It is to be expected that a large probe such as λRH52 would detect polymorphism with more than one enzyme. This is, in fact, the rationale behind the strategy of Litt *et al.* (1986a), who used large cosmid probes in a search for polymorphism. λRH52 grows poorly compared with the other phage in the study and this must be attributed to its size. If it is proved to identify a useful locus in the C.E.P.H. genetic map then the appropriate fragments should be subcloned into a plasmid.

In summary, the λRH52 probe detects base pair substitution-type variation with the enzymes *Stu*I and *Taq*I at polymorphic frequencies, and the morphs are inherited in a Mendelian fashion. The use of haplotyping, especially in the Caucasoid population in which the *Taq*I RFLP is more frequent, will make the locus more informative. In addition, a *Hinc*II variant is detected with this probe.

4.1.2 Random single-copy probe λMR3: variation detected with *Stu*I, *Mbo*I, *Hind*III, and *Pvu*II

λMR3 contains a human insert of approximately 15 kb (Table 4.1). When used as a probe it detects variation with four enzymes: *Stu*I, *Mbo*I, *Hind*III, and *Pvu*II. For each RFLP detected with the first three enzymes, all

heterozygotes are heterozygous with each enzyme, suggesting that λMR3 identifies a highly polymorphic locus similar to that for the first RFLP described, pAW101 (Wyman and White, 1980).

The polymorphic patterns detected with λMR3 in conjunction with *Stu*I, *Mbo*I, and *Hin*dIII have been seen in the three populations studied and are shown in Figure 4.6 (a), (b) and (c) respectively. The *Stu*I polymorphic fragments are 15.5 kb and (10.8 + 5.0) kb in size, those for the *Mbo*I RFLP are 5.1 kb and (3.0 + 2.4) kb, and those for the *Hin*dIII RFLP are 3.85 kb and (2.50 + 1.65) kb. The *Hin*dIII RFLP is slightly complicated by the presence of a 3.85 kb constant band which underlies the polymorphic band and results in a 'dosage' effect. Inheritance of the 'alleles' at the locus identified by λMR3 can be interpreted as fulfilling Mendelian expectations. A family study using the enzyme *Stu*I is shown in Figure 4.7.

The variation detected with the enzyme *Pvu*II has only been seen in single individuals, although 43 Negroid individuals have been screened, and cannot therefore be said to be found at polymorphic frequencies. One Negroid individual has been seen with an extra 5.5 kb fragment (Figure 4.8 (a)) and his family was not available for study of the fragment's inheritance. Another, different, variant was also seen in one

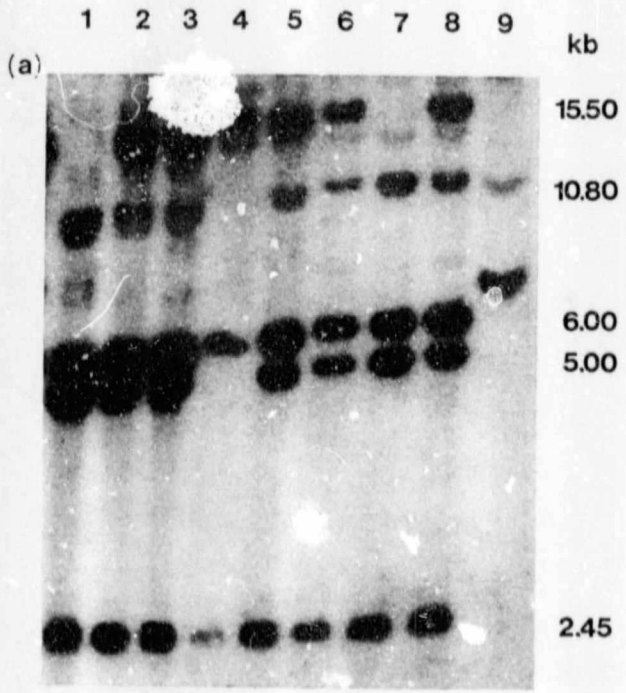

Figure 4.6 The polymorphic pattern revealed by λMR3 when hybridized to DNA from unrelated Negroid individuals digested with
(a) StuI. The polymorphic 'alleles' are 15.5 kb and (10.8 + 5.0) kb. Lane 9 contains a molecular weight marker.

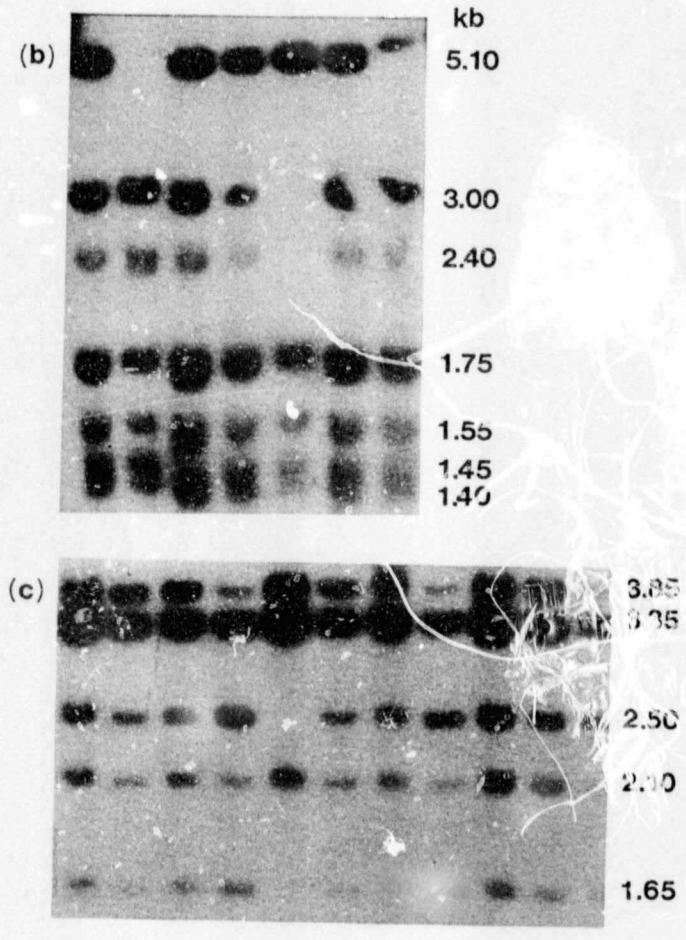

Figure 4.6 continued. λMR3 hybridized to unrelated Negroid DNAs digested with
(b) MboI. The polymorphic 'alleles' are 5.1 kb and (3.0 + 2.4) kb.
(c) HindIII. The polymorphic 'alleles' are 3.85 and (2.5 + 1.65) kb.

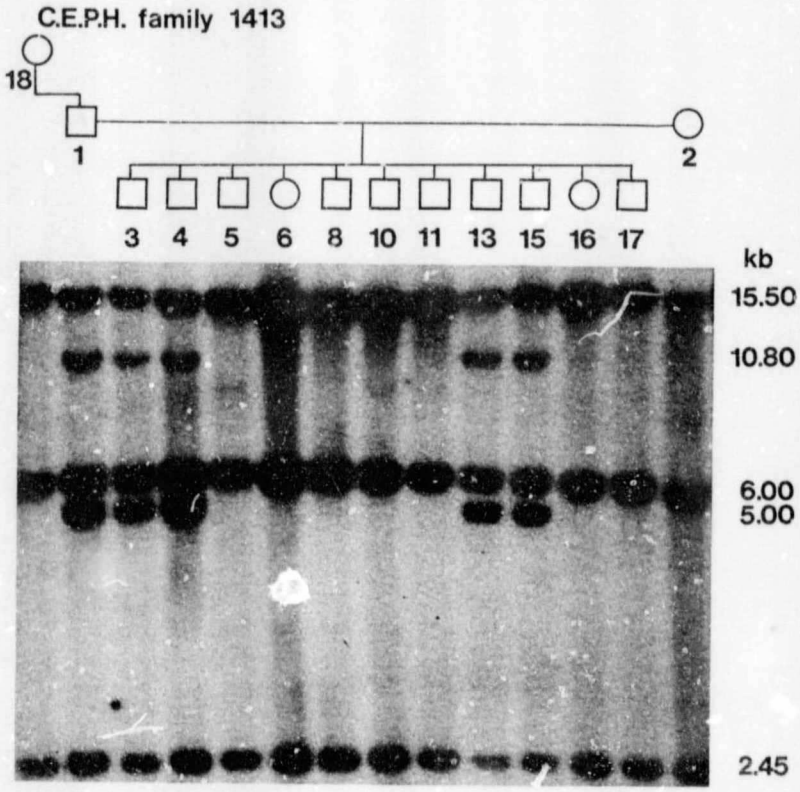

Figure 4.7 Family study showing the codominant Mendelian inheritance of the λMR3/StuI RFLP.

