Table 4.9 Comparison of South African S.A.) and C.E.P.H. Caucasoid data.

(a) Phenotypes

Phenotype	λRH52/StuI				λMR3/StuI			
	S.A. Caucasoid		C.E.P.H. Caucasoid		S.A. Caucasoid		C.E.P.H. Caucasoid	
	obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.
11	7	5.6	10	7.50	15	15.6	55	55.44
12	7	10.0	26	31.75	3	4.0	15	13.71
22	6	4.4	37	33.75	2	0.4	0	0.85
Totals (n)	(20)	(20.0)	(73)	(73.00)	(20)	(20.0)	(70)	(70.00)
χ^2 (df = 1) test for internal consistency	1.01 (p = 0.315)		2.18 (p = 0.140)		0.98 (p = 0.322)		0.97 (p = 0.325)	
χ^2 (df = 2) between two populations	5.37 (p = 0.068)				7.44 (p = 0.025)			

(b) Allele frequencies

Alleles	λRH52/StuI		λMR3/StuI	
	S.A. Caucasoid	C.E.P.H. Caucasoid	S.A. Caucasoid	C.E.P.H. Caucasoid
1	0.53 ± 0.08	0.32 ± 0.05	0.82 ± 0.08	0.89 ± 0.03
2	0.47 ± 0.08	0.68 ± 0.05	0.18 ± 0.08	0.11 ± 0.03

Chi square contingency tests show that the frequency of the λMR3/*StuI* RFLP in the southern African Caucasoid population is significantly different at the 5% level from that in the C.E.P.H. population. The chi square value for the λRH52/*StuI* RFLP is significant only at the 7% level, which is interesting but should be corroborated with data from a larger sample.

One might conclude that the southern African Caucasoid is not as close genetically to the Western European as is usually thought, and may have diverged because of random genetic drift. Another explanation may be that the two markers were found in the Negroid population and that the markers define polymorphisms which are more common in southern Africa.

4.5 Summary

Six single-copy probes that detect DNA sequence varation in southern Africa populations were isolated, and four were characterized. λRH52, λMR7, and λMR22 were found to identify base substitution-type RFLPs, and λMR3 to identify a locus at which there was a 300 bp insertion/deletion, giving rise to a rearrangement-type RFLP. All RFLPs identified by these probes were shown to be inherited in an autosomal codominant fashion. The probes were tentatively assigned by

somatic cell hybridization, to chromosomes 9, 8, 6 and 17 respectively.

The frequencies of each of the two alleles found at six different polymorphic sites were calculated in individuals from three major population groups, San, Negroid and Caucasoid. There was a good fit to Hardy-Weinberg expectations in all but one of the systems. Statistics that summarise the patterns of allele frequency variation found were presented, and the locus showing the greatest variation was that identified by λMR3.

Genetic distances between populations were calculated using the allele frequencies, and it was shown that the Negroids are more closely related to the San than either population is to the Caucasoids.

The probes λRH52 and λMR3 were included in the C.E.P.H. collaborative programme. All 'informative' families have been typed for the two RFLPs, λRH52/*Stu*I and λMR3/*Stu*I, but the results have yet to be analysed. An interesting finding was that allelic frequencies in the southern African Caucasoids were significantly different from those in the C.E.P.H. Caucasoids.

RFLPs described in this Chapter will be useful genetic markers for gene mapping and anthropological studies.

CHAPTER 5

5. A SEARCH FOR LINKAGE TO TYROSINASE-POSITIVE OCULOCUTANEOUS ALBINISM (TPOCA)

The search for linkage to TPOCA was initiated during the present study. More than fifty Negroid families with at least one affected member have been identified; over three-hundred individuals have provided peripheral blood samples for the study; and a substantial database has been established. Results presented and discussed in this chapter are preliminary, but form the basis for an ongoing linkage study in the Department.

5.1 Results

Families suitable for the linkage study continue to be identified. In the present study, a subset of forty-seven families were used in the collation of linkage data. The families are listed and their structure is given in Table 5.1 (a): the total number of individuals per family; the number of affected individuals per family; and the number of individuals per generation per family are given. The structure of the families is summarised in Table 5.1 (b).

Peripheral blood samples were obtained from all available family members, separated into plasma, red

Table 5.1 (a) Structure of the families involved in the linkage study

Family name	Total N° of individuals	N° of affected individuals	N° of individuals per generation		
			1st	2nd	3rd
ZaBA	9	3	2	7	0
ZaBB	5	2	1	4	0
ZaBC	7	3	2	5	0
ZaBD	8	3	2	6	0
ZaBE	6	3	2	4	0
ZaBF	8	3	2	6	0
ZaBG	9	3	1	8	0
ZaBH	13	2	1	11	1
ZaBI	8	2	1	7	0
ZaBJ	8	2	1	4	3
ZaBK	5	3	1	4	0
ZaBL	10	2	2	8	0
ZaBM	5	2	1	4	0
ZaBN	7	2	1	6	0
ZaBO	4	1	2	2	0
ZaBP	4	1	1	3	0
ZaBQ	7	2	1	6	0
ZaBR	4	1	1	1	2
ZaBT	5	1	1	2	2
ZaBU	7	2	1	6	0
ZaBW	6	1	2	4	0
ZaBZ	6	1	2	4	0
ZaCA	7	1	1	6	0
ZaCB	4	2	1	3	0
ZaCC	5	1	1	4	0
ZaCD	5	1	1	2	2
ZaCE	5	1	2	3	0
ZaCF	4	2	2	2	0
ZaCG	4	1	2	2	0

