Chapter Three

RESULTS

CHAPTER THREE: RESULTS

3.1 Flanking region polymorphism assays

The genotype and allele frequencies in the flanking DNA of the minisatellite MS31A within the Malaysian population were investigated in this study. A total of 310 samples were collected, of which 103 individuals were Malays, 106 Chinese and 101 Indians. The restriction endonuclease (RE) site polymorphisms analysed in this study included the *Alu*I site [four bases upstream of MS31A sequence (-4) (see Figure 5)], the *Hga*I site (-220) and the *Psp*1406I site (-108). The results were based on the presence or absence of RE sites, and were assigned the following symbols:

+/+ = cleavage of both alleles, known as homozygous positive

+/- = cleavage of one of the alleles, known as heterozygous

-/- = no cleavage, known as homozygous negative

3.1.1 *Alu*I +/- assay

PCR amplification was performed by using primers 31-TAG-A and 31A. The resultant 135 bp products were digested with *Alu*I (Section 2.6.2) to reveal the genotype based on the presence, absence, or presence and absence of the *Alu*I sites at position –4. The recognition sequence for *Alu*I is AG^CT, where '^' represents the cleavage site. Cleavage is expected for –4A alleles. *Alu*I will not cleave –4G alleles, leaving the product intact. The expected fragment sizes (band patterns) are:

i) a single band of 135 bp, representing the –4G alleles in a homozygous state, -/-;

ii) a single band of 95 bp, representing the –4A alleles in a homozygous state, +/+; and

iii) two bands of 135 and 95 bp, representing -4A/G alleles in a heterozygous state, +/-.

(Note: *Alu*I digestion produced two fragments, 95 and 40 bp. However, only the 95 bp was detected with the current electrophoresis conditions.)

As expected, three types of banding patterns were detected after electrophoresis (Figure 9). The results of *Alu*I genotype and allele analyses are given in Table 5.

The most common genotype observed was (-/-) with frequencies of 0.57, 0.55, and 0.66 for the Malays, Chinese, and Indians, respectively. The least common was (+/+) with frequencies ranging from 0.06 to 0.09. The frequencies of the A(+) allele were 0.26 for the Malays, 0.25 for the Chinese and 0.20 for the Indians, while those of the G(-) allele were 0.74, 0.75, and 0.80 for the Malays, Chinese, and Indians, respectively. These results were consistent with those reported by Koh *et al.* (1993) for the Malaysian population.

On the assumption that the distribution of the allele frequencies followed the Hardy-Weinberg equilibrium, the expected number of individuals representing each genotype was calculated. Chi-square test results were not significant, consistent with previous observations for the Malaysian population (Koh *et al.*, 1993) and for the Caucasian and Japanese populations (Neil and Jeffreys, 1993).

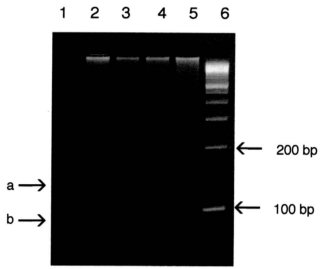

Figure 9: Ethidium bromide-stained 4% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:1) of *AluI*+/- assay in MS31A 5' flanking DNA (-4A/G polymorphisms).

Lane 1: Control (blank)
Lane 2: +/- (A/G)
Lane 3: +/+ (A/A)
Lane 4: -/- (G/G)
Lane 5: +/- (A/G)
Lane 6: 100 bp ladder
a) approximately 135 bp uncut product
b) approximately 95 bp cut product

Table 5: Distribution and frequencies of *AluI* site genotypes and alleles in Malays, Chinese and Indians in the Malaysian population.