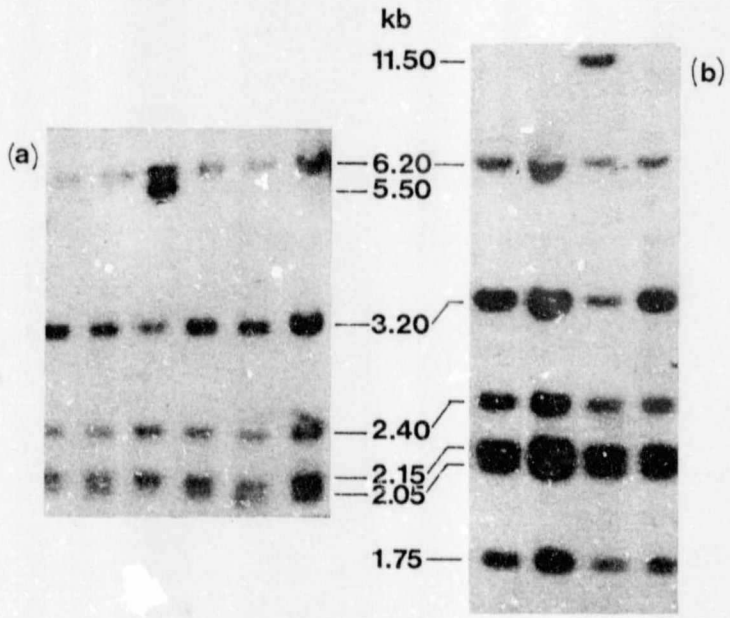

Figure 4.8 λMR3 hybridized to PvuII-digested DNA from unrelated Negroid individuals. (a) An extra 5.5 kb band is present. (b) An extra 11.5 kb band is present. The 1.75 kb band has been omitted from (a).

Negroid individual. The extra band was 11.5 kb in size (Figure 4.8 (b)), and this band was shown to be inherited (Figure 4.9).

The presence of these two isolated *Pvu*II variants at the locus identified by λMR3 suggests that the DNA in this region is subject to increased mutation. DNA rearrangement is presumably responsible for the variation detected with the three enzymes. There is a fixed length difference of 0.3 kb between the 'alleles' of each RFLP, as would be expected if the rearrangement was due to the insertion or deletion of a specific DNA sequence. A schematic interpretation of the structure of the DNA at the locus identified by λMR3 is given in Figure 4.10. Unequal crossing-over would make the locus a recombinational hotspot, and this would explain an increased mutation frequency such as is illustrated by the two *Pvu*II mutations observed.

In summary, the locus identified by λMR3 is useful as a bi-allelic genetic marker using any of three enzymes, for gene mapping and linkage studies. It is also potentially of interest as a region in which a 0.3 kb insertion/deletion has reached polymorphic frequencies in three populations. Study of such regions will help to elucidate mechanisms of recombination and may shed light on the contribution of DNA structure to its function.

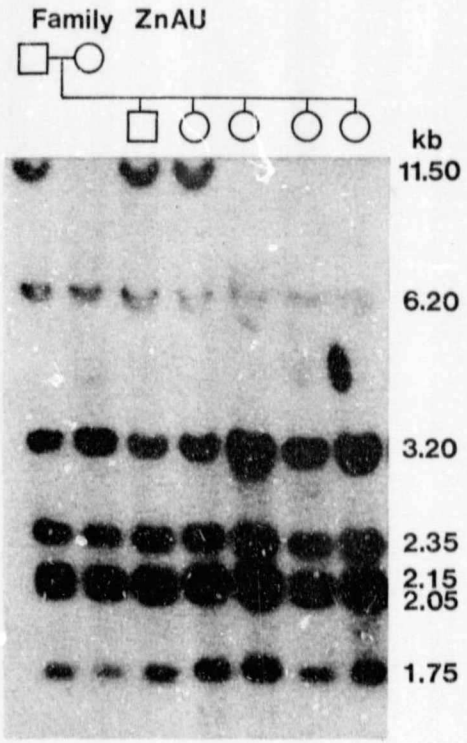

Figure 4.9 Inheritance of the 11.5 kb λMR3/PvuII fragment in a Negroid family.

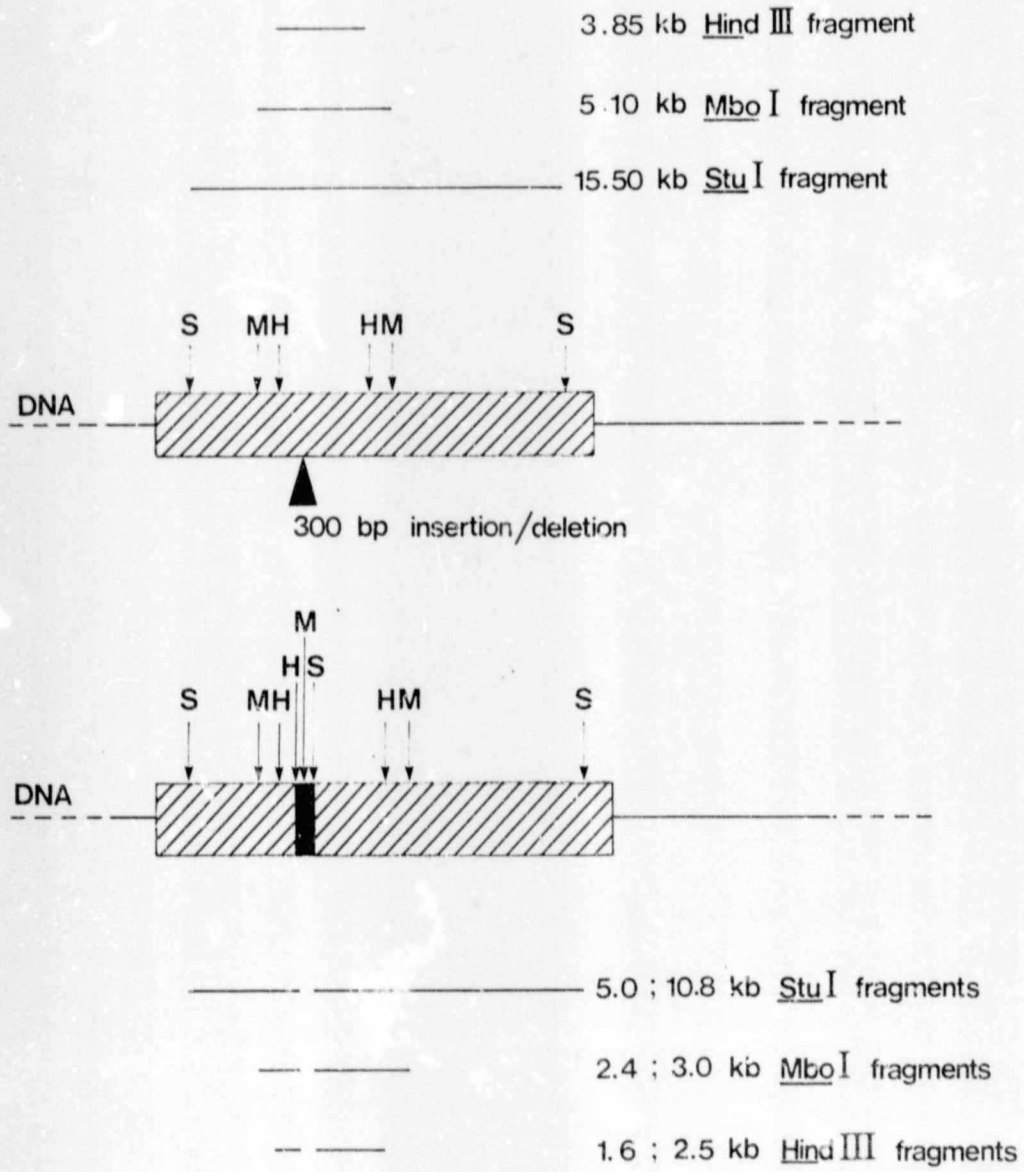

Figure 4.10 Schematic interpretation of the genomic structure at the locus identified by λMR3.