Table 5.1 (a) continued

Family name	Total N° of individuals	N° of affected individuals	N° of individuals per generation		
			1st	2nd	3rd
ZaCH	4	2	1	3	0
ZaCJ	4	1	2	2	0
ZaCK	4	1	1	1	2
ZaCL	9	2	1	8	0
ZaCM	7	1	2	5	0
ZaCN	5	1	1	4	0
ZaCO	4	1	2	2	0
ZaCP	3	1	1	2	0
ZaCR	6	1	1	5	0
ZaCS	4	1	2	2	0
ZaCU	4	1	2	2	0
ZaCY	3	1	1	2	0
ZaCZ	5	1	2	3	0
ZaDA	4	1	1	2	0
ZaDC	5	1	2	3	0
ZaDE	7	1	2	5	0
ZaDF	5	1	2	3	0
ZaDG	16	4	1	7	8

Table 5.1 (b) Summary of the structure of the families in the linkage study

Total N° of families	Total N° of individuals	N° of families with x affected individuals			N° of 2-generation families	N° of 3-generation families
		x = 3	x = 2	x = 1		
47	284	8	11	28	40	7

blood cell, serum and buffy coat fractions, and stored at -20°C or -70°C. Serogenetic markers were investigated using the plasma, serum and red blood cells; DNA markers were investigated using high-molecular-weight DNA isolated from the buffy coats. The chromosomal localization of each marker locus used in the study is given in Table 5.2 (Human gene Mapping 8 and this study).

Linkage between each polymorphic genetic marker locus and the TPOCA locus was analysed using the programme LIPED (Ott, 1974). Results of two-point linkage analysis between fifteen serogenetic markers and TPOCA are presented in Table 5.3, for values of θ = 0.05, 0.1, 0.2, 0.3 and 0.4. The number of families for which each marker was informative is also given in Table 5.3. Serogenetic markers were those being routinely used in the laboratory, and were known to be polymorphic in the South African Negro population.

DNA markers used in the study were, initially, three RFLPs isolated and described in Chapter 4: λRH52/StuI, λMR3/StuI and λMR22/BglII. These RFLPs had PIC values greater than 0.25, and hence were thought to be suitable for a linkage study.

Preliminary results from the serogenetic markers gave a positive lod score of 0.46 at θ = 0.2 for α_1-AT versus

Table 5.2 Chromosomal location of the polymorphic marker loci used in the linkage study.

Type of marker	Locus	Chromosome
Red Cell	Rh	1 p36.2 - p34
Antigen	MNSs	4 q28 - q31
	ABO	9 q34
Red Cell	PGM_1	1 p22.1
Enzyme	6PGD	1 p36.2 - p36.13
	ACP_1	2 p25 or p23
	GPX_1	3 p13 - q12
	GLO_1	6 p21.31 - p21.1
	GPT_1	8 q13 - qter or 16p
	PEP A	18 q23
Serum	Tf	3 q21 - q26.1
Protein	Gc	4 q12 - q13
	Bf	6 p21.3
	α_1 - AT	14 q32.1
	Hp	16 q22.1
DNA	ß - globin	11 p15.5
	pAW101	14 q32.1 - q32.2
	λRH52	1
	λMR3	17
	λMR22	18

Table 5.3 Results of two-point linkage analysis between 15 serogenetic markers and TPOCA

Serogenetic marker	θ 0.05	0.1	0.2	0.3	0.4	N° of informative families
Rh	-12.117	-7.217	-2.917	-1.051	-0.229	44
MNS	-6.436	-2.790	-0.426	0.075	0.061	37
ABD	-5.355	-2.095	-0.057	-0.356	-0.140	34
PGM_1	-5.942	-3.311	-1.119	-0.318	-0.054	28
6PGD	-1.437	-0.710	-0.171	-0.014	0.007	6
ACP_1	-1.458	-0.463	0.173	0.224	0.088	22
GPX_1	-1.195	-0.633	-0.178	-0.030	0.000	7
GLO_1	-5.463	-2.704	-0.710	-0.130	-0.005	27
GPT_1	-3.032	-1.746	-0.614	-0.182	-0.032	13
PEP A	-6.943	-4.075	-1.655	-0.611	-0.138	15
Tf	-0.887	-0.221	0.163	0.156	0.053	18
Gc	-8.338	-5.027	-2.051	-0.735	-0.160	26
Bf	0.607	1.575	1.510	0.819	0-223	35
α_1-AT	-2.174	-1.192	-0.378	-0.095	-0.013	11
Hp	-1.296	0.191	0.870	0.619	0.197	38

TPOCA. α_1-AT is, however, not as polymorphic in Negro populations, and some α_1-AT subtypes described in other populations are not present in the South African Negroid (Dunn *et al.*, 1986). A more polymorphic DNA marker was therefore thought to be required for more efficient linkage analysis. The α_1-AT locus has been assigned to 14q32.1 (Purrello *et al.*, 1985), and a cloned cDNA for α_1-AT has been isolated and used as a probe to search for RFLPs near the α_1-AT gene (Matterson *et al.*, 1985), but only two low frequency RFLPs were found with the cDNA probe. Three RFLPs with PIC values of 0.35 - 0.38 have been detected with a genomic probe extending into the 5' flanking region of the α_1-AT gene (Cox *et al.*, 1985), and this probe is potentially useful in a linkage study. Unfortunately, the scientist who isolated the probe did not respond to our request for it for use in this study. The probe pAW101 (Wyman and White, 1980) was available, however. This probe identifies the highly polymorphic locus D14S1 and is located at 14q32.1 - q32.2 (Balazs *et al.*, 1982; Donlon *et al.*, 1983). It was therefore used in the linkage study. The α_1-AT protein polymorphisms were further investigated at the protein level, and the lod scores obtained from these additional data are those given in Table 5.3.

pAW101 identifies eight alleles ranging in size from 14 kb to 28 kb; such large fragments are difficult to

resolve, and it was found necessary to electrophorese 0.5% agarose gels for 75 - 80 hours at 1 V/cm to do so. Figure 5.1 shows an autoradiograph of DNA from families ZaBA, ZaBB and ZaBC digested with *Eco*RI and probed with pAW101. The autoradiograph shows that fragments are resolved using the above conditions, but that the concentration of DNA loaded per lane is critical for interpretable results. Bowcock (1984) investigated pAW101 in Bantu-speaking Negroids and found that few polymorphic fragments were typically found in this population; a finding which was confirmed in the present study. In Figure 5.1, for example, only five of at least eight possible alleles were seen in the three families shown.

