.40	
D1S243	1.00	-4.76	-1.08	0.08	**0.60**	0.44	0.14	0.00-3.05
D1S214	1.06	-10.13	-4.31	-2.18	-0.63	-0.15	-0.02	2.31-29.76
D1S228	1.11	-10.18	-4.35	-2.21	-0.63	-0.14	-0.01	16.95-44.39
D1S199	1.16	-11.70	-5.18	-2.75	-0.90	-0.26	-0.05	28.53-62.08
D1S255	1.23	-12.04	-5.45	-2.96	-1.00	-0.29	-0.05	49.02-82.57
D1S220	1.31	-9.48	-4.25	-2.28	-0.77	-0.23	-0.04	75.49-102.94
D1S305*	1.56	-10.63	-4.22	-1.93	-0.37	0.01	0.03	154.76-170.01
D1S196*	1.64	-11.06	-4.61	-2.27	-0.59	-0.10	0.00	172.08-199.53
D1S210*	1.66	-3.11	-0.70	0.02	0.31	0.21	0.06	190.13-193.18
D1S242*	1.67	-13.28	-6.41	-3.70	-1.46	-0.52	-0.12	174.76-214.41
D1S191*	1.70	-14.90	-6.51	-3.46	-1.22	-0.42	-0.09	186.59-220.14
D1S202*	1.70 *PDC*	-8.66	-2.97	-1.02	0.17	0.28	0.10	198.79-207.94
D1S306	1.75 RP12	-7.20	-3.27	-1.80	-0.64	-0.21	-0.04	210.37-225.62
D1S237	1.81	-9.04	-3.33	-1.35	-0.09	0.12	0.05	227.94-243.19
D1S213	1.84	-11.67	-4.05	-1.42	0.21	0.38	0.15	236.72-251.97
D1S225	1.85	-13.23	-5.00	-2.15	-0.30	0.09	0.06	233.55-261.00
D1S251	1.85	-12.85	-4.64	-1.81	-0.02	0.25	0.10	239.65-254.90
D1S103	1.85	-8.56	-2.88	-0.94	0.23	0.33	0.12	242.70-251.85
D1S235	1.88	-6.83	-2.35	-0.80	-0.15	0.23	0.08	251.48-260.63
D1S304*	1.93	-7.58	-3.09	-1.49	-0.37	-0.06	0.00	263.06-275.26
D2S207*	2.05	-4.35	-1.23	-0.20	0.33	0.27	0.09	11.14-16.56
D2S168*	2.10	-2.43	-0.10	**0.54**	**0.65**	0.38	0.11	24.99-27.70
D2S171	2.18	-17.65	-7.59	-3.84	-1.04	-0.18	0.00	30.89-66.11
D2S119	2.25	-7.38	-2.43	-0.80	0.08	0.13	0.04	61.12-74.67
D2S134	2.32	-11.35	-4.83	-2.35	-0.50	0.01	0.05	75.09-99.48
D2S139	2.38	-9.30	-3.54	-1.50	-0.13	0.14	0.07	97.13-110.68
D2S160*	2.46	-8.47	-3.49	-1.60	-0.25	0.07	0.05	119.29-132.84
il1a (2)	2.46	-1.36	-0.62	-0.32	-0.09	-0.02	0.00	no exclusion
D2S114	2.53	-16.12	-6.49	-3.04	-0.66	-0.05	0.02	130.55-160.36
D2S141*	2.60	-14.66	-6.14	-2.98	-0.68	-0.04	0.03	149.94-179.75
D2S138	2.69	-16.82	-7.59	-4.09	-1.38	-0.43	-0.09	172.16-207.39
D2S155	2.76	-8.96	-4.28	-2.48	-1.04	-0.42	-0.10	194.26-224.07
D2S126	2.83	-16.96	-8.26	-4.84	-1.95	-0.71	-0.16	208.23-248.88
D2S206*	2.90 *SAg*	-0.70	0.41	**0.65**	**0.57**	0.31	0.09	no exclusion
D2S140*	2.98	-18.27	-8.70	-4.96	-1.88	-0.64	-0.14	249.78-274.17
D3S1270*	3.00	-4.75	-1.63	-0.58	0.03	0.10	0.04	0.00-2.37
D3S1293*	3.16	-2.02	-0.23	0.28	0.43	0.26	0.08	34.91-37.28
D3S1260*	3.25	-3.91	-1.19	-0.23	0.30	0.25	0.08	53.51-60.62
D3S11	3.26 *GNAT1*	-14.77	-6.86	-3.84	-1.39	-0.44	-0.08	41.62-77.12

Figure 5.4

Table of lod scores and exclusion regions using microsatellite markers in the set of 11 Sardinian arRP families. * markers analysed by the author

MARKER NAME	CHROMO-SOMAL POSITION	LOD SCORES AT RECOMBINATION FRACTIONS						REGION OF EXCLUSION
		0.01	0.05	0.10	0.20	0.30	0.40	
D3S1312	3.35	-2.28	-0.46	0.09	0.31	0.20	0.06	79.18-81.55
D3S1276*	3.48	-9.16	-3.96	-2.05	-0.63	-0.17	-0.03	99.99-121.32
D3S1278	3.57	-25.43	-12.12	-6.95	-2.66	-0.91	-0.19	109.11-154.14
RHO (3)	3.63 *RHO*	-3.43	-0.89	-0.04	0.41	0.33	0.11	142.01-153.86
D3S1273	3.64	-7.77	-2.12	-0.23	**0.79**	**0.67**	0.24	162.72-170.13
D3S196	3.71	-10.32	-4.46	-2.31	-0.69	-0.16	-0.02	142.05-149.16
D3S1306	3.72	-6.70	-1.60	0.10	**0.94**	**0.73**	0.25	174.55-195.88
D3S1282	3.80	-6.90	-3.42	-1.99	-0.76	-0.26	-0.05	181.46-212.27
GLUT2 (3)	3.85	-11.53	-5.45	-3.06	-1.10	-0.35	-0.07	190.78-221.59
D3S1262*	3.89	-10.92	-5.00	-2.77	-1.00	-0.33	-0.07	151.21-177.28
D3S1265	3.98	-10.72	-3.89	-1.52	-0.03	0.19	0.08	221.23-230.71
D4S179	4.01 *PDEB*	-6.36	-3.02	-1.72	-0.65	-0.22	-0.05	0.00-7.00
D4S431*	4.05	-9.34	-4.12	-2.17	-0.70	-0.20	-0.03	3.29-22.81
HOX7 (4)	4.06	**1.13**	**1.04**	**0.90**	**0.59**	0.30	0.08	no exclusion
D4S403*	4.12	-4.60	-2.01	-1.06	-0.34	-0.10	-0.02	18.14-30.34
D4S391	4.20	-6.59	-2.16	-0.66	0.23	0.27	0.10	35.12-47.32
D4S230	4.21	-8.59	-4.45	-2.73	-1.18	-0.46	-0.11	27.48-59.20
D4S174	4.27	-9.43	-3.95	-1.86	-0.34	0.05	0.05	47.53-64.61
D4S405*	4.27	-9.74	-3.95	-1.87	-0.42	-0.04	0.01	47.53-64.61
D4S428*	4.30 *CNCG1*	-5.35	-2.15	-1.02	-0.24	-0.03	0.00	56.34-68.54
D4S190	4.30	-6.72	-2.86	-1.47	-0.45	-0.11	-0.02	56.34-68.54
GRB1 (4)	4.32	-7.44	-2.93	-1.31	-0.19	0.07	0.04	60.58-72.79
D4S392	4.37	-12.95	-5.74	-3.04	-0.97	-0.26	-0.04	61.43-93.15
D4S189	4.43	-3.42	-0.82	0.08	**0.54**	0.44	0.16	86.37-93.69
ALB (4)	4.45	-4.00	-1.97	-1.16	-0.46	-0.17	-0.04	88.17-100.37
D4S407*	4.55	-8.44	-3.32	-1.51	-0.28	0.01	0.02	109.39-121.59
D4S191	4.62	-9.98	-4.19	-2.10	-0.60	-0.14	-0.02	119.36-141.32
D4S424*	4.68	-10.04	-4.23	-2.11	-0.58	-0.12	-0.01	132.10-154.06
D4S622*	4.84	-8.20	-3.65	-1.96	-0.67	-0.21	-0.04	168.49-185.57
D4S408	4.94	-9.01	-3.83	-1.94	-0.55	-0.13	-0.02	189.71-206.79
D4S171	4.97	-8.96	-4.29	-2.46	-0.95	-0.33	-0.07	203.39-212.20
D4S426*	4.99	-8.11	-3.55	-1.86	-0.58	-0.16	-0.03	200.32-212.20
D5S432	5.10	-0.49	**0.56**	**0.73**	**0.56**	0.29	0.08	no exclusion
D5S419*	5.19	-9.85	-4.04	-1.95	-0.47	-0.06	0.01	28.58-44.26
D5S407*	5.31	-12.98	-5.79	-3.10	-1.03	-0.30	-0.06	43.34-76.94
D5S424*	5.39	-15.45	-6.92	-3.71	-1.23	-0.35	-0.06	56.90-94.98
D5S401*	5.46	-7.10	-2.68	-1.16	-0.16	0.05	0.03	84.18-95.38
D5S421*	5.57	-2.94	-0.01	**0.79**	**0.92**	**0.55**	0.16	108.15-114.87
D5S399*	5.67 *PDEA*	-10.15	-4.31	-2.17	-0.60	-0.12	-0.01	118.95-143.59
D5S400*	5.82	-18.72	-8.79	-4.97	-1.87	-0.64	-0.14	139.63-182.19

Figure 5.4 (cont.)

Table of lod scores and exclusion regions using microsatellite markers in the set of 11 Sardinian arRP families. * markers analysed by the author.

MARKER NAME	CHROMO-SOMAL POSITION	LOD SCORES AT RECOMBINATION FRACTIONS						REGION OF EXCLUSION
		0.01	0.05	0.10	0.20	0.30	0.40	
D5S408*	5.92	-16.42	-7.79	-4.46	-1.72	-0.61	-0.13	161.63-184.03
D6S344*	6.01	-4.22	-1.14	-0.16	0.32	0.26	0.08	2.01-4.08
F13 (6)	6.06	-10.05	-4.65	-2.56	-0.90	-0.29	-0.06	5.86-26.56
D6S277	6.06	-15.28	-7.29	-4.18	-1.61	-0.55	-0.12	0.00-30.70
D6S202	6.11	-9.21	-3.43	-1.43	-0.18	0.06	0.03	13.84-28.33
D6S259	6.13	-11.88	-5.30	-2.82	-0.92	-0.26	-0.05	9.58-40.63
D6S289*	6.14	-7.09	-2.63	-1.06	-0.04	0.13	0.06	21.94-32.29
D6S109	6.16	-11.46	-4.89	-2.44	-0.62	-0.07	-0.02	17.69-44.60
D6S105	6.21	-11.37	-4.72	-2.23	-0.42	0.05	0.06	27.74-54.65
TNF (6)	6.22	-7.34	-2.18	-0.39	**0.63**	**0.58**	0.21	38.03-48.38
D6S291*	6.24 RP14	-7.23	-2.77	-1.20	-0.17	0.05	0.03	42.05-52.40
FTHP1 (6)	6.26	-5.38	-2.17	-1.01	-0.21	-0.01	0.01	46.08-56.43
RDS (6)	6.30 *RDS*	-10.06	-4.70	-2.60	-0.89	-0.25	-0.04	43.77-74.82
D6S294*	6.38	-14.43	-7.06	-4.16	-1.69	-0.62	-0.14	55.72-95.05
D6S275*	6.50	-10.79	-4.87	-2.63	-0.88	-0.25	-0.04	83.99-115.04
D6S268*	6.56	-4.31	-1.75	-0.84	-0.20	-0.03	0.00	106.41-116.76
ARG (6)	6.63	-2.34	-1.02	-0.53	-0.16	-0.04	-0.01	122.55-128.76
D6S292*	6.68	-5.22	-1.51	-0.32	0.30	0.26	0.09	130.54-140.89
D6S290*	6.76	-3.85	-1.35	-0.50	0.00	0.07	0.03	148.70-154.91
D6S264*	6.89	-4.56	-0.95	0.13	**0.55**	0.36	0.11	174.84-181.05
D6S281	6.99	-3.13	-0.73	-0.01	0.29	0.20	0.06	194.95-201.10
D7S531*	7.03	-14.80	-6.88	-3.85	-1.42	-0.48	-0.10	5.52-23.32
D7S513*	7.09	-11.34	-5.33	-3.01	-1.11	-0.37	-0.08	2.32-32.58
D7S493*	7.19	-10.77	-4.35	-2.06	-0.47	-0.05	0.01	22.50-45.64
D7S484*	7.30	-5.40	-1.69	-0.49	0.16	0.17	0.06	49.86-58.76
D7S506*	7.41	-18.17	-8.81	-5.15	-2.06	-0.75	-0.17	52.30-96.80
D7S502*	7.43	-6.43	-2.05	-0.58	0.25	0.27	0.09	73.78-82.68
W30 (7)	7.60	-6.54	-3.17	-1.83	-0.71	-0.24	-0.05	101.50-117.52
D7S479*	7.60 *GNGT1*	-5.14	-1.96	-0.84	-0.11	0.04	0.02	105.06-113.96
D7S480*	7.69	-4.35	-1.27	-0.27	0.23	0.20	0.06	121.62-130.52
D7S500	7.78	-16.52	-7.87	-4.52	-1.74	-0.61	-0.13	122.16-163.00
D7S505	7.89	-8.00	-3.46	-1.80	-0.56	0.01	0.01	154.86-170.88
D7S550*	7.98	-5.16	-1.98	-0.88	-0.15	0.01	0.01	174.98-182.10
D8S262*	8.03	-9.96	-4.66	-2.63	-0.99	-0.34	-0.07	0.00-18.75
D8S265	8.13	-11.33	-4.86	-2.49	-0.76	-0.20	-0.03	6.15-35.39
D8S282	8.21	-6.77	-2.88	-1.46	-0.42	-0.10	-0.01	28.06-40.10
D8S260	8.43	-9.26	-4.05	-2.12	-0.67	-0.19	-0.03	59.51-81.87
D8S273	8.56	-10.71	-4.30	-2.01	-0.41	0.01	0.04	81.14-103.50
D8S257	8.62	-9.13	-3.42	-1.44	-0.15	0.09	0.05	96.29-108.33
D8S266	8.77	-9.20	-3.96	-2.02	-0.59	-0.14	-0.02	116.09-138.45

Figure 5.4 (cont.)

Table of lod scores and exclusion regions using microsatellite markers in the set of 11 Sardinian arRP families. * markers analysed by the author.

MARKER NAME	CHROMO-SOMAL POSITION	LOD SCORES AT RECOMBINATION FRACTIONS 0.01	0.05	0.10	0.20	0.30	0.40	REGION OF EXCLUSION
D8S274	8.88	-4.62	-0.97	0.16	**0.64**	0.45	0.15	141.27-148.15
D9S54	9.03	-14.44	-6.78	-3.85	-1.49	-0.54	-0.12	0.00-22.52
D9S171*	9.25	-14.62	-6.70	-3.67	-1.27	-0.38	-0.07	22.64-59.15
D9S170*	9.75	-8.15	-3.04	-1.25	-0.07	0.14	0.06	117.58-130.72
D9S158	9.98	-12.04	-5.48	-3.01	-1.07	-0.35	-0.07	145.65-164.63
D10S213	10.31	-1.24	-0.09	0.21	0.26	0.14	0.04	no exclusion
D10S196	10.40 *RBP3*	-6.99	-2.57	-1.05	-0.10	0.07	0.03	65.44-78.11
D10S219*	10.58	-7.75	-3.24	-1.61	-0.45	-0.10	-0.01	98.15-110.82
D10S185	10.68	-8.72	-3.56	-1.71	-0.40	-0.04	0.01	114.51-130.80
D10S186*	10.91	-13.08	-5.88	-3.18	-1.09	-0.33	-0.06	147.25-181.64
D10S212*	10.99	-4.36	-1.80	-0.88	-0.24	-0.05	0.00	174.45-181.70
D11S922*	11.02	-7.39	-2.33	-0.60	0.42	0.43	0.16	1.62-7.62
TH (11)	11.02	-12.26	-5.61	-3.06	-1.04	-0.30	-0.05	0.12-18.12
D11S916*	11.51 *ROM-1*	-3.10	-0.69	0.04	0.34	0.23	0.07	76.61-81.11
D11S906*	11.54	-2.13	-0.36	0.15	0.31	0.19	0.06	82.79-84.29
D11S527	11.55	-10.31	-3.88	-1.59	-0.06	0.22	0.10	78.36-91.86
D11S901*	11.58	-5.66	-1.91	-0.67	0.06	0.13	0.05	84.54-95.04
D11S931*	11.62	**0.58**	**0.51**	0.42	0.26	0.13	0.03	no exclusion
D11S873	11.62	-8.75	-3.46	-1.51	-0.15	0.13	0.07	89.28-102.78
D11S919	11.65	-9.75	-3.99	-1.95	-0.52	-0.11	-0.01	92.47-108.97
D11S900	11.66	-14.66	-5.74	-2.60	-0.48	0.04	0.05	88.03-116.53
D11S927	11.70	-16.40	-6.70	-3.18	-0.67	-0.01	0.05	92.77-124.27
D11S939	11.76	-1.62	**0.64**	**1.19**	**1.11**	**0.64**	0.19	no exclusion
D11S912	11.88	-1.24	0.45	**0.85**	**0.79**	0.44	0.13	no exclusion
D11S968*	11.98	-8.74	-3.60	-1.75	-0.45	-0.08	0.00	145.48-154.48
D11S490	11.99	-6.48	-1.54	0.05	**0.81**	**0.63**	0.21	148.54-156.04
D12S81*	12.56	-5.62	-1.88	-0.63	0.08	0.14	0.05	88.30-99.50
D12S78	12.67	-9.80	-3.48	-1.31	0.02	0.20	0.08	105.30-119.70
D12S76*	12.81	-8.47	-3.87	-2.14	-0.78	-0.26	-0.05	124.17-148.17
PLA2 (12)	12.81	-14.07	-6.77	-3.94	-1.59	-0.59	-0.13	116.17-156.17
D12S97*	12.95	-6.29	-2.44	-1.06	-0.13	0.06	0.04	154.25-162.25
D13S175	13.06	-12.06	-5.48	-3.00	-1.05	-0.33	-0.07	5.75-22.65
D13S219*	13.26	-4.02	-1.51	-0.66	-0.12	0.00	0.01	25.35-34.45
D13S155*	13.40	-3.89	-0.83	0.12	**0.52**	0.36	0.11	43.10-49.60
D13S16	13.62	-11.72	-5.17	-2.72	-0.86	-0.24	-0.04	57.25-87.15
D14S50*	14.04	-5.46	-1.70	-0.46	0.25	0.27	0.10	2.70-8.80
D14S64*	14.13 *NRL*	-11.47	-5.43	-3.04	-1.05	-0.30	-0.05	2.08-32.58
D14S80*	14.16	-9.57	-2.69	-0.43	**0.75**	**0.63**	0.21	14.48-25.46
D14S75*	14.28	-18.41	-8.45	-4.64	-1.61	-0.50	-0.09	15.27-55.53
D14S66	14.39	-9.78	-4.42	-2.37	-0.78	-0.23	-0.04	35.51-63.57

Figure 5.4 (cont.)

Table of lod scores and exclusion regions using microsatellite markers in the set of 11 Sardinian arRP families. * markers analysed by the author.

MARKER NAME	CHROMO-SOMAL POSITION	LOD SCORES AT RECOMBINATION FRACTIONS 0.01	0.05	0.10	0.20	0.30	0.40	REGION OF EXCLUSION
D14S68	14.67	-15.15	-7.18	-4.10	-1.57	-0.54	-0.12	66.64-104.46
D14S65	14.84	-8.16	-3.07	-1.28	-0.12	0.10	0.05	100.70-114.12
D14S78	14.88	-8.97	-3.78	-1.87	0.48	-0.07	0.00	104.63-114.39
D15S122	15.06	-13.47	-6.21	-3.45	-1.24	-0.41	-0.08	5.07-23.55
D15S132*	15.41	-4.76	-0.97	0.26	**0.79**	**0.55**	0.17	40.56-48.26
D15S127	15.77	-15.86	-7.79	-4.60	-1.88	-0.70	-0.16	63.29-104.87
D15S120	15.99	-17.59	-8.28	-4.65	-1.66	-0.50	-0.09	89.08-110.20
D16S407	16.13	-11.22	-4.85	-2.44	-0.64	-0.10	0.00	15.43-29.56
D16S403*	16.33	-12.23	-5.09	-2.48	-0.60	-0.06	0.02	29.03-55.72
D16S408*	16.56 *CNCG2*	-7.60	-3.11	-1.51	-0.39	-0.07	0.00	65.40-79.53
D16S261	16.72	-5.00	-1.87	-0.82	-0.15	0.00	0.01	87.90-98.89
D17S849	17.00	-12.43	-5.73	-3.13	-1.05	-0.29	-0.05	0.00-14.56
D17S796	17.11 *RCV1 GUC2D*	-16.55	-7.43	-3.87	-1.19	-0.28	-0.03	22.48-57.84
D17S799	17.25	-11.31	-4.80	-2.39	-0.61	-0.09	0.01	17.62-44.66
D17S798	17.42	-9.47	-3.60	-1.48	-0.07	0.18	0.08	45.73-60.29
D17S791*	17.51 *PDEG*	-7.80	-2.17	-0.30	**0.68**	**0.57**	0.19	59.40-69.80
D17S807	17.68	-7.62	-1.99	-0.13	**0.86**	**0.72**	0.26	80.66-91.68
D17S785*	17.81	-2.72	-0.32	0.39	**0.62**	0.41	0.13	100.09-106.33
D17S784*	17.91	-6.62	-2.25	-0.79	0.06	0.15	0.05	110.88-119.20
D18S59*	18.00	-4.62	-1.51	-0.48	0.09	0.13	0.04	0.00-4.29
D18S62*	18.14	-7.17	-2.17	-0.50	0.43	0.41	0.15	11.61-21.62
D18S35	18.40	-7.73	-3.19	-1.52	-0.33	-0.01	0.02	42.37-55.24
D18S34	18.60	-8.90	-3.72	-1.83	-0.45	-0.05	0.01	65.70-81.13
D18S60	18.72	-7.15	-2.65	-1.05	0.01	0.18	0.07	81.99-94.86
D18S70*	18.99	-7.72	-3.20	-1.57	-0.40	-0.07	0.00	115.41-123.99
INSR (19)	19.13	-5.41	-2.59	-1.44	-0.49	-0.14	-0.02	6.89-20.21
D19S221*	19.32	-8.22	-3.05	-1.20	0.00	0.19	0.08	27.77-41.09
D19S210*	19.95	-8.66	-4.04	-2.27	-0.85	-0.29	-0.06	91.08-107.37
D20S95*	20.17	-6.30	-1.94	-0.51	0.27	0.27	0.09	10.30-21.29
D20S27	20.33	-5.51	-1.74	-0.47	0.28	0.29	0.10	26.96-35.51
D20S101*	20.50	-6.96	-2.52	-1.00	-0.03	0.12	0.05	42.15-53.13
D21S120	21.11	-8.60	-3.22	-1.27	0.01	0.22	0.09	4.27-12.26
D21S210	21.40	-8.89	-3.72	-1.83	-0.46	-0.05	0.01	15.86-30.68
D21S259*	21.68	-7.51	-3.04	-1.45	-0.36	-0.06	0.00	33.68-46.23
D21S168	21.76	-11.98	-5.45	-3.01	-1.12	-0.39	-0.09	28.19-61.26
D22S284*	22.69	-7.59	-2.57	-0.88	0.12	0.21	0.07	34.42-44.95

Figure 5.4 (cont.)

Table of lod scores and exclusion regions using microsatellite markers in the set of 11 Sardinian arRP families. * markers analysed by the author.

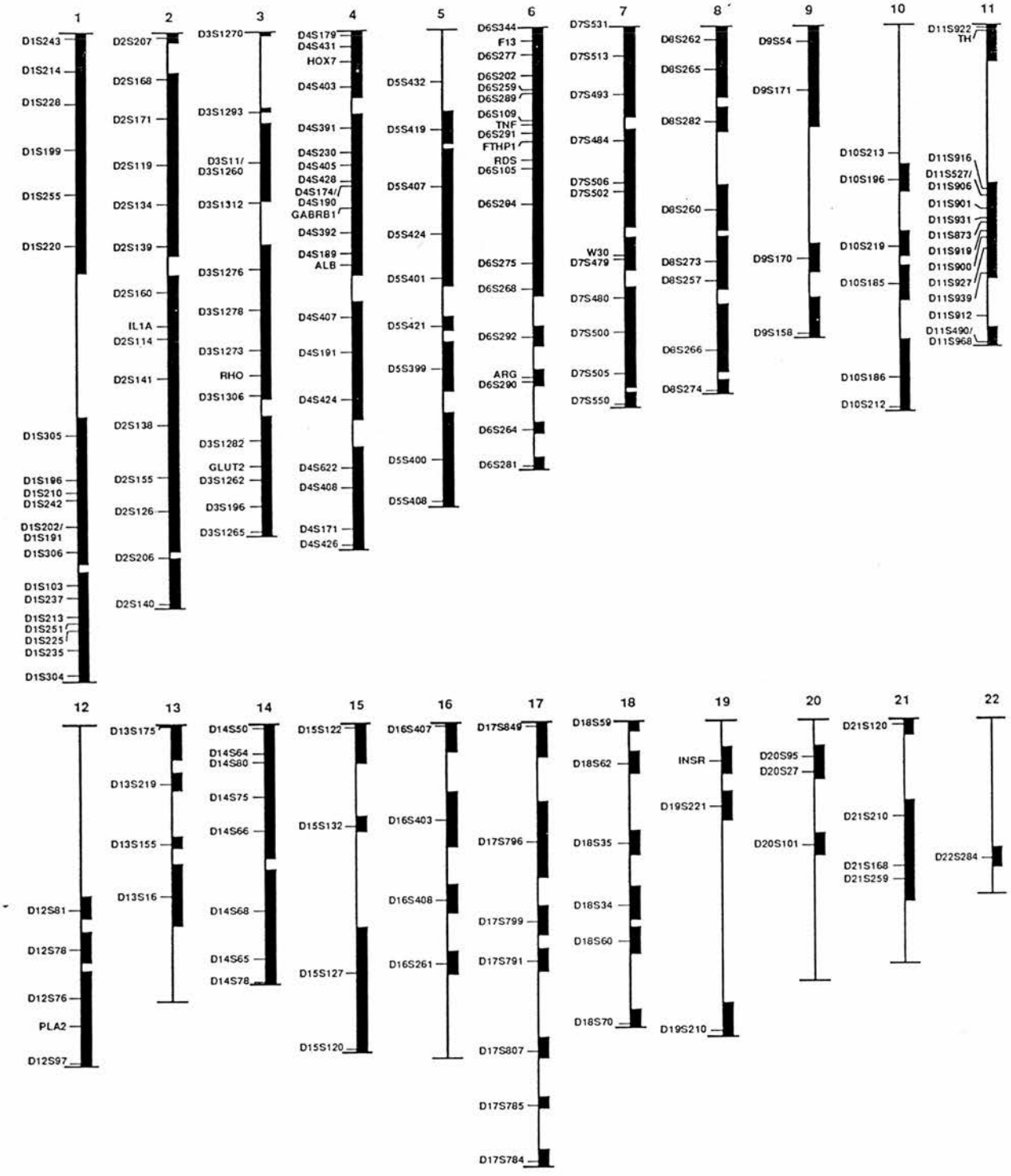


Figure 5.5 - Exclusion map

Pictorial representation of chromosomal regions excluded by the 195 microsatellite markers used in this study, based on the assumption of genetic homogeneity among the eleven families. Dark shaded regions represent regions of exclusion.

CANDIDATE GENE	GENETIC LOCATION	CLOSEST MARKER	PEDIGREE NUMBER
Phosducin	1q25-q32.1	D1S202**	**2** (0.60), **3** (1.20),**4** (0.38), **8** (0.40)
RP12	1q31-q32.1	D1S306*	**2** (0.60)
Arrestin	2q37.1	D2S206**	**1** (0.30), **4** (0.34), **8** (0.37)
Rod transducin - alpha subunit	3p21.3-p21.2	D3S11*	-
Rhodopsin	3q21.3-q24	RHO***	**7** (0.85)
Phosphodiesterase-beta subunit	4p16.3	D4S179**	**1** (0.60)
Cyclic nucleotide gated channel - alpha subunit	4p12-cen	D4S428**	**1** (0.30) **6** (0.30), **8** (0.41)
Phosphodiesterase-alpha subunit	5q31.2-q34	D5S399*	**3** (0.30), **8** (0.43)
RP14	6p21.3-p21.2	D6S291***	**2** (0.31), **6** (0.30), **8** (0.31)
Peripherin	6p21.2-cen	RDS***	**2** (0.30)
Transducin - gamma subunit	7q21.3	D7S479**	**1** (0.58), **5** (0.76)
Interphotoreceptor retinol binding protein	10q11.2	D10S196**	**1** (0.30), **2** (0.60), **4** (0.37), **5** (0.45), **6** (0.30)
ROM-1	11q13	D11S916*	**4** (0.40), **5** (0.57), **6** (0.30), **8** (0.50)
Rhodopsin kinase	13q34	No markers nearby	
Cyclic nucleotide gated channel - gamma subunit	16q13	D16S408**	**1** (0.30), **3** (0.30), **4** (0.46), **13** (0.31)
Recoverin Guanylate cyclase	17p13-p12 17p13.1	D17S796* D17S796**	**1** (0.73)
Phosphodiesterase-gamma subunit	17q21.1	D17S791**	**2** (0.30), **3** (1.20), **5** (1.01)

Figure 5.6

Table showing individual families with positive lod scores > 0.30 (in brackets) for markers close to specific candidate loci. *** within 1 cM of the candidate gene; ** within 5 cM; *within 10 cM.

gene in question, to establish if the linkage is real, or has occurred by chance with the initial marker studied. If the results still suggest linkage in any families, SSCP analysis could be used to scan the exons of the candidate gene in the obligate carriers and affected members of the family.