▨ region of probe homology

S, M, H = StuI, MboI and HindIII restriction sites

4.1.3 Random single-copy probe λMR22: variation detected with *Bgl*II and *Pst*I

λMR22 is a recombinant phage containing a relatively small insert of approximately 10 kb. When used as a probe it detects a two-allele *Bgl*II RFLP. The morphs are 7.5 and 6.0 kb in size (Figure 4.11) and they are inherited in a fashion which can be interpreted as Mendelian (Figure 4.12). The RFLP has been seen at high frequencies in the three populations studied, should be a useful polymorphic genetic marker.

λMR22 also detects a variant with the enzyme *Pst*I. The usual restriction pattern has four bands (6.1, 4.7, 3.0 and 1.3 kb), but DNA from one Negroid individual has been seen in which the 6.1 kb fragment is absent. DNA from a further twenty Negroid and twenty San individuals has been examined for this variant, but it has not been seen again. Although a family study is not possible, repeat digests with DNA from the individual consistently show only three bands. The variant is either present in the population at a very low frequency, or is a mutation which has occurred in one individual.

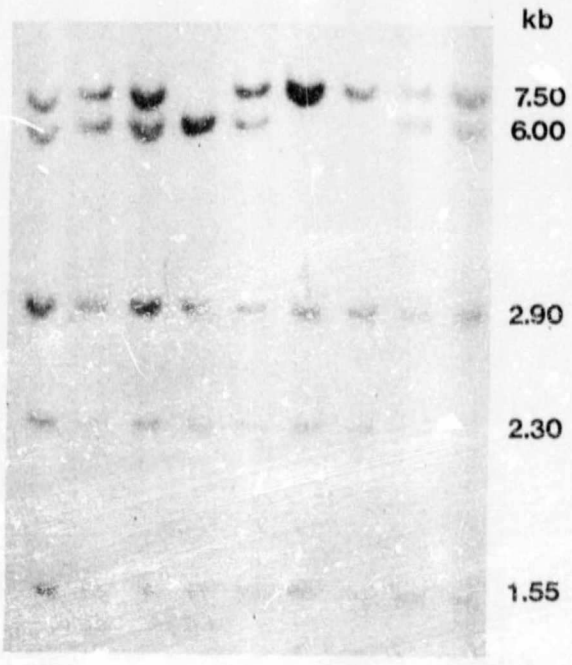

Figure 4.11 The polymorphic pattern revealed by λMR22 when hybridized to BglII-digested DNA from unrelated Caucasoid individuals. Polymorphic 'alleles' of 7.5 kb and 6 kb can be seen.

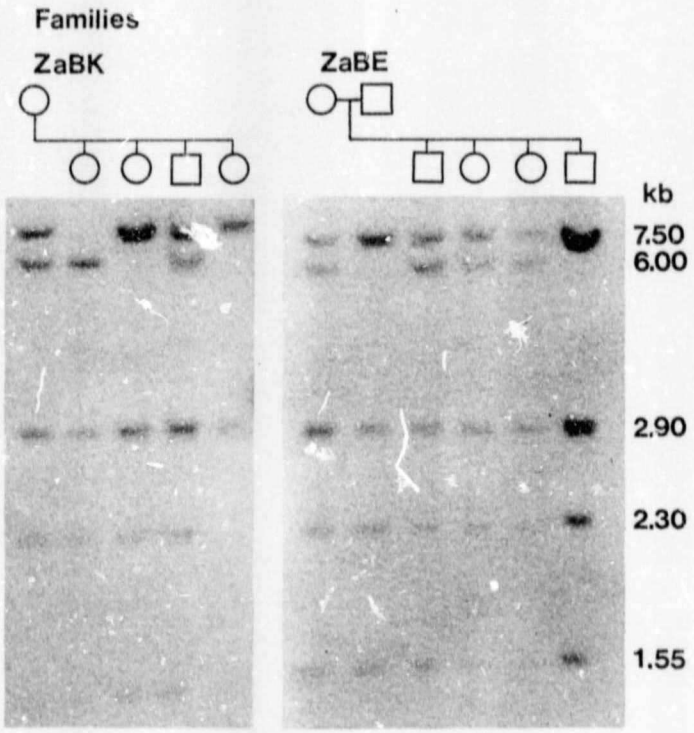

Figure 4.12 Study of two Negroid families showing the codominant Mendelian inheritance of the λMR22/BglII RFLP.

4.1.4 Random single-copy probe λMR7: variation detected with *Hinc*II and *Taq*I

The λMR7 recombinant phage contains a human insert of approximately 14 kb (Table 4.1). When used as a probe it identifies a two-allele polymorphism with the enzyme *Hinc*II, in three populations. The polymorphic fragments are 17 kb and 13.5 kb in size (Figure 4.13 (a)) and are inherited in a Mendelian fashion (Figure 4.13 (b)).

4.2 Assignment of λRH52, λMR3, λMR7, and λMR22

The localization of an arbitrary single-copy probe is a further step towards its characterization. In a first effort to localize some of the RFLP loci reported in this study, the technique of *in situ* hybridization was used. Several attempts have not been successful, however, whereas the use of somatic cell hybrids has proved more promising. A crude panel of somatic cell hybrids has been constructed by colleagues in the Cytogenetics Unit of the Department from hybrid lines which have not yet reached stability. The lines therefore contain some chromosomes in low percentages. In addition, the analysis of their chromosome content is not complete in some cases. The lines have nevertheless been useful as an indication of which chromosomes the probes may be derived from. The panel

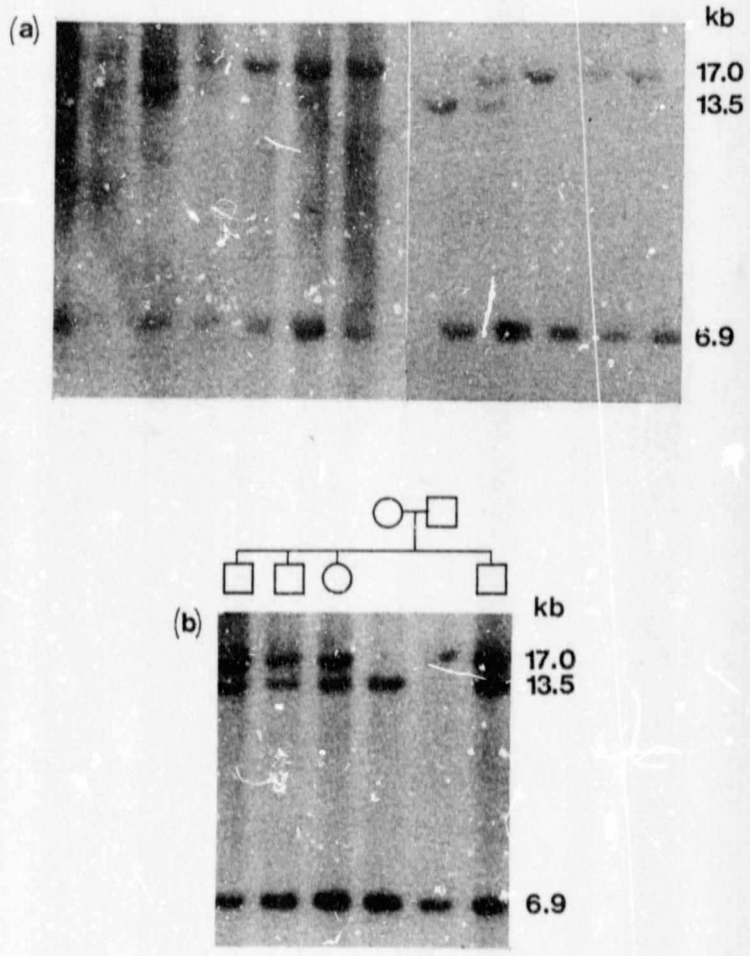

Figure 4.13 (a) The polymorphic pattern revealed by λMR7 when hybridized to HincII-digested DNA from unrelated Negroid individuals. The polymorphic 'alleles' are 17.0 kb and 13.5 kb in size.