		Malays	Chinese	Indians
Obsd.+/+(A/A) (f)		9 (0.09)	7 (0.07)	6 (0.06)
Obsd.+/- (A/G) (f)		35 (0.34)	40 (0.38)	28 (0.28)
Obsd. - /- (G/G) (f)		59 (0.57)	59 (0.55)	67 (0.66)
Total no. of individuals		103	106	101
Expd. +/+ (A/A)		6.96	6.63	4.04
Expd. - /+ (A/G)		39.64	39.75	32.32
Expd. - /- (G/G)		56.40	59.62	64.64
Total no.		103	106	101
$\frac{(Obsd.-Expd)^2}{Expd.}$	+/+	0.60	0.02	0.95
	-/+	0.54	0.002	0.58
	-/-	0.12	0.006	0.09
Chi-square value		1.26	0.028	1.62
p, df= 1		0.30>p>0.25	0.90>p>0.80	0.25>p>0.20
No. of + or A (Freq.)		53 (0.26)	54 (0.25)	40 (0.20)
No. of - or G (Freq.)		153 (0.74)	158 (0.75)	162 (0.80)
Total no. of alleles		206	212	202
f^2	+/+	0.008	0.005	0.004
	-/+	0.116	0.144	0.078
	- /-	0.325	0.303	0.436
Total f^2		0.449	0.452	0.518
Pd		0.551	0.548	0.482
Heterozygosity (h)		0.345	0.377	0.319

Keys: Obsd. = Observed number
f = Observed frequency
Expd. = Expected number
p = Probability
df = Degree of freedom
Pd = Power of discrimination

For the Malay, Chinese, and Indian samples, the **h** values at the *Alu*I site were 0.345, 0.377, and 0.319, respectively. The Pd for the Malay, Chinese, and Indian samples were 0.551, 0.548, and 0.482, respectively. Pd shows the probability that two unrelated individuals drawn at random from the same breeding population have different genotypes at this *Alu*I site.

3.1.2 *Hga*I +/- assay

The PCR amplification with 31F and 31RsaI primers produced a 205 bp fragment. The 31RsaI primer with a base mismatch was used to force a point mutation into the DNA adjacent to the polymorphic site in the PCR product. This creates a restriction site for *Rsa*I since the use of *Hga*I is precluded by the cost. The resulting product was digested with *Rsa*I as described in Section 2.6.2.

Digestion will reveal a genotype based on the presence, absence, or presence and absence of the *Rsa*I sites, hence the *Hga*I sites, at position –220. The recognition sequence for *Rsa*I is GT^AC, where '^' represents the cleavage site. Cleavage is expected for -220G alleles. *Rsa*I will not cleave –220C alleles, leaving the product intact. The expected fragment sizes (band patterns) are:

i) a single band of 205 bp product, representing the –220C alleles in a homozygous state, -/-;

ii) a single band of 182 bp product, representing the –220G alleles in a homozygous state, +/+; and

iii) two bands of 205 and 182 bp, representing -220C/G alleles in a heterozygous state, +/-.

It was noted that for heterozygous samples in this assay, the lower band is less intense than the upper. This has been suggested to be due to uncleavable heteroduplexes (Neil and Jeffreys, 1993).

As expected, three types of banding patterns were detected after electrophoresis (Figure 10). The results of *Rsa*I (*Hga*I) genotype and allele analyses are given in Table 6.

Interestingly, the digestion of PCR products in this assay produced two additional banding patterns. In seven of the samples, two distinct patterns were obtained that did not correlate with the expected sizes from heterozygous and homozygous individuals. These samples were from the Malays and Indians, while no irregular pattern was obtained among the Chinese. Electrophoresis results show the existence of fragments shorter than 182 bp (Figure 10).

The nature of the pattern shift was investigated further by sequencing these samples (see 3.3 Sequencing).

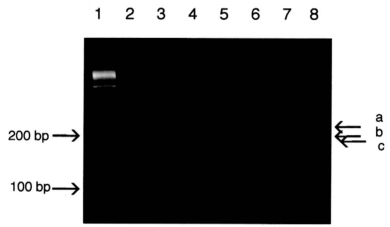

Figure 10: Ethidium bromide-stained 5% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:2) of *Hgal*+/- assay in MS31A 5' flanking DNA (-220 G/C polymorphisms).