Recently, work has begun examining the phosphodiesterase gamma subunit gene on chromosome 17q21.1. Individual lod scores in families (see **fig. 5.6**) highlighted this as a possible region of interest, with two families achieving lod scores > 1.0 for the closest marker studied, D17S791. Unfortunately this gene has not been precisely mapped, so an approximate genetic location was taken from the Genetic Location Database (LDB), and placed the gene within 5 cM of D17S791. Multipoint lod scores were calculated by P. Teague, using the three closest markers studied, D17S798, D17S791 and D17S807. The lod scores are shown in **figure 5.7**, with the maximum likelihood location given as a distance from the *PDEG* gene. Dr. F. Manson (MRC Human Genetics Unit) is examining this gene, which codes for 87 amino-acids, in families 2, 3, 5, 7, 11 and 12. As the exact gene location is not known and only three markers have been used in the linkage analysis, any families showing signs of linkage within 42 cM of the approximate gene location are being studied. This gene is a strong candidate for arRP (see section **1.6.1**) but previous SSCP analysis in 142 arRP patients failed to reveal any mutations (Hahn, Berson and Dryja, 1994).

5.5 - CHROMOSOME 14q11

Close examination of the results of the 195 markers studied revealed two chromosomal regions where further investigation was required. The first of these was a region of chromosome 11q, where exclusions were not possible due to low positive lod scores for several microsatellite markers, including D11S902, D11S914 and D11S903. Several retinal disorders, including Usher syndrome type I, Bardet-Biedl syndrome and Best's vitelliform macular dystrophy, have been linked to this region, and linkage was investigated using further microsatellite markers, by Dr. D. Mansfield (results not shown). However, no significant results were obtained when further markers were studied in each of the families.

The other interesting result was obtained with microsatellite marker D14S80, in chromosomal region 14q11.1-q11.2. When the lod scores for each individual family were

PEDIGREE NUMBER	MAXIMUM LOD SCORE	APPROXIMATE DISTANCE FROM PDEG
1	0.72	56 cM proximal
2*	0.60	18 cM proximal
3*	1.17	0 cM
4	0.23	110 cM proximal
5*	1.21	18 cM proximal
6	0.00	
7*	2.05	42 cM distal
8	0.40	96 cM distal
11*	0.85	42 cM distal
12*	0.59	42 cM distal
13	0.35	68 cM distal

Figure 5.7

Table showing maximum multipoint lod scores, and their location, for each family, using markers flanking the *PDEG* locus. Direct DNA sequencing of this gene is being carried out in families marked with an asterisk.

studied, it was clear that a subset of five of the eleven families were linked to this marker, with no recombinants, while the remainder appeared unlinked. The summed lod score of these five families gave a significant result, Z = 4.17, at $\theta = 0.00$. Heterogeneity analysis, using the HOMOG program (Ott, version 3.3) confirmed that a subset of families ($\alpha = 0.65$) were linked to this locus ($\chi^2 = 7.156$, p= 0.03).

Further Genethon microsatellite markers in the flanking region (D14S264, D14S64, D14S275, D14S262 and D14S252) (see **fig 5.8**), spanning a total of 6 cM, were studied in the set of eleven families, and the results analysed in the form of haplotypes, and multipoint lod scores (as calculated by P. Teague). Linkage to this region was excluded in six of the families, with support for linkage heterogeneity ($\alpha = 0.24$) at borderline significance (p=0.065). Haplotype analysis revealed a region of homozygosity in family ARRP1, suggesting identity-by-descent (IBD), despite no report of consanguinity in this pedigree. None of the affected individuals from other pedigrees that appeared linked to this region showed homozygosity for these markers. However, further investigation of this family by Dr. Fossarello was able to confirm the presence of consanguinity, as the parents are second cousins (co-efficient of inbreeding, F=0.0156). This suggested that the two affected sibs studied (BB and BD) had inherited an IBD region from a common ancestor. This region of homozygosity, shown in **figure 5.8**, spans four of the markers (D14S64 - 4 cM - D14S80/D14S275 - 2 cM - D14S262, Gyapay *et al*, 1994), covering a total of 6 cM. In affected children of a consanguineous mating, several cM surrounding the disease gene are usually identical-by-descent. Other regions in the genome will also be homozygous by chance, but will vary between affected and unaffected children.

In family ARRP1, the coefficient of inbreeding is 0.0156, meaning that at 1.56% of autosomal loci the offspring will be predicted to have inherited two identical copies of a gene from each parent, both inherited from a common ancestor (Bittles *et al*, 1991). As the region of homozygosity around D14S80 spans 6 cM, representing 0.16% of the genome, this accounts for approximately 10% of the loci where IBD is expected in the offspring of this family. This region of homozygosity could have occurred by chance, as the homozygous alleles present are commonly found within the Caucasian population. However, based on the allele frequencies from the six Sardinian families tested with these markers, the expected proportion of individuals homozygous for all four alleles is less than 0.01%, assuming no linkage disequilibrium.

Interest in this region of the genome increased when it was discovered that a candidate gene for retinal disorders, the neural retina leucine zipper gene (*NRL*) (introduced in **section**

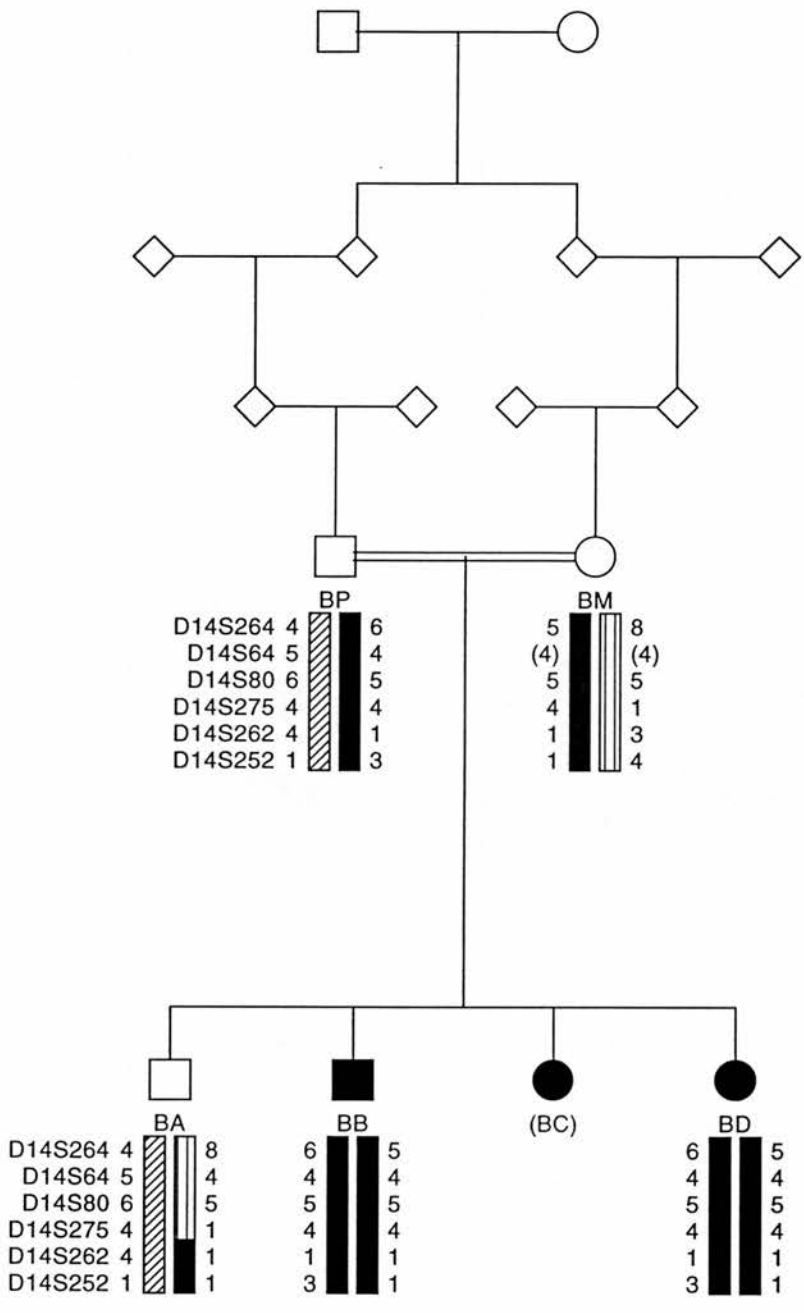


Figure 5.8

Pedigree of family ARRP1, showing haplotypes for markers on chromosome 14q.

1.5.6), had been mapped to the same cosmid as the microsatellite marker D14S64, in chromosomal bands 14q11.1-q11.2 (A. Swaroop, personal communication). The highly conserved, retina-expressed *NRL* gene, which may have a role in the regulation of rhodopsin expression (Rehemtulla *et al*, 1996), comprises three exons, the first of which is untranslated, covering a total coding region of around 700 base pairs. The comparatively small size of this gene made it ideal for screening for mutations in this affected family. This gene, and the results of candidate gene screening using SSCP and heteroduplex analysis, and direct DNA sequencing, are detailed in **Chapter Six**.

5.6 - HOMOZYGOSITY MAPPING

At the start of this study no DNA was available from the third affected sib of family ARRP1, individual BC. When interest increased in this family we were able to obtain DNA for individual BC. Initial analysis with the microsatellite markers shown in **fig. 5.8** revealed that homozygosity was also present in this individual for markers D14S64, D14S80 and D14S262 (see **fig. 5.9a**). D14S275 proved difficult to PCR amplify in this individual, and two different results were originally obtained. However, following resynthesis of the primers for this marker, the PCR reaction was successful, and appeared to show that BC was not homozygous for D14S275. This result was repeated and confirmed, and suggested three possible explanations; (i) that the marker order was incorrect, and that D14S275 lay outwith the region of homozygosity; (ii) that a double crossover had occurred, which was highly unlikely in the small interval involved; or (iii) that the homozygosity seen in the other two individuals was not due to identity-by-descent. The maximum multipoint lod score for family ARRP1 (including BC) for the markers studied was 2.4, at a locus defined by the microsatellite markers D14S262 and D14S252 (see **fig. 5.10**). This result still suggested linkage, although the region of interest had shifted to D14S262 and D14S252, distal to D14S64 and the *NRL* candidate gene (assuming the published order of markers is correct).

To investigate this region in more detail further Genethon markers were studied, following the publication of their sequences in 1996 by Dib *et al*. The markers chosen were D14S972, D14S1032, D14S1041 and D14S1042 (see **fig. 5.9b**). This publication also

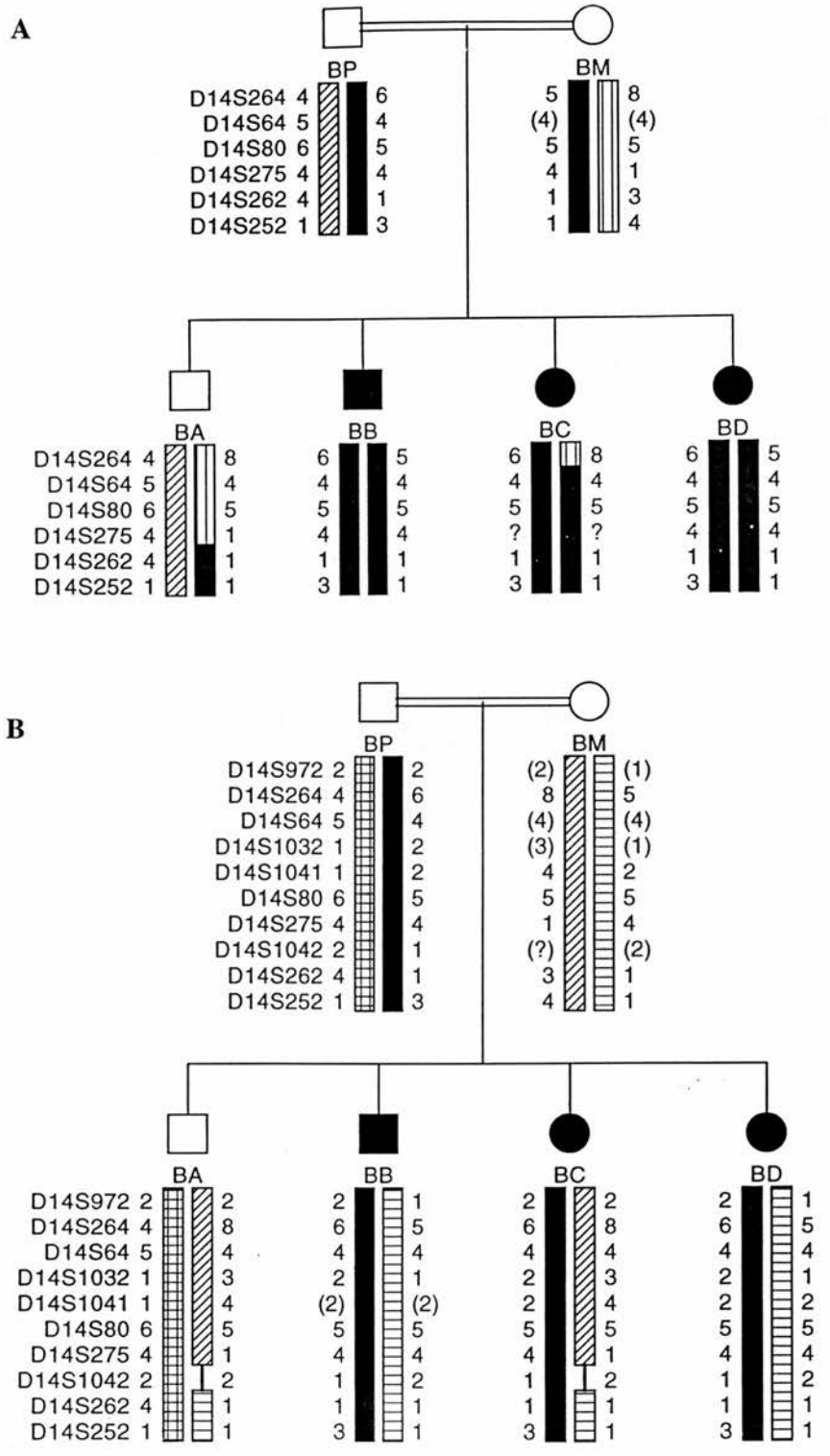


Figure 5.9

A - Initial haplotype analysis for family ARRP1, including individual BC

B - Haplotype analysis of family ARRP1 with further markers on chromosome 14q, showing lack of homozygosity in individual BC

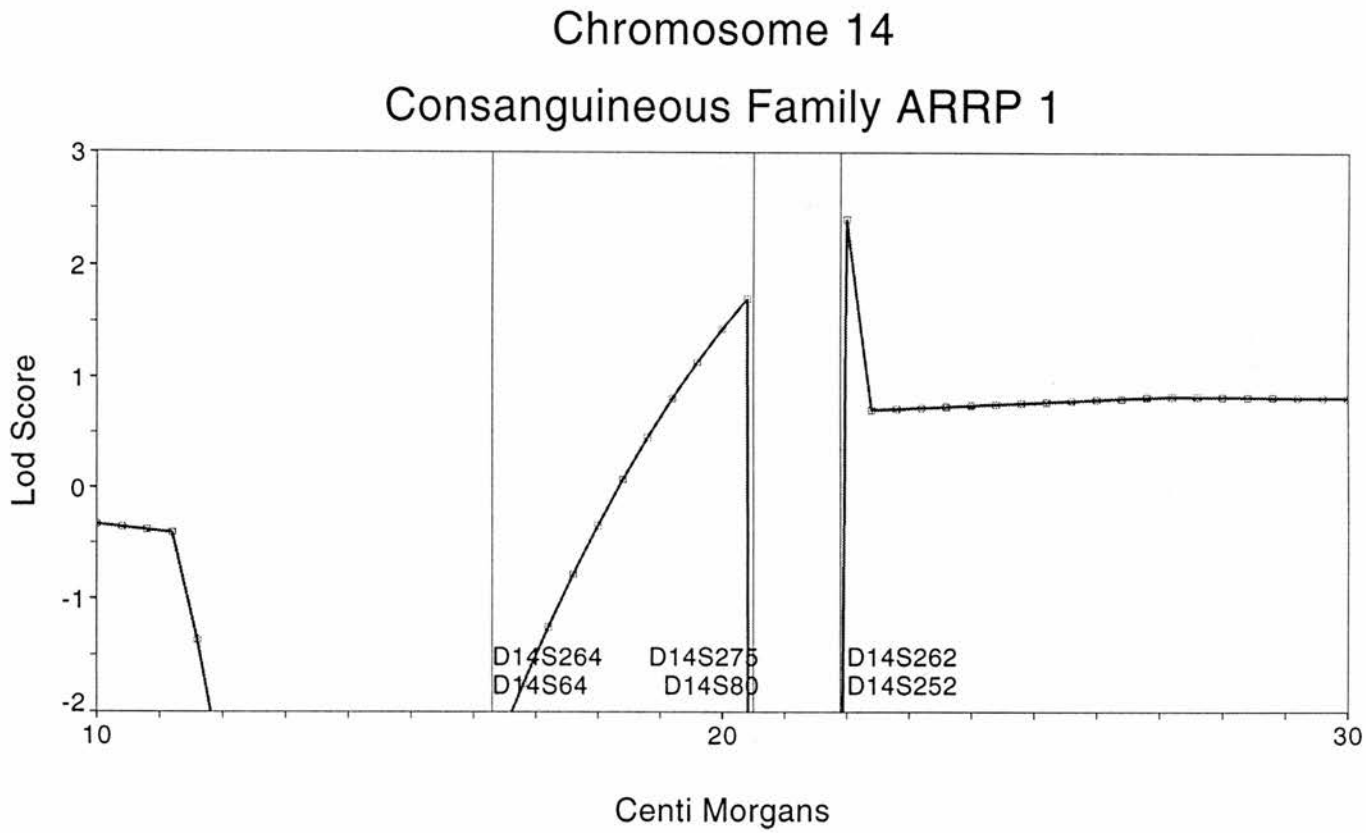


Figure 5.10 - Multipoint analysis for family ARRP1 with markers as indicated.

Marker positions, represented by vertical lines, correspond to the distances of the markers from the centromere of chromosome 14.

confirmed the order and distances between the markers (all at odds >1000:1) as shown below, with the exception of D14S252, which was not included in the 1996 map:

D14S972 - 1.2 cM - D14S264/D14S64 - 0.6 cM - D14S1032 - 0.1 cM - D14S1041 - 3.6 cM - D14S80/D14S275 - 1.3 cM - D14S1042 - 0.1 cM - D14S262

The results of these markers indicated that the apparent homozygosity in BC was due to uninformative alleles from D14S64 and D14S80 in the mother, although the family remains linked to markers below D14S275. Indeed, BC has inherited a different maternal haplotype from BB and BD proximal to D14S1042, but this only became clear after analysis of fully informative microsatellite markers in this region. However, this family still shows no recombination with markers distal to D14S275 in all three affecteds, suggesting that another gene on chromosome 14q could be causing the arRP in this family.

5.7 - DISCUSSION

This project has attempted to map genes causing arRP within a genetically isolated population. The eleven families chosen for study were from south-central Sardinia, a genetic outlier with a high proportion of arRP, and comparative and atypical distribution of homogeneity for other genetic diseases. For example, over 95% of all Sardinian β-thalassaemia chromosomes are accounted for by one mutation (β^{39}) (Rosatelli *et al,* 1992), and only five common haplotypes (one of which accounts for over half of all Sardinian mutations) have been identified for Wilson's disease (Figus *et al*, 1995). The Sardinian population is also thought to be comparatively homogeneous with respect to glucose-6-phosphate dehydrogenase (G6PD) mutations, and has the lowest frequency of Rhesus negative genes in the Mediterranean (Bernardinelli *et al*, 1994). Diseases associated with the unique distribution of HLA-types in this population are also particularly prevalent, including insulin-dependent diabetes mellitus associated with HLA-B18-DR3, and multiple sclerosis, associated with DR4 and DQw3 (Bernardinelli *et al*, 1994). Based on this data, the assumption was that fewer arRP mutations would be present within this population, hence increasing the likelihood of detecting linkage to specific loci.

The unique genetic composition of Sardinia is probably largely due to random genetic drift. A small founder population, estimated at 700-1800 at saturation in the late Paleolithic period around 9,000 years ago expanded to 150,000 by 1,500 B.C., and was augmented with colonisation by Phoenicians and Carthagians from North Africa about 2,000 years ago (Bernardinelli *et al*, 1994). The geographic isolation of the island, with little migration or

immigration in the last two millenia, has not diluted the effects of drift, which has altered the gene frequencies within this population - some alleles in the founder population have been enriched, while rarer mutations have been lost. A combination of drift and selection (due to heterozygote advantage) have influenced the prevalence of disorders such as β-thalassaemia and G6PD.

A similar genetic situation is seen in Finland. Archaeological evidence shows that the country has been inhabited for up to 10,000 years, but the present population of 5 million are thought to be largely descended from a small number of founders (including Germanic and Baltic immigrants), approximately 2,000 years ago (de la Chapelle, 1993). Subsequent immigration has had little effect on this isolated population. In Finland there is a high incidence of over thirty (mostly autosomal recessive) genetic diseases which are rare elsewhere, including Usher syndrome type III, aspartylglucosaminuria and diastrophic dysplasia (de la Chapelle, 1993). Where the defect has been identified, most of these conditions have been shown to be highly homogeneous. Again, this is probably due to genetic drift, enriching the population for specific genetic mutations. This is further confirmed by the low prevalence of other diseases that are comparatively common elsewhere, such as cystic fibrosis and phenylketonuria, where the mutations have been lost in this population through drift.

Within the Finnish population, linkage disequilibrium - the non-random association of alleles at linked loci - has allowed for finer localisation of causative genes than the expected 1-2 cM in conventional linkage mapping. Linkage disequilibrium is influenced by the number of generations the mutation has passed through in the population, and the recombination frequency of the region the gene lies in. When this information is known, calculations can be made regarding the position of the gene relative to flanking markers (Jorde, 1995). Indeed, linkage disequlibrium is a powerful tool for fine-mapping disease loci, and is most likely to be successful in isolated and well-defined populations, such as Sardinia and Finland. Other isolate populations, including the Old Order Amish, French Canadians, Ashkenazi and Yemenite Jews, and Louisiana Acadians have been used for such studies (Zlotogora, 1994).

Genetic mapping with eleven Sardinian arRP families, using 195 genome-wide microsatellite markers, was unable to identify any consistent linkage. Large exclusions (62% of the genome) were possible on the assumption of genetic homogeneity. Although some positive lod scores were obtained, none were statistically significant ($Z > 3.0$) with all eleven families, suggesting nonallelic heterogeneity within this population. This may be

due to the genetic heterogeneity observed between communities within Sardinia, especially between highland and lowland villages (Workman *et al*, 1976). High consanguinity rates, especially among the upper classes, and in highland communities (Workman *et al*, 1976), in combination with a founder effect and genetic drift will have contributed to this local differentiation. Thus, the population could be thought of as a being composed of many smaller, previously isolated communities.

One region of interest, in chromosomal region 14q11, was noted from the results with microsatellite marker D14S80. This marker appeared linked in a subset of five families, and was investigated using further flanking markers. The results initially revealed a region of homozygosity covering 6 cM in two affected members of a consanguineous family. As this region is thought to include the *NRL* gene, a candidate for retinal dystrophies, it was decided to investigate *NRL* within this family, using mutation analysis. This gene, and the results of the analysis, are described and discussed in **Chapter Six**. However, when DNA became available for analysis from the third affected individual in this pedigree, the region of homozygosity was decreased, and analysis with further flanking markers confirmed that the original homozygosity was due to uninformative parental markers. Nevertheless, this family still show no recombination with markers distal to D14S275 (see **fig. 5.9b** and **5.10**), suggesting that another gene in this region could be causing the arRP in this family. One possible candidate is the *Chx10* gene at 14q24.3 (Burmeister *et al*, 1996), which is briefly discussed in **Chapter Six.**

Following the lack of linkage with genome-wide microsatellite markers, a closer examination of the individual family lod scores was undertaken for each marker. As many candidate genes for retinal dystrophies have been mapped, careful consideration was given to microsatellite markers within 10 cM of such loci. Initial indications suggest that several families merit further investigation for each locus and work has already begun examining the *PDEG* gene in several of the Sardinian families. More closely linked markers are now available for many of the loci, and these could be studied in families where exclusion is not possible, to confirm or refute any tentative linkage. This positional candidate approach could also be used for other candidate genes, such as rhodopsin kinase, recently mapped to 13q34, a region where no microsatellite markers have previously been studied in the Sardinian families. This method does not rely on genetic homogeneity, but considers each family individually. However, as most of the families available for study are small, and not consanguineous, the lod scores obtained will be low.

For further analysis of data with the set of families used in this study, it may be useful to consider the power of the study - that is, to estimate the probability of detecting linkage with a given set of families and markers. Several computer simulation programs are available for this purpose, including SLINK and SIMLINK (discussed by Terwilliger and Ott, 1994). This information could be particularily useful when looking at a subset of families, as the lod scores may not be significant ($Z > 3.0$), but are still unlikely to have occurred by chance alone. The subject of significance of linkage results has recently been debated by Lander and Kruglyak (1995, 1996) with reference to multifactorial traits, but several of the points can also be applied to Mendelian traits (Curtis, 1996). In particular, they point out the importance of accurately reporting, and following-up, results even if they are not highly significant, as there is no way of distinguishing real results from false positives until they have been further investigated. Furthermore, the possibility of oligogenic inheritance for RP highlights the importance of careful interpretation of results from linkage studies.

The increase in density of marker maps, such as the Genethon human genetic linkage map, with over 5,000 microsatellites and an average interval size of 1.6 cM (Dib *et al*, 1996), facilitates gene localisation. The location of the gene can subsequently be narrowed using information from linkage disequilibrium studies, homozygosity mapping, animal models, and transcript maps. However, this method can still be time-consuming. Candidate gene screening is another way to identify causative genes, but care must be taken in choosing the type of screening technique employed, especially when dealing with a limited number of samples. In a population such as the Sardinian families used in this study, it is very possible that there is a limited number of mutations, which may be rare in other populations. This can also be seen as a disadvantage of working with a founder population, as common mutations in outbred populations, which are of more use in genetic testing and diagnosis, may be under-represented in the founder population.

This project has highlighted the problem of genetic heterogeneity in arRP, even within a genetically isolated population, and emphasises the increased power of consanguineous pedigrees in linkage mapping. Since the start of this work, DNA from a further five Sardinian families with arRP has become available. These include a family where the parents are second cousins, and a large, extended, multi-generational family. The latter could be used for DNA pooling, a method employed to identify linkage in Bardet-Biedl syndrome by Carmi *et al* (1995a). This approach involves combining DNA samples from affected members of the pedigree, and comparing the alleles observed in this group (for a specific marker) with the alleles observed in a pooled group of unaffected members. If the

alleles in the affected group show a shift towards homozygosity, this suggests linkage disequilibrium with the disease phenotype, and hence linkage to the disease.

This kind of approach could also be used in more homogeneous populations, such as the island of Tristan de Cunha, where most of the inhabitants are direct descendants of 15 original settlers (Roberts, 1980). The estimated carrier frequency of arRP in this population is as high as 1 in 6, making it ideal for linkage mapping of arRP. However, even in a single large arRP kindred, genetic heterogeneity can be present, as identified by van Soest *et al* (1994) and Leutelt *et al* (1995). Both groups studied extended pedigrees, and found genetic and clinical heterogeneity between separate branches of each family. They found that the PPRPE phenotype correlated with linkage to chromosome 1q, and that the branches with normal RP were not linked to this region, suggesting that two different arRP genes are segregating within these families. These studies emphasise the need for precise clinical evaluation of patients, which may also help establish correlations between genotype and phenotype.

CHAPTER SIX

MUTATION DETECTION IN *NRL*

6.1 - INTRODUCTION

Chapter five detailed the results from genome-wide linkage mapping, using eleven arRP families from south-central Sardinia. No significantly positive lod scores ($Z > 3.0$) were obtained, but closer examination of the results indicated a positive result on chromosome 14q, for marker D14S80, in a subset of families. Further microsatellite markers in the surrounding region revealed the possibility of a region of homozygosity-by-descent, in one consanguineous family, ARRP1. A candidate gene for retinal degenerations, the neural retina leucine zipper gene (*NRL*), lies close to D14S80, and was considered a good candidate for arRP in this inbred family. This chapter describes work carried out to identify *NRL* mutations in this family, using three methods of mutation detection. The *NRL* gene is also discussed, with reference to its possible role in the retina, and regulatory mechanisms.

6.2 - NRL

In higher eukaryotes, gene expression is partially controlled through regulation of transcription, involving *cis*-acting regulatory sequences recognised by *trans*-acting DNA-binding proteins. At least ten classes of such nuclear transcription factor proteins are known, including the paired-box (PAX), homeobox (HOX), and basic domain-leucine zipper (bZIP) families (for review see Mitchell and Tijan, 1989). Mutations in the genes encoding some of these factors are known to result in specific human diseases. For example, mutations in *PAX-3* and *PAX-6* cause Waardenburg's Syndrome (Tassabehji *et al*, 1992) and aniridia respectively (Ton *et al*, 1991), and synpolydactyly is associated with mutations at the *HOXD13* locus (Muragaki *et al*, 1996).

Using a subtractive hybridisation technique, Swaroop *et al* (1992) identified a new member of the bZIP family of transcription factors. A human cDNA (AS321) was isolated from an enriched adult retinal library, and was found to be exclusively and highly expressed in all cells of the adult neural retina, particularly in the photoreceptor cells. This cDNA was used to isolate additional clones from a retinal library, most of which had 1.3 kb transcripts, although one clone of 2 kb was found, due to an extended 3' untranslated region. Another cDNA isolated, designated DD10, contained an in-frame deletion of 105 amino-acids from the full-length open-reading frame (ORF) of 237 residues, and has since been shown to be present in the human retina (Rehemtulla *et al*, 1996). The deduced polypeptide sequence from the normal ORF contained several conserved phosphorylation sites, and a periodic

repeat of leucine residues preceded by a putative basic DNA-binding region. The C-terminal region showed strong sequence similarity to the jun/fos proto-oncogenes of the bZIP family, with 75% similarity in one exon to v-maf, suggesting functional similarity in this region.

The gene was also shown to be evolutionarily conserved, with homologous transcripts detected in the retina of several species from baboons to frogs, and low-stringency Southern blotting identified homologues in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. Swaroop *et al* proposed the name *NRL* (Neural Retina Leucine zipper) for this gene, which was later mapped to 14q11.1 - q11.2 by *in situ* hybridisation to chromosome spreads (Yang-Feng and Swaroop, 1992). Analysis of the murine homologue, *nrl*, showed around 90% homology to the human gene in the coding region, with significant homology in the 5' and 3' untranslated regions (Farjo *et al*, 1993). The basic motifs and leucine zipper are identical, with strong conservation of the Proline/Serine/Threonine-rich putative transactivation domain (see **fig 6.1**), which is also seen in the jun family of proteins. In mouse, the 3' untranslated region contains a polymorphic AGG repeat, which shows variation between different strains. A reduced form of this repeat is also seen in the human gene (Swaroop *et al*, 1992).