Gamma-globin RFLPs from the β-globin gene cluster were also used in the linkage study, since there is evidence for a conserved linkage group in lower mammals which includes albino loci and the haemoglobin-β chain locus. An example of families whose DNA was digested with *Hin*dIII and probed with the $^{A}\gamma$ probe is given in Figure 5.2. Haplotypes were constructed with the $^{A}\gamma$- and $^{G}\gamma$-RFLPs in order to maximise linkage information from the β-globin gene cluster.

Results of two-point linkage analysis between the five DNA markers described above and the TPOCA locus are

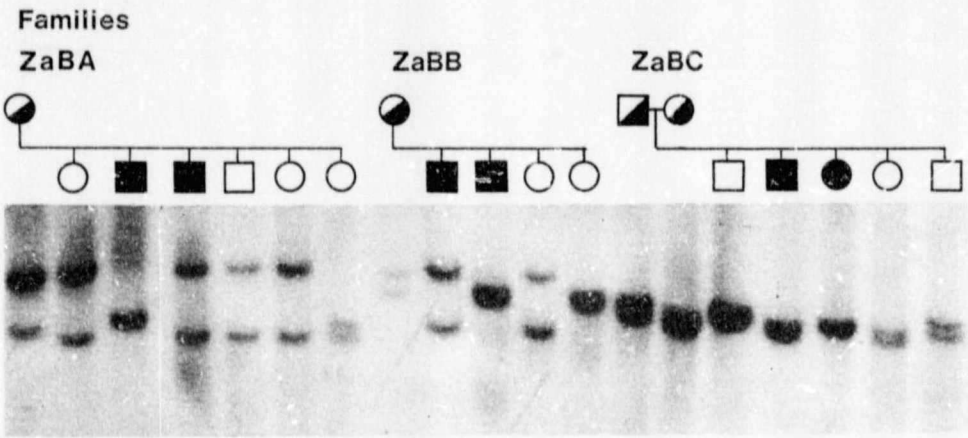

Figure 5.1 Autoradiograph of DNA extracted from families in the linkage study, digested with EcoRI and probed with pAW101. The 0.5% agarose gel was electrophoresed for 68 hours at 2V/cm.

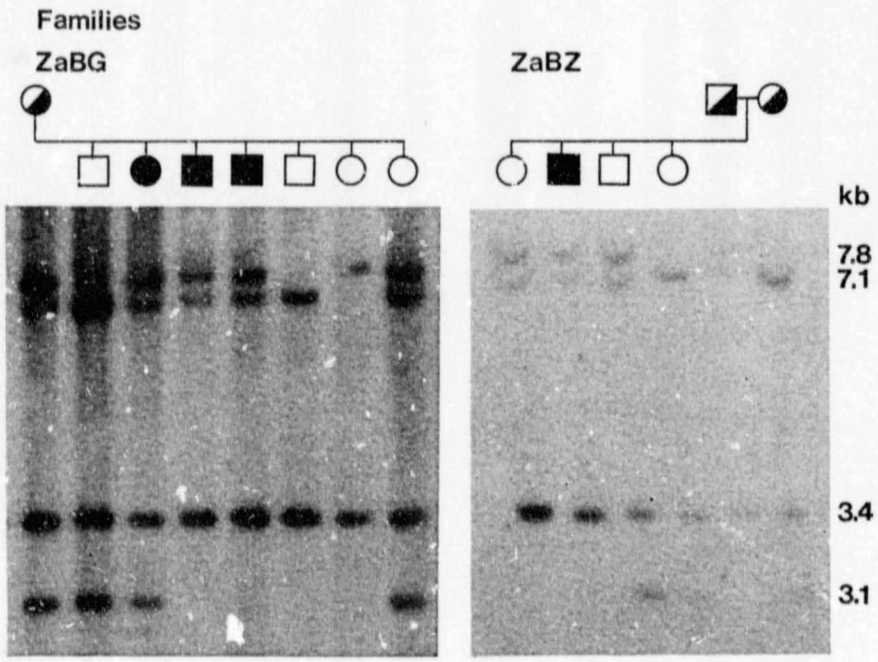

Figure 5.2 Autoradiographs of DNA extracted from families in the linkage study, digested with HindIII and hybridized to the $^{A}\gamma$ probe. The 7.1 and 7.8 kb bands constitute the $^{G}\gamma$ RFLP, and the 3.4 and 3.1 kb bands constitute the $^{A}\gamma$ RFLP.

given in Table 5.4, for values of θ = 0.05, 0.1, 0.2, 0.3 and 0.4.

Table 5.4 Results of two-point linkage analysis between 5 DNA markers and TPOCA

DNA marker	θ 0.05	0.1	0.2	0.3	0.4	N° of informative families
ß-globin cluster	-9.847	-5.514	-2.043	-0.687	-0.141	20
pAW101	-1.299	0.023	0.591	0.405	0.120	9
λRH52	-3.632	-1.900	0.489	-0.042	0.023	21
λMR3	-1.139	-0.344	0.738	0.093	0.032	12
λMR22	-3.321	-1.283	0.001	0.205	0.093	19

5.2 Discussion

5.2.1 Serogenetic markers

The results of two-point linkage analysis between fifteen serogenetic markers and TPOCA (Table 5.3) show no evidence of close linkage between the TPOCA locus and the Rh, MNS, ABO, PGM_1, GLO_1, PEP A and Gc loci, at θ = 0.1. At θ = 0.2, linkage can be excluded between the TPOCA locus and the Rh and Gc loci. Lod scores for the remaining loci (excluding Bf) tend towards negative values, suggesting that there is no linkage to TPOCA at

these loci. Additional informative families will probably confirm this trend.