Lane 1: 100 bp ladder
Lane 2:+/- (G/C) normal bands
Lane 3: +/+ (G/G) normal bands
Lane 4: -/- (C/C) normal bands
Lane 5: +/- (G/C) irregular bands
Lane 6: +/+ (G/G) irrregular bands
Lane 7: +/- (G/C) normal bands
Lane 8: Control (blank)
a) approximately 205 bp uncut product
b) approximately 182 bp cut product
c) slightly less than 182 bp cut product

Table 6: Distribution and frequencies of *Hgal* site genotypes and alleles in Malays, Chinese and Indians in the Malaysian population.

		Malays	Chinese	Indians
Obsd.+/+(G/G) (f)		40 (0.39)	51 (0.48)	29 (0.29)
Obsd.+/- (C/G) (f)		45 (0.44)	47 (0.44)	52 (0.51)
Obsd. - /- (C/C) (f)		18 (0.17)	8 (0.08)	20 (0.20)
Total no. of individuals		103	106	101
Expd. +/+ (G/G)		38.33	51.94	29.45
Expd. - /+ (C/G)		49.01	44.52	50.18
Expd. - /- (C/C)		15.66	9.54	21.37
Total no. of individuals		103	106	101
$\frac{(Obsd.-Expd)^2}{Expd.}$	+/+	0.07	0.02	0.007
	-/+	0.33	0.14	0.066
	-/-	0.35	0.25	0.088
Chi-square value		0.75	0.41	0.161
p, df= 1		0.50>p>0.30	0.70>p<0.50	0.70>p>0.50
No. of + or G (Freq.)		125 (0.61)	149 (0.70)	110 (0.54)
No. of - or C (Freq.)		81 (0.39)	63 (0.30)	92 (0.46)
Total no. of alleles		206	212	202
$(f)^2$	+/+	0.15	0.23	0.08
	-/+	0.19	0.20	0.27
	- /-	0.03	0.01	0.04
Total f^2		0.37	0.44	0.39
Pd		0.63	0.56	0.61
Heterozygosity (h)		0.48	0.42	0.50

Keys: Obsd. = Observed number
f = Observed frequency
Expd. = Expected number
p = Probability
df = Degree of freedom
Pd = Power of discrimination

The least common genotype was (-/-) with frequencies of 0.17, 0.08, and 0.20 for the Malays, Chinese, and Indians, respectively. The most common was (+/-), with frequencies ranging from 0.44 to 0.51. The frequencies of the G(+) allele were 0.61 for the Malays, 0.70 for Chinese and 0.54 for the Indians, while those of the C(-) allele were 0.39, 0.30 and 0.46 for the Malays, Chinese, and Indians, respectively. These results were consistent with those reported by Koh *et al.* (1994) for the Malaysian population.

On the assumption that the distribution of the allele frequencies followed the Hardy-Weinberg equilibrium, the expected number of individuals representing each genotype was calculated. Chi-square test results were not significant, consistent with previous observations in the Malaysian population (Koh *et al.*, 1994) and for the Caucasian and Japanese populations (Neil and Jeffreys, 1993).

For the Malay, Chinese, and Indian samples, the expected **h** at the *Hga*I site were 0.48, 0.42, and 0.50, respectively. The Pd for the Malay, Chinese, and Indian samples were 0.63, 0.56, and 0.61, respectively. Pd shows the probability that two unrelated individuals drawn at random from the same breeding population have different genotypes at this *Hga*I site.

3.1.3 *Psp*1406I +/- assay

PCR amplification was performed by using primers 31E and 31F. The resultant 406 bp products were digested with *Psp*1406I (Section 2.6.2) to reveal the genotype based on the presence, absence, or presence and absence of the *Psp*1406I site at position –108. The recognition sequence for *Psp*1406I is AA^CGTT, where '^' represents the

cleavage site. Cleavage is expected for –108C alleles. *Psp*1406I should not cleave –108T alleles, leaving the product intact. The expected fragment sizes (band patterns) are:

i) a single band of 406 bp, representing the –108T alleles in a homozygous state, -/-;

ii) two bands of 296 and 110 bp, representing the –108C alleles in a homozygous state, +/+; and

iii) three bands of 406, 296, and 110 bp, representing –108C/T alleles in a heterozygous state, +/-.