The *NRL* gene is composed of three exons covering around 6 kb, the first of which is untranslated. The 5' upstream region contains a consensus CAP sequence, a CCAAT box, binding sites for AP-2, AP-1 and two helix-loop-helix binding proteins, suggesting complex regulation of this gene (Farjo *et al*, 1993).

6.3 - bZIP PROTEINS

In bZIP proteins, the leucine zipper region allows homo- or hetero-dimerisation, via hydrophobic interaction between the two parallel α-helices. The adjacent basic region is thought to extend the α-helices and make sequence-specific contact with the phosphodiester backbone of DNA (Ellenberger *et al*, 1992). This dimerisation allows increased specificity and combined control of transcriptional responses.

Proteins in the bZIP family include Fos, Jun and related proteins, which, when dimerised, can bind to cAMP response element (CRE)-like sites, and activator protein-1 (AP-1) sites (Kerppola and Curran, 1994a). Nrl and c-maf (the cellular equivalent of the viral maf transforming proto-oncogene) are designated as members of a new sub-family of bZIP proteins, along with MafK, MafF, and *kr* (the kreisler gene, involved in segmentation) (Kerppola and Curran, 1994a). Kerppola and Curran (1994a) showed that truncated Maf or

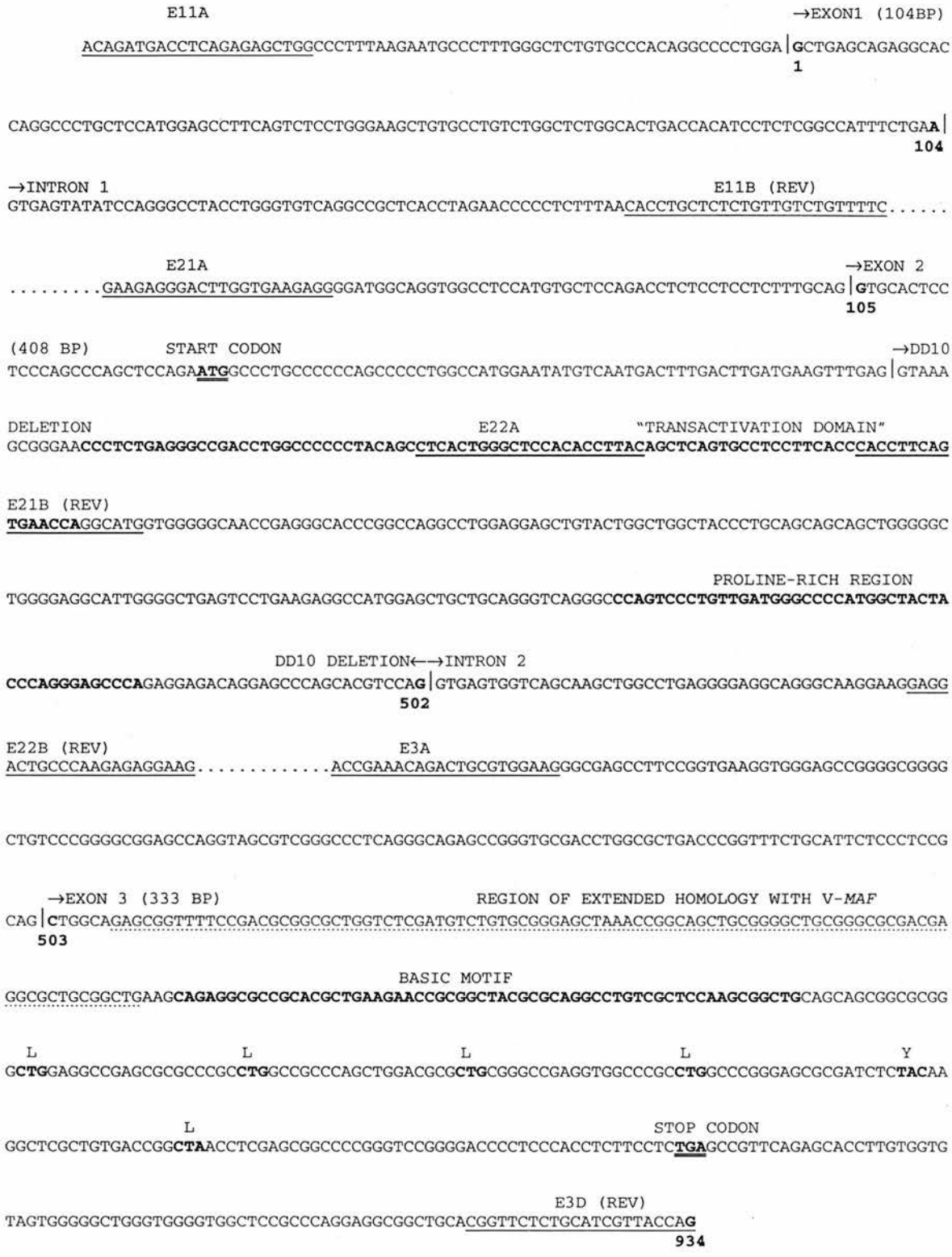


Figure 6.1 - *NRL* coding sequence and exon/intron boundaries. PCR primer sequences are underlined with continuous lines, specific domains and codons are in bold. The region of extended homology with v-*maf* is underlined with a dotted line. The polyadenylation addition signal (not shown) is located 289 bp 3' to the E3D primer. Numbers below bold single nucleotides refer to nucleotide numbers in the Genbank Human NRL gene product mRNA sequence, accession number M81840.

Nrl proteins (containing the leucine zipper and basic regions) can homodimerise and bind to a range of AP-1 and CRE sites. These peptides have also been shown to associate with each other *in vitro*, and with jun and fos, but not with more distantly related members of the bZIP family. The resultant dimers can then bind to AP-1 sites with varying affinities. The consensus sequence that Nrl homo- and hetero-dimers bind to, designated the NRE - Nrl response element - is shown below, where the central section is similar to an AP-1 site or CRE-site (Kerppola and Curran, 1994a):

TGC(TGA(G)CTCA)GCA

Rehemtulla *et al* (1996) identified an NRE-like sequence, among other DNA-binding sites, in the promoter region of the rhodopsin gene. This NRE is highly conserved in human, bovine and mouse rhodopsin. This suggested that Nrl could have a role in the transcriptional regulation of this photoreceptor-specific protein. Electrophoretic mobility shift assays showed that retinal nuclear extract contains proteins that interact with the rhodopsin NRE, and addition of Nrl-specific antiserum revealed that Nrl (or an antigenically similar protein) was one of the proteins binding to this site (Rehemtulla *et al*, 1996).

Co-transfection of rhodopsin and Nrl in CV-1 cells, which do not naturally express these proteins, has shown that Nrl can transactivate the rhodopsin promoter (Rehemtulla *et al*, 1996). However, co-transfection with the naturally occurring truncated form of Nrl (DD10) did not cause expression. The presence of this alternatively spliced *NRL* mRNA in the human retina has been shown using reverse-transcription and sequence analysis (Rehemtulla *et al*, 1996). When DD10 was transfected with normal Nrl, the DD10 protein actively inhibited Nrl-mediated transactivation of the rhodopsin promoter, probably by competing with Nrl homodimers for binding to the NRE (Rehemtulla *et al*, 1996).

The presence of an alternatively spliced isoform is a common strategy for regulation of transcription factor activity (reviewed by Lopez, 1995). A similar situation is seen with another member of the bZIP family, the cAMP response element modulator (CREM), of which three isoforms are known to exist. As with the truncated form of Nrl, two of these isoforms lack the transactivator domain, and inhibit binding of the normal protein, by binding to the CRE as homodimers or heterodimers (Foulkes and Sassone-Corsi, 1992). It seems likely that the truncated isoform of Nrl acts as a negative regulator of wild-type Nrl.

Liu *et al* (1996) have analysed the expression of murine Nrl (nrl) by *in situ* hybridisation of an antisense nrl probe to embryonic, neonatal and adult mouse sections, and reverse transcription PCR (RT-PCR). Expression was first detected in 12.5 day embryos in the brainstem, spinal cord and cephalic mesenchyme. This expression expanded throughout the

nervous system during neurogenesis, but was always restricted to specific differentiated neurons that had migrated to their final positions. Postnatally, nrl was down-regulated in the brain and spinal cord, and three months after birth only a weak signal was detected in the forebrain by RT-PCR analysis. However, high levels were detected in all nuclear layers of the adult retina and the photoreceptor inner segments. These results suggest that nrl has a role in the maturation and/or establishment of the differentiated neuronal phenotype, and that in the retina it may maintain the differentiated state of cells by regulating the expression of specific genes.

6.4 - PCR OF *NRL* EXONS

The *NRL* gene is composed of three exons, the first of which is not translated. The start codon is situated near the beginning of the second exon, which is 408 bp in length, encoding 127 amino-acid residues. This exon also contains the proline/serine/threonine-rich putative transactivation domain, and the proline-rich activation domain, both of which are deleted in the naturally occurring truncated form of Nrl mRNA (DD10). The third exon, comprising 333 bp from the splice site to the stop codon, coding for 110 residues, contains three specific regions. The first of these is the region of extended homology with v-maf, the next is the basic motif, thought to be involved in DNA binding, and the final region is the leucine zipper, containing five highly conserved leucine residues (see **fig. 6.1**).
The PCR primers designed to amplify each exon, for mutation analysis and DNA sequencing, are shown in **fig 6.1**. They produce fragments of 254 bp and 573 bp respectively, for exons 1 and 3. Exon 2 was amplified in two overlapping fragments, 2a (271 bp) and 2b (337 bp). For mutation analysis by SSCP, the 2b fragment was amplified and analysed electrophoretically by PAGE, both without digestion and following digestion with *Pvu*II, producing fragments of 133 bp and 204 bp. The exon 3 fragment was analysed both without digestion and following digestion with *Taq*I, giving fragments of 186 bp, 254 bp and 133 bp.
Initially, problems were encountered with PCR amplification of exon 3. This was probably due to the very high GC-content of this region (74%), as stretches of DNA with a high GC-content are often difficult to amplify, due to the formation of complex secondary and tertiary structures which inhibit the primer binding to the template. The PCR primers were redesigned, and different magnesium concentrations, annealing temperatures and primer

concentrations used, but the best results were achieved when 10% dimethylsulphoxide (DMSO) was included in the reaction. DMSO is thought to improve PCR by eliminating non-specific amplification, altering the thermal activity profile of the DNA polymerase, or destabilising secondary structures (Sidhu, Liao and Rashidbaigi, 1996), and its use with *Taq* polymerase was first reported by Shen and Hohn (1992).

6.5 - SSCP MUTATION DETECTION

The single-stranded conformation polymorphism technique (SSCP, introduced in section **1.7**) developed by Orita *et al* (1989) has proved effective at detecting polymorphisms and mutations in many genes, including the cyclic nucleotide gated channel (*CNCG1*, Dryja *et al*, 1995), peripherin/*RDS* (Farrar *et al*, 1991) and the alpha subunit of phosphodiesterase (*PDEA*, Huang *et al*, 1995) in arRP. SSCP detects alterations in single-stranded DNA (from denatured PCR products) as a result of their differing mobilities in a gel matrix. The mobility is altered by the nucleotide sequence of the fragment, which in turn alters the intramolecular interactions that produce the three-dimensional folded structure. This alteration in mobility is very difficult to predict, even for a known mutation, as it seems to depend upon the surrounding base sequence, and not the overall proportion of each nucleotide within the fragment (Glavac and Dean, 1993). As the changes in mobility may be very small, the ability to detect them depends upon the size of the DNA fragment, and the gel matrix in which the fragments are electrophoretically separated.

SSCP has a reported detection rate of between 35% and close to 100% (Glavac and Dean, 1993). However, various experimental conditions can be altered, such as the length and temperature of electrophoresis, the type of gel matrix and running buffer, and additions to the gel such as glycerol. Whilst optimisation of these parameters is ideal for each fragment, this is not always possible or practicable when using SSCP as a scanning method to search for unknown mutations, and detracts from the simplicity and speed of this method which makes it so useful. In this project, only one set of SSCP conditions was used, as heteroduplex analysis (HA) and direct DNA sequencing were being used together with SSCP. Together, HA and SSCP are thought to have a detection rate close to 100% (White *et al*, 1992). We chose to analyse the *NRL* gene using a specialised type of gel matrix, the Mutation Detection Enhancement (MDE) gel, closely related to Hydrolink (manufactured by AT Biochem), which is also suitable for use in HA. The gels were run at 7W at room

temperature for 14 hours. Glavac and Dean (1993) have reported that gels run at room temperature at 10W gave similar separation to gels run at 4°C.

6.6- HETERODUPLEX MUTATION DETECTION

Unlike SSCP, heteroduplex analysis (HA) does not rely on the formation of secondary structures within single-stranded DNA, but rather involves direct detection of heteroduplexes between wild-type and mutant sequences, due to their slower migration in an acrylamide (or similar vinyl polymer) gel. These heteroduplexes can be formed during a PCR reaction (as in this project), or post-PCR, by mixing the PCR products from "normal" and "mutant" DNA, allowing them to denature and re-anneal slowly, forming hetero- and homo-duplexes. For recessive disorders, the heteroduplexes form during PCR of DNA from carriers, such as the parents of affecteds, as they have one normal allele and one mutant allele. If the affected members have inherited the same mutation from each parent, their DNA should form mutant homoduplexes, which may also migrate differently from "wild-type" homoduplexes. For the consanguineous family studied in this project, we would expect this to be the case.

It has been estimated that HA detects approximately 80% of mutations, but that HA and SSCP together may detect nearly all mutations (Cotton, 1993). This method was first reported by Keen *et al*, (1991) using a specific form of acrylamide designed for detecting mutations, Hydrolink (AT Biochem). Subsequently, normal polyacrylamide gels have also been used, both denaturing and non-denaturing, although White *et al* (1992) have suggested that 15% urea is optimal for HA analysis, as it increases separation between homoduplexes and heteroduplexes, eliminates doublets that can form in some homoduplex samples, and helps sharpen bands without denaturing the double-stranded DNA. Since 1992, numerous genes have been screened for mutations using HA, including the rhodopsin and peripherin/*RDS* genes (Inglehearn *et al*, 1992; Meins *et al*, 1993). We chose to use the MDE matrix (also used for SSCP analysis), with 15% urea added. The gels were run at room temperature, at 600 V constant voltage (20V/cm), for 15 - 20 hours, depending on the size of fragments to be resolved. As HA with MDE gel is suitable for PCR products of between 200 and 600 basepairs (MDE Heteroduplex Kit Protocol, AT Biochem), the larger exon fragments (2b and 3) were not digested.

6.7 - RESULTS OF MUTATION DETECTION

Primers were chosen which flank each of the *NRL* exons (see **fig. 6.1**), and used to amplify each exon (fragments 1, 2a, 2b, and 3) in all available individuals from family ARRP1 (BP, BM, BA, BB, BC, BD), and a normal control (CEPH DNA). Each primer lies at least 67 basepairs from the closest splice site, so that every product contained at least 67 bases of intronic sequence. The PCR fragments for exons 2b (337bp) and 3 (573bp) were digested with restriction enzymes to produce fragments of less than 300 bp for SSCP analysis. Radioactive PCR reactions and enzyme digestion were carried out by K. Porter. Both heteroduplex analysis and SSCP techniques were used to screen each exon in all the individuals. No evidence of any sequence alteration was observed in any of the family members or unaffected controls (see **figs. 6.2** and **6.3** - fragment 2b is not shown for HA). This analysis covered the whole coding region, all splice junctions, and included 527 basepairs of intronic sequence.

6.8 - DNA SEQUENCING

In mutation detection, DNA sequencing is often regarded as a final step, following the use of scanning methods to localise putative mutations. In a gene with a large coding region, and many exons, the use of direct DNA sequencing is impractical for the whole gene, and scanning methods such as SSCP and heteroduplex analysis can be used to narrow down the region of search. However, these alternative, quicker methods are not always 100% effective, and may miss mutations that can be detected using DNA sequencing. Sequencing also has the advantage that it precisely defines the nature and position of the defect in the gene. The most commonly used form of DNA sequencing involves the chain-termination method first described by Sanger *et al* (1977). This can be used to sequence both PCR products and DNA cloned into appropriate vectors.

Problems can often be encountered when attempting to sequence double-stranded GC-rich regions of DNA. Typical artifacts encoutered include ambiguous signals, premature stops and high background, possibly due to the formation of secondary structures, and the higher melting temperature of GC-rich DNA. The *NRL* gene is very GC-rich, especially in exon three, which is also thought to code for the functionally important leucine zipper and basic domains. Overall, the coding region covering exons two and three is 68.5% guanine and

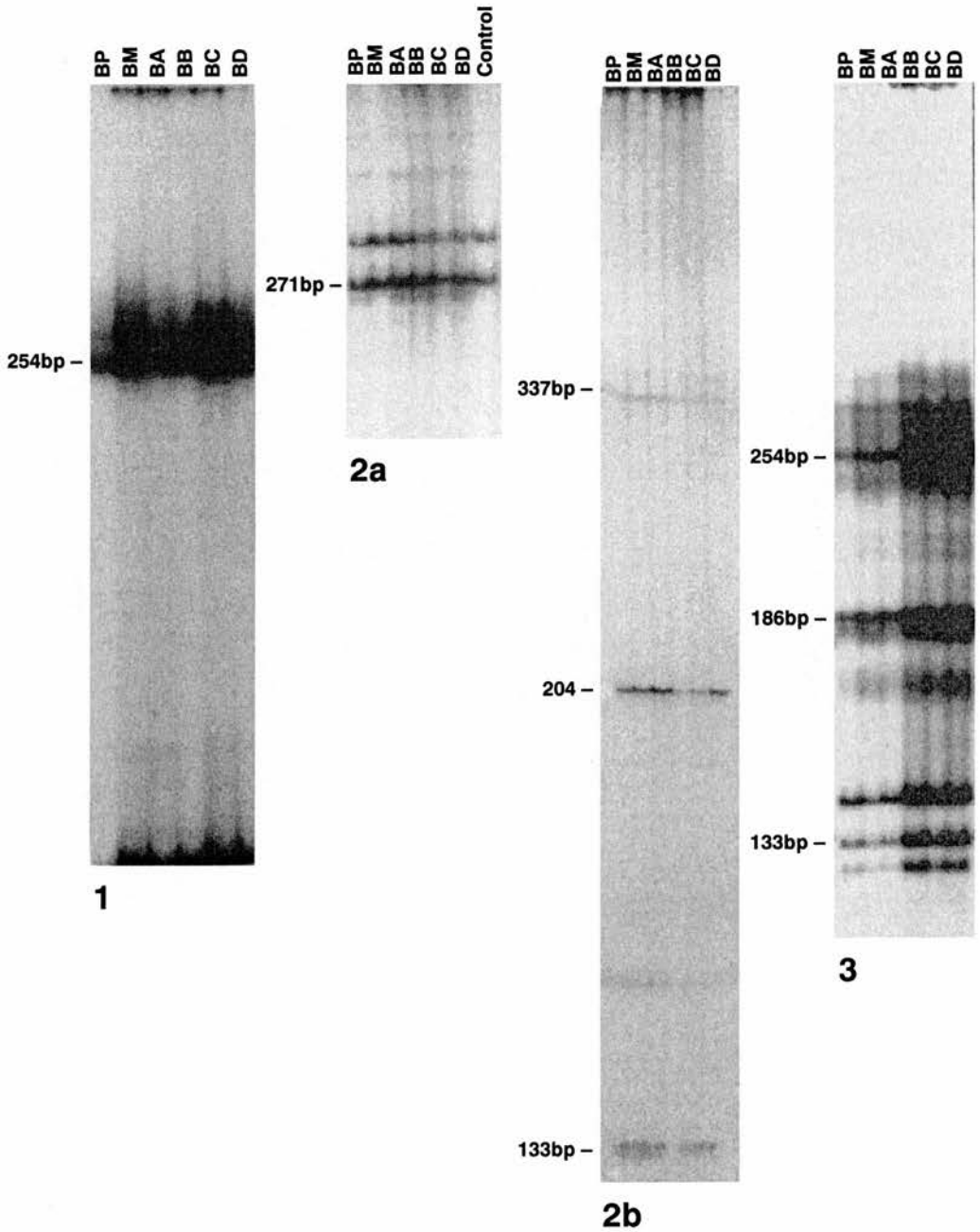


Figure 6.2

SSCP analysis of samples from all members of family ARRP1, for exon fragments 1, 2a, 2b and 3. Fragment sizes are indicated for the appropriate band.

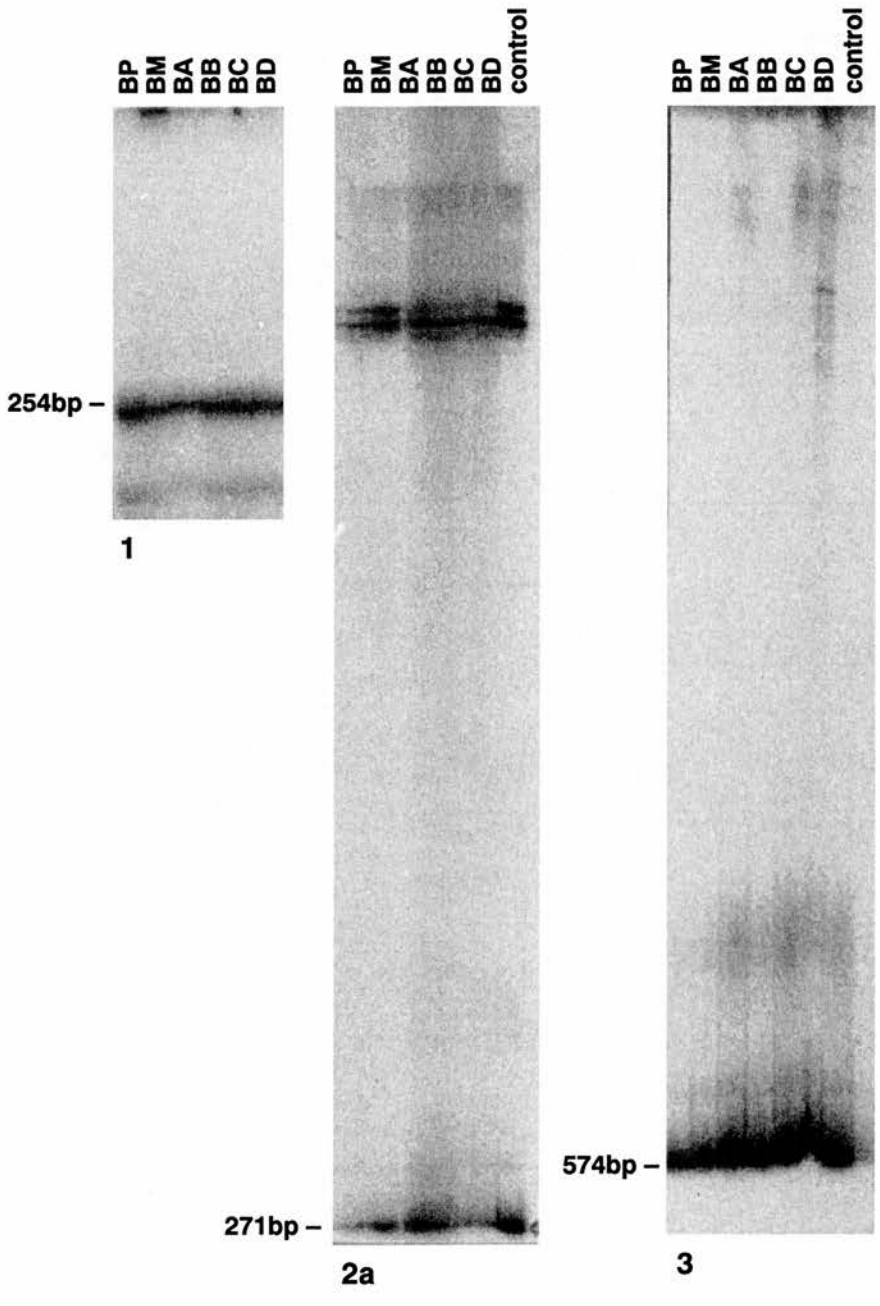


Figure 6.3

Heteroduplex analysis for all members of family ARRP1, for exonic fragments 1, 2a and 3. Fragment sizes are indicted for the appropriate band.

cytosine residues, with exon two having a G-C content of 63.8%, and exon three a G-C content of 73.9%. This led to problems in sequencing these regions, especially due to high background and "stops" - bands in all four reaction lanes.

Several methods have been proposed to reduce or eliminate these problems. Woodford, Weitzmann and Usdin (1995) have suggested the use of buffers free from monovalent cations, especially potassium ions, to eliminate premature chain termination in guanine-rich stretches of DNA. The reason for the success of this method may be that these ions usually aid the stability of DNA secondary structures that can form in G-rich regions. A suitable buffer, containing only Tris-HCl and magnesium chloride, is supplied with the IsoTherm DNA sequencing kit (Cambio). This kit also contains 7-deaza-GTP instead of dGTP, to reduce band compressions during electrophoresis, and a thermostable form of DNA polymerase, cloned from the thermophilic bacterium *Bacillus stearothermophilus*. This enzyme has an optimal reaction temperature of 65°C, thus reducing the problems of secondary structure encountered at the 37°C reaction temperature used with the Sequenase sequencing kit (USB). Templates were also boiled for at least three minutes and snap-frozen to ensure complete denaturation. The use of DMSO (used in the PCR of GC-rich exon 3), which was incorporated into the annealing reaction at 10% concentration (see section **2.6.4**), has been suggested to eliminate background and increase signal intensity (Winship, 1989). For very GC-rich regions, the best results were achieved when formamide, which weakens hydrogen bonding between nucleotides, was added to the annealing reaction. The amount of formamide was determined by the percentage of GC nucleotides in the PCR product (% formamide = 0.7 [GC% - 50]) (Zhang, Hu and Deisseroth, 1991). This succeeded in eliminating most of the stops in the sequence, and increased band intensity, enabling us to read far more of the sequence.

6.9 - RESULTS OF DNA SEQUENCING

DNA sequencing was carried out in each exon of the *NRL* gene, using DNA from affected individuals BB and BD, and their father BP. All the coding region and at least 50 basepairs upstream from both 5' and 3' splice sites was sequenced in at least one affected individual. DNA sequencing also included reverse transcribed cDNA PCR products from individual BB (and a normal control) that covered each exon/exon boundary in order to exclude mutations affecting RNA splicing (see **fig. 6.4**). This approach also allowed us to check for alternatively transcribed gene products. As cells naturally expressing high levels of Nrl

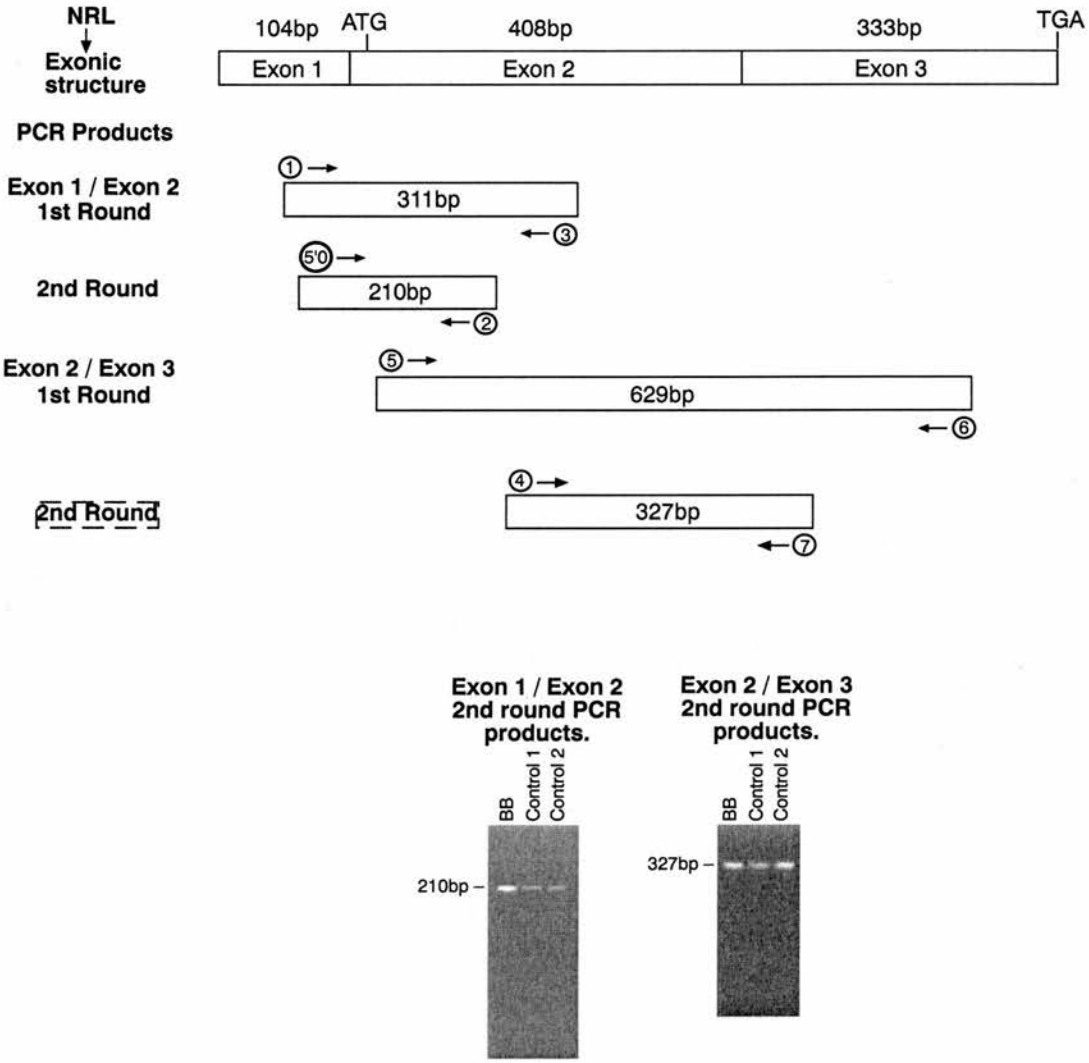


Figure 6.4 - Intragenic mutation analysis of Nrl

Diagram illustrating the strategy used to amplify nested PCR fragments from a solid-phase cDNA library. Also pictured are the final PCR products containing exon/exon junctions from affected member BB, and two normal controls, showing no variation from the expected band size.