The lod score for α_1-AT versus TPOCA, which was slightly positive when the data from a few informative families were analysed, became negative with the addition of data from more families. This suggests that TPOCA is not linked to the α_1-AT locus on chromosome 14. Lod scores for the pAW101 locus, which has been assigned to the same chromosomal region as the α_1-AT locus, are slightly positive at higher θ values (Table 5.4). More data need to be collated in order to determine whether the TPOCA locus is in or near the 14q32 region, but these preliminary results suggest that this is not likely.

The lod scores for TPOCA versus Bf are 'interesting' and suggest that the TPOCA locus may be on chromosome 6p. Bf has been assigned to 6p21.3 and is within the class III HLA linkage group (Olaisen *et al.*, 1983; Carroll *et al.*, 1984). Using data from 35 informative families and the programme LINKAGE (Lathrop *et al.*, 1984), the maximum lod score calculated for Bf versus TPOCA is 1.916, at recombination fractions of 0.067 in females and 0.174 in males. This result is intriguing, because reported differences in recombination rates between male and female meioses have led to the supposition that there is a generally higher rate of

recombination in females. Studies of chromosomes 6, 13 and 19 (Weitkamp *et al.*, 1971; Leppert *et al.*, 1986; Sherman *et al.*, 1985; Naylor *et al.*, 1985) have shown this in particular, and the recent linkage map of the human genome proposed by Donis-Keller *et al.* (1987) indicates that the genetic map of the autosomes is roughly 90% longer in females than in males. However, a higher rate of recombination in females may not be constant or even consistent over the entire human genome: a significant excess of recombination in males over females has been found in a cluster of four loci on the short arm of chromosome 11 (White *et al.*, 1985). In the present study, there seems to be a two-fold higher rate of recombination in males than in females at the proposed Bf - TPOCA interval. In contrast, Leach *et al.* (1986) studied the same region of chromosome 6 using RFLPs at the HLA locus and the protein marker GLO_1, and found that sex differences in recombination frequency were not significant.

There is no suggestion of close linkage between the GLO_1 locus, also localised to 6p21, and the TPOCA locus (see Table 5.3: lod score of -2.704 at $\theta = 0.1$). Lod scores for Bf versus GLO_1 were calculated using data obtained in the present study, however, and the scores given below in Table 5.5 are suggestive of linkage. The programme MLINK from the LINKAGE package (Lathrop *et al.*, 1984) was used to calculate which of the three

Table 5.5 Results of two-point linkage analysis between Bf and GLO_1

θ					N° of informative families
0.05	0.1	0.2	0.3	0.4	
0.488	1.353	1.994	1.920	1.109	11

possible orders for the three loci Bf, GLO_1 and TPOCA was most likely: the most probable order turned out to be GLO_1 - Bf - TPOCA. It is thus feasible that the TPOCA and Bf loci could be linked whereas there may be no detectable linkage between the TPOCA and GLO_1 loci. Figure 5.3 shows a genetic map of chromosome 6 based on male recombination fractions (after Human Gene Mapping 8, p 144). A putative map position for TPOCA is distal to Bf.

The HLA loci are highly informative markers, and an investigation of the HLA polymorphisms in the TPOCA families has been initiated. This should confirm whether this search for linkage to TPOCA has in fact been successful or if a new set of markers needs to be investigated.

There is a precedent for the detection of linkage with a marker from a small subset of markers in an initial screen for linkage: Gusella *et al.* (1983) found linkage

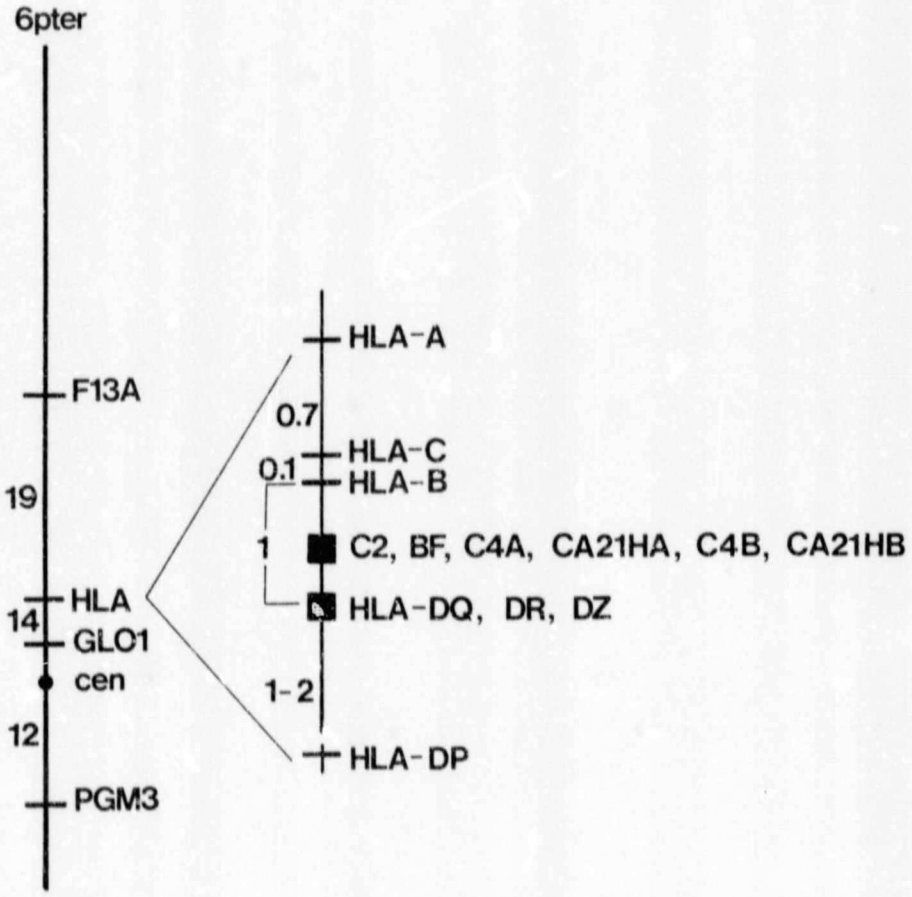

Figure 5.3 Genetic map of chromosome 6 (after Human Gene Mapping 8, pg 144).