[Note: For homozygous (+/+) samples in this assay, a faint upper band might exist, owing to partial digestion.]

As expected, three types of banding patterns were detected from the samples analysed after electrophoresis (Figure 11). The result of *Psp*1406I genotype and allele analyses are given in Table 7.

The least common genotype was (+/+) with frequencies of 0.16, 0.18 and 0.07 for the Malays, Chinese, and Indians, respectively. The most common was (+/-), with frequencies ranging from 0.43 to 0.51. The frequencies of the C(+) allele were 0.38 for the Malays, 0.43 for the Chinese, and 0.29 for the Indians, while those of the T(-) allele were 0.62, 0.57, and 0.71 for the Malays, Chinese, and Indians, respectively.

Conformation of the population data to Hardy-Weinberg equilibrium expectations means that the genotype frequencies can be reliably calculated from allele frequencies, since the χ^2 values showed no significant differences for all three races.

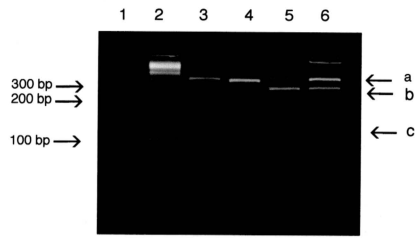

Figure 11: Ethidium bromide-stained 4% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:1) of *Psp*1406I +/- assay in MS31A 5' flanking DNA (-108 C/T polymorphisms).

Lane 1: Control (blank)
Lane 2: 100 bp ladder
Lane 3: +/- (C/T)
Lane 4: -/- (T/T)
Lane 5: +/+ (C/C) (with a faint 406 bp band)
Lane 6: +/- (C/T)
a) approximately 406 bp uncut product
b) approximately 269 bp cut product
c) slightly less than 110 bp cut product

Table 7: Disribution and frequencies of *Psp*1406I site genotypes and alleles in Malays, Chinese and Indians in the Malaysian population.

		Malays	Chinese	Indians
Obsd. +/+ (C/C) (f)		17 (0.16)	19 (0.18)	7 (0.07)
Obsd. +/- (C/T) (f)		44 (0.43)	54 (0.51)	44 (0.44)
Obsd. - /- (T/T) (f)		42 (0.41)	33 (0.31)	50 (0.49)
Total no. of individuals		103	106	101
Expd. +/+ (C/C)		14.80	19.97	8.32
Expd. - /+ (C/T)		48.48	52.08	41.34
Expd. - /- (T/T)		39.72	33.96	51.35
Total no. of individuals		103	106	101
$\frac{(Obsd.-Expd.)^2}{Expd.}$	+/+	0.33	0.05	0.21
	+/-	0.41	0.07	0.17
	-/-	0.13	0.03	0.04
Chi-square value		0.87	0.15	0.42
p, df= 1		0.50>p>0.30	0.70>p>0.50	0.70>p>0.50
No. of + or C(Freq.)		78 (0.38)	92 (0.43)	58 (0.29)
No. of - or T (Freq.)		128 (0.62)	120 (0.57)	144 (0.71)
Total no. of alleles		206	212	202
$(f)^2$	+/+	0.03	0.03	0.005
	+/-	0.18	0.26	0.19
	- /-	0.17	0.10	0.24
Total no. f^2		0.38	0.39	0.435
Pd		0.62	0.61	0.565
Heterozygosity (h)		0.473	0.494	0.411

Keys: Obsd. = Observed number
f = Observed frequency
Expd. = Expected number
p = Probability
df = Degree of freedom
Pd = Power of discrimination

For the Malay, Chinese, and Indian samples, the observed **h** at the *Psp*1406I site were 0.473, 0.494, and 0.411, respectively. The Pd for the Malay, Chinese, and Indian samples were 0.62, 0.61, and 0.565, respectively. Pd shows the probability that two unrelated individuals drawn at random from the same breeding population have different genotypes at this *Psp*1406I site.

3.2 Haplotype assays

Samples that were double heterozygous from the *Alu*I and *Hga*I assays and double heterozygous from the *Hga*I and *Psp*1406I assays were included in haplotype analyses to determine the phase for these three restriction endonuclease sites.