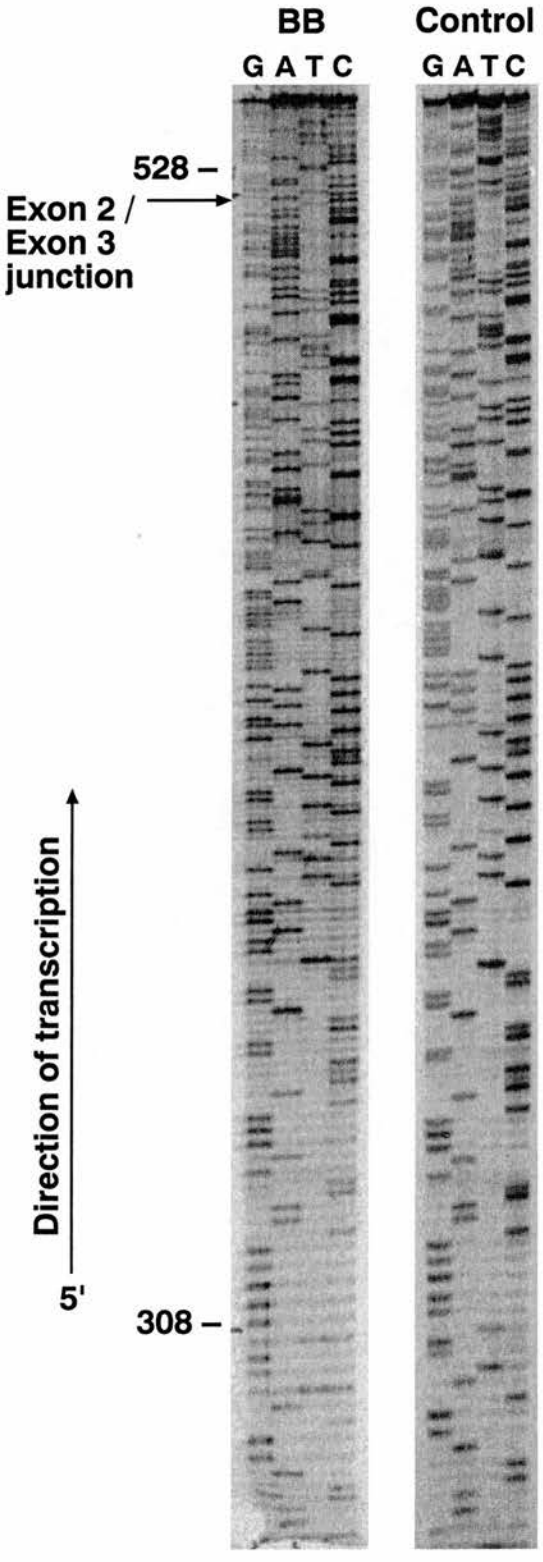


Figure 6.5 - DNA sequencing of exon/exon junctions

Template DNA from a nested PCR reaction (see **Fig. 6.4**) including the exon2/exon3 junction was amplified from a solid phase cDNA template from individual BB, and a normal control. Nucleotide positions correspond to the numbering used in the Genbank Human NRL gene product mRNA sequence, accession number M81840

were not available, lymphoblastoid cell lines derived from blood samples were used to create a solid-phase cDNA template. Nested PCR products were detectable from this template due to illegitimate transcription - the transcription of tissue specific genes at very low levels in inappropriate cell types (Kaplan, Kahn and Chelly, 1992). The promoter region was not analysed because this is a less common site of mutation. We found that no mutations were present in the coding sequence of the *NRL* gene, or intron/exon boundaries of this gene, in the affected individuals when compared to the sequence obtained from a normal control (CEPH DNA) (see **fig. 6.5**), and the sequence of the *NRL* gene as determined by Dr. A. Swaroop's group in Michigan (personal communication). This result was also confirmed by Dr. A. Swaroop and co-workers, who were sent the PCR primers and DNA from BB and BD. We also detected only bands of the expected size, and no alternatively transcribed products, following nested PCR from the solid-phase cDNA template (see **fig. 6.4**).

6.10 - DISCUSSION

The possible presence of mutations in the *NRL* gene in individuals with autosomal recessive RP has been investigated in an inbred family from Sardinia, ARRP1. This family showed homozygosity of three adjacent markers in three affected members suggesting a homozygosity-by-descent region in which the *NRL* gene has been mapped. This gene was analysed for mutations in family ARRP1. Individual *NRL* exons, including flanking intronic sequence, were PCR amplified and subjected to mutation analysis by SSCP, heteroduplex analysis and direct DNA sequencing. No variations from the genomic sequence determined by Swaroop *et al* were identified in any of the samples investigated (including normal controls). Furthermore, this result was confirmed by Swaroop *et al*, who also failed to find any sequence alterations in two of the affected individuals from family ARRP1.

The DNA sequencing included at least 50 basepairs of intronic sequence upstream from the splice sites. As well as enabling us to exclude splice site mutations, this region is predicted to contain the branch splice site, the point where base pairing occurs between the small nuclear ribonucleoprotein particle (snRNP) U2, one of the snRNPs involved in the formation of an active splicing complex (Norton, 1994). Mutations in this site have been shown to cause partial androgen insensitivity in the androgen receptor gene (Reifenstein Syndrome) (Ris-Stalpers *et al*, 1994) and familial hypercholesterolaemia in the LDL

receptor gene (Webb *et al*, 1996), but can probably be discounted in this family for the *NRL* gene. We were also able to exclude cryptic splice site mutations, as no alternatively spliced products were detected following nested PCR. However, other intronic regions, some of which may be important in maintaining mRNA stability have not been analysed. Distal and proximal promoter or enhancer mutations cannot be excluded either, although 67 basepairs of sequence proximal to the first exon, and 85 basepairs distal to the stop codon were included in the SSCP and heteroduplex analysis and no variant bands were detected. In this consanguineous arRP family we would expect affected children to have inherited two copies of the same mutation. As no visible difference was seen between the size of exonic PCR products for affected and unaffected individuals, this suggests that large intragenic deletions can be excluded as a cause of mutation. In summary, most common types of mutation have been excluded in the *NRL* gene in this family with autosomal recessive RP.

The highly conserved *NRL* gene is thought to be a good candidate for involvement in retinal diseases. It is highly expressed in the adult retina, where it may have a role in the transcriptional regulation of rhodopsin, and is very likely to have many other target retinal protein genes (Rehemtulla *et al*, 1996). The *NRL* mouse homologue has also been shown to be widely expressed throughout neurogenesis, in differentiated cells of the central and peripheral nervous system. It is downregulated in the brain after birth, with very low levels detected in the neocortex and brainstem in the adult. High expression levels are detected in the adult retina, especially in the three nuclear layers and the inner segments of the photoreceptors (Liu *et al*, 1996).

This expression pattern is similar to that of a member of the POU family of transcription factors, Pit-1. The Pit-1 protein is exclusively expressed in the adult pituitary, but is also expressed throughout embryogenesis in the developing neural tube. *Pit-1* has been shown to be mutated in combined pituitary hormone deficiency (Radovick *et al*, 1992), where the mutant protein can still bind to its DNA-binding site, preventing the normal protein from binding, but cannot activate transcription. This causes the disease to be dominantly inherited, as one mutant copy of the gene is sufficient to inhibit gene activation. A similar biphasic expression pattern is also seen with CHX10, encoded by a member of the homeobox gene family also located on chromosome 14 (14q24.3) (Ploder *et al*, 1995). This gene is expressed in the developing brain, spinal cord, and neuroblasts in the eye, but is restricted to the inner nuclear layer of the mature retina (Liu *et al*, 1994). Finally, another member of the bZIP family, *c-fos*, is also expressed in the nervous system during mouse development, and may form a transcriptional regulatory complex with Nrl in neural cells (Caubet, 1989).

As transcription factors are critical in gene expression, and one transcription factor can interact with many genes, inactivation of a gene coding for such a protein may be lethal. Nrl is expressed throughout development of the nervous system, suggesting that lack of this protein could cause significant and potentially severe disruption of neuronal differentiation. However, Pit-1 is also expressed in the developing neural tube, but mutations in this gene are compatible with survival. A possible reason for this may be an overlap in expression patterns between transcription factors, so that other proteins are able to replace the function of Pit-1 in the developing neural tube, but in the adult, the high tissue-specific levels of expression of the mutant protein result in a discernible lack of function for target genes. Likewise, mutation of the *NRL* gene could result in a retinal disorder due to high expression levels in the adult retina. The predicted pattern of inheritance of such a disorder depends upon how many functional copies of *NRL* are required for sufficient transcriptional activation. Many transcription factor diseases result from mutation of one allele, causing either a dominant negative mutation, as seen with Pit-1 or a dominant loss-of-function mutation (haplo-insufficiency) as with Pax-6, although this type of mutation can also be recessively inherited if one copy of the gene can produce enough protein to induce transcription (Latchman, 1996). The ability of Nrl to form dimers suggests that mutations in this gene could cause a dominant negative effect, as the mutant protein encoded by one copy of the gene may interfere with wild-type Nrl, and other proteins that Nrl can dimerise with.

What other genes, in addition to rhodopsin, are targets for Nrl? The tissue specificity of expression in the adult suggests that most targets will be found in the three nuclear layers of the retina and the inner segments of photoreceptors. As the Nrl recognition element (NRE) sequence is conserved in the promoter of all mammalian opsins, other retinal opsin genes may be targets for Nrl, such as the blue-, green- and red- pigment opsins present in cone cells. Fos, another member of the bZIP family, is induced in ganglion cells by light. As this protein cannot homodimerise, it is possible that it may dimerise with Nrl, hence responding to light by increasing the expression of opsin (Kerppola and Curran, 1994b). Further putative target genes include glutathione-S-transferase, and NAD(P)H:quinone reductase, due to the presence of a suitable recognition site for dimers formed by Nrl, Maf, Fos and Jun (Kerppola and Curran, 1994b). A database search for genes with similar recognition sequence motifs could reveal other potential targets for these proteins.

In conclusion, no mutations have been identified in the *NRL* gene in a consanguineous family with autosomal recessive RP. This finding correlates with the results of further linkage analysis in this family, which showed the suggested identity-by-descent in affected

members to be due to uninformative markers (see **Chapter Five**). The expression pattern of the *NRL* gene, and targets of the transcribed protein, could reveal diseases in which this gene may be involved. Mutated *NRL* is a possible candidate for developmental defects of the nervous system due to expression during neurogenesis, but high levels of expression in the adult retina suggest that it has many target genes in this tissue, including rhodopsin. Thus, Nrl remains a good candidate gene for retinal disorders, and could be considered in mutation analysis for patients with RP and related disorders. The family studied, ARRP1, may still be linked to chromosome 14q, despite a lack of homozygosity, as a maximum multipoint lod score of 2.4 was obtained at a location defined by markers D14S252 and D14S262, distal to the *NRL* gene. Thus, other retinally expressed genes that lie in this region may be causing the arRP seen in this family. One candidate is the previously mentioned homeobox gene coding for CHX10, as a premature stop codon in this gene can cause microphthalmia, a thin hypocellular retina and optic nerve aplasia in the mouse (Burmeister *et al*, 1996). This gene has previously been examined using SSCP analysis in a set of patients, 17 with microphthalmia, 11 with anophthalmia, 64 with Leber's congenital amaurosis, and 12 with arRP, but no mutations were identified (Ploder *et al*, 1995). However, this gene is still a candidate for retinal disease, and may merit further investigation in this family.

CHAPTER SEVEN

DISCUSSION

Retinitis pigmentosa is a common genetic cause of blindness, affecting around 1 in 4000 individuals worldwide. The autosomal recessive form is one of the most severe and prevalent types of RP, whose true frequency may often be underestimated. No proven cure or preventative treatment is currently available, although high doses of vitamin A (18,000 i.u. daily, close to a potentially toxic dose) have been reported to slow the loss of a cone electroretinogram response (Berson *et al*, 1993). Identification of genes responsible for RP increases our understanding of the mechanisms of cellular dysfunction, allows for accurate diagnosis, and may eventually lead to successful treatments.

Seven genes have been identified that cause simplex RP when mutated. Of these, four have been shown to cause arRP - rhodopsin (Rosenfeld *et al*, 1992), the alpha and beta subunits of cGMP phosphodiesterase (Huang *et al*, 1995; McLaughlin *et al*, 1995), and the alpha subunit of the cyclic nucleotide gated channel (Dryja *et al*, 1995). All of the proteins coded for by these genes are involved in the phototransduction cascade that takes place within the rod outer segment. However, together these four loci are thought to account for less than 10% of all arRP cases (Dryja and Li, 1995). Linkage for arRP has also been reported to 6p21.3 (RP14) in a large pedigree from the Dominican Republic (Knowles *et al*, 1994), and to 1q31-q32.1 (RP12) for a relatively uncommon specific type of RP, PPRPE, in extended kindreds from the Netherlands and Pakistan (van Soest *et al*, 1994; Leutelt *et al*, 1995). Overall, this suggests that there are still a large number of genes causing arRP that await identification.

One aim of this project was to localise arRP genes in a set of eleven families from south-central Sardinia, using linkage mapping. As arRP is known to be nonallelically heterogeneous, the Sardinian population was chosen for its reported genetic homogeneity for adRP (Fossarello *et al*, 1993) and other genetic diseases. Also, this population has a very high proportion of arRP cases (21.5% arRP compared to 5% adRP and 1.3% xlRP), possibly due to increased levels of inbreeding, and religious and economic factors have encouraged the establishment of large families.

Using an automated laser fluorescence DNA sequencer (ALF, Pharmacia) 195 microsatellite markers were typed in cooperation with Dr. D. Mansfield in all eleven families, but failed to detect any significant linkage. While it is possible that a major causative gene lies outwith the regions covered by the markers used in this study, this result also suggests that there is genetic heterogeneity within this group of families. A possible reason for heterogeneity may be that highland and lowland communities were highly

isolated until this century, so that the whole Sardinian population can be considered as being composed of many small isolates, each having experienced its own founder effect, and undergoing genetic drift. This suggestion correlates with geographical, historical, linguistic and archaeological evidence, and studies on polymorphisms amongst Sardinian villages support this prediction, by demonstrating significant heterogeneities, especially between lowland and highland settlements (Workman *et al*, 1976).

Previous work by Dr. Mansfield had excluded some of the families for linkage to genes such as rhodopsin and peripherin using closely linked markers. It is still possible that the families not yet excluded may have mutations in these, and other strong candidate genes. However, the *PDEB* gene has been examined by Dr. D. Farber and colleagues (University of California) in the eleven Sardinian families used in this study. They have not detected any mutations, although they did detect several variants that did not segregate with the disease. Work is also currently underway by Dr. F. Manson (MRC Human Genetics Unit) to investigate the gene coding for the gamma subunit of PDE, with which some of our families showed no recombination. No mutations have previously been reported in this gene, even though it is a strong candidate for retinal degenerations, as mutations have been identified in the two other subunits that comprise the active cGMP phosphodiesterase enzyme. Work by Tsang *et al* (1996) has also shown that targeted disruption of the gamma PDE gene in mice results in a rapid retinal degeneration, and that this subunit is necessary for activity and stability of PDE *in vivo*.

Close examination of individual lod scores has suggested other regions worth investigation in particular families. The main problem with further linkage work in the individual families is that most are small two generation pedigrees, so that high lod scores providing definite linkage to, or exclusion from, specific loci is not possible unless many densely-spaced markers are used. However, any families not showing recombination with markers closely linked to a candidate gene may merit further investigation, perhaps by screening the candidate gene for mutations in an affected individual and an obligate carrier. The suggestion of genetic heterogeneity has led us to consider the candidate gene approach, as successfully used by Dryja *et al* to identify mutations in *PDEB*, *PDEA* and *CNCG1*. A large number of candidates for arRP - with a structural or biochemical role in rod photoreceptors - have been cloned, and their genomic structure elucidated. In addition, many have been localised to specific chromosomal regions. In the original set of eleven two-generational Sardinian arRP pedigrees only two were consanguineous, but we have since obtained another consanguineous pedigree and two large extended kindreds. These will be more powerful for use in linkage mapping, initially to candidate regions, and if

required, in a full genome scan. Any promising candidate genes could then be tested in the smaller pedigrees for mutation.

A combination of positional cloning and candidate gene analysis, the positional candidate approach, using initial linkage mapping followed by candidate gene screening in the region of interest, was adopted with one consanguineous family, ARRP1. An interesting result was noted with marker D14S80 in 14q11, with a subset of five of the families showing no recombination. Flanking markers were typed in these families, and appeared to reveal a region of homozygosity in four markers (D14S64, D14S80, D14S275 and D14S262) covering 6 cM in two affected individuals from this one family, ARRP1, suggesting identity-by-descent for this region. A good candidate gene for RP, coding for the neural retina leucine zipper protein (Nrl) was mapped to the same region (14q11.1-q11.2) by Yang-Feng and Swaroop (1992), and was found to be on the same cosmid clone as one of the homozygous markers (D14S64). This gene codes for a conserved member of the bZIP family of transcription factors, which is highly expressed in the adult retina, with transient expression in the developing nervous system (Liu *et al*, 1996). Furthermore, the promoter region of the rhodopsin gene has been shown to contain a conserved binding site for the Nrl protein, suggesting a role for Nrl in regulation of this and other retina-specific proteins (Rehemtulla *et al*, 1996). It was decided to screen this gene in the ARRP1 family using heteroduplex analysis and SSCP. No band shifts were identified, and DNA sequencing of the whole coding region, and adjacent splice sites, also failed to reveal any base changes in this family. Thus, this gene has been excluded for the common causes of mutation in this pedigree. Further linkage mapping with another affected member of ARRP1 has since confirmed this result, by showing that the suspected homozygosity in this individual was due to uninformative parental alleles. While Nrl is still a good candidate for retinal degenerations, its expression pattern during embryogenesis suggests that mutations in this gene could have more widespread effects than a retinal degeneration alone. The family still shows no recombination with three markers in this region, distal to D14S275, so that another gene on chromosome 14 may be responsible for the disease. However, it is also likely that this is a chance association between the markers and the disease, as only one of the markers studied is homozygous in this consanguineous family.

The results of linkage analysis with this one inbred family did however prove the power of working with consanguineous families. Homozygosity mapping, a form of linkage mapping, relies on the principal that each affected member of a consanguineous union will have inherited two copies of the mutated gene from a common ancestor, and hence will have inherited identical alleles at loci closely flanking this gene. This greatly increases the

power of mapping, by considering all the meioses in past generations in which recombination could have occurred between the gene and flanking markers. This approach has been used to map several genes for rare autosomal recessive traits, as fewer families are required to achieve a significant lod score. For example, Hirschsprung disease was mapped using a large Mennonite kindred (Puffenberger *et al*, 1994) and three genes causing Bardet-Biedl syndrome have been mapped using large Bedouin kindreds (Kwitek-Black *et al,* 1993; Sheffield *et al*, 1994; Carmi *et al* , 1995a).

Another form of genetic mapping which could be used in an isolated genetic population such as Sardinia is linkage disequilibrium mapping. This relies on a similar principle to homozygosity mapping, but considers the spread of an ancestral haplotype (carrying a mutation) throughout a whole enclosed community, thus greatly increasing the available sampling population. This approach has been successfully used in several populations, including the Finnish, the Ashkenazim and the Old Order Amish, to map genes responsible for rare autosomal recessive disorders. This method is also able to narrow the region in which the causative gene must lie, by at least an order of magnitude compared to the results of normal linkage mapping (de la Chapelle, 1993).

In summary, in specific inbred or isolated groups, specialised types of genetic mapping can provide a powerful means of localising the region in which a gene for an autosomal recessive disorder may lie. However, in the small non-consanguineous Sardinian pedigrees we have studied there is no clear sign of high levels of homogeneity, leading us to consider candidate gene analysis. A potential limitation of the candidate gene approach is the reliability of the method of mutation detection employed. SSCP, the most commonly used technique, has been reported to detect as few as 35% of mutations, although this detection rate can be increased by employing a variety of electrophoresis conditions (Glavac and Dean, 1993). Heteroduplex analysis (HA), another widely used technique, has the advantage that it can easily be used without the need for radioactive detection, as protocols are available for silver-staining and ethidium bromide staining. HA in combination with SSCP has been reported to detect nearly all mutations (Glavac and Dean ,1995), but this is often not reliable enough, especially if dealing with small sample numbers where only one mutation may be present. Many of the other more reliable techniques, such as HOT, are laborious and involve toxic chemicals.

Part of this project attempted to produce a tool for mutation detection, by designing a fusion protein for MutS. MutS is a bacterial mismatch binding protein involved in the methyl-directed mismatch repair system in *E. coli* and *S. typhimurium*. Research has shown MutS to be highly effective for *in vitro* detection of single base mismatches and deletions or

insertions of up to four bases (Smith and Modrich, 1996). Although time constraints prevented development of a soluble fusion protein, a protocol has subsequently been published detailing production of a His_6-MutS fusion protein (Feng and Winkler, 1995). This provides a unique opportunity for exploitation in the field of routine candidate gene screening, although more research will be required to refine the optimal working conditions for this protein.

Linkage mapping for Bardet-Biedl syndrome, a syndromic form of RP, was also carried out in this project. In addition to RP, other cardinal features of this condition are polydactyly, hypogenitalism, obesity and mental retardation, although all are highly variable both within and between families. Further commonly associated features include renal anomalies and syn- or brachy-dactyly. All of these clinical features were observed among the set of 29 families collected worldwide that were used in this study. Linkage has been reported to four loci for BBS - 11q13 (*BBS1*), 16q21 (*BBS2*), 3p13-p12 (*BBS3*) and 15q22.3-q23 (*BBS4*) (Leppert *et al*, 1994; Kwitek-Black *et al*, 1993; Sheffield *et al*, 1994; Carmi *et al*, 1995a). Carmi *et al* used a form of DNA pooling for the initial linkage work, by pooling the affected members, and unaffected individuals (controls), of the pedigree and comparing the number of alleles observed for an individual marker. Any markers that showed a shift towards homozygosity in the affected group (and not the controls) were then investigated with further flanking markers. This method may be suitable for use in any large inbred families. We studied markers closely linked to all four loci in the set of BBS families, using conventional linkage analysis. No linkage was identified for *BBS3* in our set of families, but signs of linkage were found to the three other BBS loci.

A multipoint lod score of 6.26 was achieved for the *BBS1* locus, assuming genetic heterogeneity. This analysis places the locus in a region overlapping with the results of homozygosity mapping in Puerto Rican kindreds (Cornier *et al*, 1995). In total, eight families showed evidence of linkage to this locus, by combining lod scores and haplotype analysis. For the *BBS4* locus on chromosome 15, a lod score of 6.10 was reached, again assuming genetic heterogeneity. Seven families, three of which were consanguineous, appeared linked to this locus when lod scores and haplotypes were combined. Two of the consanguineous families showing homozygosity for the *BBS4* region contained informative crossovers that helped us to narrow the region in which this gene must lie. In family BB18, a crossover was identified with D15S125, which is therefore a flanking marker, as found by Carmi *et al* (1995a) in the original report of linkage to this region. In family BB31, two informative crossovers have allowed us to narrow the region even further, by showing that D15S131 and D15S114 are both flanking markers. Previously reported flanking markers

D15S131 and GATA89D04 (D15S823) placed the *BBS4* locus in an interval of 4.3 cM (Haider *et al*, 1996). However, results in family BB31 suggest that the locus must lie in an interval of around 1.6 cM, probably corresponding to a physical distance of less than two megabases. This size of interval should be suitable for a positional cloning project, but it may be possible to reduce this region by typing more markers between D15S131 and D15S114, in the hope of identifying the site of each recombinational event more precisely. There are currently 14 polymorphic markers reported to lie in this interval (Dib *et al*, 1996), and these could be tested for homozygosity in this family.

Weak signs of linkage were also found for *BBS2* on chromosome 16, with a lod score of 1.09 assuming two BBS loci. Only three families appear linked to this region when lod scores and haplotypes are examined, but one of these (BB13) is consanguineous, and homozygous for the three markers studied. Thus, typing further markers with this family may identify the limits of this region of identity-by-descent, hence defining the limits within which the *BBS2* locus must lie. It is interesting to note that at least three of the BBS families studied show no sign of linkage to any of the reported loci. While it is possible that further work with markers surrounding the *BBS3* locus may reveal linkage in these families, preliminary results from this locus did not suggest linkage in any of the families. These families could be linked to a fifth, as yet unidentifed, BBS locus, as suggested by Bonneau *et al* (1996) and Sunden *et al* (1995), due to lack of linkage in some BBS families from France, North Africa, Portugal and Canada. This level of heterogeneity is unexpected for such a specific disorder, and as yet no clear phenotype/genotype correlations have been found in our families.

Whilst work with transgenic mice and cultured cells containing RP mutations have suggested possible causes of the retinal degeneration, there remains considerable debate about the relationship between a functional deficit and the resultant degeneration. Various suggested mechanisms include the equivalent light hypothesis, increase in intracellular cGMP, an accumulation of misfolded protein, and abnormal photoreceptor disc structure (reviewed by Lisman and Fain, 1995). Cell death usually occurs through two pathways - necrotic cell death (often due to injury), which is commonly accompanied by inflammation, or apoptosis, an internally controlled suicide program. Apoptosis affects isolated cells, which are usually phagocytosed by neighbouring cells. The nuclear DNA is also fragmented by endonucleases, providing a simple method of detection (Thompson, 1995). Apoptosis has been shown to occur during normal neuronal development in all layers of the retina, including the photoreceptors. Research with animal models of RP and transgenic mice has suggested that photoreceptor cell death in RP is due to apoptosis (Portera-Cailliau

et al, 1994). Indeed, increased apoptosis has been associated with other neurodegenerative disorders such as Alzheimer's disease and cerebellar degeneration (Thompson, 1995). However, the actual mechanism by which RP mutations re-activate apoptosis is still unknown. Nevertheless, it seems plausible that treatments that disrupt the apoptotic pathway could halt or slow the gradual degeneration of photoreceptors that occurs in RP.

Bcl-2 and related proteins naturally repress apoptosis, and it has been suggested that overexpression of these proteins may slow or prevent photoreceptor degeneration. Flannery *et al* (1995) have studied the offspring of transgenic mice overexpressing Bcl-2 crossed with mice with inherited retinal degenerations, and have suggested that Bcl-2 can slow disease progression in mice with opsin mutations. However, Joseph and Li (1996) have since provided conflicting evidence that overexpression of *bcl-2* and *bcl-*X_L does not halt the apoptosis in photoreceptors in mice with opsin or PDEB mutations.

Another possible future treatment for RP is the use of gene therapy. Ocular tissues are a suitable target for this treatment, as most cells are comparatively accessible. Anatomical boundaries within the eye allow targeted gene transfer, and no other effective therapy has yet been developed. Bennett *et al* (1996) have used a recombinant replication-defective adenovirus containing the cDNA for *PDEB* in *rd* mice, and have observed transgene expression in the retina at least 100 days after injection. However, potential problems with toxicity of adenovirus-mediated transfer have been avoided by Ali *et al* (1996) who have studied an adeno-associated viral vector, as this vector is deleted for all virally-encoded proteins. They have reported efficient and non-toxic gene transfer into the murine retina. Although much more work is required to identify safe and effective gene delivery systems, this approach may ultimately be able to arrest the degeneration seen in RP (and allied disorders), hence preserving the sight of thousands of people.

REFERENCES

Abe, T., Kikuchi, T. and Shinohara, T. (1994). The sequence of the human phosducin gene (*PDC*) and its 5'-flanking region. *Genomics* **19**, 369-372.

Al-Maghtheh, M., Inglehearn, C.F., Jeen, T.J,. Evans, K., Moore, A.T., Jay, M., Bird, A.C. and Bhattacharya, S.S. (1994). Identification of a sixth locus for autosomal dominant retinitis pigmentosa on chromosome 19. *Hum. Molec. Genet.* **3**, 351-354.

Ali, R.R., Reichel, M.B., Thrasher, A.J., levinsky, R.J., Kinnon, C., Kanuga, N., Hunt, D.M. and Bhattacharya, S.S. (1996). Gene transfer into the mouse retina mediated by an adeno-associated viral vector. *Hum. Molec. Genet.* **5**, 591-594.

Altherr, M.R., Wasmuth, J.J., Seldin, M.F., Nadeau, J.H., Baehr, W. and Pittler, S.J. (1992). Chromosome mapping of the rod photoreceptor cGMP phosphodiesterase beta-subunit gene in mouse and human: tight linkage to the Huntington disease region (4p16.3). *Genomics* **12**, 750-754.

Ammann, F. (1970). Investigations cliniques et genetiques sur le syndrome de Bardet-Biedl en Suisse. *J. Genet. Hum.* **18** (suppl.), 1-310.

Ammann, F., Klein, D. and Franceschetti, A. (1965). Genetic and epidemiological investigation of pigmentary degeneration of the retina and allied disorders in Switzerland. *J. Neurol. Sci.* **2**, 183-196.

Ardell, M.D., Makhija, A.K., Oliviera, L., Miniou, P., Viegas-Pequignot, E. and Pittler, S.J. (1995). cDNA, gene structure and chromosomal localization of human GAR1 (*CNCG3L*), a homolog of the third subunit of bovine photoreceptor cGMP-gated channel. *Genomics* **28**, 32-38.

Ardell, M.D., Aragon, I., Oliveira, L., Porche, G.E., Burke, E. and Pittler, S.J. (1996). The beta-subunit of human rod photoreceptor cGMP-gated cation channel is generated from a complex transcription unit. *FEBS Letters* **389**, 213-218.

Au, K.G., Clark, S., Miller, J.H. and Modrich, P. (1989). *Escherichia coli mutY* gene encodes an adenine glycosylase active on G-A mispairs. *Proc. Natl. Acad. Sci.. USA* **86**, 8877-8881.

Au, K.G., Welsh, K. and Modrich, P. (1992). Initiation of methyl-directed mismatch repair. *J. Biol. Chem.* **267**, 12142-12148.

Babon, J.J., Youil, R. and Cotton, R.G.H. (1995). Improved strategy for mutation detection - a modification to the enzyme mismatch cleavage method. *Nucleic Acids Res.* **23**, 5082-5084.

Banfi, S., Borsani, G., Rossi, E., Bernard, L., Guffanti, A., Rubboli, F., Marchitiello, A., Giglio, S., Coluccia, E., Zollo, M., Zuffardi, O. and Ballabio, A. (1996). Identification and mapping of human cDNAs homologous to *Drosophila* mutant-genes through EST database searching. *Nature Genet.* **8**, 70-76.