to the Huntington's disease gene with one of only twelve DNA markers tested. Elston and Lange (1975) have computed the *a priori* probability of detecting linkage (at $\theta \leq 0.40$) for 20 random markers as 0.35; for 30 markers this probability rises to 0.50. A systematic approach has now been proposed by Donis-Keller *et al.* (1987), suggesting that a panel of 100 - 200 selected probes will be sufficient to locate a disease locus, provided that it lies within the 95% of the genome this group claim is now linked.

5.2.2 DNA markers

The three anonymous probes isolated in the present study, λRH52, λMR3 and λMR22, were tested on DNA extracted from families ZaBA - ZaCC and ZaDG. Linkage to the TPOCA locus was excluded for λRH52 and for λMR22 at distances of up to 7-8 cm, but definitive exclusion of linkage for λRH52 was not possible (Table 5.4). The use of these three probes has demonstrated that two-allele polymorphisms are not very useful in linkage analyses, particularly when the families in the study are very small, so that there are relatively few informative meioses. In addition, because TPOCA is a recessive disorder in which heterozygotes cannot be distinguished, the linkage information obtained from each family is less than it would have been if heterozygotes were identifiable. A large number of families must be tested using these DNA markers in order to

obtain sufficient data and achieve significant lod scores. It is questionable whether the expense involved is justifiable; other markers should perhaps be tested first if they are more polymorphic or if there is some evidence that they could be linked to TPOCA.

Lod scores obtained using the probe pAW101 have also not reached significance, in part because of the quality of the families, and in part because the probe does not appear to be as polymorphic in the Negroid population as it is in the Caucasoid population in which the polymorphism was originally described. It seems possible that the results are an indication of genetic heterogeneity and this is discussed below in section 5.2.3. In contrast, lod scores obtained for the ß-globin locus versus TPOCA have reached significance, because haplotyping markedly increased the informativeness of the ß-globin RFLPs at the ß-globin locus.

The lod scores for ß-globin versus TPOCA provide strong evidence against measurable linkage between these two loci. Linkage of the *c* (albino) locus to the *p* (pink-eyed dilution) locus and the structural loci for the haemoglobin beta chain (*Hbb*) has been reported in the rat and the deermouse (Cramer, 1984; Stolc and Gill, 1983). Similarly, the *c* locus has been assigned to the

mouse chromosome 7, tightly linked to *Hbb*, *Ldh-1* (lactate dehydrogenase-1) and *p* (Roderick and Davisson, 1984). More recently, O'Brien *et al.* (1986) reported the assignment of the *Hbb-c* linkage group to the D1 chromosome of the domestic (Siamese) cat, and Sandberg and Andersson (1987) have found the same linkage group in the rabbit. Other loci mapped to feline D1 are *Ldh-1* and *ACP2* (acid-phosphatase-2) (O'Brien, 1984). This provides evidence for a conserved mammalian linkage group, but it is not clear whether the OCA loci in man are syntenic with this group. With the exception of the *p* and *c* albino loci, which are not mapped in man, each human locus homologous to loci in the cat, rat, mouse and deermouse are also syntenic in man on chromosome 11 (Human Gene Mapping 8, 1985).

The *c* locus in lower mammals has been considered equivalent to the human TYROCA locus, while features of the *p* locus mutation show a marked resemblance to TPOCA in man (Witkop, 1985). Phenotypic similarities between TPOCA and *p* include no impairment of the tyrosinase system, as well as an apparent limitation in the amount of tyrosinase available for melanin synthesis (Sidman and Pearlstein, 1965; Witkop *et al.*, 1983).

If it is reasonable to extrapolate from mouse to man, and if the TPOCA and *p* loci can be equated, then these results demonstrate that the TPOCA locus is not

included in the human linkage group. Conversely, the results can be interpreted as evidence against extrapolation from lower mammals to man. Other human OCA loci may yet prove to be linked to the β-globin gene. In order to exclude the possibility that DNA rearrangement on chromosome 11 has disrupted the linkage group, in a search for linkage to TPOCA it would be worthwhile investigating linkage between the TPOCA locus and other chromosome 11 markers.

5.2.3 Heterogeneity

A useful application of linkage analysis, besides gene mapping, is the possible demonstration of genetic heterogeneity. When some families show linkage of the 'same' disease to a genetic marker while others do not, it is probable that the disease phenotype, despite appearing homogeneous, is in fact due to different underlying genetic defects.

Analysis of lod scores obtained for individual families in the present study has proved to be interesting. Four of the families tested with pAW101 give a combined negative lod score of -1.224 at $\theta = 0.2$, whereas the five remaining families tested at this locus give a combined lod score of +1.815. The interval is +0.591. Although this value is not significant in terms of linkage, the possibility remains that there is

genetic heterogeneity among the families in the study and that, given sufficient data, the disease locus in subsets of the families may prove to show linkage to different marker loci.

The writer considers that genetic heterogeneity among the South African Negroid individuals thus far diagnosed as having the TPOCA phenotype is likely. The differential diagnosis of albinism by both the hairbulb incubation test and by clinical examination has not been entirely successful in distinguishing TNOCA from TPOCA individuals. It is not impossible that both albinism types are present in this study. Furthermore, the biochemical defect responsible for the TPOCA phenotype is not known. The wide tyrosinase activity range found in these individuals may be evidence for genetic heterogeneity within the TPOCA classification itself. Clinical and biochemical analyses have not resolved these problems: linkage analysis promises to do so, but if families in the linkage study are genetically heterogeneous then a large number will be required to prove linkage to one or more loci.