(b) Study of a Negroid family showing co-dominant Mendelian inheritance of the λMR7/ HincII RFLP.

has been supplemented with better characterized hybrids kindly donated by Drs A.R. Retief, U. Francke and T. Mohandas.

4.2.1 Results

The lines used in the present study are listed in Table 4.2, together with their chromosomal constitutions. Autoradiographs showing the hybridization of the probes λMR3, λMR7, λMR22 and λRH52 to individual lines are shown in Figures 4.14 - 4.17 and the hybridization patterns are summarized in Table 4.2. Analysis of the marker segregation with human chromosomes in the hybrid lines is given in Table 4.3.

In order to use all available data towards the assignment of the markers, linkage analysis was performed where possible, using information from the linkage study described in Chapter 5. Known markers on the chromosomes suggested by somatic hybrid analysis were investigated for linkage against the probes λRH52, λMR3 and λMR22. The latter probes had been included in the linkage study. These results are given in Table 4.4.

Table 4.5 (page 166) represents a summary of the tentative chromosomal assignments.

Table 4.2 Segregation of DNA polymorphic markers with human chromosomes in somatic cell hybrid lines

Hybrid cell line	Source of line	Ref N°	Human Chromosome Complement																							Probe			
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	λMR3	λMR22	λMR7	λMR52
DMW5II-c	AR	1	/	-	-	/	/	100	40	90	/	60	100	80	90	30	60	90	/	-	80	90	/	80	90	-	-	NT	NT	NT
RBC13H-c	AR	2	-	-	30	-	-	-	30	-	-	-	50	-	-	-	60	-	70	80	-	40	-	100	-	-	+	NT	NT	NT
[illegible]	AR	3	+	-	-	-	-	100	-	-	-	-	100	-	90	-	-	-	-	-	-	-	-	-	100	-	-	NT	NT	NT
MDW5.5-c	AR	4	-	-	-	+	+	-	+	-	+	+	+	-	-	+	+	+	-	-	-	+	+	+	+[a]	-	-	NT	NT	NT
24-2-3	TM	5	-	-	-	-	-	-	-	-	-	-	-	86	95	/	-	-	-	-	-	-	-	-	-	-	-	-	-	+
24-2-10	TM	6	-	-	-	-	-	-	-	-	48	-	91	-	87	-	-	-	39	-	-	-	-	-	-	-	+	-	-	+
XI-4A-1D-E	UF	7	-	-	+	-	-	-	-	-	-	-	+	-	-	+[c]	-	+	-	+	-	-	-	-	+[c]	+	-	+	-	-
XII-2D-1d	UF	8	-	-	+	-	-	-	-	-	-	-	-	-	+	+	+	+[d]	-	-	-	-	+	-	-	-	-	-	-	-
XII-2D-1f	UF	9	+	+	+	+	-	+	-	+	+	-	-	+	+	+	+	+	-	+	+	-	+	+	+[c]	-	-	+	+	+
JNB B4.1	CU	10	-	-	-	88	-	-	-	88	-	-	-	84	72	84	-	-	92	-	-	-	84	-	-	-	+	-	NT	-
JNB B2	CU	11	-	/	-	-	-	-	93	-	/	-	-	82	36	64	25	/	86	29	/	46	93	68	-	-	+	-	NT	-
JNB B4.3	CU	12	-	28	-	72	-	-	-	78	/	61	-	83	44	61	-	-	89	/	-	-	83	33	-	-	+	-	+	-
JNB B4.5	CU	13	-	/	-	77	-	-	-	88	-	77	-	88	59	65	-	-	94	-	-	-	82	71	-	-	+	-	NT	-
CRR 1C5.1	CU	14	47	-	-	29	71	-	53	-	35	82	-	-	/	-	-	65	35	-	-	-	47	-	47	-	+	-	-	+
CRR 1C5.2	CU	15	59	-	-	35	71	-	88	-	53	82	-	-	/	-	-	71	-	-	-	-	?	-	59	-	-	-	-	+
CRR 2G6	CU	16	/	-	-	-	82	55	?	-	-	36	73	-	27	55	45	91	-	64	/	73	100	27	73	-	-	+	NT	+
SPR 1E1.1	CU	17	-	/	-	83	-	-	-	72	-	-	-	78	61	78	-	-	-	-	-	-	89	-	-	-	NT	-	+	-
SPR 1J1.1	CU	18	-	/	-	89	-	-	-	-	-	-	-	74	37	42	-	-	95	-	-	-	100	-	-	-	NT	-	-	-
FBR 3B1.2	CU	19	-	-	?	-	?	-	100	-	?	100	-	-	-	-	-	-	73	-	-	-	100	-	40	-	-	-	-	-

AR = A Retief ; TM = T Mohandas ; UF = U Francke ; CU = Cytogenetics Unit
+ = chromosome present in >15% of cells (all UF lines >80% of cells) ; - = chromosome not observed ; / = chromosome observed in <15% of cells ; ? = presence of chromosome not confirmed ; NT = not tested
Partial chromosomes : a = Xq22 -Xqter ; b = 14qter - 14q21 ; c = Xp22 - Xpter ; d = 16qter - 16q23
Numbers refer to the percentage of cells in which the chromosomes are present