Bardien, S., Ebenezer, N., Greenberg, J., Inglehearn, C.F., Bartmann, L., Goliath, R., Beighton, P., Ramesar, R. and Bhattacharya, S.S. (1995). An 8th locus for autosomal-dominant retinitis pigmentosa is linked to chromosome 17q. *Hum. Molec. Genet.* **4**, *1459-1462*

Bascom, R., Manara, S., Gallie, B., Willard, H., Kalnins, V. and McInnes, R.R. (1989). Identification of a new mammalian photoreceptor specific gene family. *Am. J. Hum. Genet.* **45**, A172 (abstract).

Bascom, R.A., Manara, S., Collins, L., Molday, R.S., Kalnins, V.I. and McInnes, R.R. (1992a). Cloning of the cDNA for a novel photoreceptor membrane protein (rom-1) identifies a disk rim protein family implicated in human retinopathies. *Neuron* **8**, 1171-1184.

Bascom, R.A., Garcia-Heras, J., Hsieh, C.-L., Gerhard, D.S., Jones, C., Francke, U., Willard, H.F., Ledbetter, D.H. and McInnes, R.R. (1992b). Localization of the photoreceptor gene *ROM-1* to human chromosome 11 and mouse chromosome 19: sublocalization to human 11q13 between *PGA* and *PYGM*. *Am. J. Hum. Genet.* **51**, 1028-1035.

Bascom, R.A., Liu, L., Humphries, P., Fishman, G.A., Murray, J.C. and McInnes, R.R. (1993). Polymorphisms and rare sequence variants at the *ROM-1* locus. *Hum. Molec. Genet.* **2**, 1975-1977.

Bascom, R.A., Liu, L., Heckenlively, J.R., Stone, E.M. and McInnes, R.R. (1995). Mutation analysis of the *ROM-1* gene in retinitis pigmentosa. *Hum. Molec. Genet.* **4**, 1895-1902.

Bates, G. and Lehrach, H. (1994). Trinucleotide repeat expansions and human genetic disease. *Bioessays* **16**, 277-284.

Bayes, M., Giordano, M., Balcells, S., Grinberg, D., Vilageliu, L., Martinez, I., Ayuso, C., Benitez, J., Ramos-Arroyo, M.A., Chivelet, P., Solans, T., Valverde, D., Amselem, S., Goossens, M., Baiget, M., Gonzalez-Duarte, R. and Besmond, C. (1995a). Homozygous tandem duplication within the gene encoding the beta-subunit of rod phosphodiesterase as a cause for autosomal recessive retinitis pigmentosa. *Hum. Mutat.* **5**, 228-234.

Bayes, M., Valverde, D., Balcells, S., Grinberg, D., Vilageliu, L., Benitez, J., Ayuso, C., Beneyto, M., Baiget, M. and Gonzalez-Duarte, R. (1995b). Evidence against involvement of recoverin in autosomal recessive retinitis pigmentosa. *Hum. Genet.* **96**, 89-94.

Bennett, J., Tanabe, T., Sun, D., Zeng, Y., Kjeldbye, H., Gouras, P. and Maguire, A.M. (1996). Photoreceptor cell rescue in retinal degeneration (*rd*) mice by *in vivo* gene therapy. *Nat. Med.* **2**, 649-654.

Bernardinelli, L., Maida, A., Morinoni, A., Clayton, D., Romano, G., Montomoli, C., Fadda, D., Solinas, G., Castiglia, P., Cocco, P.L., Ghislandi, M., Berzuini, C., Pascutto, C., Nerini, M., Styles, B., Capocaccia, R., Lispi, L. and Mallardo, E. (1994). *Atlas of Cancer Mortality in Sardinia 1983-1987.* Progetto Finalizzato FATMA a cura della Direzione, Rome.

Berson, E.L., Rosner, B., Sandberg, M.A., Hayes, K.C., Nicholson, B.W., Weigeldifranco, C. and Willet, W. (1993). A randomised trial of vitamin A and vitamin E supplementation for retinitis pigmentosa. *Arch. Ophthalmol.* **111**, 1463-1465.

Bird, A.C. (1992). Investigation of disease mechanisms in retinitis pigmentosa. *Ophthalm. Paediat. Genet.* **13**, 57-66.

Bird, A.C. and Jay, B. (1994). Diagnosis in inherited retinal disorders. *In*: A.F. Wright and B. Jay (eds.) *Molecular Genetics of Inherited Eye Disorders*, pp.53-88, Harwood Academic Publishers, Chur, Switzerland.

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

Blanton, S.H., Heckenlively, J.R., Cottingham, A.W., Friedman, J., Sadler, L.A., Wagner, M., Friedman, L.H. and Daiger, S.P. (1991). Linkage mapping of autosomal dominant retinitis pigmentosa (*RP1*) to the pericentric region of human chromosome 8. *Genomics* **11**, 857-869.

Blatt, C., Eversole-Cire, P., Cohn, V.H., Zollman, S., Favnier, R.E., Mohandas, L.T., Nesbitt, M., Lugo, T., Jones, D.T. and Reed, R.R. (1988). Chromosomal localisation of genes encoding guanine nucleotide-binding protein subunits in mouse and human. *Proc. Natl. Acad. Sci. USA* **85**, 7642-7646.

Bonneau, D., Kaplan, J., Girard, G. and Dufier, J.L. (1992). Autosomal inheritance of "senile" retinitis pigmentosa: a report of a family with consanguinity. *Clin. Genet.* **42**, 199-200.

Bonneau, D., Amati, P., Chomel, J.C., Philip, N., Kaplan, J., Bitoun, P. and Munnich, A. (1996). Genetic heterogeneity of Bardet-Biedl syndrome. *Am. J. Hum. Genet.* **59**, A213 - Abstract 1222.

Boughman, J.A., Conneally, P.M. and Nance, W.E. (1980). Population genetic studies of retinitis pigmentosa. *Am. J. Hum. Genet.* **32,** 223-235.

Boughman, J.A. and Fishman, G.A. (1983). A genetic analysis of retinitis pigmentosa. *Br. J. Ophthalmol.* **67**, 449-454.

Brown, M.D., Lott, M.T. and Wallace, D.C. (1994). Mitochondrial DNA mutations and the eye. *In*: A.F. Wright and B. Jay (eds.) *Molecular Genetics of Inherited Eye Disorders*, pp.469-490, Harwood Academic Publishers, Chur, Switzerland.

Bullock , W.O., Fernandez, J.M. and Short, J.M. (1987). XL1-Blue - A high-efficiency plasmid transforming recA *Escherichia coli* strain with beta-galactosidase selection. *Biotechniques* **5**, 376.

Bundey, S. and Crews, S.J. (1984). A study of retinitis pigmentosa in the city of Birmingham: I. Prevalence. *J. Med. Genet.* **21,** 417-420.

Bunker, C.H., Berson, E.L., Bromley, W.C., Hayes, R.P. and Roderick, T.H. (1984). Prevalence of retinitis pigmentosa in Maine. *Am. J. Ophthalmol.* **97**, 357-365.

Burmeister, M., Novak, T., Liang, M.Y., Basu, S., Ploder, L., Hawes, N.L., Vidgen, D., Hoover, F., Goldman, D., Kalnins, V.I., Roderick, T.H., Taylor, B.A., Hankn, M.H. and McInnes, R.R. (1996). Ocular retardation mouse caused by CHX10 homeobox null allele - impaired retinal progenitor proliferation and bipolar cell differentiation. *Nature Genet.* **12**, 376-384.

Cappello, N., Rendine, S., Griffo, R., Mameli, G.E., Succa, V., Vona, G. and Piazza, A. (1996). Genetic analysis of Sardinia: I. Data on 12 polymorphisms in 21 linguistic doamins. *Ann. Hum. Genet.* **60**, 125-141.

Carmi, R., Rokhlina, T., Kwitek-Black, A.E., Elbedour, K., Nishimura, D., Stone, E.M and Sheffield, V.C. (1995a). Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. *Hum. Molec. Genet.* **4**, 9-13.

Carmi, R., Elbedour, K., Stone, E.M. and Sheffield, V.C. (1995b). Phenotypic differences among patients with Bardet-Biedl syndrome linked to three different chromosome loci. *Am. J. Med. Genet.* **59**, 199-203.

Caubet, J.F. (1989). C-fos proto-oncogene expression in the nervous system during mouse development. *Mol. Cell Biol.* **9**, 2269-2272.

Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1994). *The History and Geography of Human Genes,* Princeton University Press, New Jersey.

Cavalli-Sforza, L.L. and Piazza, A. (1993). Human genomic diversity in Europe: a summary of recent research and prospects for the future. *Eur. J. Hum. Genet.* **1**, 3-18.

Chen, C.K., Inglese, J., Lefkowitz, R.J. and Hurley, J.B. (1995). Ca^{2+} -dependent interaction of recoverin with rhodopsin kinase. *J. Biol. Chem.* **270**, 18060-18066.

Chen, H., Charlat, O., Tartaglia, L.A., Woolf, E.A., Weng, X., Ellis, S.J., Lakey, N.D., Culpepper, J., Moore, V.J., Breitbart, R.E., Duyk, G.M., Tepper, R.I. and Morgenstern, J.P. (1996). Evidence that the diabetes gene encodes the leptin receptor - identification of a mutation in the leptin receptor gene in db/db mice. *Cell* **84**, 491-495.

Chen, T.-Y., Illing, M., Molday, L.L., Hsu, Y.-T., Yau, K.-W. and Molday, R.S. (1994). Subunit 2 (or β) of retinal rod cGMP-gated cation channel is a component of the 240-kDa channel-associated protein and mediates Ca^{2+}-calmodulin modulation. *Proc. Natl. Acad. Sci. USA* **91**, 11757-11761.

Chrunyk, B.A., Evans, J., Lillquist, J., Young, P. and Wetzel, R. (1993). Inclusion body formation and protein stability in sequnece variants of interleukin-1β. *J. Biol. Chem.* **268**, 18053-18061.

Chung, W.K., Goldberg-Berman, J., Power-Kehoe, L. and Leibel, R.L. (1996). Molecular mapping of the tubby (*tub*) mutation on mouse chromosome 7. *Genomics* **32**, 210-217.

Claverys, J.-P. and Lacks, S.A. (1986). Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. *Microbiol. Rev.* **50**, 133-165.

Coleman, D.L. and Eicher, E.M. (1990) Fat (*fat*) and tubby (*tub*): two autosomal recessive mutations causing obesity syndromes in the mouse. *J. Hered.* **81**, 424-427.

Coleman, M., Bhattacharya, S., Lindsay, S., Wright, A., Jay, M., Litt, M., Craig, I. and Davies, K. (1990). Localization of the microsatellite probe DXS426 between DXS7 and DXS255 on Xp and linkage to X-linked retinitis pigmentosa. *Am. J. Hum. Genet.* **47**, 935-940.

Colley, N.J., Cassill, J.A., Baker, E.K. and Zuker, C.S. (1995). Defective intracellular transport is the molecular basis of rhodopsin-dependent dominant retinal degeneration. *Proc. Natl. Acad. Sci. USA* **92**, 3070-3074.

Collins, F.S. (1995). Positional cloning moves from peridítional to traditional. *Nature Genet.* **9**, 347-350.

Conlan, R.A. (1995). Retinoic acid and pattern formation in vertebrates. *Trends Genet.* **11**, 314-319.

Conneally, P.M., Edwards, J.H., Kidd, K.K., Lalouel, J.M., Morton, N.E., Ott, J. and White, R. (1985). Report of the committee on methods of linkage analysis and reporting. *Cytogenet. Cell Genet.* **40**, 356-359.

Connell, G.J. and Molday, R.S. (1990). Molecular cloning, primary structure, and orientation of the vertebrate photoreceptor cell protein peripherin in the rod outer segment disk membrane. *Biochemistry* **29**, 4691-4698.

Cornier, A.S., Fulton, A.B., Rokhlina, T., Nishimura, D., Stone, E.M., Sheffield, V.C., Whiteman, D.A.H., and Cox, G.F. (1995). Homozygosity mapping of a Bardet-Biedl syndrome gene in inbred families of Puerto-Rican ancestry confirms the existence of a chromosome 11 locus (Abstract). *Am. J. Hum. Genet.* **57**, A189.

Cotran, P.R., Ringens, P.J., Crabb, J.W., Berson, E.L. and Dryja, T.P. (1990). Analysis of the DNA of patients with retinitis pigmentosa with a cellular retinaldehyde binding protein cDNA. *Exp. Eye Res.* **51**, 15-19.

Cotton, R.G.H. (1993). Current methods of mutation detection. *Mutat. Res.* **285**, 125-144.

Cotton, R.G.H., Rodrigues, N.R. and Campbell, R.D. (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. USA* **85**, 4397-4401.

Crino, A., Tonini, G., Vido, L., Balsamo, A., Desimone, M., Detoni, T., Bosio, L., Iughetti, L., Livieri, C., Pasquino, A.M., Carusonicoletti, M., Ciampalini, P., Digilio, M.C., Pinello, L., Vignutelii, L., Sarni, P., Bernasconi, S., Fazzi, E., Pomella, S., Lopresti, D., Bosco, D., Beccaria, L. and Chiumello, G. (1994). Bardet-Biedl syndrome - an Italian multicentric study. *Italian J. Paediatrics* **20**, 530-536.

Croft, J.B., Morrell, D., Chase, C.L. and Swift, M. (1995). Obesity in heterozygous carriers of the gene for the Bardet-Biedl syndrome. *Am. J. Med. Genet.* **55**, 12-15.

Curtis, D. (1996). Genetic dissection of complex traits - reply. *Nat. Genet* **12**, 356-357.

de la Chapelle, A. (1993). Disease gene mapping in isolated human populations: the example of Finland. *J. Med. Genet.* **30**, 857-865.

Demmer, L.A., Birkenmeier, E.H., Sweetser, D.A., Levin, M.S., Zollman, S., Sparkes, R.S., Mohandan, T., Lusis, A.J. and Gordon, J.I. (1987). The cellular retinol-binding protein II gene: sequence analysis of the rat gene, chromosomal localisation in mice and humans, and documentation of its close linkage to the CRBP gene. *J. Biol. Chem.* **262**, 2458-2467.

Denton, M.J., Chen, J.-D., Serravalle, S., Colley, P., Halliday, F.B. and Donald, J. (1988). Analysis of linkage relationships of X-linked retinitis pigmentosa with the following Xp loci: L1.28, OTC, 754, XJ-1.1, pERT87 and C7. *Hum. Genet.* **70**, 60-64.

Deutmann, A.F. (1977). Rod-cone dystrophy: primary, hereditary, pigmentary retinopathy, RP. *In*: A.E. Krill (ed.) *Hereditary Retinal and Choroidal Diseases*, pp.479-576, Harper and Row Publishers Inc., Hagerston, Maryland, USA.

Dhallan, R.S., Macke, J.P., Eddy, R.L., Shows, T.B., Reed, R.R., Yau, K.-W. and Nathans, J. (1992). Human rod photoreceptor cGMP-gated channel: amino acid sequence, gene structure, and functional expression. *J. Neurosci.* **12**, 3248-3256.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morissette, J. and Weissenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152-154.

Dickson, C. and Peters, G. (1987). Potential oncogene product related to growth factors. *Nature* **326**, 833.

Dryja, T.P. (1992). Doyne Lecture: Rhodopsin and autosomal dominant retinitis pigmentosa. *Eye* **6**, 1-10.

Dryja, T.P., McGee, T.L., Reichel, E., Hahn, L.B., Cowley, G.S., Yandell, D.W., Sandberg, M.A. and Berson, E.L. (1990). A point mutation of the rhodopsin gene in one form of retinitis pigmentosa. *Nature* **343**, 364-366.

Dryja, T.P., Berson, E.L., Rao, V.R. and Oprian, D.D. (1993). Heterozygous missense mutation in the rhodopsin gene as a cause of congenital stationary night blindness. *Nature Genet.* **4**, 280-283.

Dryja, T.P. and Li, T. (1995). Molecular genetics of retinitis pigmentosa. *Hum. Molec. Genet.* **4**, 1739-1743.

Dryja, T.P., Finn, J.T., Peng, Y.-W., McGee, T.L., Berson, E.L. and Yau, K.-W. (1995). Mutations in the gene encoding the alpha-subunit of the rod cGMP-gated

channel in autosomal recessive retinitis pigmentosa. *Proc. Natl. Acad. Sci. USA* **92**, 10177-10181.

Dryja, T.P., Hahn, L.B., Reboul, T. and Arnaud, B. (1996). Missense mutation in the gene encoding the alpha subunit of rod transducin in the Nougaret for of congenital stationary night blindness. *Nature Genet.* **13**, 358-360.

El-Amraoui, A., Sahly, I., Picaud, S., Sahel, J., Abitbol, M. and Petit, C. (1996). Human Usher 1B/Mouse shaker-1 - the retinal phenotype discrepancy explained by the presence/absence of myosin VIIA in the photoreceptor cells. *Hum. Molec. Genet.* **5**, 1171-1178.

Elbedour, K., Zucker, N., Zalzstein, E., Barki, Y. and Carmi, R. (1994). Cardiac abnormalities in the Bardet-Biedl syndrome - echocardiographic studies of 22 patients. *Am. J. Med. Genet.* **52**, 164-169.

Elder, J.T., Astrom, A., Petterson, U., Voorhees, J.J. and Trent, J.M. (1992). Assignment of the human CRABP II gene to chromosome 1q21 by non-isotopic *in situ* hybridisation. *Hum. Genet.* **89**, 487-490.

Ellenberger, T.E., Brandl, C.J., Struhl, K. and Harrison, S.C. (1992). The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted apla-helices - crystal-structure of the protein-DNA complex. *Cell* **71**, 1223-1237.

Ellis, L.A., Taylor, G.R., Banks, R. and Baumberg, S. (1994). MutS binding protects heteroduplex DNA from exonuclease digestion *in vitro*: a simple method for detecting mutations. *Nucleic Acids Res.* **22**, 2710-2711.

Enfors, S.O. (1992). Control of *in vivo* proteolysis in the production of recombinant proteins. *Trends Biotech.* **10**, 310-315.

Ernst, W. and Moore, A.T. (1988) Heterogeneity, anomalous adaptation and incomplete penetrance in autosomal dominant retinitis pigmentosa. *In*: E. Zrenner, H. Krastel, H.-H. Goebel (eds.) *Research in Retinitis Pigmentosa*, pp.115-120, Pergammon Press, Oxford, England.

Falls, H.F. and Cotterman, C.W. (1948). Choroidoretinal degenerations: a sex-linked form in which heterozygous women exhibit a tapetal-like retinal reflex. *Arch. Ophthalmol.* **40**, 685-703.

Fannemel, M., Riise, R., Lofterod, B. and Tommerup, N. (1993). High-resolution chromosome analysis in autosomal recessive disorders: Laurence-Moon-Bardet-Biedl syndrome (Letter). *Clin. Genet.* **43**, 111-112.

Farag , T.I. and Teebi, A.S. (1989). High incidence of Bardet-Biedl syndrome among the Bedouin. (Letter) *Clin. Genet.* **36**, 463-465.

Farber, D.B. and Danciger, M. (1994). Inherited retinal degenerations in the mouse. *In*: A.F. Wright and B. Jay (eds.) *Molecular Genetics of Inherited Eye Disorders*, pp.123-151, Harwood Academic Publishers, Chur, Switzerland.

Farjo, Q., Jackson, A.U., Xu, J., Gryzenia, M., Skolnock, C., Agarwal, N. and Swaroop, A. (1993). Molecular characterisation of the murine neural retina leucine zipper gene, *Nrl. Genomics* **18**, 216-222.

Farrall, M. (1993). Homozygosity mapping - familiarity breeds debility. *Nat. Genet.* **5**, 107-108.

Farrar, G.J., Kenna, P., Jordan, S.A., Kumar-Singh, R., Humphries, M.M., Sharp, E.M., Sheils, D.M. and Humphries, P. (1991). A three-base-pair deletion in the peripherin-*RDS* gene in one form of retinitis pigmentosa. *Nature* **354**, 478-480.

Fawcett, D., Pasceri, P., Fraser, R., Colbert, M., Rossant, J. and Giguere, V. (1995). Postaxial polydactyly in forelimbs of *CRABP-II* mutant mice. *Development* **121**, 671-679.

Feng, G. and Winkler, M.E. (1995). Single-step purification of His_6-MutH, His_6-MutL and His_6-MutS repair proteins of *Escherichia coli* K-12. *Biotechniques* **19**, 956-965.

Figus, A., Angius, A., Loudianos, G., Bertini, C., Dessi, V., Loi, A., Delana, M., Lovicu, M., Olla, N., Sole, G., De Virgillis, S., Lilliu, F., Farci, A., Nurchi, A., Giacchino, R., Barabino, A., Marassi, M.G., Zancan, L., Greggio, N.A., Marcellini, M., Solinas, A., Deplano, A., Barbera, C., Devoto, M., Ozsoylu, S., Kocak, N., Akar, N., Karayaclin, S., Mokini, V., Cullufi, P., Balestrieri, A., Cao, A. and Pirastu, M. (1995). Molecular pathology and haplotype analysis of Wilson disease in Mediterranean populations. *Am. J. Hum. Genet.* **57**, 1318-1324.

Fitzgibbon, J., Pilz, A., Gayther, S., Appukuttan, B., Dulai, K.S., Delhanty, J.D.A., Helmkamp, G.M., Yarbrough, L.R. and Hunt, D.M. (1994). Localisation of the gene encoding human phosphatidylinositol transfer protein (*PITPN*) to 17p13.3 - a gene showing homology to the *Drosophila* retinal degeneration B gene (*rdgB*). *Cytogenetics Cell Genet.* **67**, 205-207.

Fitzgibbon, J., Katsanis, N., Wells, D., Delhanty, J., Vallins, W. and Hunt, D.M. (1996). Human guanylate kinase (*GUK1*) - cDNA sequence, expression and chromosomal localization. *FEBS Letters* **385**, 185-188.

Flannery, J.G., Chen, J., Xu, J. and Simon, M.I. (1995). Overexpression of Bcl-2 can interfere with apoptosis in retinal degeneration. *Invest. Ophthalmol. Vis. Sci.* **36**, S615 (Abstract 2842).

Fleck, O., Michael, H. and Heim, L. (1992). The *swi4*$^+$ gene of *Schizosaccharomyces pombe* encodes a homologue of mismatch repair enzymes. *Nucleic Acids Res.* **20**, 2271-2278.

Fodde, R. and Losekoot, M. (1994). Mutation detection by denaturing gradient gel electrophoresis (DGGE). *Hum. Mutat.* **3**, 83-94.

Fossarello, M., Serra, A., Mansfield, D., Wright, A., Loudianos, J., Pirastu, M and Orzalesi, N. (1993). Genetic and epidemiological study of autosomal dominant retinitis pigmentosa (ADRP) and autosomal recessive (ARRP) retinitis pigmentosa in

Sardinia. *In*: J.G. Hollyfield, R.E. Anderson, N. Orzalesi (eds.) *Retinal Degeneration*, pp.79-90, Plenum Press, New York, USA.

Foulkes, N.S. and Sassone-Corsi, P. (1992). More is better: activators and repressors from the same gene. *Cell* **68**, 411-414.

Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M. and Dryja T.P. (1986). A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643-646.

Fuchs, S., Nakazawa, M., Maw, M., Tamai, M., Oguchi, Y. and Gal, A. (1995). A homozygous one-base pair deletion in the arrestin gene is a frequent cause of Oguchi disease in Japanese. *Nature Genet.* **10**, 360-362.

Gal, A., Orth, U., Baehr, W., Schwinger, E. and Rosenberg, T. (1994). Heterozygous missense mutation in the rod cGMP phosphodiesterase beta-subunit gene in autosomal dominant stationary night blindness. *Nature Genet.* **7**, 64-68. Correction: *Nature Genet.* **7**, 551.

GDB - Online human genome database - http://www.hgmp.mrc.ac.uk/gdb/

Geisow, M.J. (1991). Both bane and blessing - inclusion bodies. *Trends Biotech.* **11**, 368-369.

Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K.A., Antonio, M., Beisel, K.W., Steel, K.P. and Brown, S.D.M. (1995). A type VII myosin encoded by the mouse deafness gene *shaker-1*. *Nature* **374,** 62-64.

Gilliam, T.C., Brzustowicz, L.M., Castillo, L.H., Lehner, T., Penchaszadeh, G.K., Daniels, R.J., Byth, B.C., Knowles, J., Hislop, J.E., Shapira, Y., Dubowtiz, V., Munsat, T., Ott, J. and Davies, K. (1990). Genetic homogeneity between acute and chronic forms of spinal muscular atrophy. *Nature* **345**, 823-825.

Glavac, D. and Dean, M. (1993). Optimization of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. *Hum. Mutat.* **2**, 404-414.

Glavac, D. and Dean, M. (1995). Applications of heteroduplex analysis for mutation detection in disease genes. *Hum. Mutat.* **6**, 281-287.

Goldberg, A.F.X., Moritz, O.L. and Molday, R.S. (1995). Heterologous expression of photoreceptor peripherin/RDS and rom-1 in *cos-1* cells: assembly, interactions and localisation of multi-subunit complexes. *Biochemistry* **34**, 14213-14219.

Gorczyca, W.A., Polans, A.S., Surgucheva, I.G., Subbaraya, I., Baehr, W. and Palczewski, K. (1995). Guanylyl cyclase-activating protein - a calcium sensitive regulator of phototransduction. *J. Biol. Chem.* **270**, 22029-22036.

Gough, J.A. and Murray, N.E. (1983). Sequence diversity among related genes for recognition of specific target molecules. *J. Mol. Biol.* **166**, 1-19.

Graff, C., Forsman, K., Larsson, C., Nordstrom, S., Lind, L., Johansson, K., Sandgren, O., Weissenbach, J., Holmgren, G., Gustavson, K.-H. and Wadelius, C. (1994). Fine mapping of Best's macular dystrophy localizes the gene in close proximity but distinct from the D11S480/*ROM-1* loci. *Genomics* **24**, 425-434.

Green, J.S., Parfrey, P.S., Harnett, J.D., Farid, N.R., Cramer, B.C., Johnson, G., Heath, O., McManamon, P.J., O'Leary, E. and Pryse-Phillips, W. (1989). The cardinal manifestations of Bardet-Biedl syndrome, a form of Laurence-Moon-Biedl syndrome. *N. Engl. J. Med.* **321**, 1002-1009.

Greenberg, J., Goliath, R., Beighton, P. and Ramesar, R. (1994). A new locus for autosomal dominant retinitis pigmentosa on the short arm of chromosome 17. *Hum. Molec. Genet.* **3**, 915-918.

Grondahl, J. (1986). Tapeto-retinal degeneration in four Norwegian counties: II. Diagnostic evaluation of 407 relatives and genetic evaluation of 87 families. *Clin. Genet.* **29**, 17-41.

Grondahl, J. (1987a). Autosomal recessive inheritance in "senile" retinitis pigmentosa. *Acta Ophthalmol.* **65**, 231-236.

Grondahl, J. (1987b). Estimation of prognosis and prevalence of retinitis pigmentosa and Usher syndrome in Norway. *Clin. Genet.* **31**, 255-264.

Guerts van Kessel, A., de Leeuw, H., Dekker, E.J., Rijks, L., Spurr, N., Ledbetter, D., Kootwijk, E. and Vaessen, M.J. (1991). Localization of the cellular retinoic acid-binding protein (CRABP) gene relative to the acute promyelocytic leukaemia-associated breakpoint on human chromosome 15. *Hum. Genet.* **87**, 201-204.

Guilford, P., Ayadi, H., Blanchard, S., Chaib, H., Le Paslier, D., Weissenbach, J., Drira, M. and Petit, C. (1994). A human gene responsible for neurosensory, non-syndromic recessive deafness is a candidate homologue of the mouse *sh-1* gene. *Hum. Mol. Genet.* **3**, 989-993.

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M. and Weissenbach, J. (1994). The 1993-94 Genethon human genetic linkage map. *Nature Genet.* **7**, 243-339.

Haber, L.T. and Walker, G.C. (1991). Altering the conserved nucleotide binding motif in the *Salmonella typhimurium* MutS mismatch repair protein affects both its ATPase and mismatch binding activities. *EMBO J.* **10**, 2707-2715.

Haber, L.T., Pang, P.P., Sobell, D.I., Mankovich, J.A. and Walker, G.C. (1988). Nucleotide sequence of the *Salmonella typhimurium mutS* gene required for mismatch repair: homology of MutS and HexA of *Streptococcus pneumoniae*. *J. Bacteriol.* **170**, 197-202.

Hahn, L.B., Berson, E.L. and Dryja, T.P. (1994). Evaluation of the gene encoding the gamma subunit of rod phosphodiesterase in retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **35**, 1077-1082.

Haider, N., Kwitek-Black, A.E., Iannaccone, A., Carmi, R., K., Elbedour, Nishimura, D.Y., Stone, E.M., and Sheffield, V.C. (1996). Genetic and physical fine mapping of the chromosome 15 Bardet-Biedl syndrome (BBS4) locus. *Am. J. Hum. Genet.* **59** , A220 - Abstract 1261.

Haim, M. (1992). Prevalence of retinitis pigmentosa and allied disorders in Denmark: 3. Hereditary pattern. *Acta Ophthalmol.* **70**, 615-624.

Hammerberg, B. (1990). *J. Biotechnol.* **14**, 423-438.

Hanson, I.M., Fletcher, J.M., Jordan, T., Brown, A., Taylor, D., Adams, R.J., Punnett, H.H. and van Heyningen, V. (1994). Mutations at the *PAX6* locus are found in heterogeneous anterior segment malformations including Peters' anomaly. *Nature Genet.* **6**, 168-173.

Harnett, J.D., Green , J.S., Cramer, B.C., Johnson, G., Chafe, L., McManamon, P., Farid, N., Pryse-Phillips, W. and Parfrey, P.S. (1988). The spectrum of renal disease in Laurence-Moon-Biedl syndrome. *N. Engl. J. Med.* **319**, 615-618.

Hasson, T., Heintzelman, M.B., Santos-Sacchi, J., Corey, D.P. and Mooseker, M.S. (1995). Expression in cochlea and retina of myosin VIIA, the gene product defective in Usher syndrome type 1B. *Proc. Natl. Acad. Sci. USA* **92**, 9815-9819.

Hawkins, R.K., Jansen, H.G. and Sanyal, S. (1985). Development and degeneration of retina in *rds* mutant mice: photoreceptor abnormalities in the heterozygotes. *Exp. Eye Res.* **41**, 701-720.