For *Alu*I-*Hga*I haplotype assay, primer 31HgaI+t was specifically designed to amplify –220G allele (allele specific primer). A fragment of 296 bp was produced when primers 31-TAG-A and 31HgaI+t were used for PCR of genomic DNA. The products were subsequently digested with *Alu*I. A positive cleavage of the PCR product indicates the presence of the -4A allele and a negative result the -4G allele. The genotypes at both *Alu*I and *Hga*I were determined for the samples selected for this assay. The samples having the -220G(+) allele (*Hga*I site) and the -4A(+) allele (*Alu*I site) were designated the (++/--) genotype, whilst those having the –220G(+) allele (*Hga*I site) and the -4G(-) allele (*Alu*I site) were designated the (+-/-+) genotype (Figure 12).

Similarly, haplotype analysis was performed for the *Hga*I and *Psp*1406I flanking polymorphism assays. PCR was performed by using the same primer set (31HgaI+t and 31-TAG-A), and the resulting products were digested with *Psp*1406I. A

positive cleavage indicates the presence of the -108C(+) allele and a negative result the –108T(-) allele (Figure 13).

The samples with the -220G(+) allele (*Hga*I site) and the –108C(+) allele (*Psp*1406I site) were denoted the (++/--) genotype, whilst those with the –220G(+) allele (*Hga*I site) and the –108T(-) allele (*Psp*1406I site) were given the (+-/-+) genotype (Figure 13).

3.2.1 *Hga*I- *Alu*I assay

The *Hga*I-*Alu*I haplotype assay was carried out on 26 double heterozygous individuals with *Alu*I(+/-) and *Hga*I(+/-) to determine the phase of the respective allelic pairs. The Chi-square 2x2 contingency test was performed to investigate the association between *Alu*I site and *Hga*I site polymorphisms (Table 8) for 310 individuals.

3.2.2 *Hga*I- *Psp*1406I assay

The *Hga*I-*Psp*1406I haplotype assay was carried out on 64 double heterozygous individuals with *Hga*I(+/-) and *Psp*1406I(+/-) to determine the phase of the respective allelic pairs. The Chi-square 2x2 contingency test was performed to investigate the association between *Hga*I site and *Psp*1406I site polymorphisms (Table 9) for 310 individuals.

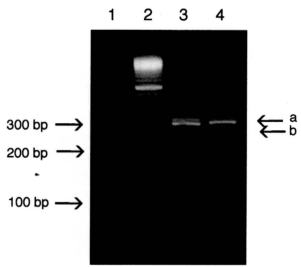

Figure 12: Ethidium bromide-stained 5% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:2) of *Hga*I-*Alu* I haplotype assay.

Lane 1: Control (blank)
Lane 2: 100 bp ladder
Lane 3: +-/-+
Lane 4: ++/--
a) approximately 296 bp uncut product
b) approximately 284 bp cut product

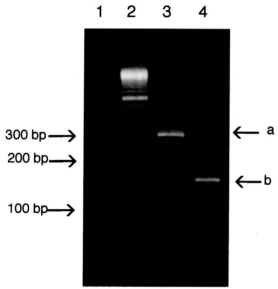

Figure 13: Ethidium bromide-stained 5% (w/v) agarose gel (NuSieve GTG: SeaKem LE, 3:2) of *Hga*I-*Psp*1406I haplotype assay.

Lane 1: Control (blank)
Lane 2: 100 bp ladder
Lane 3: +-/-+
Lane 4: ++/--
a) approximately 296 bp uncut product
b) approximately 150 bp cut product

Table 8 : 2x2 contingency test for *Hga*I-*Alu*I MS31A 5' flanking haplotype

	*Alu*I+	*Alu*I-	Total
*Hga*I+	103	286	389
*Hga*I-	31	200	231
Total	134	486	620

f	F	f-F
103 *Hga*I+ *Alu*I+	84.07	18.93
286 *Hga*I+ *Alu*I-	304.93	-18.93
31 *Hga*I- *Alu*I+	49.93	-18.93
200 *Hga*I- *Alu*I-	181.07	18.93

E.g., f = 103; F= (134)(389)/620 = 84.07; f-F = 18.93

Where, f = observed no.