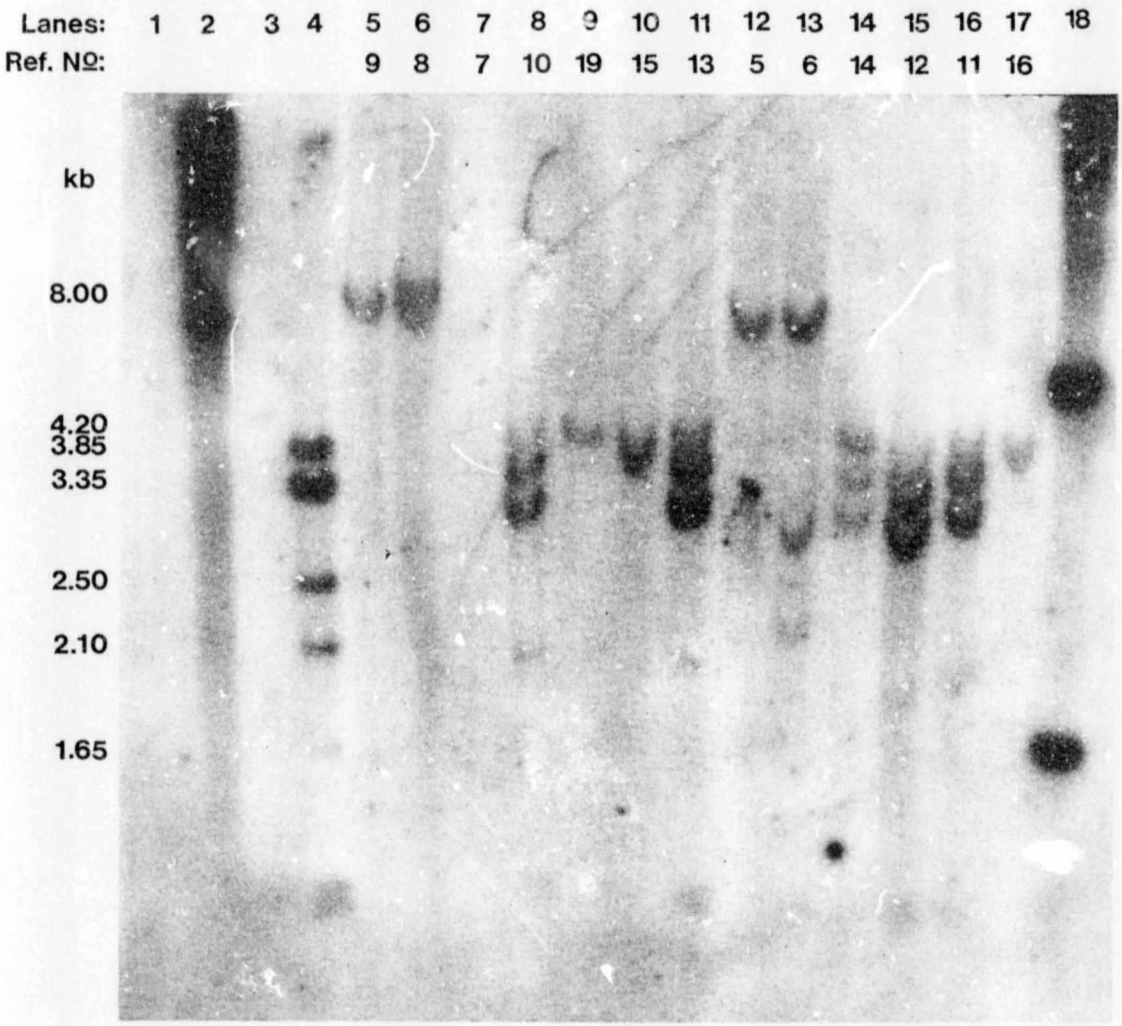

Figure 4.14 λMR3 hybridized to DNA from somatic cell hybrid lines (lanes 5 - 17). The reference numbers shown for each line are those used in Table 4.2.

DNA from chinese hamster cell line (lane 1).
Lambda molecular-weight markers (lanes 2 and 19).
DNA from mouse cell line (lane 3).
Human male DNA (lane 4).

All DNA was digested with HindIII.

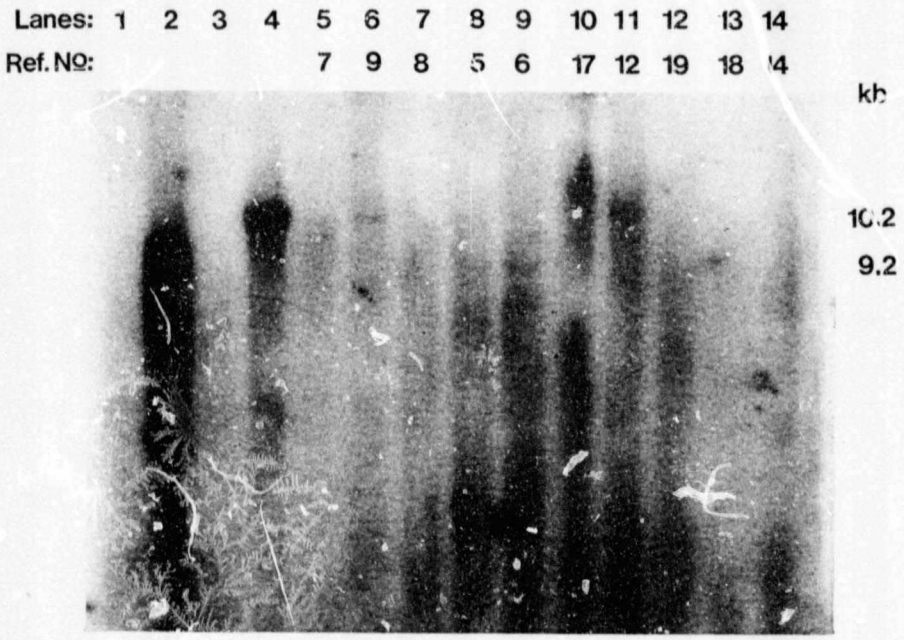

Figure 4.15 λMR7 hybridized to DNA from somatic cell hybrid lines (lanes 5 - 14). The reference numbers shown for each line are those used in Table 4.2.

DNA from chinese hamster cell line (lane 1).
Lambda molecular-weight marker (lane 2).
DNA from mouse cell line (lane 3).
Human male DNA (lane 4).

All DNA was digested with HindIII.

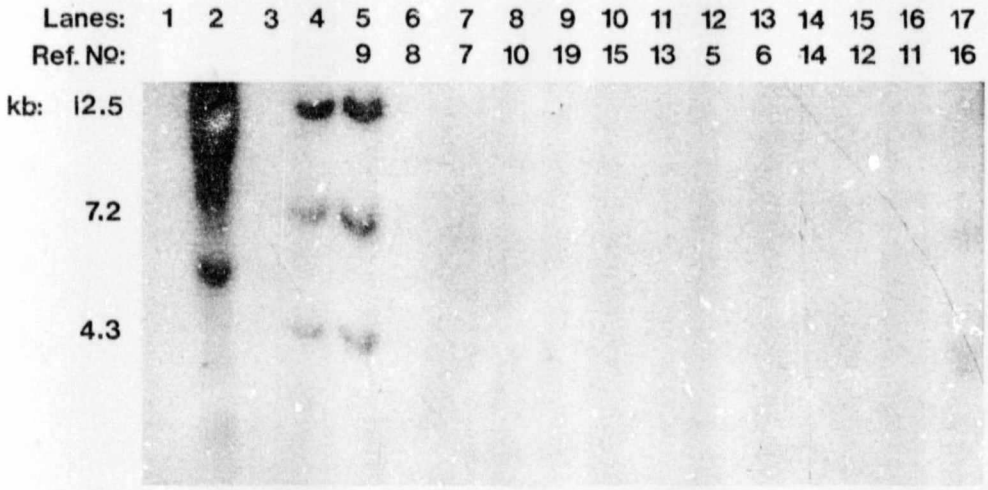

Figure 4.16 λMR22 hybridized to DNA from somatic cell hybrid lines (lanes 5 - 17). The reference numbers shown for each line are those used in Table 4.2.

DNA from chinese hamster cell line (lane 1).
Lambda molecular-weight markers (lane 2).
DNA from mouse cell line (lane 3).
Human male DNA (lane 4).

All DNA was digested with HindIII.

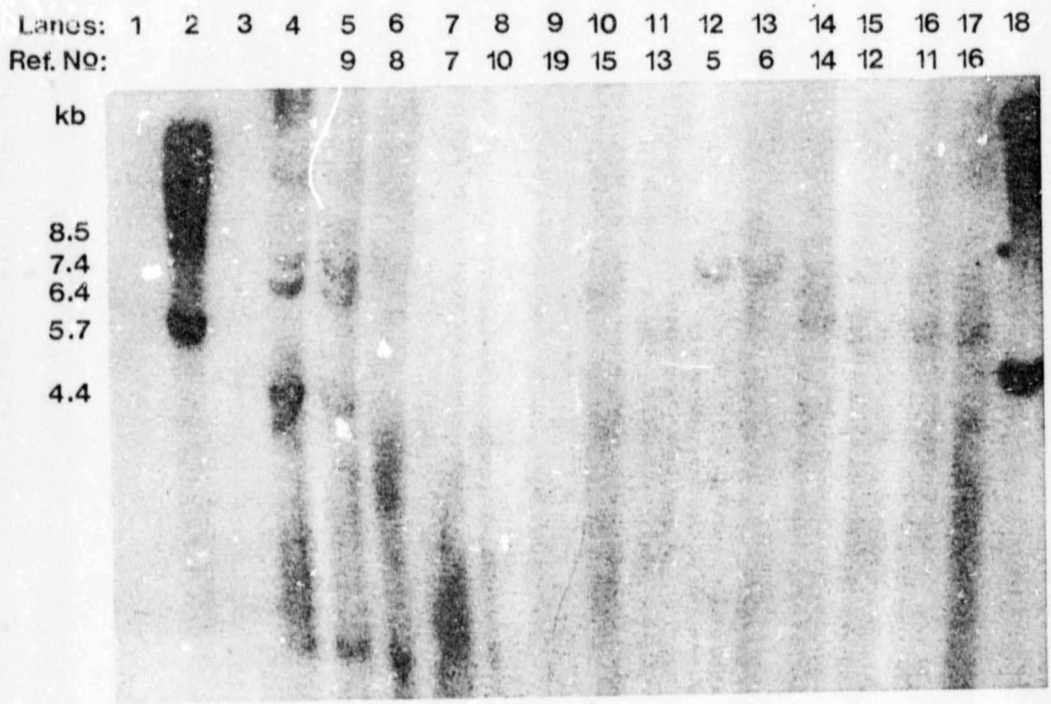

Figure 4.17 λRH52 hybridized to DNA from somatic cell hybrid lines (lanes 5 - 17). The reference numbers shown for each line are those used in Table 4.2.