Heckenlively, J.R. (1988a). The diagnosis and classification of retinitis pigmentosa. *In*: J.R. Heckenlively (ed.) *Retinitis Pigmentosa*, pp.6-24, J.B. Lippincott Co., Philadelphia, USA.

Heckenlively, J.R. (1988b). Clinical findings in retinitis pigmentosa. *In*: J.R. Heckenlively (ed.) *Retinitis Pigmentosa*, pp.68-89, J.B. Lippincott Co., Philadelphia, USA.

Heckenlively, J.R. (1988c). Autosomal recessive retinitis pigmentosa. *In*: J.R. Heckenlively (ed.) *Retinitis Pigmentosa*, pp.150-161, J.B. Lippincott Co., Philadelphia, USA.

Heckenlively, J.R. (1988d). RP Syndromes. *In*: J.R. Heckenlively (ed.) *Retinitis Pigmentosa,* pp.221-252, J.B. Lippincott Co., Philadelphia, USA.

Heckenlively, J.R., Chang, B., Erway, L.C., Peng, C., Hawes, N.L., Hageman, G.S. and Roderick, T.H. (1995). Mouse model for Usher syndrome: Linkage mapping suggests homology to Usher type I reported at human chromosome 11p15. *Proc. Natl. Acad. Sci. USA* **92**, 11100-11104.

Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974). Analysis of R. *Eco*RI fragments of DNA from lambdoid bacteriophages and other viruses by agarose gel electrophoresis. *J. Virol.* **14**, 1235.

Hockney, R.C. (1994). Recent developments in heterologous protein production in *Escherichia coli*. *Trends Biotech.***12**, 456-463.

Hoffmann, A. and Roeder, R.G. (1991). Purification of His-tagged proteins in nondenaturing conditions suggests a convenient method for protein-interaction studies. *Nucleic Acids Res.* **19**, 6337-6338.

Holt, I.J., Harding, A.E., Petty, R.K.H. and Morgan-Hughes, J.A. (1990). A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *Am. J. Hum. Genet.* **46**. 428-433.

Hu, D. (1982). Genetic aspects of retinitis pigmentosa in China. *Am. J. Med. Genet.* **12**, 51-56.

Hu, D. (1987). Prevalence and mode of inheritance of major genetic eye diseases in China. *J. Med. Genet.* **24**, 584-588

Huang, S.H., Pittler, S.J., Huang, X., Oliveira, L., Berson, E.L. and Dryja, T.P. (1995). Autosomal recessive retinitis pigmentosa caused by mutations in the alpha-subunit of rod cGMP phosphodiesterase. *Nature Genet.* **11**, 468-471.

Hui, C.C. and Joyner, A.L. (1993). A mouse model of Greig cephalopolysyndactyly syndrome - the extra-toes (J) mutation contains an intragenic deletion of the *GLI3* gene. *Nature Genet.* **3**, 241-246.

Humphries, P., Kenna, P. and Farrar, G.J. (1992). On the molecular genetics of retinitis pigmentosa. *Science* **256**, 804-808.

Huq, L., McLachlan, T., Hammer, H.M., Bedford, D., Packard, C.J., Shepherd, J., *et al* (1993). An increased incidence of apolipoprotein E2/E2 and E4/E4 in retinitis pigmentosa. *Lipids* **28**, 995-998.

Inglehearn, C.F., Keen, T.J., Bashir, R., Jay, M., Fitzke, F., Bird, A.C., Crombie, A. and Bhattacharya, S.S. (1992). A completed screen for mutations of the rhodopsin gene in a panel of patients with ADRP. *Hum. Mol. Genet.* **1**, 41-45.

Inglehearn, C.F., Keen, T.J., Al-Maghtheh, M., Gregory, C.Y,. Jay, M.R., Moore, A.T., Bird, A.C. and Bhattacharya, S.S. (1994). Further refinement of the location for autosomal dominant retinitis pigmentosa on chromosome 7p (*RP9*). *Am. J. Hum. Genet.* **54**, 675-680.

Itakura, K., Rossi, J.J. and Wallace, R.B. (1984). Synthesis and use of synthetic oligonucleotides. *Annu. Rev. Biochem.* **53**, 323-356.

James, M.R., Richard, C.W., Schott, J.J., Yousry, C., Clark, K., Bell, J., Terwilliger, J.D., Hazan, J., Dubay, C., Vignal, A., Agrapart, M., Imai, T., Nakamura, Y., Polymeropoulos, M., Weissenbach, J., Cox D.R. and Lathrop, G.M. (1994). A radiation hybrid map of 506 STS markers spanning human chromosome 11. *Nature Genet.* **8**, 70-76.

Jay, M. (1982). On the heredity of retinitis pigmentosa. *Br. J. Ophthalmol.* **66**, 405-416.

Jorde, L.B. (1995). Linkage disequilibrium as a gene mapping tool. *Am. J. Hum. Genet.* **56**, 11-14.

Joseph, R.M. and Li, T. (1996). Overexpression of Bcl-2 or Bcl-X_L transgenes and photreceptor degeneration.

Kajiwara, K., Hahn, L.B., Mukai, S., Travis, G.H., Berson, E.L. and Dryja, T.P. (1991). Mutations in the human retinal degeneration slow gene in autosomal dominant retinitis pigmentosa. *Nature* **354**, 480-483.

Kajiwara, K., Sandberg, M.A., Berson, E.L. and Dryja, T.P. (1993). A *null* mutation in the human peripherin/*RDS* gene in a family with autosomal dominant retinitis punctata albescens. *Nature Genet.* **3**, 208-212.

Kajiwara, K., Berson, E.L. and Dryja, T.P. (1994). Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/*RDS* and *ROM1* loci. *Science* **264**, 1604-1608.

Kaplan, J., Bonneau, D., Frezal, J., Munnich, A. and Dufier, J.L. (1990). Clinical and genetic heterogeneity in retinitis pigmentosa. *Hum. Genet.* **85**, 635-642.

Kaplan, J., Gerber, S., Bonneau, D., Rozet, J.M., Delrieu, O., Briard, M.L., Dollfus, H., Ghazi, I., Dufier, J.L., Frezal, J. and Munnich, A. (1992). A gene for Usher syndrome type I (*USH1A*) maps to chromosome 14q. *Genomics* **14**, 979-987.

Kaplan, J.C., Kahn, A. and Chelly, J. (1992). Illegitimate transcription: its use in the study of inherited disease. *Hum. Mutat.* **1**, 357-360.

Keats, B.J.B., Nouri, N., Pelias, M.Z., Deininger, P.L. and Litt, M. (1994). Tightly linked flanking microsatellite markers for the Usher type I locus on the short arm of chromosome 11. *Am. J. Hum. Genet.* **54**, 681-686.

Keen, J., Lester, D., Inglehearn, C., Curtis, A. and Bhattacharya, S. (1991). Rapid detection of single base mismatches as heteroduplexes on Hydrolink gels. *Trends Genet.* **7**, 5.

Kenealy, W.R. (1987). *Dev. Ind. Microbiol.* **28**, 45-52.

Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M. and Tsui, L.C. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073-1080.

Kerppola, T.K. and Curran, T. (1994a). Maf and Nrl can bind to AP-1 sites and form heterodimers with Fos and Jun. *Oncogene* **9**, 675-684.

Kerppola, T.K. and Curran, T. (1994b). A conserved region adjacent to the basic domain is required for recognition of an extended DNA binding site by Maf/Nrl family proteins. *Oncogene* **9**, 3149-3158.

Khani, S.C., Abitbol, M., Yamamoto, S., Maravic-Magovcevic, I. and Dryja, T.P. (1996). Characterisation and chromosomal localisation of the gene for rhodopsin kinase. *Genomics* **35**, 571-576.

Kikawa, E., Nakazawa, M., Chida, Y., Shiono, T. and Tamai, M. (1994). A novel mutation (asn244-to-lys) in the peripherin/*RDS* gene causing autosomal dominant retinitis pigmentosa associated with bull's-eye maculopathy detected by nonradioisotopic SSCP. *Genomics* **20**, 137-139.

Kimberling, W.J., Moller, C.G., Davenport, S., Priluck, I.A., Beighton, P.H., Greenberg, J., Reardon, W., Weston, M.D., Kenyon, J.B., Grunkemeyer, J.A., Pieke Dahl, S., Overbeck, L.D., Blackwood, D.J., Brower, A.M., Hoover, D.M., Rowland, P. and Smith, R.J.H. (1992). Linkage of Usher syndrome type I gene (*USH1B*) to the long arm of chromosome 11. *Genomics* **14**, 988-994.

Kimberling, W.J., Weston, M. and Moller, C. (1994). Clinical and genetic heterogeneity of Usher syndrome. *In*: A.F. Wright and B. Jay (eds.) *Molecular Genetics of Inherited Eye Disorders*, pp.359-381, Harwood Academic Publishers, Chur, Switzerland.

Kimberling, W.J., Weston, M.D., Moller, C., van Aarem, A., Cremers, C.W.R., Sugemi, J., Ing, P.S., Connolly, C., Martini, A., Milani, M., Tamayo, M.L., Bernal, J., Greenberg, J. and Ayuso, C. (1995). Gene mapping of Usher syndrome type IIa: localization of the gene to a 2.1-cM segment on chromosome 1q41. *Am. J. Hum. Genet.* **56**, 216-223.

Kleyn, P.W., Fan, W., Kovats, S.G., Lee, J.J., Pulido, J.C., Wu, Y., Berkemeier, L.R., Misumi, D.J., Holmgren, L., Charlat, O., Woolf, E.A., Tayber, O., Brody, T., Shu, P., Hawkins, F., Kennedy, B., Baldini, L., Ebeling, C., Alperin, G.D., Deeds, J., Lakey, N.D., Culpepper, J., Chon, H., Glucksmann-Kuls, M.A., Carlson, G.A., Duyk, G.M. and Moore, K.J. (1996). Identification and characterisation of the mouse obesity gene *tubby*: a member of a novel gene family. *Cell* **85**, 1-20.

Knowles, J.A., Shugart, Y., Banerjee, P., Gilliam, T.C., Lewis, C.A., Jacobson, S.G. and Ott, J. (1994). Identification of a locus distinct from *RDS*-peripherin, for autosomal recessive retinitis pigmentosa on chromosome 6p. *Hum. Mol. Genet.* **3**, 1401-1403.

Kojis, T.L., Heinzmann, C., Flodman, P., Ngo, J.T., Sparkes, R.S., Spence, M.A., Bateman, J.B. and Heckenlively, J.R. (1996). Map refinement of locus *RP13* to human chromosome 17p13.3 in a second family with autosomal dominant retinitis pigmentosa. *Am. J. Hum. Genet.* **58**, 347-355.

Krill, A.E. (1972). Evaluation of night vision: dark adaptation. *In*: A.E. Krill (ed.) *Hereditary Retinal and Choroidal Diseases, Volume I - Evaluation,* pp.189-226. Harper and Row, Hagerston, Maryland, USA.

Kumar-Singh, R., Farrar, G.J., Mansergh, F., Kenna, P., Bhattacharya, S., Gal, A. and Humphries, P. (1993). Exclusion of the involvement of all known retinitis pigmentosa loci in the disease present in a family of Irish origin provides evidence for a sixth autosomal dominant locus (*RP8*). *Hum. Molec. Genet.* **2,** 875-878.

Kumaramanickavel, G., Maw, M., Denton, M.J., John, S., Srisailapathy Srikumari, C.R., Orth, U., Oehlmann, R. and Gal, A. (1994). Missense rhodopsin mutation in a family with recessive RP. *Nature Genet.* **8**, 10-11.

Kusters, J.G., Jager, E.J. and van der Zeijst, B.A.M. (1989). Improvement of the cloning linker of the bacterial expression vector pEX. *Nucleic Acids Res.* **17**, 8007.

Kwitek-Black, A.E., Carmi, R., Duyk, G.M., Buetow, K.H., Elbedour, K., Parvari, R., Yandava, C.N., Stone, E.M. and Sheffield, V.C. (1993). Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic genetic heterogeneity. *Nature Genet.* **5**, 392-396.

Kwitek-Black, A.E., Doggett, N.A., Carmi, R., Goodwin, L., Charlat, O.,. Stone, E.M. and Sheffield, V.C. (1996). Genetic and physical fine-mapping of the locus causing Bardet-Biedl syndrome in an inbred Bedouin family (Abstract). *Cytogenet. Cell Genet.* **72**, 291.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Lammer, E.J., Chen, D.T., Hoar, R.M., Agnish, N.D., Benke, P.J., Braun, J.T., Curry, C.J., Fernhoff, P.M., Grix, A.W., Lott, I.T., Richard, J.M. and Shyan, C.S. (1985). Retinoic acid embryopathy. *N. Engl. J. Med* **313**, 837-841.

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.

Lander, E.S. and Kruglyak, L. (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting results. *Nat. Genet.* **11**, 241-247.

Lander, E.S. and Kruglyak, L. (1996). Genetic dissection of complex traits: guidelines for interpreting and reporting results - in reply. *Nat. Genet.* **12**, 357-358.

Latchman, D.S. (1996). transcription factor mutations and disease. *N. Engl. J. Med.* **334**, 28-33.

Lathrop, G.M., Lalouel, J.M., Julier, C. and Ott, J. (1984). Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81**, 3443-3446.

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L.A., Nystrom-Lahti, M., Guan, X.-Y., Zhang, J., Meltzer, P.S., Yu, J.-W., Kao, F.-T., Chen, D.J., Cerosaletti, K.M., Fournier, R.E.K., Todd, S., Lewis, T., Leach, R.J., Naylor, S.L., Weissenbach, J., Mecklin, J.-P., Jarvinen, H., Petersen, G.M., Hamilton, S.R., Green, J., Jass, J., Watson, P., Lynch, H.T., Trent, J.M., de la Chapelle, A., Kinzler, K.W. and Vogelstein, B. (1993). Mutations of a *mutS* homolog in hereditary nonpolyposis colorectal cancer. *Cell* **75**, 1215-1225.

Lem, J., Flannery, J.G., Li, T., Applebury, M.L., Farber, D.B. and Simon, M.I. (1992). Retinal degeneration is rescued in transgenic *rd* mice by expression of the cGMP phosphodiesterase beta subunit. *Proc. Natl. Acad. Sci. USA* **89**, 4422-4426.

Leppert, M., Baird, L., Anderson, K.L., Otterund, B. Lupski, J.R and Lewis, R.A. (1994). Bardet-Biedl syndrome is linked to DNA markers on chromosome 11q and is genetically heterogeneous. *Nature Genet.* **7**, 108-112.

Leutelt, J., Oehlmann, R., Younus, F., van den Born, L.I., Weber, J.L., Denton, M.J., Mehdi, S.Q. and Gal, A. (1995). Autosomal recessive retinitis pigmentosa maps on chromosome 1q in a large consanguineous family from Pakistan. *Clin. Genet.* **47**, 122-124.

Lishanski, A., Ostrander, E.A. and Rine, J. (1994). Mutation detection by mismatch binding protein, MutS, in amplified DNA: application to the cystic fibrosis gene. *Proc. Natl. Acad. Sci. USA* **91**, 2674-2678.

Liou, G.I., Fong, S.L., Gosden, J., Vantuinen, P., Ledbetter, D.H., Christie, S., Rout, D., Bhattacharya, S., Cook, R.G., Li, Y., Wang, C. and Bridges, C.D.B. (1987). Human interstitial retinol-binding protein (IRBP) - cloning, partial sequence, and chromosomal localization. *Somat. Cell Molec. Genet.* **13**, 315-323.

Lisman, J. and Fain, G. (1995). Support for the equivalent light hypothesis for RP. *Nat. Med.* **1**, 1254-1255.

Liu, Q., Ji, X., Breitman, M.L., Hitchcock, P.F. and Swaroop, A. (1996). Expression of the bZIP transcription factor gene *Nrl* in the developing nervous system. *Oncogene* **12**, 207-211.

Lonnqvist, F., Arner, P., Nordfors, L. and Schalling, M.(1995). Overexpression of the obese (Ob) gene in adipose tissue of human obese subjects. *Nat. Med.* **1**, 950-953.

Lopez, A.J. (1995). Developmental role of transcription factor isoforms generated by alternative splicing. *Developmental Biol.* **172**, 396-411.

Lu, A.-L. and Chang, D.-Y. (1988). A novel nucleotide excision repair for the conversion of an A/G mismatch to C/G base pair in *E.Coli*. *Cell* **54**, 805-812.

Lu, A.-L. and Hsu, I.-C. (1992). Detection of single DNA base mutations with mismatch repair enzymes. *Genomics* **14**, 249-255.

Lui, I.S.C., Chen, J., Ploder, L., Vidgen, D., van der Kooy, D., Kalnins, V.I. and McInnes, R.R. (1994). Developmental expression of a novel murine homoebox gene (*CHX10*): evidence for roles in determination of the neuroretina and inner nuclear layer. *Neuron* **13**, 377-393.

Lyness, A.L., Ernst, W., Quinlan, M.P., Clover, G.M., Arden, G.B., Carter, R.M., Bird, A.C. and Parker, J.A. (1985). A clinical, psychophysical and electroretinographic survey of patients with autosomal dominant retinitis pigmentosa. *Br. J. Ophthalmol.* **69**, 326-339.

Mansfield, D.C., Brown, A.F., Green, D.K., Carothers, A.D., Morris, S.W., Evans, H.J. and Wright, A.F. (1994). Automation of genetic linkage analysis using fluorescent microsatellite markers. *Genomics* **24**, 225-233.

Massof, R.W. and Finkelstein, D. (1981). Two forms of autosomal dominant primary retinitis pigmentosa. *Doc. Ophthalmol.* **51**, 289-346.

Maw, M.A., John, S., Jablonka, S., Muller, B., Kumaramanickavel, G., Oehlmann, R., Denton, M.J. and Gal, A. (1995). Oguchi disease: suggestion of linkage to markers on chromosome 2q. *J. Med. Genet.* **32**, 396-398.

Mazuruk, K., Schoen, T.J., Chader, G.J. and Rodriguez, I.R. (1995). Structural organisation and expression of the human phophatidylinositol-specific phospholipase C beta-3 gene. *Biochem. Biophys. Res. Commun.* **212**, 190-195.

McGuire, R.E., Gannon, A.M., Sullivan, L.S., Rodriguez, J.A. and Daiger, S.P (1995a). Evidence for a major gene (*RP10*) for autosomal dominant retinitis pigmentosa on chromosome 7q: linkage mapping in a second, unrelated family. *Hum. Genet.* **95**, 71-74.

McGuire, R.E., Sullivan, L.S., Blanton, S.H., Church, M.W., Heckenlively, J.R. and Daiger, S.P. (1995b). X-linked dominant cone-rod degeneration: linkage mapping of a new locus for retinitis pigmentosa (*RP15*) to Xp22.13-p22.11. *Am. J. Hum. Genet.* **57**, 87-94.

McLaughlin, M.E., Sandberg, M.A., Berson, E.L. and Dryja, T.P. (1993). Recessive mutations in the gene encoding the beta-subunit of rod phosphodiesterase in patients with retinitis pigmentosa. *Nature Genet.* **4**, 130-134.

McLaughlin, M.E., Ehrhart, T.L., Berson, E.L. and Dryja, T.P. (1995). Mutation spectrum of the gene encoding the beta-subunit of rod phosphodiesterase among patients with autosomal recessive retinitis pigmentosa. *Proc. Natl. Acad. Sci. USA* **92**, 3249-3253.

McWilliam, P., Farrar, G.J., Kenna, P., Bradley, D.G., Humphries, M.M., Sharp, E.M., McConnell, D.J., Lawler, M., Sheils, D., Ryan, C., Stevens, K., Daiger, S.P. and Humphries, P. (1989). Autosomal dominant retinitis pigmentosa (ADRP): localization of an ADRP gene to the long arm of chromosome 3. *Genomics* **5**, 619-622.

Meindl, A., Dry, K., Herrmann, K., Manson, F., Ciccodicola, A., Edgar, A., Carvalho, M.R.S., Achatz, H., Hellebrand, H., Lennon, A., Migliaccio, C., Porter, K., Zrenner, E., Bird, A., Jay, M., Lorenz, B., Wittwer, B., D'Urso, M., Meitinger, T. and Wright, A. (1996). A gene (*RPGR*) with homolgy to the *RCC1* guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nature Genet.* **13**, 35-42.

Meins, M., Gruning, G., Blankenangel, A., Krastel, H., Reck, B., Fuchs, S., Schwinger, E. and Gal, A. (1993). Heterozygous null allele mutation in the human peripherin/*RDS* gene. *Hum. Molec. Genet.* **2**, 2181-2182.

Menozzi, P., Piazza, A. and Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science* **201**, 786-792.

Merin, S. and Auerbach, E. (1976). Retinitis pigmentosa. *Surv. Ophthalmol.* **20**, 303-346.

Messer, A., Plummer, J., Wong, V. and Lavail, M.M. (1993). Retinal degeneration in motor-neuron degeneration (*mnd*) mutant mice. *Exp. Eye Res.* **57**, 637-641.

Mitchell, P.J. and Tijan, R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245**, 371-378.

Modrich, P. (1987). DNA mismatch correction. *Ann. Rev. Biochem.* **56**, 435-466.

Monaco, A.P., Neve, R.L., Colletti-Feener, C., Bertelson, C.J., Kurnit, D.M. and Kunkel, I.M. (1986). Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* **323**, 646-650.

Morgan, B.A., Izpisua-Belmonte, J.C., Duboule, D. and Tabin, C.J. (1992). Targeted misexpression of Hox-4.6 in the avian limb bud causes apparent homeotic transformation. *Nature* **358**, 236.

Moritz, O.L. and Molday, R.S. (1996). Molecular cloning, membrane topology, and localisation of bovine rom-1 in rod and cone photoreceptor cells. *Invest Ophthalmol Vis Sci* **37**, 352-362.

Morris, T.A. and Fong, S.L. (1993). Characterisation of the gene encoding human cone transducin alpha-subunit (*GNAT2*). *Genomics* **17**, 442-448.

Muenke, M. and Schell, U. (1995). Fibroblast-growth-factor receptor mutations in human skeletal disorders. *Trends Genet.* **11**, 308-13.

Muragaki, Y., Mundlos, S., Upton, J. and Olsen, B.R. (1996). Altered growth and branching patterns in synpolydactyly caused by mutations in *HOXD13*. *Science* **272**, 548-551.

Musarella, M.A., Anson-Cartwright, L., Burghes, A., Worton, R.G., Lesko, J.G. and Nussbaum, R.L. (1989). Linkage analysis of a large Latin-American family with X-linked retinitis pigmentosa and a metallic sheen in the heterozygous carrier. *Genomics* **4**, 601-605.

Myers, R.M., Larin, Z. and Maniatis, T. (1985a). Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes. *Science* **230**, 1242-1246.

Myers, R.M., Lumelsky, N., Lerman, L.S. and Maniatis, T. (1985b). Detection of single base substitutions in total genomic DNA. *Nature* **313**, 495-498.

Nakamura, F., Sasaki, H., Kajihara, H. and Yamanoue, M. (1990). Laurence-Moon-Biedl syndrome accompanied by congenital hepatic fibrosis. *J. Gastroent. Hepatol.* **5**, 206-210.

Narcisi, T.M.E., Shoulders, C.C., Chester, S.A., Read, J., Brett, D.J., Harrison, G.B., Grantham, T.T., Fox, M.F., Povey, S., Bevruin, T.W.A., Erkelens, D.W., Muller, D.P.R., Lloyd, J.K. and Scott, J. (1995). Mutations of the microsomal triglyceride-transfer-protein gene in abetalipoproteinemia. *Am. J. Hum. Genet.* **57**, 1298-1310.

Nathans, J. (1992). Rhodopsin: structure, function, and genetics. *Biochemistry* **31**, 4923-4931.

Nathans, J. and Hogness, D.S. (1984). Isolation and nucleotide sequence of the gene encoding human rhodopsin. *Proc. Nat. Acad. Sci. USA* **81**, 4851-4855.

Nelson, S.F., McCusker, J.H., Sander, M.A., Kee, Y., Modrich, P. and Brown, P.O. (1993). Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nature Genet.* **4**, 11-18.

Nichols, B.E., Sheffield, V.C., Vandenburgh, K., Drack, A.V., Kimura, A.E. and Stone, E.M. (1993). Butterfly-shaped pigment dystrophy of the fovea caused by a point mutation in codon 167 of the *RDS* gene. *Nature Genet.* **3**, 202-207.

Nichols, B.E., Bascom, R., Litt, M., McInnes, R., Sheffield, V.C. and Stone, E.M. (1994). Refining the locus for Best vitelliform macular dystrophy and mutation analysis of the candidate gene *ROM-1*. *Am. J. Hum. Genet.* **54**, 95-103.

Nicoletti, A., Wong, D.J., Kawase, K., Gibson, L.H., Yang-Feng, T.l., Richards, J.E. and Thompson, D.A. (1995). Molecular characterisation of the human gene encoding an abundant 61kDa protein specific to the retinal pigment epithelium. *Hum. Molec. Genet.* **4**, 641-649.

Nilsson, B. and Abrahmsen, L. (1990). Fusions to Staphylococcal protein A. *Methods Enzymol.* **185**, 144-161.

Noben-Trauth, K., Naggert, J.K., North, M.A. and Nishina, P.M. (1996). A candidate gene for the mouse mutation *tubby*. *Nature* **380**, 534-538.

Norton, P.A. (1994). Polypyrimidine tract sequences direct selection of laternative branch sites and influence protein binding. *Nucleic Acids Res.***22**, 3854-3860.

Nussbaum, R.L., Lewis, R.A., Lesko, J.G. and Ferrell, R. (1985). Mapping X-linked ophthalmic diseases: II. Linkage relationship of X-linked retinitis pigmentosa to X chromosome short arm markers. *Hum. Genet.* **70**, 45-50.

Nygren, P.A., Stahl, S. and Uhlen, M. (1994). Engineering proteins to facilitate bioprocessing. *Trends Biotech.* **12**, 184-188.

Odea, D., Parfrey, P.S., Harnett, J.D., Hefferton, D., Cramer, B.C. and Green, J. (1996). The importance of renal impairment in the natural history of Bardet-Biedl syndrome. *Am. J. Kidney Diseases* **27**, 776-783.

Ohlemiller, K.K., Hughes, R.M., Mosinger-Ogilvie, J., Speck, J.D., Grosof, D.H. and Silverman, M.S. (1995). Cochlear and retinal degeneration in the *tubby* mouse. *Neuroreport* **6**, 845-849.

Oliviera, L., Miniou, P., Viegaspequignot, E., Rozet, J.M., Dollfus, H. and Pittler, S. (1994). Human retinal guanylate cyclase (GUC2D) maps to chromosome 17p13.1. *Genomics* **22**, 478-481.

OMIM - Online Mendelian Inheritance in Man (Dr. Victor McKusick, Johns Hopkins University, Baltimore, USA) - http://www.hgmp.mrc.ac.uk/omim/

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. and Sekiya, T. (1989). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**, 2766-2770.

Ott, J. (1985). *Analysis of human genetic linkage*. Johns Hopkins University Press, Baltimore, USA.

Pagon, R.A. (1988). Retinitis pigmentosa. *Surv. Ophthalmol.* **33**, 137-177.

Pagon, R.A., Haas, J.E., Bunt, A.H. and Rodaway, K.A. (1982). Hepatic involvement in the Bardet-Biedl syndrome. *Am. J. Med. Genet.* **13**, 373-381.

Pak, W.L. (1995) Drosophila in vision research - the Friedenwald lecture. *Invest. Ophthalmol. Vis. Sci.* **36**, 2340-2357.

Palczewski, K. (1994). Is vertebrate phototransduction solved? New insights into the molecular mechanism of phototransduction. *Invest.Ophthalmol. Vis. Sci.* **35**, 3577-3581.

Parker, B.O. and Marinus, M.G. (1992). Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proc. Natl.Acad.Sci. USA* **89**, 1730-1734.

Peng, Y.-W., Robishaw, J.D., Levine, M.A. and Yau, K.-W. (1992). Retinal rods and cones have distinct G protein beta and gamma subunits. *Proc. Natl. Acad. Sci. USA* **89**, 10882-10886.

Perrault, I., Rozet, J.M., Calvas, P., Gerber, S., Camuzat, A., Dollfus, H., Chatelin, S., Souied, E., Ghazi, I., Leowski, C., Bonnemaison, M., Le Paslier, D., Frezal, J., Dufier, J.-L., Pittler, S., Munnich, A. and Kaplan, J. (1996). Retinal-specific guanylate cyclase gene mutations in Leber's congenital amaurosis. *Nature Genet.* **14**, 461-464.

Peters, G., Brookes, S., Smith, R. and Dickson, C. (1983). Tumourigenesis by mouse mammary-tumour-virus - evidence for a common region for provirus integration in mammary tumours. *Cell* **33**, 364-377.

Piazza, A., van Loghem, E., de Lange, G., Curtoni, E.S., Ulizzi, L. and Terrenato, L. (1976). Immunoglobulin allotypes in Sardinia. *Am. J. Hum. Genet.* **28**, 77-86.

Pittler, S.J., Lee, A.K., Altherr, M.R., Howard, T.A., Seldin, M.F., Hurwitz, R.L., Wasmuth, J.J. and Baehr, W. (1992). Primary structure and chromosomal localisation of human and mouse rod photoreceptor cGMP-gated cation channel. *J. Biol. Chem.* **267**, 6257-6262.

Ploder, L., Liu, L., Dechen, J., Duncan, A., Nguyen, V., Cox, D.W., traboulsi, E., Levin, A. and McInnes R.R. (1995). Cloning of the human CHX10 gene and mutation screening of candidate disease. *Am. J. Hum. Genet.* **57**, 772 (Abstract).

Pollak, M.R., Wu Chou, Y.-H., Cerda, J.J., Steinmann, B., La Du, B.N., Seidman, J.G. and Seidman, C.E. (1993). Homozygosity mapping of the gene for alkaptonuria to chromosome 3q2. *Nature Genet.* **5**, 201-204.

Portera-Cailliau, C., Sung, C.-H., Nathans, J. and Adler, R. (1994). Apoptotic photoreceptor cell death in mouse models of retinitis pigmentosa. *Proc. Natl. Acad. Sci. USA* **91**, 974-978.

Prosser, J. (1993). Detecting single-base mutations. *Trends Biotech.* **11**, 238-246.

Puffenberger, E.G., Kauffman, E.R., Bolk, S., Matise, T.C., Washington, S.S., Angrist, M., Weiseenbach, J., Garver, K.L., Mascari, M., Ladda, R., Slaugenhaupt, S.A. and Chakravati, A. (1994). Identity-by-descent and association mapping of a recessive gene for Hirschsprung disease on human chromosome 13q22. *Hum. Molec. Genet.* **3**, 1217-1225.