5.3 Summary

Twenty polymorphic markers have been investigated for linkage to TPOCA. This represents approximately five per cent of the genome, if it is accepted that linkage

is detectable up to a distance of 0.4 cM, and that the genome is approximately 33 M in size (Renwick, 1971).

Close linkage to the TPOCA locus has been excluded for the Rh, Gc, and β-globin loci. The latter finding indicates that the β-globin and albino syntenic group found in lower mammals may not have been evolutionarily conserved in man, as indirect evidence had previously suggested. There is no suggestion of linkage to MNS, ABO, PGM_1, 6PGD, ACP_1, GPX_1, GLO_1, GPT_1, PEP A, Tf, Gc, α_1-AT, Hp, λRH52, λMR3, or λMR22. A slightly positive lod score obtained for pAW101 versus TPOCA may reflect genetic heterogeneity, or may not be significant. Additional data will improve lod scores for each marker locus versus TPOCA until significance is reached.

The lod score for Bf versus TPOCA is 'interesting' and suggests the possible assignment of TPOCA to chromosome 6. This possibility is being investigated; HLA typing of the TPOCA families has been initiated.

Preliminary results presented in this chapter form the basis for an ongoing linkage study. The results have given indications for future work: markers at the HLA locus, markers on chromosome 11, and markers at the 14q32 locus should be investigated, followed by highly polymorphic markers chosen at random from all chromo-

somes so that maximum information can be obtained from the relatively small families available in the study.

CHAPTER 6

6. GENERAL DISCUSSION

The discussion of specific topics has already been dealt with in appropriate parts of the thesis, in particular in the Results sections. In this chapter genetic variation at the molecular level will be discussed, as will be the usefulness of DNA sequence polymorphisms in anthropological studies and in gene mapping.

6.1 Molecular variation

DNA sequence polymorphisms, first identified in man in the late 1970's, have demonstrated genetic diversity to an extent not previously apparent. Variation seen at the protein level has often not reflected variation at the corresponding structural gene locus. Murray *et al.* (1984), for example, found numerous RFLPs around the albumin gene, although the albumin protein shows very little variation.

The search for RFLPs has made rapid progress since the proposal for their use in the construction of genetic linkage maps was put forward by Botstein *et al.* (1980). Sufficient numbers have now been found for partial linkage maps of chromosomes (O'Connell *et al.*, 1985;

Kittur *et al.*, 1985; Holm *et al.*, 1985; Drayna and White, 1985; Leach *et al.*, 1986; Leppert *et al.*, 1986) to have been reported. One group has claimed to have succeeded in producing a map of the human genome (Donis-Keller *et al.*, 1987) although such a claim has been considered to be premature by others (Barinaga, 1987). The map presented by Donis-Keller *et al.* (1987) used 403 polymorphic loci to cover an estimated 95 per cent of the human genome.

Strategies for the detection of RFLPs have improved, techniques have been developed and with better understanding of the genome. It is now clear that hypervariable regions are highly polymorphic and constitute the most useful genetic markers. Recombinant phage containing such regions which were obtained from genomic libraries constructed in *RecA*$^-$ hosts, are now known to be unstable (Leach and Stahl, 1983; Wyman *et al.*, 1984; Wyman *et al.*, 1985; Donlon *et al.*, 1986), enabling better strategies to be developed when searching for probes that detect multiple allelic RFLPs. The use of the 'core' sequence from the myoglobin hypervariable region and similar candidate genes or synthetic sequences has also enabled the identification of unique VNTR (Variable Number of Tandem Repeat) loci (Wong *et al.*, 1986b; Nakamura *et al.*, 1987a, 1987b).

The majority of RFLPs thus far detected are, however, bi-allelic, and these constitute approximately one third of the markers used in the construction of the putative first human linkage map (Donis-Keller *et al.*, 1987). One benefit of using simple polymorphic markers is ease of interpretation. Two of the markers described in the present study have been included in the C.E.P.H. programme and should be useful, either in the complete linkage map or, if they are in close proximity to a disease locus, in prenatal or preclinical diagnosis of the disease.

RFLPs were first detected using single-copy probes, but the screening procedure for single-copy phage may have been unnecessarily stringent. Repetitive probes can, in fact, identify unique loci and hence serve as genetic markers. The function of single-copy regions is still obscure, but it seems increasingly unlikely that they exist due to chance, as 'junk' DNA, but may have a structural role that in itself is functional.

The origin and maintenance of repetitive sequences is not yet clear. It is possible that only certain regions are 'fluid' (Young, 1979) and become hypervariable. Their maintenance in different populations would suggest that they are of structural or functional significance. It may be that an ancestral sequence is the progenitor of many

hypervariable regions, with the length variation being attributed to unequal crossing-over. Hypervariable regions are commonly near structural genes, possibly suggesting that the regions away from genes are more conserved. The polymorphism may be maintained by selection pressures or by hitchhiking effects; one possibility is that DNA polymorphism evolved to facilitate the packing of eukaryote DNA into chromatin (Keene and Elgin, 1984). The latter idea is based on evidence that there is a distinctive organization of nucleolytic cleavage sites in eukaryote genomes. The pattern of chromatin organization has been shown to follow this distinctive pattern of DNA structural organization in Drosophila, and selective pressures may have maintained this sequence organization (Keene and Elgin, 1984). If polymorphic sequences are selected for in this way, then they cannot be described as neutral.