DNA from chinese hamster cell line (lane 1).

Lambda molecular-weight markers (lanes 2 and 18).

DNA from mouse cell line (lane 3).

Human male DNA (lane 4).

All DNA was digested with HindIII.

Table 4.3 Analysis of DNA marker segregation with human chromosomes in somatic cell hybrid lines

Probe	Discordancy	Chromosome																						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
λMR3	Fraction	9/15	5/15	10/16	6/16	9/15	11/17	8/16	6/17	6/13	9/17	10/17	6/17	7/15	8/15	10/17	12/15	0/16	7/16	8/15	8/17	7/15	7/17	11/14	8/17
	Percentage	60	33	63	23	60	67	50	35	46	53	59	35	47	53	59	80	0	44	53	47	47	41	79	47
λMR7	Fraction	4/11	0/9	4/10	3/11	5/10	2/11	6/11	0/11	4/9	5/11	5/11	2/11	4/9	2/9	3/11	6/10	5/10	2/10	2/11	3/11	4/10	1/11	5/9	4/11
	Percentage	36	0	40	27	50	18	55	0	44	45	45	18	44	22	27	60	50	20	18	27	40	9	55	36
λMk22	Fraction	3/14	3/11	2/14	9/15	4/14	1/15	6/14	6/15	5/12	7/15	2/15	9/15	10/13	7/13	3/15	3/13	10/15	1/14	1/13	3/15	10/14	4/15	3/13	2/13
	Percentage	21	27	14	60	29	6	43	40	42	47	13	60	77	54	20	23	67	7	8	20	71	27	23	15
λRH52	Fraction	0/14	3/11	4/	5/15	2/14	3/15	3/14	6/15	1/12	5/15	6/15	9/15	10/13	10/13	5/15	3/13	7/15	5/14	2/13	5/15	9/14	6/15	2/13	4/15
	Percentage	0	27	29	33	14	20	21	40	8	33	40	60	77	77	33	23	47	36	15	33	71	40	15	27

Table 4.4 Results of two-point linkage analysis between new DNA markers and known serogenetic markers

Two-point analysis	Chromosomal location of serogenetic marker	θ					Number of informative families
		0.05	0.1	0.2	0.3	0.4	
λRH52 vs Rh	1	-4.437	-2.329	-0.689	-0.138	-0.012	13
λRH52 vs PGM_1	1	-2.425	-1.360	-0.513	-0.173	-0.034	9
λRH52 vs PGD	1	-1.898	-1.104	-0.479	-0.206	-0.053	4
λRH52 vs ABO	9	-3.869	-1.019	0.783	0.851	0.308	20
λMR22 vs GLO_1	6	-6.191	-3.076	-0.793	-0.095	0.034	16
λMR22 vs Bf	6	-6.627	-3.480	-1.032	-0.215	-0.033	16
λMR22 vs PEP A	18	-6./18	-2.679	-0.253	0.294	0.158	16

4.2.2 Discussion

λMR3

Figure 4.14 shows the autoradiograph obtained from the hybridization of λMR3 to somatic cell hybrid DNA. When interpreting this autoradiograph it is necessary to differentiate between the fragment sizes that can be seen, because the probe shows cross-hybridization to both chinese hamster DNA and mouse DNA. The three U. Francke lines (7, 8 and 9) and the two T. Mohandas lines (5 and 6) used chinese hamster cells for somatic fusions. These five lines all show hybridization to an 8.0 kb band which is not seen in the human male control (H). Although the band is not seen in the chinese hamster DNA control lane, it is possible that the chinese hamster cell line obtained from the Cytogenetics Unit was not the same line used for the fusion. In all the remaining lines, which were constructed by fusion with the mouse cell line RAG, a 4.2 kb band is visible and this band is faintly visible in the mouse DNA control but not in the human DNA control. The probe is polymorphic with *Hin*dIII (see section 4.1.2) and bands seen in lines 6, 10, 11, 12, 13 and 14 are consistent with the polymorphism. Hybridization was also obtained to line 2 (not shown). These, then, are the lines designated '+' in Table 4.2.

Table 4.3 gives the discordancy fractions and the corresponding percentages for each chromosome to which the probes tested on the hybrid lines in Table 4.2 may have hybridized. This analysis shows that λMR3 segregates with chromosome 17 in all lines.

No markers on chromosome 17 were included in the linkage study, therefore no linkage analysis was possible.

λMR7

The autoradiograph showing the hybridization of λMR7 to *Hind*III-digested somatic hybrid lines is presented in Figure 4.15. There is no cross-hybridization to either mouse or chinese hamster DNA.

It is evident from the analysis in Table 4.3 that λMR7 segregates with chromosomes 2, 8 and 22, and that there is one discordancy with respect to chromosome 22. The lines containing chromosome 8 are much better characterized than those containing chromosome 2, but no conclusion can be drawn as to whether λMR7 is assigned to chromosomes 2 or 8 .

λMR22

Figure 4.16 shows the autoradiograph obtained after hybridization of λMR22 to somatic cell hybrid DNA digested with *Hin*dIII. There is strong hybridization to line 9, weak hybridization to lines 7 and 16, and no cross-hybridization to either chinese hamster or mouse DNA.

From the analysis in Table 4.3, λMR22 segregates with chromosome 6 in 15 lines, with one discordancy, and with chromosomes 18 and 19 in 13 and 14 lines respectively, again with one discordancy each. According to this analysis λMR22 is derived from chromosome 6, but this conclusion is tentative.

Linkage analysis was performed with λMR22 against markers on chromosomes 6 and 18 (Table 4.4). There is no suggestion of linkage to GLO or to Bf, on chromosome 6, or to PEP A on chromosome 18.

λRH52

The hybridization of λRH52 to *Hin*dIII-digested somatic cell hybrid DNA is shown in Figure 4.17. There appears to be cross-hybridization to a chinese hamster DNA fragment of 8.5 kb in lines 5, 6, 7, 8 and 9; and cross-hybridization to a 5.7 kb mouse DNA fragment in

the remaining lines. Hybridization to bands seen in the human control has taken place with lines 9, 14 and 15. Hybridization to 15 is faint in this photograph but was confirmed in a separate experiment.

According to the analysis in Table 4.3, λRH52 segregates with chromosome 1 and also with chromosome 9, with one discordancy in the latter segregation. The tentative localization of λRH52 is, then, chromosome 1.

λRH52 was tested against three markers on chromosome 1, Rh, PGM and PGD, and one marker on chromosome 9, ABO (Table 4.4). There is no suggestion of linkage to the markers on chromosome 1, but a slightly positive lod score was obtained for linkage between λRH52 and ABO (0.85 at $\theta = 0.3$). The score is not significant but is nonetheless interesting and may constitute evidence for the probe being from chromosome 9 rather than 1.

In conclusion, from the results of hybrid analysis (modified by linkage analysis in the case of λRH52), λMR3, λMR7, λMR22 and λRH52 have been tentatively localized to chromosomes 17, 8, 6 and 9, respectively, as shown in Table 4.5 below.