Radovick, S., Nations, M., Du, Y.F., Berg, L.A., Weintraub, B.D. and Wondisford, F.E. (1992). A mutation in the POU-homeodomain of Pit-1 responsible for combined pituitary hormone deficiency. *Science* **257**, 1115-1118.

Reed, D.R., Ding, Y., Xu, W., Cather, C. and Price, R.A. (1995). Human obesity does not segregate with the chromosomal regions of Prader-Willi, Bardet-Biedl, Cohen, Borjeson or Wilson-Turner syndromes. *Int. J. Obesity* **19**, 599-603.

Reenan, R.A.G. and Kolodner, R.D. (1992). Isolation and characterisation of two *Saccharomyces cerevisiae* genes encoding homologs of the bacterial HexA and MutS mismatch repair proteins. *Genetics* **132**, 963-973.

Rehemtulla, A., Warwar, R., Kumar, R., Ji, X., Zack, D.J. and Swaroop, A. (1996). The basic motif-leucine zipper transcription factor Nrl can positively regulate rhodopsin gene expression. *Proc. Natl. Acad. Sci. USA* **93**, 191-195.

Reig, C., Serra, A., Gean, E., Vidal, M., Arumi, J., De la Calzada, M.D., Antich, J. and Carballo, M. (1995). A point mutation in the *RDS*-peripherin gene in a Spanish family with central areolar choroidal dystrophy. *Ophthalm. Genet.* **16**, 39-44.

Riess, O., Noerremoelle, A., Collins, C., Mah, D., Weber, B. and Hayden, M.R. (1992a). Exclusion of DNA changes in the beta-subunit of the cGMP phosphodiesterase gene as the cause for Huntington disease. *Nature Genet.* **1**, 104-108.

Riess, O., Noerremoelle, A., Weber, B., Musarella, M.A. and Hayden, M.R. (1992b). The search for mutations in the gene for the beta-subunit of the cGMP phosphodiesterase (*PDEB*) in patients with autosomal recessive retinitis pigmentosa. *Am. J. Hum. Genet.* **51**, 755-762.

Riise, R. (1987). Visual function in Laurence-Moon-Bardet-Biedl syndrome: a survey of 26 cases. *Acta Ophthalmol.* **65**, 128-131.

Riise, R. (1996). The cause of death in Laurence-Moon-Bardet-Biedl syndrome. *Acta Ophthalmol. Scandinavica* **74**, 45-47.

Ringens, P.J., Fang, M., Shinahara, T., Bridges, C.D., Lerea, C.L., Berson, E.L. and Dryja, T.P. (1990). Analysis of the gene coding for S-antigen, interstitial retinol binding protein, and the alpha subunit of cone transducin in pateints with retinitis pigmentosa. *Invest. Ophthalmol. Vis. Sci.* **31**, 1421-1426.

Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., Drumm, M.L., Iannuzzi, M.C., Collins, F.S. and Tsui, L.C. (1989). Identification of the cystic fibrosis gene: cloning and characteristaion of complementary DNA. *Science* **245**, 1066-1073.

Ris-Satlpers, C., Verleunmooijman, M.C.T., deblaei, T.J.P., Degenhart, H.J., Trapman, J. and Brinkmann, A.O. (1994). Differential splicing of human androgen receptor pre-messenger-RNA in X-Linked Reifenstein syndrome, because of a deletion involving a putative branch site. *Am. J. Hum. Genet.* **54**, 609-617.

Roberts, D.F. (1980). Genetic structure and the pathology of an isolated population. *In:* A.W. Eriksson (ed.) *Population Structure and Genetic Disorders*, pp. 7-26, Academic Press, London.

Roberts, R.G., Bobrow, M. and Bentley, D.R. (1992). Point mutations in the dystrophin gene. *Proc. Natl. Acad. Sci. USA* **89**, 2331-2335.

Rocchi, M., Covone, A., Romeo, G., Faraonio, R. and Colantuoni, V. (1989). Regional mapping of RBP4 to 1q23-q24 and RBP1 to 3q21-q22 in man. *Somat. Cell Molec. Genet.* **15**, 185-190.

Rommens, J.M., Iannuzzi, M.C., Kerem, B., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.l., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J.R., Tsui, L.C and Collins, F.S. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**, 1059-1065.

Rosatelli, M.C., Dozy, A., Faa, V., Meloni, A., Sardu, R., Saba, L., Kan, Y.W. (1992). Molecular characterisation of β-thalassaemia in the Sardinian population. *Am. J. Hum. Genet.* **50**, 422-426.

Rosenfeld, P.J., Cowley, G.S., McGee, T.L., Sandberg, M.A., Berson, E.L. and Dryja, T.P. (1992). A null mutation in the rhodopsin gene causes rod photoreceptor dysfunction and autosomal recessive retinitis pigmentosa. *Nature Genet.* **1**, 209-213.

Ruiz-Avila, L., McLaughlin, S.K., Wildman, D., McKinnon, P.J., Robichon, A., Spickofsky, N. and Margolskee, R.F. (1995). Coupling of bitter receptor to phosphodiesterase through transducin in taste receptor cells. *Nature* **376**, 80-85.

Rychlik, W. and Rhoads, R.E. (1989). A computer-program for choosing optimal oligonucleotides for filter hybridisation, sequencing and in vitro amplification of DNA. *Nuc. Acids. Res.* **17**, 8543-8551.

Sakuma, H., Inana, G., Murakami, A., Yajima, T., Weleber, R.G., Murphey, W.H., Gass, J.D.M., Hotta, Y., Hayakawa, M., Fujiki, K., Gao, Y.Q., Dancinger, M., Farber, D., Cideciyan, A.V. and Jacobson, S.G. (1995). A heterozygous putative null mutation in *ROM-1* without a mutation in peripherin/*RDS* in a family with retinitis pigmentosa. *Genomics* **27**, 384-386.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual, 2nd Edition.* Cold Spring Harbor Laboratory Press, USA.

Sanger, F., Niklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Sankila, E.M., Pakarinen, L., Kaariainen, H., Aittomaki, K., Karjalainen, S., Sistanen, P. and de-la-Chapelle, A. (1995). Assignment of an Usher syndrome type III (*USH3*) gene to chromosome 3q. *Hum. Molec. Genet.* **4**, 93-98.

Schachat, A.P. and Maumenee, I.H. (1982). Bardet-Biedl syndrome and related disorders. *Arch. Ophthalmol.* **100**, 285-288.

Schein, C.H. and Noteborn, M.H.M. (1988). Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperatures. *Biotechnology* **6**, 291-294.

Schlensog, V. and Bock, A. (1991). The *Escherichia Coli fdv* gene probably encodes MutS and is located at minute 58.8 adjacent to the *hyc-hyp* gene cluster. *J. Bacteriol.* **173**, 7414-7415.

Sheffield, V.C., Beck, J.S., Nichols, B., Cousineau, A., Lidral, A.C. and Stone, E.M. (1989). Attachment of a 40-base pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. USA* **86**, 232-236.

Sheffield, V.C., Carmi, R., Kwitek-Black, A., Rokhlina, T., Nishimura, D., Duyk, G.M., Elbedour, K., Sunden, S.L. and Stone, E.M. (1994). Identification of a Bardet-Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. *Hum. Molec. Genet.* **3**, 1331-1335.

Shen, W.H. and Hohn, B. (1992). DMSO improves PCR amplification of DNA with complex secondary structure. *Trends Genet.* **8**, 227.

Shugart, Y.Y., Banerjee, P., Knowles, J.A., Lewis, C.A., Jacobson, S.G., Matise, T.C., Penchaszadeh, G., Gilliam, T.C. and Ott, J. (1995). Fine genetic mapping of a gene for autosomal recessive retinitis pigmentosa on chromosome 6p21. *Am. J. Hum. Genet.* **57**, 499-502.

Sidhu, M.K., Liao, M.J. and Rashidbaigi, A. (1996). Dimethyl-sulfoxide improves RNA amplification. *Biotechniques* **21**, 44-47.

Sieving, P.A., Richards, J.E., Naarendorp, F., Bingham, E.L, Scott, K. and Alpern, M. (1995). Dark-light model for nightblindness from the human rhodopsin gly-90-asp mutation. *Proc. Natl. Acad. Sci. USA* **92**, 880-884.

Sinke, R.J. and Geurts van Kessel, A. (1995). Localization of the human phosphatidylinositol-specific phospholipase C β_3 gene (PLCB3) within chromosome band 11q13. *Genomics* **25**, 568-569.

Smith, J. and Modrich, P. (1996). Mutation detection with MutH, MutL, and MutS mismatch repair proteins. *Proc. Natl.Acad. Sci. USA* **93**, 4374-4379.

Smith, R.J.H., Lee, E.C., Kimberling, W.J., Daiger, S.P., Pelias, M.Z., Keats, B.J.B., Jay, M., Bird, A., Reardon, W., Guest, M., Attagari, R. and Hejtmancik, J.F. (1992).

Localisation of two genes for Usher syndrome type-1 to chromosome 11. *Genomics* **14**, 995-1002.

Souied, E., Soubrane, G., Benlian, P., Coscas, G.J., Gerber, S., Munnich, A. and Kaplan, J (1996). Retinitis punctata albescens associated with the arg135trp mutation in the rhodopsin gene. *Am. J. Ophthalmol.* **121**, 19-25.

Sparkes, R.S., Klisak, I., Kaufman, D., Mohandas, T., Tobin, A.J. and McGinnis, J. (1986). Assignment of the rhodopsin gene to human chromosome 3, region 3q21-3q24 by *in situ* hybridisation studies. *Curr. Eye Res.* **5**, 797-798.

Sparkes, R.S., Heinzmann, C., Goldflam, S., Kojis, T., Saari, J.C., Mohandas, T., Klisak, I., Bateman, J.B. and Crabb, J.W. (1992). Assignment of the gene (RLBP1) for cellular retinaldehyde-binding protein (CRALBP) to human chromosome 15q26 and mouse chromosome 7. *Genomics* **12**, 58-62.

Stoler, J.M., Herrin, J.T. and Holmes, L.B. (1995). Genital abnormalities in females with Bardet-Biedl syndrome. *Am. J. Med. Genet.* **55**, 276-278.

Stone, E.M., Nichols, B.E., Streb, L.M., Kimura, A.E. and Sheffield, V.C. (1992). Genetic linkage of vitelliform macular degeneration (Best's disease) to chromosome 11q13. *Nature Genet.* **1**, 246-250.

Su, S.-H. and Modrich, P. (1986). *Escherichia coli mutS*-encoded protein binds to mismatched DNA base pairs. *Proc. Natl. Acad.Sci USA* **83**, 5057-5061.

Subbaraya, I., Ruiz, C.C., Helekar, B.S., Zhao, X., Gorczyca, W.A., Pettenati, M.J., Rao, P.N., Palczweski, K. and Baehr, W. (1994). Molecular characterisation of human and mouse photoreceptor guanylate cyclase-activating protein (GCAP) and chromosomal localization of the human gene. *J. Biol. Chem.* **269**, 31080-31089.

Suber, M.L., Pittler, S.J., Qin, N., Wright, G.C., Holcombe, V., Lee, R.H., Craft, C.M., Lolley, R.N., Baehr, W. and Hurwitz, R. (1993). Irish setter dogs affected with rod/cone dysplasia contain a nonsense mutation in the rod cGMP phosphodiesterase beta-subunit gene. *Proc. Natl. Acad. Sci. USA* **90**, 3968-3972.

Sunden, S.L.F., Heon, E., Kwitek-Black, A.E., Green, J., Parfrey, P., Musarella, M.A., Nishimura, D., Harnett, J., Hefferton, D., Stone, E.M. and Sheffield, V.C. (1995). Genetic heterogeneity in 23 Canadian Bardet-Biedl families and evidence for a fifth BBS locus. *Invest. Ophthalmol. Vis. Sci* **36**, Abstract 3582.

Sung, C.-H., Makino, C., Baylor, D. and Nathans, J. (1994). A rhodopsin gene mutation responsible for autosomal dominant retinitis pigmentosa results in a protein that is defective in localization to the photoreceptor outer segment. *J. Neurosci.* **14**, 5818-5833.

Surguchev, A., Ruiz, C., Rao, N., Palczewski, K. and Baehr, W. (1996). Cloning, expression and chromosomal localisation of the human and mouse GCAP2 gene. *Invest. Ophthalmol. Vis. Sci.* **37**, Abstract 1538.

Swaroop, A., Xu, J., Agarwal, N and Weissman, S.M. (1991). A simple and efficient cDNA library subtraction procedure: isolation of human retina-specific cDNA clones. *Nucleic. Acids Res.* **19**, 1954.

Swaroop, A., Xu, J., Pawar, H., Jackson, A., Skolnick, C. and Agarwal, N. (1992). A conserved retina-specific gene encodes a basic motif/leucine zipper domain. *Proc. Natl. Acad. Sci. USA* **89**, 266-270.

Tabin, C.J. (1992). Why we have (only) 5 fingers per hand - *Hox* genes and the evolution of paired limbs. *Development* **116**, 289-296.

Tassabehji, M., Read, A.P., Newton, V.E., Harris, R., Balling, R., Gruss, P. and Strachan, T. (1992). Waardenburg's syndrome patients have mutations in the human homologue of the *Pax-3* paired-box gene. *Nature* **355**, 635-636.

Ton, C.C.T., Hirvonen, H., Miwa, H., Weil, M.M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N.D., Meijersheijboer, H., Dreschler, M., Royerpokora, B., Collins, F., Swaroop, A., Strong, L.C. and Saunders, G.F. (1991). Positional cloning and characterisation of a paired box- and homeobox-containing gene from the aniridia region. *Cell* **67**, 1059-1074.

Teague, P.W., Aldred, M.A., Jay, M., Dempster, M., Harrison, C., Carothers, A.D., Hardwick, L.J., Evans, H.J., Strain, L., Brock, D.J.H., Bundey, S., Jay, B., Bird, A.C., Bhattacharya, S.S. and Wright, A.F. (1994). Heterogeneity analysis in 40 X-linked retinitis pigmentosa families. *Am. J. Hum. Genet.* **55**, 105-111.

Terwillinger, J.D. and Ott, J. (1994). *Handbook of human genetic linkage.* John Hopkins University Press, Baltimore, Maryland.

Thompson, C.B. (1995). Apopotosis in the pathogenesis and treatment of disease. *Science* **267**, 1456-1462.

Travis, G.H., Christerson, L., Danielson, P.E., Klisak, I., Sparkes, R.S., Hahn, L.B., Dryja, T.P. and Sutcliffe, J.G. (1991). The human *retinal degeneration slow (RDS)* gene: chromosome assignment and structure of the mRNA. *Genomics* **10**, 733-739.

Tsang, S.H., Gouras, P., Yamashita, C.K., Kjeldbye, H., Fisher, J., Farber, D.B. and Goff, S.P. (1996). Retinal degeneration in mice lacking the gamma subunit of the rod cGMP phosphodiesterase. *Science* **272**, 1026-1029.

van den Born, L.I., Bergen, A.A.B. and Bleekers-Wagemakers, E.M. (1992). A retrospective study of registered retinitis pigmentosa patients in the Netherlands. *Ophthalm. Paediat. Genet.* **13**, 227-236.

van Soest, S., van den Born, L.I., Gal, A., Farrar, G.J., Bleeker-Wagemakers, L.M., Westerveld, A., Humphries, P., Sandkuijl, L.A. and Bergen, A.A.B. (1994). Assignment of a gene for autosomal recessive retinitis pigmentosa (*RP12*) to chromosome 1q31-q32.1 in an inbred and genetically heterogeneous disease population. *Genomics* **22**, 499-504.

van Soest, S., Nuehuis, S.T., van den Born, L.I., Bleeker-Wagemakers, L.M., Sankuijl, L.A., Westerveld, A. and Bergen, A.A.B. (1996). Fine mapping of the

autosomal recessive retinitis pigmentosa locus (*RP12*) on chromosome 1q - exclusion of the phosducin gene (*PDC*). *Cytogenet. Cell Genet.* **73**, 81-85.

Vihtelic, T.S., Hyde, D.R. and Otousa, J.E. (1991). Isolation and characterisation of a the *Drosophila* retinal degeneration B (rdgB) gene. *Genetics* **127**, 761-768.

Vortkamp, A., Gessler, M. and Grzeschik, K.H. (1991). *GLI3* zinc-finger gene interuppted by translocations in Greig syndrome patients. *Nature* **352**, 539-540.

Wagner, R., Debbie, P. and Radman, M. (1995). Mutation detection using immobilised mismatch binding protein (MutS). *Nucleic Acids Res.* **23**, 3944-3948.

Wayne, S., Der Kaloustian, V.M., Schloss, M., Polomeno, R., Scott, D.A., Hejtmancik, J.F., Sheffield, V.C. and Smith, R.J.H. (1996). Localization of the Usher syndrome type ID gene (Ush1D) to chromosome 10. *Hum. Molec. Genet.* **5**, 1689-1692.

Webb, J.C., Patel, D.D., Shoulders, C.C., Knight, B.L. and Soutar, A.K. (1996). Genetic variation at a splicing branch point in intron 9 of the low-density-lipoprotein (LDL)-receptor gene - a rare mutation that disrupts messenger RNA splicing in a patient with familial hypercholesterolemia and a common polymorphism. *Hum. Molec. Genet.* **5**, 1325-1331.

Weber, B., Riess, O., Hutchinson, G., Collins, C., Lin, B., Kowbel, D., Andrew, S., Schappert, K. and Hayden, M.R. (1991). Genomic organisation and complete sequence of the human gene encoding the beta-subunit of the cGMP phosphodiesterase and its localisation to 4p16.3. *Nucleic Acids Res.* **19**, 6263-6268.

Weil, D., Blanchard, S., Kaplan, J., Guilford, P., Gibson, F., Walsh, J., Mburu, P., Varela, A., Levilliers, J., Weston, M.D., Kelley, P.M., Kimberling, W.J., Wagenaar, M., Levi-Acobas, F., Larget-Piet, D., Munnich, A., Steel, K.P., Brown, S.D.M. and Petit, C. (1995). Defective myosin VIIA gene responsible for Usher syndrome type 1B. *Nature* **374**, 60-61.

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

Weleber, R.G., Carr, R.E., Murphey, W.H., Sheffield, V.C. and Stone, E.M. (1993). Phenotypic variation including retinitis pigmentosa, pattern dystrophy, and fundus flavimaculatus in a single family with a deletion of codon 153 or 154 of the peripherin/*RDS* gene. *Arch. Ophthalmol.* **111**, 1531-1542.

Wells, J., Wroblewski, J., Keen, J., Inglehearn, C., Jubb, C., Eckstien, A., Jay, M., Arden, G., Bhattacharya, S., Fitzke, F. and Bird, A. (1993). Mutations in the human retinal degeneration slow (*RDS*) gene can cause either retinitis pigmentosa or macular dystrophy. *Nature Genet.* **3**, 213-218.

White, M.B., Carvalho, M., Derse, D., O'Brien, S.J. and Dean, M. (1992). Detecting single base substitutions as heteroduplex polymorphisms. *Genomics* **12**, 301-306.

Wilkinson, D.G., Bhatt, S. and McMahon, A.P. (1989). Expression pattern of the FGF-related proto-oncogene *int-2* suggests multiple roles in fetal development. *Development* **105**, 131-136.

Winship, P.R. (1989). An improved method for directly sequencing PCR amplified material using dimethyl-sulfoxide. *Nuc. Acids Res.* **17**, 1266.

Winter, E.M., Yamamoto, F, Almoguera, C. and Perucho, M. (1985). A method to detect and characterise point mutations in the transcribed genes: amplification and overexpression of the mutant c-Kis-*ras* allele in human tumor cells. *Proc. Natl. Acad. Sci. USA* **82**, 7575-7579.

Woodford, K., Weitzmann, M.N. and Usdin, K. (1995). The use of K+ - free buffers eliminates a common cause of premature chain termination in PCR and PCR sequencing. *Nuc. Acids Res.* **23**, 539.

Workman, P.L., Lucarelli, P., Agostino, R., Scarabino, R., Scacchi, R., Carapella, E., Palmarino, R. and Bottini, E. (1976). Genetic differentiation among Sardinian villages. *Am. J. Phys. Anthrop.* **43**, 165-176.

Wright, A.F. (1992). New insights into genetic eye disease. *Trends Genet.* **8**, 85-91.

Wright, A.F. (1994). Degeneration of the retina. *In*: G. Jolles, J.M. Stutzmann (eds.) *Neurodegenerative Diseases,* pp183-203. Academic Press, London, UK.

Wright, A.F., Bhattacharya, S.S., Clayton, J.F., Dempster, M., Tippett, P., McKeown, C.M.E., Jay, M., Jay, B. and Bird, A.C. (1987). Linkage relationships between X-linked retinitis pigmentosa and nine short-arm markers: exclusion of the disease locus from Xp21 and localisation to between DXS7 and DXS14. *Am. J. Hum. Genet.* **41**, 635-644.

Xu, S.Y., Schwartz, M., Rosenberg, T. and Gal, A. (1996). A 9th locus for autosomal dominant retinitis pigmentosa maps in the pericentromeric region of chromosome 1. *Hum. Molec. Genet.* **5**, 1193-1197.

Yang-Feng, T.L. and Swaroop, A. (1992). Neural retina-specific leucine zipper gene *NRL* (D14S46E) maps to human chromosome 14q11.1-q11.2. *Genomics* **14**, 491-492.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains - nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

Yau, K.W. (1994). Phototransduction mechanism in retinal rods and cones - the Friedenwald Lecture. *Invest. Ophthalmol. Vis Sci.* **35**, 9-32.

Zhang, W., Hu, G. and Deisseroth, A. (1991). Improvement of PCR sequencing by formamide. *Nucleic Acids Res.* **19**, 6649.

Zlotogora, J. (1994). High frequencies of human genetic diseases: founder effect with genetic drift or selection? *Am. J. Med. Genet.* **49**, 10-13.

Zueco, J. and Boyd, A. (1992). Protein A fusion vectors for use in combination with pEX vectors in the production and affinity purification of specific antibodies. *Gene* **121**, 181-182.

GENOMICS 41, 93–99 (1997)
ARTICLE NO. GE974613

Linkage Mapping in 29 Bardet–Biedl Syndrome Families Confirms Loci in Chromosomal Regions 11q13, 15q22.3–q23, and 16q21

E. A. BRUFORD,[1] R. RIISE,[2] P. W. TEAGUE,[1] K. PORTER,[1] K. L. THOMSON,[1] A. T. MOORE,[3] M. JAY,[3] M. WARBURG,[4] A. SCHINZEL,[5] N. TOMMERUP,[6] K. TORNQVIST,[7] T. ROSENBERG,[8] M. PATTON,[9] D. C. MANSFIELD,[1] AND A. F. WRIGHT[1,10,*]

[1]*MRC Human Genetics Unit and* [10]*Department of Medicine, Western General Hospital Trust, Crewe Road, Edinburgh EH4 2XU, United Kingdom;* [2]*Department of Ophthalmology, Central Hospital of Hedmark, N-2300 Hamar, Norway;* [3]*Department of Clinical Ophthalmology, Institute of Ophthalmology, University College London, Bath Street, London WC1E 6BT, United Kingdom;* [4]*Department of Ophthalmology, Division of Paediatric Ophthalmology and Handicaps, Faculty of Medicine, University of Copenhagen, 2820 Gentofte, Denmark;* [5]*Institute of Medical Genetics, University of Zurich, 8006 Zurich, Switzerland;* [6]*Department of Medical Genetics, The John F. Kennedy Institute, Glostrup, DK 2600, Denmark;* [7]*Department of Ophthalmology, Eye Hospital, University Hospital of Lund, S-221 00 Lund, Sweden;* [8]*National Eye Clinic for the Visually Impaired, Copenhagen DK 2900, Denmark; and* [9]*Department of Clinical Genetics, St. George's Hospital Medical School, Cranmer Terrace, London SW17 ORE, United Kingdom*

Received July 22, 1996; accepted January 14, 1997

Bardet–Biedl syndrome (BBS) is a clinically and genetically heterogeneous autosomal recessive disorder characterized by retinitis pigmentosa, polydactyly, obesity, hypogenitalism, mental retardation, and renal anomalies. To detect linkage to BBS loci, 29 BBS families, of mixed but predominantly European ethnic origin, were typed with 37 microsatellite markers on chromosomes 2, 3, 11, 15, 16, and 17. The results show that an estimated 36–56% of the families are linked to the 11q13 chromosomal site (BBS1) previously described by M. Leppert *et al.* (1994, *Nature Genet.* 7, 108–112), with the gene order cen–D11S480–5 cM–BBS1–3 cM–D11S913/D11S987–qter. A further 32–35% of the families are linked to the BBS4 locus, reported by R. Carmi *et al.* (1995, *Hum. Mol. Genet.* 4, 9–13) in chromosomal region 15q22.3–q23, with the gene order cen–D15S125–5 cM–BBS4–2 cM–D15S131/D15S204–qter. Three consanguineous BBS families are homozygous for three adjacent chromosome 15 markers, consistent with identity by descent for this region. In one of these families haplotype analysis supports a localization for BBS4 between D15S131 and D15S114, a distance of about 2 cM. Weak evidence of linkage to the 16q21 (BBS2) region reported by A. E. Kwitek-Black *et al.* (1993, *Nature Genet.* 5, 392–396) was observed in 24–27% of families with the gene order cen–D16S408–2 cM–BBS2–5 cM–D16S400. A fourth group of families, estimated at 8%, are unlinked to all three of the above loci, showing that at least one other BBS locus remains to be found. No evidence of linkage was found to markers on chromosome 3, corresponding to the BBS3 locus, reported by V. C. Sheffield *et al.* (1994, *Hum. Mol. Genet.* 3, 1331–1335), or on chromosome 2 or 17, arguing against the involvement of a BBS locus in a patient with a t(2;17) translocation. © 1997 Academic Press

* To whom correspondence should be addressed at MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK.

INTRODUCTION

Bardet–Biedl syndrome (BBS; MIM 209900) is an autosomal recessive condition in which the five cardinal features of retinal degeneration, polydactyly, obesity, hypogenitalism, and mental retardation are manifest to varying degrees within and between families (Schachat and Maumenee, 1982; Green *et al.*, 1989). Renal anomalies are another common manifestation, often causing significant morbidity, so that this is increasingly regarded as a cardinal sign (Harnett *et al.*, 1988; Green *et al.*, 1989). The population prevalence varies from 1 in 17,500 to 1 in 160,000 in different populations (Green *et al.*, 1989; Klein and Ammann, 1969), with high local frequencies within certain Arab groups associated with a high frequency of consanguineous marriages (Farag and Teebi, 1989). Linkage has been reported to distinct BBS loci in 3 large consanguineous Bedouin kindreds at chromosomal regions 3p13–p12 (BBS3), 15q22.3–q23 (BBS4), and 16q21 (BBS2) (Sheffield *et al.*, 1994; Carmi *et al.*, 1995a; Kwitek-Black *et al.*, 1993) and in a heterogeneous group of 31 North American BBS families to chromosomal region 11q13 (BBS1) (Leppert *et al.*, 1994). A clinical study of 11 Scandinavian families with Bardet–Biedl syndrome, all of whom are included in the present study, emphasized the relatively broad phenotypic manifestations of the disorder within and between families and confirmed the increased prevalence of associated features including dental anomalies, diabetes mel-

0888-7543/97 $25.00

litus, deafness, and small stature (Riise *et al.,* 1997). The most consistent clinical findings in affected individuals consist of a severe retinal dystrophy, often leading to blindness by the end of the second decade; obesity; hypogenitalism, which is most evident in males; digital anomalies such as polydactyly, often with brachydactyly and/or syndactyly; mental retardation; and renal anomalies.

In this study, DNA samples were obtained from 29 families, each containing at least two affected members, from nine different countries, to confirm the reported linkages and to determine the relative frequencies of the different loci.

METHODS

Subjects and diagnostic criteria. The ethnic origins of the families are shown in Table 2. The families are from the United Kingdom (38%), Norway (21%), Denmark (14%), Switzerland (7%), Sweden (7%), Egypt (3%), Italy (3%), Somalia (3%), and Brazil (3%). The criteria for inclusion of families were the presence of at least 2 affected members, at least 1 of whom met the diagnostic criteria for three of the five cardinal signs [retinal dystrophy with subnormal ERG, polydactyly, obesity (body mass index >28 kg/m^2), hypogenitalism, and mental retardation]. Sibs of affected individuals were required to meet two instead of three of the cardinal signs. The total sample consists of 69 affected members from 29 kindreds, 2 of which contain 2 families with affected children. Parents and unaffected siblings (N = 78) were also sampled wherever available, making a total sample of 147 individuals. Parental consanguinity was present in 6 families (BB7, BB13, BB18, BB20, BB25, and BB31).

Genotyping. Subjects were sampled by venopuncture after ethical approval and informed consent were obtained. DNA was extracted by standard methods (Aldred *et al.,* 1994), and polymorphic microsatellites were analyzed using a semiautomated method of linkage analysis with an automated sequencer (ALF, Pharmacia) and single fluorescein-labeled primer, as described (Mansfield *et al.,* 1994).

Linkage and heterogeneity analyses. Linkage analysis was carried out by multipoint analysis using the LINKAGE program package (LINKMAP) (Lathrop *et al.,* 1984). LINKMAP was used to obtain multipoint lod scores by means of a series of three-point analyses on adjacent loci moving along one locus at a time. A linear genetic map over equally spaced fixed points on each chromosome was derived from these lod scores and used for the analysis. The genetic distances of the loci were taken from published data. Heterogeneity analysis was carried out using either the HOMOG program package (Terwilliger and Ott, 1994) or a Fortran program that located the maximum of the total lod score using the formulas employed by HOMOG. The following markers were typed: D2S126, D2S207, D2S139, D2S138, D2S119, D2S114, D2S141, D2S206, D2S168, D2S171, D2S134, D2S155, D2S140, D2S160, D3S1271, D3S1276, D3S1251, D11S480, D11S987, D11S913, D11S916, D11S527, D11S906, D11S901, D15S125, D15S131, D15S204, D15S114, D16S408, D16S400, D16S421, D17S799, D17S784, D17S796, D17S798, D17S807, and D17S849.