Intriguing data suggests that the rate of evolution of many non-coding regions is much higher than that of coding regions (van der Berg *et al.*, 1978; Poncz *et al.*, 1983). It could be argued that this is to be expected because these regions are not likely to have functional significance and would be changed by random genetic drift. A fast rate of evolution seems unlikely for functionally insignificant or neutral regions, however, since population genetic theory indicates a slow rate of evolution by random genetic drift (Bodmer,

hypervariable regions, with the length variation being attributed to unequal crossing-over. Hypervariable regions are commonly near structural genes, possibly suggesting that the regions away from genes are more conserved. The polymorphism may be maintained by selection pressures or by hitchhiking effects; one possibility is that DNA polymorphism evolved to facilitate the packing of eukaryote DNA into chromatin (Keene and Elgin, 1984). The latter idea is based on evidence that there is a distinctive organization of nucleolytic cleavage sites in eukaryote genomes. The pattern of chromatin organization has been shown to follow this distinctive pattern of DNA structural organization in Drosophila, and selective pressures may have maintained this sequence organization (Keene and Elgin, 1984). If polymorphic sequences are selected for in this way, hen they cannot be described as neutral.

Intriguing data suggests that the rate of evolution of many non-coding regions is much higher than that of coding regions (van der Berg *et al.*, 1978; Poncz *et al.*, 1983). It could be argued that this is to be expected because these regions are not likely to have functional significance and would be changed by random genetic drift. A fast rate of evolution seems unlikely for functionally insignificant or neutral regions, however, since population genetic theory indicates a slow rate of evolution by random genetic drift (Bodmer,

1981). Furthermore, whether these regions are indeed neutral or not is unknown. It remains unclear how such non-coding regions diverge rapidly during evolution if neither selection leading to rapid changes nor neutral random genetic drift are adequate to account for their rate of evolution. Mutation pressure, even if mutation rates are high, is not an answer since at the molecular level each mutation is essentially a unique event.

A possible explanation for a high rate of evolution of non-coding regions could be asymmetrical gene conversion in heterozygotes during meiosis. Perhaps heterozygotes for complex mutational events, particularly involving insertion or deletion and so looping out of the DNA of one parent strand during meiotic pairing, might lead to preferential excision of material from one or other parental strand. This could be a form of directed meiotic drive for some kinds of new mutation. The changes would occur at a fairly high rate, especially if unopposed by natural selection.

The neutrality of DNA polymorphism can be tested by examining the distribution of RFLPs with respect to coding and non-coding regions, or by comparing the degree of polymorphism in a specified region between individuals. A random distribution would imply that the polymorphism is neutral. Adams and Rothman (1982) examined restriction enzyme sites in human mitochon-

drial DNA and in viral DNA and found their distribution to be non-random. This finding may not reflect on any resulting polymorphisms however, since the pallindromes in restriction enzyme recognition sites may be of structural or functional significance and be non-randomly distributed for this reason. The use of neutral polymorphisms will give unbiased data for the inference of evolutionary relationships between populations, but with the present state of knowledge, the neutrality of DNA sequence polymorphisms can only be speculated upon.

If all DNA sequence changes are important to the organism for either structure or function, as seems increasingly likely, then no DNA sequence should be termed 'junk'. The study of restriction site polymorphisms represents an accurate and non-invasive monitor of human genetic variability and mutational load. The number of DNA markers is virtually limitless and their value as genetic markers clearly apparent. Furthermore, information on any biological function of the marker under investigation can potentially be derived. Their usefulness is not limited to anthropological studies and gene mapping, but their further investigation will give evidence as to the behaviour of chromosomes during meiosis, and information on genomic organization and function.

6.2 Anthropological applications of DNA markers

The origins of man

Boyd (1950) used ABO, MN and Rh blood group data to elucidate the origins and affinities of certain populations. Few loci are sufficiently informative for drawing anthropological conclusions, however, and information from many loci must be accumulated before the evolutionary history of a group of populations can be reconstructed. Cavalli-Sforza and Edwards (1964) described a tree of descent of human ethnic groups constructed using five serogenetic (blood group) loci in 15 populations. The origin, or root, of the tree was placed midway between the two most remote populations: Bantu-speaking Africans and Australian aborigines, thus dividing the world into Europe and Africa ('Eurafrica') on one side, and Asia, America and Australia ('Greater Asia') on the other. More recently, when the comparisons were extended to new enzyme markers and restricted to representative populations in Europe, East Asia and Africa, results were obtained which were different from those obtained using blood groups (Nei and Roychoudhury, 1982). The new results favoured a primary split between the Africans on one side and Eurasia on the other.

Once DNA markers became available, the number of markers that could be studied increased. Results obtained from DNA polymorphisms, both from mitochondrial variants (Johnson *et al.*, 1983) and from RFLPs (Wainscoat *et al.*, 1986a; Cann *et al.*, 1987; Bowcock *et al.*, 1987), have been consistent with an early Africa/Eurasia split.

Data presented in this study are consistent with those obtained using other DNA markers. The genetic distance between the 'African' populations (San and Negroid) is less than the distance between either of these populations and the 'Eurasian ' population (the southern African Caucasoid, which is north-Western European in origin). This is consistent with a primary African/Eurasian split even though it has not been possible to include Asian populations in the study.

The discrepancy between information obtained from blood groups and that obtained from enzyme and DNA data is possibly due to the limited number of loci that were available for blood group data, and it may be that collated information derived from a larger number of loci is more accurate.

Polymorphisms may not always prove to be conclusive on the question of the origin of peoples, although they can be useful in identifying certain individuals or

communities. Problems arise for several reasons. Evolution may occur in 'jumps' or punctuated equilibria, which would explain the absence of hypothetical intermediate forms in the fossil record. It is possible that transposable elements have played a role in evolution, resulting in the immediate acquisition of a specific character instead of its gradual evolution by a series of point mutations. If a region of DNA is conserved then it cannot have an evolutionary timetable or history that is interpretable. There is also the complication of parallel evolution. Antonarakis *et al.* (1984) and Pagnier *et al.* (1984) have used RFLP data to show that the sickle cell mutation arose at least four times in Africa. It would seem, then, that RFLPs can give information about populations although not necessarily about evolutionary timescales.