Table 4.5 Tentative chromosomal localization of DNA markers

Probe	Chromosome
λMR3	17
λMR7	8 (or 2)
λMR22	6 (or 18 or 19)
λRH52	9 (or 1)

The above somatic cell hybrid analyses can be improved by including more hybrid lines, and should be complemented by in situ hybridization or definitive linkage results. It would obviously have been advisable to have used the four Retief lines, but, unfortunately, they were no longer available.

The current state of the human linkage map is such that the chromosomal localization of every probe contributing to the map need not be known. The putative map reported by Donis-Keller *et al.* (1967) was constructed using a combination of mathematical linkage analysis and physical localization of selected loci, and of the 360 RFLPs included in the map, only 54 (15%) constituted physically localized 'anchor' loci.

If the probes reported here are included in the human linkage map, then these tentative assignments will be of value for their inclusion in specified linkage groups.

4.3 Population studies

Four arbitrary single-copy probes described above have been used to obtain data on three major southern African populations. The probes represent cloned DNA segments of unknown function, which are thought to show lesser effects of natural selection, on average, than do clones of known genes. Their particular usefulness in anthropological studies lies in the untestable but perhaps reasonable assumption that anonymous marker systems are neutral. Neutral markers give better information about the evolutionary history of a group of populations than do markers under selection, which better reveal the ecological history of populations.

Sampling difficulties often result in a small number of individuals being tested. This can and should be compensated for by an increase in the number of loci tested.

In this section, population data are given and analysed in 4.3.1, and discussed in 4.3.2.

4.3.1 Results

Four single-copy DNA probes identifying six two-allele polymorphisms were investigated in the San, Caucasoid and Negroid populations. Phenotypic frequencies are

given in Table 4.6, as are the chi square values obtained when testing each marker for the validity of Hardy-Weinberg equilibrium. The number of chromosomes studied for each locus in Table 4.6 averaged 93 for the Negroids, 40 for the Caucasoids and 48 for the San.

Table 4.7 shows the allelic frequencies and PIC (Polymorphism Information Content) values for each marker. The PIC value is an estimator of polymorphic variation, first defined by Botstein *et al.* (1980), that takes into account both the number and the frequency of alleles for each marker. It gives the probability that offspring will be informative for allelic segregation, assuming a Hardy-Weinberg distribution of alleles within the population.

$$PIC = 1 - \left[\sum_{i=1}^{n} P_i^2 \right] - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2P_i^2 \times P_j^2$$

where n = number of alleles

P_i = frequency of ith allele

The PIC value lies between 0 and 1.0, where 0 describes a totally uninformative locus and 1.0 describes a highly informative locus.

Table 4.8 gives the variation within and between populations. The F_{ST} value is a measure of absolute heterozygosity for each marker and was calculated

Table 4.6 Phenotype data on 6 polymorphisms in 3 populations and Hardy-Weinberg tests for internal consistency.

Probe/enzyme		Negroid						Caucasoid						San					
			Phenotype			χ^2			Phenotype			χ^2			Phenotype			χ^2	
		n	11	12	22	df=1	P	n	11	12	22	df=1	P	n	11	12	22	df=1	P
λRH52/StuI	obs.	51	8	20	23	1.01	0.315	20	7	7	6	2.21	0.137	21	17	4	0	--	--
	exp.		6.3	23.2	21.5				5.6	10	4.4				17	3.8	0.2		
λRH52/TaqI	obs.	38	0	11	27	--	--	20	1	16	3	7.59	0.006	37	0	[illegible]	33	--	--
	exp.		0.75	9.15	28				4.05	9.9	6.05				0.1	3.5	33.4		
λRH52/HincII	obs.	26	24	2	0	--	--	20	20	0	0	--	--	20	20	0	0	--	--
	exp.		23.96	2	0.04				--	--	--				--	--	--		
StuI	obs.	77	14	39	24	0.48	0.841	20	15	3	2	0.98	0.322	22	0	4	16	--	--
	exp.		16.3	36.6	24.1				15.6	4.0	0.4				0.2	3.6	16.2		
λMR22/BglII	obs.	64	24	33	7	0.76	0.383	20	19	1	0	--	--	22	19	3	0	--	--
	exp.		25.4	29.8	8.8				18.87	1.16	0.02				19.02	2.86	0.12		
λMR3/HincII	obs.	23	16	5	1	0.56	0.415	22	10	11	0	--	--	22	4	18	0	--	--
	exp.		15.9	5.6	0.5				10.6	8.6	1.8				7.6	10.7	3.7		

n = number of individuals

P = probability that sample conforms to Hardy-Weinberg equilibrium

Table 4.7 Allele frequencies and Polymorphism Information Content (PIC) values for six polymorphisms in three populations.

Probe/enzyme	Allele (sizes in kb)	Allele frequencies (PIC)		
		Negroid	Caucasoid	San
λRH52/StuI	1 (5.8)	0.35 ± 0.05 (0.35)	0.53 ± 0.08 (0.37)	0.90 ± 0.05 (0.16)
	2 (4.0;1.8)	0.65 ± 0.05	0.47 ± 0.08	0.10 ± 0.05
λRH52/TaqI	1 (20.05)	0.14 ± 0.04 (0.21)	0.45 ± 0.08 (0.37)	0.05 ± 0.02 (0.09)
	2 (13.10;6.95)	0.86 ± 0.04	0.55 ± 0.08	0.95 ± 0.02
λRH52/HincII	1 (9.40)	0.96 ± 0.03	1.00 (0.00)	1.00 (0.00)
	2 (8.85)	0.04 ± 0.03	0.00	0.00
λMR3/StuI	1 (15.5)	0.40 ± 0.05 (0.37)	0.82 ± 0.08 (0.25)	0.10 ± 0.04 (0.18)
	2 (10.8;5.0)	0.60 ± 0.05	0.18 ± 0.08	0.90 ± 0.04
λMR22/BglII	1 (7.6)	0.63 ± 0.04 (0.36)	0.97 ± 0.03 (0.06)	0.93 ± 0.04 (0.12)
	2 (5.0)	0.37 ± 0.04	0.03 ± 0.03	0.07 ± 0.04
λMR7/HincII	1 (17.0)	0.85 ± 0.05 (0.22)	0.71 ± 0.07 (0.32)	0.59 ± 0.07 (0.37)
	2 (13.5)	0.15 ± 0.05	0.29 ± 0.07	0.41 ± 0.07

Table 4.8 Absolute and relative heterozygosities in three populations

Probe/ enzyme	Absolute hetero- zygosity (F_{ST} x1000)	Relative heterozygosity (dx1000)		
		Caucasoid	Negroid	San
λMR3/StuI	354	295	480	180
λMR7/HincII	56	412	255	484
λMR22/BglII	171	58	466	130
λRH52/StuI	217	498	455	180
λRH52/TaqI	177	495	241	95
λRH52/HincII	27	0	77	0
		D = 293	D = 329	D = 178

$$F_{ST} = \left[\frac{1}{3} \sum_{i=1}^{n} (p_i - \bar{p})^2 \right] / \bar{p}\bar{q}$$

where $\bar{p}$ is $\frac{1}{n} \sum_{i=1}^{n} p_i$ and $\bar{q}$ is $\frac{1}{n} \sum_{i=1}^{n} q_i$

$$d = 1 - \sum_{k=1}^{t} p_k^2$$

where t is the number of alleles
and p_k is the frequency of the kth allele

$$D = \frac{1}{n} \sum_{i=1}^{n} d_i$$

according to the Wahlund approach, by computing the averages $\bar{p}$ and $\bar{q}$, and the variance of p_i [variance = $1/3 \sum_{i=1}^{n} (p_i - \bar{p})^2$, where n is the number of populations], and then by dividing the variance by $\bar{p}\bar{q}$.

The average (unweighted) F_{ST} is 0.167.