RESULTS

Chromosome 11 (BBS1) Linkage and Heterogeneity Analyses

The results of multipoint linkage analysis in 29 Bardet–Biedl syndrome families, using seven microsatellites reported to be linked to the BBS1 locus by Leppert *et al.* (1994), are shown in Fig. 1. The markers used were D11S480, D11S913, D11S987, D11S916, D11S527, D11S906, and D11S901. The multipoint lod scores from the combined family data provide statistically significant evidence of linkage, with a maximum likelihood lod score of 6.26 at a genetic location 5 cM distal to D11S480 and 3 cM proximal to D11S913/D11S987 on chromosome 11q. The order of the loci is cen–D11S480–5 cM–BBS1–3 cM–D11S913/D11S987–11 cM–D11S916–5 cM–D11S527/D11S906–7 cM–D11S901–qter.

Heterogeneity analysis using the multipoint data and HOMOG programs (Terwilliger and Ott, 1994) provides statistically significant evidence of genetic heterogeneity (Table 1), with the maximum lod score of 6.26 obtained under the hypothesis of two loci, one of which is linked (H2), being 28 times greater than the lod score, 4.82, assuming a single linked locus (H1) (χ^2 = 6.64, P = 0.005). The maximum likelihood value occurs with 56% of the families linked to chromosome 11 when it is considered singly, or with 36% of families linked to this locus when all four regions (11, 15, 16, and unlinked) are considered simultaneously (Table 1). Haplotype analysis was consistent with the estimated conditional probabilities of linkage in each kindred, showing that 12 families had a probability >0.5 of linkage to the BBS1 locus (Table 2).

Chromosome 15 (BBS4) Linkage and Heterogeneity Analyses

The results of multipoint linkage analysis in 29 Bardet–Biedl syndrome families, using four microsatellites (D15S125, D15S131, D15S204, and D15S114) from the chromosome 15 region shown to be linked to BBS (BBS4) by Carmi *et al.* (1995a), are shown in Fig. 1. The maximum likelihood multipoint lod score is 6.10 at a genetic location 5 cM distal to D15S125 and 2 cM proximal to D15S131/D15S204 (Fig. 1). The order of loci is cen–D15S125–5 cM–BBS4–2 cM–D15S131/D15S204–2.5 cM–D15S114–qter.

Heterogeneity analysis was then carried out and showed statistically significant genetic heterogeneity (Table 1). The maximum lod score obtained under the hypothesis of two loci (H2) is 6.10 compared with 3.84 on the hypothesis (H1) of a single linked locus (180:1, χ^2 = 10.39, P = 0.0006). The maximum likelihood occurs with 35% of families linked to the above locus when chromosome 15 linkage is considered alone, or 32% of families when all four possible loci are considered together. Eight families show a conditional probability >0.5 of linkage to BBS4 on the basis of multipoint lod scores (Table 2), which is consistent with the haplotype analysis (not shown). Three of these families (BB7, BB18, and BB31) contain the offspring of consanguineous marriages, and in each case there is marker homozygosity spanning three adjacent chromosome 15 markers (D15S125, D15S204, and D15S131), indicating identity by descent and resulting in significantly

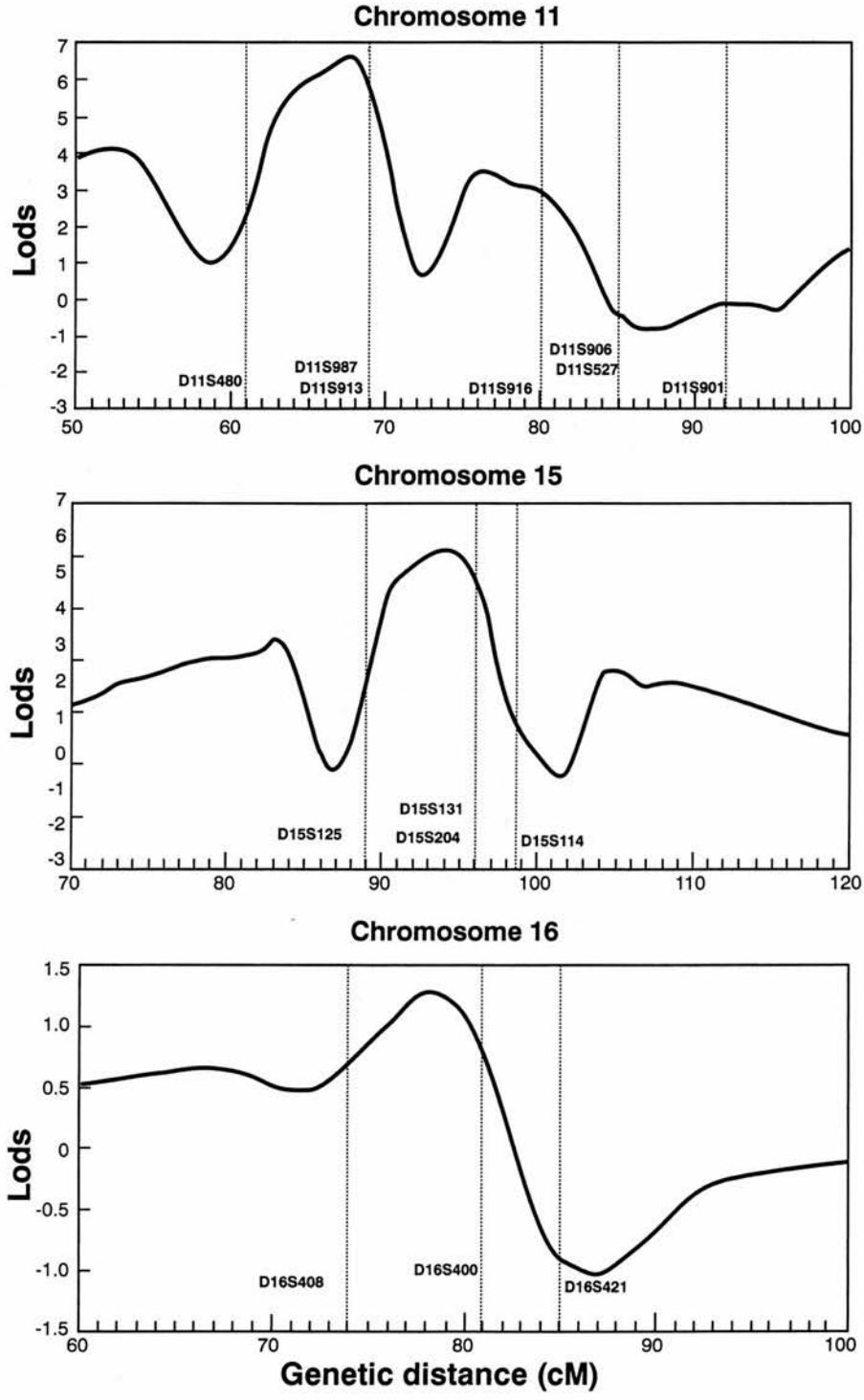


FIG. 1. Multipoint lod scores in Bardet–Biedl families corresponding to different possible genetic locations (in centimorgans from the distal short arm) for the disease locus in relation to markers from chromosomes 11, 15, and 16.

higher lod scores (1.69, 2.37, and 2.85) than in nonconsanguineous families. Haplotype analysis in consanguineous family BB31 (Fig. 2) shows two informative crossovers that help to localize the disease locus further. A crossover in affected male 31A places BBS4 proximal to D15S114; a second crossover in unaffected female 31B places BBS4 distal to D15S131. The distance between D15S114 and D15S131 is about 2 cM (Dib *et al.,* 1996).

The average lod score for the six consanguineous families, taken at the maximum likelihood chromosomal location, was 1.59 compared with 0.57 for the remaining families (excluding only BB8, which showed negative lod scores at each BBS locus). This empha-

TABLE 1

Summary of Linkage Heterogeneity Analysis in Bardet–Biedl Syndrome Families

Locus	Proportion of families linked, α	Genetic location (max likelihood lod score)	Linkage hypothesis (χ^2; P value)
Chromosome 11	0.36 (0.56)	66 cM (6.26)	H2 vs H0 ($\chi^2 = 28.84$;$P = 0.0000$) H2 vs H1 ($\chi^2 = 6.64$; $P = 0.005$)
Chromosome 15	0.32 (0.35)	94 cM (6.10)	H2 vs H0 ($\chi^2 = 28.07$, $P = 0.0000$); H2 vs H1 ($\chi^2 = 10.39$, $P = 0.0006$)
Chromosome 16	0.24 (0.27)	76 cM (1.09)	H2 vs H0 ($\chi^2 = 5.02$; $P = 0.041$) H2 vs H1 ($\chi^2 = 3.50$; $P = 0.031$)

Note. The linkage hypotheses are H2—two disease loci, only one of which is linked to the markers; H1—one disease locus linked to the markers; and H0—null hypothesis of no linkage. The genetic location is shown in centimorgans (cM) from the distal short arm of the chromosome based on the location of markers described by Gyapay *et al.* (1994). The proportion of linked families was calculated by testing for linkage to all three loci versus nonlinkage simultaneously or individually (in parentheses).

sizes the greater utility of consanguineous families for mapping recessive disorders.

Chromosome 16 (BBS2) Linkage and Heterogeneity Analyses

The results of linkage analysis in 29 BBS families, using markers D16S408, D16S400, and D16S421, which are close to the BBS2 locus (Kwitek-Black *et al.*, 1993), are shown in Fig. 1. The maximum likelihood lod score is 1.09 under the hypothesis of genetic heterogeneity (H2) ($\chi^2 = 5.02, P = 0.04$), a result of borderline significance (Table 1). Under the hypothesis of heterogeneity, 27% of families are estimated to be linked to this site when chromosome 16 linkage is considered alone, compared with 24% when all four locations are considered jointly. The estimated gene location is cen–D16S408–2 cM–BBS2–5 cM–D16S400–4 cM–D16S421–qter.

Haplotype and conditional probability analyses show that only two families have a probability >0.5 of linkage to BBS2 (Table 2), although two additional families have probabilities of linkage >0.4 for chromosome 16, which is higher than that for any other site.

Chromosomal Exclusions

One family, BB8, shows negative lod scores at each BBS locus, with lod scores of −2.31, −2.09, and −1.4 at the maximum likelihood locations for the chromosome 11, 15, and 16 marker regions. Three other families show lod scores ≤0.10 for all three regions. These results indicate that at least one other BBS locus remains to be mapped.

Sixteen markers distributed on chromosomes 2, 3, and 17 were analyzed for linkage to BBS. Strongly negative lod scores were obtained (data not shown) in each case, excluding the BBS3 locus and a proposed site on either chromosome 2 or 17 corresponding to the breakpoints of a balanced translocation in a BBS patient (Dallapiccola, 1971).

DISCUSSION

Bardet–Biedl syndrome has proved to be genetically very heterogeneous for such a comparatively rare disorder. The initial linkage of BBS to markers from chromosomal region 16q21 (BBS2) was made in a single inbred Bedouin kindred by Kwitek-Black *et al.* (1993). Other inbred families from the Middle East failed to show linkage, and the disease in these families was later mapped to loci in chromosomal regions 3p13–p12 (BBS3) and 15q22.3–q23 (BBS4) (Sheffield *et al.*, 1994; Carmi *et al.*, 1995a). The only reported linkage study of BBS in a mixed population of families is that of Leppert *et al.* (1994), who found evidence of genetic heterogeneity with linkage to markers from chromosomal region 11q13 (BBS1) in 17 of 31 North American families, which were similar in size and ethnic origin to those used in this study. The results of the present study from an ethnically diverse, predominantly European group of 29 families are very similar to those of Leppert *et al.* (1994) with regard to the 11q13 locus. Our data confirm the BBS1 locus as the most common site of mutation in Bardet–Biedl syndrome, with an observed 36–56% of families having mutations at this locus (Table 1).

The precise genetic location of BBS1 is difficult to establish with certainty since recombination events in individual small families, which might refine the localization, cannot be distinguished from nonlinkage due to heterogeneity. Nevertheless, the location of BBS1 identified in this study is very similar to that of Leppert *et al.* (1994). The maximum likelihood location of the BBS1 locus is 3 cM proximal to D11S913 (Fig. 1), which compares well with the estimate of 5 cM from D11S913 in a subset of 17 PYGM-linked families. The Utah/Houston group found the PYGM marker to be apparently closer than D11S913 to BBS1, although this probably occurred only because this marker is less informative than D11S913, which is estimated to be 3 cM distal to PYGM (Quackenbush *et al.*, 1995) and was found to segregate together with the latter in 39 families (Leppert *et al.*, 1994). The statistical significance of our BBS1 linkage

TABLE 2

Exclusion or Inclusion of Linkage to Bardet–Biedl Syndrome Loci on Chromosomes 11 (BBS1), 15 (BBS4), and 16 (BBS2) in Individual Families Based on the Conditional Probabilities of Linkage Calculated by the HOMOG Program from Multipoint Linkage Data

Family	Ethnic origin	Linkage exclusion $P < 0.10$	Linkage inclusion $P > 0.5$
BB1	UK	15 (0.04)	11 (0.75)
BB2	UK	11 (0.03)	15 (0.52)
BB4	UK	11 (0.03)	16 (0.50)
BB5	UK		11 (0.51)
BB6	UK	15 (0.00)	11 (0.97)
		16 (0.01)	
BB7*	UK	11 (0.00)	15 (0.98)
		16 (0.01)	
BB8	Brazil	15 (0.06)	
		16 (0.09)	
BB9	UK		
BB10	Switzerland		
BB11	Switzerland		
BB12	Norway		11 (0.54)
BB13[a]	Norway	16 (0.10)	11 (0.72)
BB14	Norway	16 (0.01)	11 (0.54)
BB15	UK	16 (0.03)	11 (0.51)
BB17	UK	15 (0.01)	11 (0.87)
		16 (0.04)	
BB18[a]	Norway	11 (0.04)	15 (0.95)
		16 (0.00)	
BB19	UK	16 (0.03)	15 (0.64)
BB20[a]	Egypt	15 (0.10)	11 (0.78)
		16 (0.02)	
BB21	UK	15 (0.02)	11 (0.85)
		16 (0.05)	
BB22	Sweden	11 (0.03)	16 (0.55)
BB23	Denmark	16 (0.07)	11 (0.62)
BB24	Denmark	15 (0.02)	11 (0.61)
BB25/26[a]	Denmark	11 (0.01)	
		16 (0.04)	
BB27	Denmark	16 (0.03)	
BB28	Norway	11 (0.08)	
BB29	Italy	11 (0.01)	15 (0.97)
		16 (0.01)	
BB30	Sweden	11 (0.05)	15 (0.83)
		16 (0.07)	
BB31[a]	Norway	11 (0.06)	15 (0.94)
		16 (0.00)	
BB32	Somalia	11 (0.08)	15 (0.57)

Note. The probability of linkage to the chromosome indicated is given in parentheses.

[a] A consanguineous family.

data (lod score 6.26, $\chi^2 = 28.84$, 2 *df*; $P < 0.0001$) provides convincing evidence of a locus in this region. A very large series of BBS families (~100) could provide narrower confidence limits, but the identification of further consanguineous pedigrees linked to this region provides the most promising approach. A report by Cornier *et al.* (1995) describes two large inbred Puerto Rican families that show homozygosity for chromosome 11 markers and share a common haplotype between the PYGM and the GATA8A08 loci, which are, respectively, approximately 2 and 12 cM proximal to D11S913.

Possible candidate loci in the 11q13 region include ROM1, estimated to be about 6 cM proximal to our location for BBS1 (Bascom *et al.,* 1995); the human homologue of the *Drosophila* retinal degeneration locus, *rdgB* (Banfi *et al.,* 1996); and the phospholipase C $\beta 3$ gene (Lagercrantz *et al.,* 1995). Twelve BBS families show results consistent with linkage to the BBS1 locus (Table 2) on which further studies of candidate genes could be focused.

A locus for BBS on chromosome 15 was also identified in an observed 32–35% of the 29 BBS families, using markers shown by Carmi *et al.* (1995a) to be linked to BBS4 (Table 1). The two markers showing tightest linkage to BBS4 in the consanguineous pedigree of Carmi *et al.* (1995a) were D15S131 and D15S204, neither of which showed any recombination. In the present study, multipoint heterogeneity analysis supports a locus for BBS4 at 2 cM proximal to these markers (Fig. 1). The statistical support for a locus on chromosome 15 in addition to one or more unlinked loci (lod score 6.10, $\chi^2 = 28.07$, 2 *df*; $P < 0.0001$) confirms the reported linkage to this chromosomal region. Eight BBS families show results consistent with linkage to the BBS4 locus (Table 2), 3 of which (BB7, BB18, and BB31) are homozygous for three adjacent chromosome 15 markers, indicating identity by descent spanning an estimated 7.5-cM region (Carmi *et al.,* 1995a; Dib *et al.,* 1996). Haplotype analysis in 1 of the 3 consanguineous families (BB31, Fig. 2) shows two informative recombinants placing BBS1 between the D15S131 and the D15S114 loci, a distance of about 2 cM, substantially narrowing the location of the gene from the 9-cM interval (D15S125–D15S114) reported by Carmi *et al.* (1995a).

An observed 24–27% of the families have a BBS locus in chromosomal region 16q21 (BBS2) (Table 1). The maximum multipoint lod score is found at a genetic location 2 cM distal to D16S408, between this marker and D16S400 (Fig. 1). This is similar to the findings of Kwitek-Black *et al.* (1993), who found no evidence of recombination within an 18-cM region around D16S408 between the proximal marker D16S419 and the distal marker D16S265. The statistical support for this locus ($\chi^2 = 5.02$, 2 *df*; $P = 0.041$) is, however, much less convincing than that for the other two loci. This finding must therefore be regarded as tentative confirmation of the BBS2 locus at a borderline significance level. This is borne out by the fact that only four families show better evidence for chromosome 16 linkage than for other chromosomal sites (Table 2).

Are there clinical features that serve to distinguish the different genetic forms of BBS? The first problem is to classify such comparatively small families satisfactorily. Haplotype analysis and conditional probabilities of linkage derived from the heterogeneity analysis yield similar results with regard to classifying individual families (Table 2). A lack of recombination in small families, many with only two affected individuals—potentially four meioses, two of which determine the parental linkage phase—provides a rather unsatisfac-

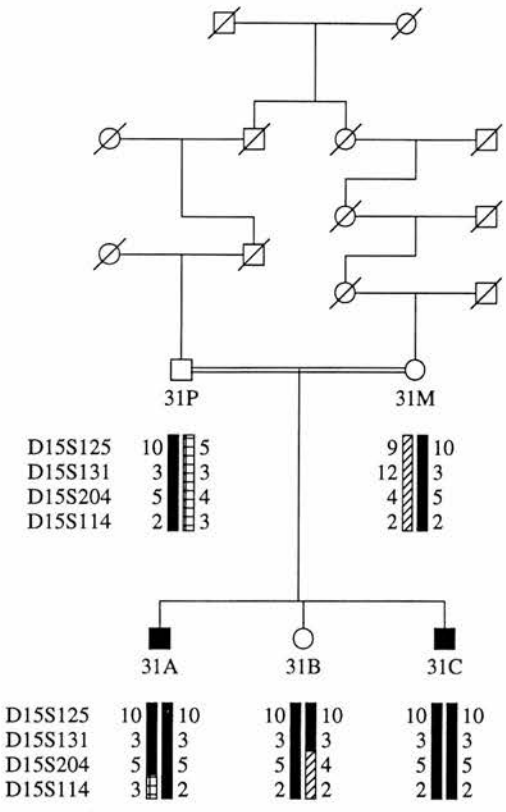


FIG. 2. Marker haplotypes in consanguineous family BBS31 in which the affected sons are homozygous for markers D15S125, D15S131, and D15S204, indicating possible identity by descent. The recombination events in 31A and 31B support a location for the BBS4 locus between D15S131 and D15S114, a distance of about 2 cM.

tory means of classification. However, the presence of recombination, particularly if it involves flanking markers, can effectively exclude a locus. On this basis, the chromosome 11 locus (BBS1) is excluded in 11 families, the chromosome 15 locus (BBS4) is excluded in 6 families, and the chromosome 16 locus is excluded in 16 families. In one family from Brazil (BB8), all three loci are excluded, consistent with the view that at least one other BBS locus remains to be found.

The clinical features in this family do not differ from those found in families linked to BBS1, BBS2, or BBS4. Similarly, when two of three loci are excluded and there is evidence for linkage to a third locus, four families appear to be linked to BBS1 (BB6, BB17, BB20, and BB21) and six families to BBS4 (BB7, BB18, BB25/26, BB29, BB30, and BB31). Again, no clear clinical distinctions are apparent between the BBS1- and BBS4-linked families. On the basis of careful examination of the available data, there are no reliable clinical distinctions among the BBS1, BBS2, and BBS4 loci or among the unlinked families. Carmi *et al.* (1995b) suggested that mutations at the BBS4 locus result in postaxial polydactyly predominantly of the upper limbs, whereas BBS3 (chromosome 3) mutations are associated with postaxial polydactyly of hands and feet. Similarly, early-onset morbid obesity was more evident in a large BBS4-linked family than in BBS3 or BBS2 families. These observations could not be confirmed in this set of families, suggesting that the findings of Carmi *et al.* (1995b) could relate more to the specific mutations or to the inbred genetic background than to the BBS locus per se.

The genetic diversity of Bardet–Biedl syndrome begs the question of how to identify specifically the responsible loci in the presence of such a high degree of genetic diversity. One possibility is to identify a patient with a cytogenetic abnormality involving a BBS locus. A systematic cytogenetic survey by Fannemel *et al.* (1993) failed to identify chromosomal rearrangements in 22 unrelated BBS patients. The report of a BBS patient with a balanced translocation t(2p−;17p+) (Dallapiccola, 1971) suggests a possible locus in one of these regions. However, we typed 14 markers spanning the length of chromosome 2 and 6 markers spanning chromosome 17 without finding evidence of linkage to BBS. The most promising approach, in the absence of such a rearrangement, is to narrow the region of search for each BBS locus in extended consanguineous pedigrees. These pedigrees provide large numbers of meioses that can be tracked indirectly from the common ancestor by homozygosity mapping, so that the identification of a small region of homozygosity in the affected child of distantly related parents can potentially narrow the interval to a few megabases of DNA, close enough for an attempt at positional cloning. Even with the small consanguineous families analyzed in this study, loss of homozygosity for 2 of the markers provides reasonable grounds for narrowing the search interval to about 2 cM, sufficiently small for a positional cloning effort. This provides further validation of the approach taken in this study, namely, a systematic linkage analysis of unselected BBS families with at least two affected members. This approach not only is essential for determining the relative frequencies of the different genetic forms but also can provide convincing evidence for linkage to a specific locus, particularly within consanguineous families, hence improving the ability to refine the gene localization. Further efforts to collect consanguineous BBS families with only two or three affected members, which are considerably more frequent than extended consanguineous kindreds from isolate populations, would be an effective means of fine genetic mapping these loci. Haplotyping of apparently unrelated affected individuals from a small geographic isolate is a third means of refining the genetic mapping of BBS loci, by making use of the large numbers of meioses since the presumed common ancestor. Finally, the mutational analysis of candidate genes in the vicinity of each locus is another possible route to the identification of BBS genes, but the high density of identified genes such as expressed sequence-tagged cDNAs shows that this approach also needs to be guided by good genetic mapping.

ACKNOWLEDGMENTS

This study was supported by grants from the British Retinitis Pigmentosa Society, the Gift of Thomas Pocklington, the Swiss Reti-

nitis Pigmentosa Association, and the International Retinitis Pigmentosa Society, which are gratefully acknowledged. E.A.B. was in receipt of a SERC Case Studentship. We also thank Norman Davidson and staff for the artwork and Dr. Jeff Haywood, Professor Veronica van Heyningen, and Professor Nick Hastie for their support. We are grateful to Professor M. Warburg and Professor S. Bernasconi for referring families BB32 and BB29, respectively.

REFERENCES

Aldred, M. A., Dry, K. L., Knight-Jones, E. B., Hardwick, L. J., Teague, P. W., Lester, D. H., Brown, J., Spowart, G., Carothers, A. D., Raeburn, J. A., Bird, A. C., Fielder, A. R., and Wright, A. F. (1994). Genetic analysis of a kindred with X-linked mental handicap and retinitis pigmentosa. *Am. J. Hum. Genet.* **55:** 916–922.

Banfi, S., Borsani, G., Rossi, E., Bernard, L., Guffanti, A., Rubboli, F., Marchitiello, A., Giglio, S., Coluccia, E., Zollo, M., Zuffardi, O., and Ballabio, A. (1996). Identification and mapping of human cDNAs homologous to Drosophila mutant genes through EST database searching. *Nature Genet.* **13:** 167–174.

Bascom, R. A., Liu, L., Heckenlively, J. R., Stone, E. M., and McInnes, R. R. (1995). Mutation analysis of the ROM1 gene in retinitis pigmentosa. *Hum. Mol. Genet.* **4:** 1895–1902.

Carmi, R., Elbedour, K., Stone, E. M., and Sheffield, V. C. (1995b). Phenotypic differences among patients with Bardet–Biedl syndrome linked to three different chromosomal loci. *Am. J. Med. Genet.* **59:** 199–203.

Carmi, R., Rokhlina, T., Kwitek-Black, A. E., Elbedour, K., Nishimura, D., Stone, E. M., and Sheffield, V. C. (1995a). Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. *Hum. Mol. Genet.* **4:** 9–13.

Cornier, A. S., Fulton, A. B., Rokhlina, T., Nishimura, D., Stone, E. M., Sheffield, V. C., Whiteman, D. A. H., and Cox, G. F. (1995). Homozygosity mapping of a Bardet–Biedl syndrome gene in inbred families of Puerto Rican ancestry confirms the existence of a chromosome 11 locus. *Am. J. Hum. Genet.* **57:** A189.

Dallapiccola, B. (1971). Familial translocation t(2p−;17p+). *Ann. Genet.* **14:** 153–155.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, D., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morisette, J., and Weissenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380:** 152–154.

Fannemel, M., Riise, R., Lofterod, B., and Tommerup, N. (1993). High-resolution chromosome analysis in autosomal recessive disorders: Laurence–Moon–Bardet–Biedl syndrome. *Clin. Genet.* **43:** 111–112.

Farag, T. I., and Teebi, A. S. (1989). High incidence of Bardet-Biedl syndrome among the Bedouin. *Clin. Genet.* **36:** 463–465.

Green, J. S., Parfrey, P. S., Harnett, J. D., Farid, N. R., Cramer, B. C., Johnson, G., Heath, O., McManamon, P. J., O'Leary, E., and Pryse-Phillips, W. (1989). The cardinal manifestations of Bardet–Biedl syndrome, a form of Laurence–Moon–Bardet–Biedl syndrome. *N. Engl. J. Med.* **321:** 1002–1009.

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., and Weissenbach, J. (1994) The 1993–94 Généthon human genetic linkage map. *Nature Genet.* **7** (Special issue): 246–339.

Harnett, J. D., Green, J. S., Cramer, B. C., Johnson, G., Chafe, L., McManamon, P., Farid, N., Pryse-Phillips, W., and Parfrey, P. S. (1988). The spectrum of renal disease in Laurence-Moon-Biedl syndrome. *N. Engl. J. Med.* **319:** 615–618.

James, M. R., Richard, C. W., III, Schott, J.-J., Yousry, C., Clark, K., Bell, J., Terwilliger, J. D., Hazan, J., Dubay, C., Vignal, A., Agrapart, M., Imai, T., Nakamura, Y., Polymeropoulos, M., Weissenbach, J., Cox, D. R., and Lathrop, G. M. (1994). A radiation hybrid map of 506 STS markers spanning human chromosome 11. *Nature Genet.* **8:** 70–76.

Klein, D. and Ammann, F. (1969). The syndrome of Laurence-Moon-Bardet-Biedl and allied diseases in Switzerland. *J. Neurol. Sci.* **9:** 479–513.

Kwitek-Black, A. E., Carmi, R., Duyk, G. M., Buetow, K. H., Elbedour, K., Parvari, R., Yandava, C. N., Stone, E. M., and Sheffield, V. C. (1993). Linkage of Bardet–Biedl syndrome to chromosome 16q and evidence for non-allelic genetic heterogeneity. *Nature Genet.* **5:** 392–396.

Lagercrantz, J., Carson, E., Phelan, C., Grimmond, S., Rosen, A., Dare, E., Nordenskjold, M., Hayward, N. K., Larsson, C., and Weber, G. (1995). Genomic organization and complete cDNA sequence of the human phosphoinositide-specific phospholipase C $\beta 3$ gene (PLCB3). *Genomics* **26:** 467–472.

Lathrop, G. M., Lalouel, J. M., Julier, C. and Ott, J. (1984). Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* **81:** 3443–3446.

Leppert, M., Baird, L., Anderson, K. L., Otterud, B., Lupski, J. R., and Lewis, R. A. (1994). Bardet–Biedl syndrome is linked to DNA markers on chromosome 11q and is genetically heterogeneous. *Nature Genet.* **7:** 108–112.

Mansfield, D. C., Brown, A. F., Green, D. K., Carothers, A. D., Morris, S. W., Evans, H. J. and Wright, A. F. (1994). Automation of genetic linkage analysis using fluorescent microsatellite markers. *Genomics* **24:** 225–233.

Quackenbush, J., Davies, C., Bailis, J. M., Khristich, J. V., Diggle, K., Marchuck, Y., Tobin, J., Clark, S. P., Rodkins, A., Marcano, S., Churukian, A. C., Hutchinson, J. S., Probst, S., Romberg, L., Wei, Y. H., Nowack, N. J., Garner, H. R., Smith, M. W., Selleri, L., and Evans, G. A. (1995). An STS content map of human chromosome 11—Localization of 910 YAC clones in 109 islands. *Genomics* **29:** 512–525.

Riise, R., Andreasson, S., Borgstrom, M. K., Wright, A. F., Tommerup, N., Rosenberg, T., and Tornqvist, K. (1997). Intrafamilial variation of the phenotype in Bardet–Biedl syndrome. *Br. J. Ophthalmol.*, in press.

Schachat, A. P., and Maumenee, I. H. (1982). Bardet–Biedl syndrome and related disorders. *Arch. Ophthalmol.* **100:** 285–288.

Sheffield, V. C., Carmi, R., Kwitek-Black, A., Rokhlina, T., Nishimura, D., Duyk, G. M., Elbedour, K., Sunden, S. L., and Stone, E. M. (1994). Identification of a Bardet–Biedl syndrome locus on chromosome 3 and evaluation of an efficient approach to homozygosity mapping. *Hum. Mol. Genet.* **3:** 1331–1335.

Terwilliger, J. D., and Ott, J. (1994). "Handbook of Human Genetic Linkage," Johns Hopkins Press, Baltimore/London.