Racial divergence in Africa

Southern Africa is a region of particular interest to the human biologist, not least because early hominids lived here. The subcontinent is also the home of the Khoisan, who are perhaps the smallest and most distinctive of the major races of man, and of two large immigrant populations which evolved in distant ecosystems: Negroes, who migrated here over the last two millenia and Caucasoids, who have arrived over the past three centuries. The descendants of the Negroes and

Caucasoids have remained essentially unmixed and have also produced new hybrid populations by miscegenation with modern or ancient neighbours. It is possible to study the effects of racial admixture on a variety of traits in a range of different ecosystems, in southern Africa.

If the Khoisan and Negro descended from a common stock, which seems likely, then divergence had already taken place at least 20 000 - 30 000 years ago, outside southern Africa. Further local differentiation of the Khoisan stock into San (hunter-gatherer) and Khoi (pastoralist) subdivisions took place in southern Africa, and must have occurred relatively late, judging by their modern close resemblance at both the morphological and genetic levels (Nurse *et al.*, 1985). The differentiation probably began when one group began practising pastoralism.

Results in the present study show that the San and Negroid populations are closer, genetically, than either is to the Caucasoid population, and this is consistent with the Caucasoids having diverged from African populations well before the divergence within Africa of the San and Negro peoples. Unfortunately, it did not prove possible to include a Khoi sample in the study.

6.3 Linkage studies

Knowledge of the chromosomal linkage map gives insight into evolution, chromosomal organization in relation to genetic control mechanisms, and the pathogenesis of neoplasms and malformations. It is also useful in the preclinical or prenatal diagnosis of hereditary diseases, and may produce information fundamental to the understanding of the mechanisms underlying disease susceptibility and pathophysiology.

It is only twenty years since the first autosomal gene (Donahue *et al.*, 1968) was mapped. Technological advances have since enabled rapid progress in the field of gene mapping. Two important advances were the development of somatic cell hybrids and the identification of DNA sequence polymorphisms at unique loci. Advances in the analysis of data by computer have substantially decreased the time taken to manipulate the large amount of data needed for linkage analysis, and have increased the accuracy of such calculations.

An important application of linkage data is the identification of loci whose mutant alleles can cause genetic disease. Localization has several benefits. Heterogeneity can be determined if separate loci are found to be associated with a given disorder, linked

markers can be used to diagnose the genetic status of family members, and the gene itself can potentially be identified and isolated. Pulsed field gel electrophoresis (PFGE) enables the separation of large fragments of DNA (Schwartz and Cantor, 1984; Smith and Cantor, 1986a; Carle *et al.*, 1986) and constitutes a faster technique for cloning the disease locus after a linked RFLP has been found. The argument that PFGE is too sophisticated for routine use cannot hold: according to historical precedent, no convenient diagnostic technique is not eventually adopted for clinical use.

Newer linkage strategies have been propounded. These include the possibility of mapping rare recessive diseases (Lander and Botstein, 1987), by investigating affected children from consanguinous marriages. Using this method the disease locus is detected by virtue of the fact that the adjacent region will preferentially be homozygous by descent in inbred children. Homozygosity mapping is, however, only practical with the advent of a human genetic map and will be most efficient with a high resolution map.

The use of candidate genes, identified by characterizing gene products in a metabolic pathway, or by analysing relevant biological parameters, is also a

good approach to specific linkage analyses and should be used whenever possible.

Carrying out all possible pairwise tests between a new locus and other loci already investigated would be both wasteful and inefficient. Alternatives to this approach include analysis of haplotype distribution patterns in informative families, thus identifying the general region to which a locus belongs (White *et al.*, 1985), and attempting to detect or exclude linkage between the new locus and a whole linkage group, on the assumption that the current map of a region is sufficiently accurate (Lathrop *et al.*, 1984).

The first putative genomic linkage map reported in man (Donis-Keller *et al.*, 1987) represents a first step towards a 'complete' map. Its average spacing is approximately 10 cM, with some regions being relatively poorly covered. A map consisting of markers at 5 cM intervals is a worthwhile goal, but it may not be easy to attain this goal, even with combined data from the C.E.P.H. collaborative study. Certain regions of the genome are 'cold' and appear to lack polymorphic sites, while others remain intractable to cloning. It remains to be seen whether or when this goal will be reached.

6.4 Future studies

The single-copy probes, λMR12 and λMR17, were found to identify variation at the DNA level. They should be further investigated in terms of inheritance and frequency in order to determine their usefulness as genetic markers.

Only λMR3 has been clearly localized, to chromosome 17, by somatic cell hybridization. Three other probes have been tentatively assigned using the same technique. *In situ* hybridization or formal linkage analysis using family studies should be performed in order to regionally localize the probes.

It would be interesting to extend the population data by examining additional markers in southern African populations.

Lastly, the search for linkage to tyrosinase-positive oculocutaneous albinism should be continued. The candidate gene approach may become feasible as genes in the melanin pathway become known. A cDNA clone containing a putative human tyrosinase gene has recently been described (Kwon *et al.*, 1987). The use of homozygosity mapping may solve problems of heterogeneity if several loci are found at which at least one is homozygous by descent in most of the inbred children

Author Heim Ruth Alice
Name of thesis Restriction fragment length polymorphisms in Southern African populations and a search for linkage to Tyrosinase-positive Oculocutaneous Albinism. 1988

PUBLISHER:
University of the Witwatersrand, Johannesburg
©2013

LEGAL NOTICES:

Copyright Notice: All materials on the University of the Witwatersrand, Johannesburg Library website are protected by South African copyright law and may not be distributed, transmitted, displayed, or otherwise published in any format, without the prior written permission of the copyright owner.

Disclaimer and Terms of Use: Provided that you maintain all copyright and other notices contained therein, you may download material (one machine readable copy and one print copy per page) for your personal and/or educational non-commercial use only.

The University of the Witwatersrand, Johannesburg, is not responsible for any errors or omissions and excludes any and all liability for any errors in or omissions from the information on the Library website.