The relative heterozygosities (d) of the populations at each locus, given in Table 4.8, were calculated as:

$$d = 1 - \sum_{k=1}^{t} p_k^2$$

where t is the number of alleles at the locus and p_k is the frequency of the k^{th} allele. The heterozygosity (D) of each population over n loci was calculated as the mean over all loci:

$$D = 1/n \sum_{i=1}^{n} d_i$$

Genetic distances between the three populations were computed, using the frequency data in Table 4.7, according to the method of Edwards and Cavalli-Sforza (1972) as follows:

$$\chi^2 = \sum_{k,l} \left[\frac{(P_{i,k,l} - P_{j,k,l})^2}{P_{i,k,l} + P_{j,k,l}} \right] / \sum_k d_k$$

where $P_{i,k,l}$ is the frequency of the allele l at locus k in population i and d_k is the number of degrees of freedom at locus k (Steinberg *et al.*, 1967).

The choice of a distance measure was not crucial since the clustering proceedure depends only on the rank

ordering of the pairwise distances. The clustering implied by the distances is given in Figure 4.18.

4.3.2 Discussion

Phenotypic and allelic frequencies, as well as statistics that summarise the patterns of frequency variation found, were presented in Tables 4.6, 4.7 and 4.8.

Table 4.6 shows the values obtained after chi square testing of the validity of Hardy-Weinberg equilibrium. Chi square was not calculated when only two phenotypes were observed, as the degree of freedom in this instance is zero. The distribution of chi square values for the seven markers that could be tested is as follows:

Chi square values	Probability	No. Obs.	No. Exp.
$\chi^2 < 3.84$	$P > 0.95$	6	6.65
$6.64 > \chi^2 > 3.84$	$0.1 < P < 0.95$	0	0.27
$0.84 > \chi^2 > 6.64$	$0.01 < P < 0.1$	1	0.07
$\chi^2 > 10.8$	$P < 0.01$	0	0.01
	Totals	7	7.00

There is excellent agreement with Hardy-Weinberg for almost all markers in the three populations. In the case of λRH52/*Taq*I the probability is less than 0.1, and the population may not be in Hardy-Weinberg equili-

Genetic distances between populations			
	Caucasoid	Negroid	San
Caucasoid	-	0.095	0.165
Negroid	-	-	0.071

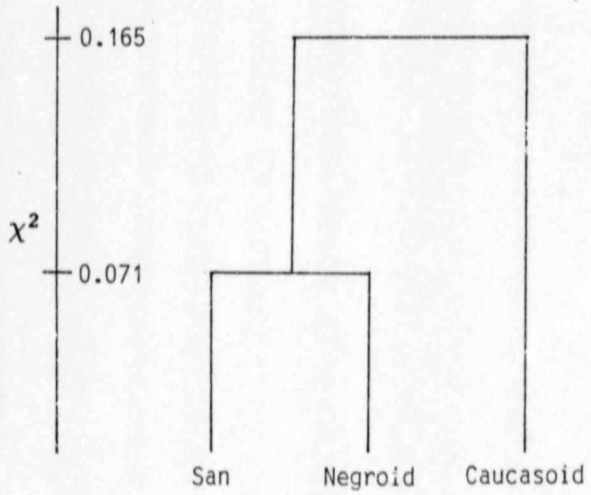

Figure 4.18 Clustering implied by gene frequencies at six loci in three southern African populations. The genetic distances between populations are shown in the table above, and a dendrogram derived from the distances is shown below.

brium. In this case there are more heterozygotes than would be expected, and this also seems to be the case for λMR7/*Hinc*II. The sample populations are very small, however, and the chi square test may not be an accurate indicator of equilibrium. Both λRH52/*Taq*I and λMR7/*Hinc*II must be tested in new, independent samples from the Caucasoid and San populations respectively, before it can be claimed that the populations are not in Hardy-Weinberg equilibrium for these systems.

The PIC values given in Table 4.7 are an indication of the usefulness of the RFLPS as genetic markers. In practice, RFLPs with values over 0.25 are said to be the most useful for linkage studies (Skolnick and White, 1982). The maximum value attainable for a two allele system is 0.37, and in the three populations of the present study there are 8 markers with values between 0.25 and 0.37. Generally, the PIC values in the San population are low, whereas those in the Negroids and Caucasoids are higher. This reflects the difference in allele frequencies between populations, shown statistically by the D value in Table 4.8.

F_{ST}, the standardised measure of gene frequencies, was calculated for each locus and given in Table 4.8. The distribution of F_{ST} is as follows:

F_{ST}	No of markers
0-0.1	2
-0.2	2
-0.3	1
-0.4	0
-0.5	0
0.5	0
Total	6

The locus showing the highest variation is that identified by λMR3/*Stu*I. This marker identifies a polymorphic site containing an insertion/deletion of 300 bp. Rearrangement due to unequal crossing-over may contribute to the variation between populations at this locus. λMR7/*Hinc*II has a low F_{ST} value, but the large relative heterozygosities (d) obtained in each population at this locus show that the lack of discrimination is not due to lack of polymorphism.

Taking each locus to be independent, the average heterozygosities for each population (D) are given in Table 4.8. These data show that heterozygosities are greatest in Negroids, probably reflecting the fact that the loci were identified by screening in Negroids. It is perhaps more significant to note that the San, who represent a small, isolated population, still show an appreciable level of heterozygosity. This is

consistent with data obtained using red cell enzyme systems (Jenkins *et al.*, 1971). Only one of the six sites polymorphic in the Negroids was monomorphic in the San, and the same site was monomorphic in the Caucasoids as well.

The presence of a polymorphism in a limited number of populations may reflect its origin. For example, it seems likely that the sickle cell mutation and some G6PD variants were selected for in hyper-endemic malaria areas, although this may not have been the only factor accounting for their distribution (Jenkins, 1982).

Of the DNA polymorphisms reported here, only one (λRH52/*Hin*cII) is monomorphic, in the Caucasoid and San populations. The remaining RFLPs are polymorphic in all three populations. Two of the probes (λMR3/*Stu*I and λMR22/*Bgl*II) are present at significantly higher frequencies in the Negroids than in the Caucasoids or the San. This probably reflects the fact that the polymorphisms were selected for in the Negroid population. The λRH52/*Taq*I RFLP, however, was found at highest frequencies in the Caucasoids, and the λMR7/*Hin*cII RFLP at highest frequencies in the San. One could postulate that a selective constraint acting at these sites or acting on DNA adjacent to the sites in each population maintains the polymorphisms.

The present study has shown that the usefulness of DNA polymorphisms found in Negroids can be extended to other populations, although, not unexpectedly, the degree of heterozygosity varies.

The value of gene frequency data in anthropology becomes apparent when genetic distances between populations are calculated using the allele frequencies. The clusterings derived from these distances can then be interpreted in the light of tradition, linguistics and the archaeological record. Jenkins (1972), for example, used genetic distance measurements to show convincingly that the San populations cluster together, while the Khoikhoi are more closely related to the 'Coloured' or hybrid populations of the region.

Results from the present study (Figure 4.18) show that the Negroids are more closely related to the San than either population is to the Caucasoids. This result is consistent with other evidence that the Khoisan and Negroid populations are descended from a common stock (Nurse *et al.*, 1985). The divergence probably took place outside South Africa, at least 20 000 - 30 000 years ago. The divergence between the Khoi and San must have occurred later, judging by their modern close

resemblance at both the morphological and genetic levels, and probably took place in southern Africa.

4.4 Contribution to the C.E.P.H. database

It is anticipated that collaborative studies with C.E.P.H will be an efficient means of creating a human linkage map. Of the probes isolated in the present study, two detected polymorphism in the Caucasoid population at frequencies suitable for linkage studies, λRH52 and λMR3. The λRH52/*Stu*I and the λMR3/*Stu*I RFLPs have been typed on the C.E.P.H. families: 25 families were 'informative' with the former RFLP and 17 with the latter. All individuals in these families have been typed. The results have not yet been analysed however, because the continued waging of an Academic boycott against South Africa has resulted in difficulties in obtaining the updated version of the LINKAGE programme.

The C.E.P.H. families come from Utah, France and Venezuala. At the C.E.P.H. meeting in Paris, 1986, it was agreed that the population data on random individuals (parents) from the French and Utah populations would probably be representative of a Western European population and could therefore be pooled. When these data were pooled and compared with southern African Caucasoid data (shown in Table 4.9), however, the respective allele frequencies seemed to be different.

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