F = expected no.

$$\chi 2 = (f\text{-}F)^2 \sum_{i=1}^{n} 1/F$$

$$= (358.3449)\,[1/84.07 + 1/304.93 + 1/49.93 + 1/181.07\,]$$

$$= 14.52$$

Analysis by the Chi-square 2x2 contingency test revealed that the χ^2 value was 14.52. This χ^2 value is greater than the χ^2 values at both 5 and 1% levels of significance, df=1. There is therefore a significant association between the variables; hence, the two restriction endonuclease sites were not independent of each other.

Table 9: 2x2 contingency test for *Hga*I-*Psp*1406I MS31A 5' flanking haplotype.

	*Psp*1406I+	*Psp*1406I-	Total
*Hga*I+	188	201	389
*Hga*I-	43	188	231
Total	231	389	620

f	F	f-F
188 *Hga*I+ *Psp*1406I+	144.93	43.07
201 *Hga*I+ *Psp*1406I-	244.07	-43.07
43 *Hga*I- *Psp*1406I+	86.07	-43.07
188 *Hga*I- *Psp*1406I-	144.93	43.07

E.g., f = 188; F = (231)(389)/ 620 = 144.93; f-F = 43.07

Where, f = observed no.

F = expected no.

$$\chi 2 = (f\text{-}F)^2 \sum_{i=1}^{n} 1/F$$

$$= (1855.0249)\ (1/144.93 + 1/244.07 + 1/86.07 + 1/144.93]$$

$$= 55.65$$

Analysis by the Chi-square 2x2 contingency test revealed that the χ^2 value was 55.65. This χ^2 value is greater than the χ^2 values at both 5 and 1% levels of significance, df=1. There is therefore a significant association between the variables; hence, the two restriction endonuclease sites were not independent of each other.

3.3 Sequencing

During the *Hga*I assay, normal and irregular bands were observed (Figure 10). DNAs from 11 individuals were randomly chosen for sequencing. The samples consisted of 2 normal heterozygous and 2 normal homozygous patterns as controls, 5 samples with irregular heterozygous pattern (+/-) and 2 samples with irregular homozygous pattern (+/+) (Figure 14).

3.3.1 Irregular heterozygous samples

Each of these samples shows one uncut PCR product and a cut PCR product of irregular size that was shorter than that from the control heterozygous sample (Figure 14). Direct sequencing of these samples revealed that a deletion of 12 nucleotides, from -230A to -241C, occurred in the irregular band. This explained the shorter PCR product after treatment with *Rsa*I. The sequencing results are shown in Figures 15 and 16.

3.3.2 Irregular homozygous samples

Results from gel electrophoresis showed two different *Rsa*I digested PCR products (Figure14). Normal homozygous (+/+) samples yielded one band only, whereas irregular homozygous samples yielded two bands that migrated close together. The absence of the 205 bp uncut product indicated that both alleles possess the *Rsa*I site. However, the presence of an additional (shorter) band implied that the cleaved products were present in two different sizes. These samples were also sequenced and results are shown in Figures 15 and 17.

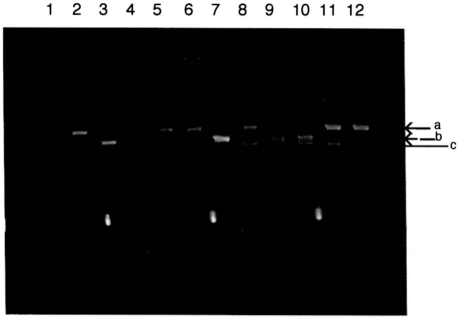

Figure 14: Samples chosen from *Hga*I assay for sequencing

Lane 1: Control (blank)
Lane 2: UH461 normal heterozygous +/- (positive control)
Lane 3: UH455 normal homozygous +/+
Lane 4: UH542 irregular heterozygous +/-
Lane 5: UH560 irregular heterozygous +/-
Lane 6: UH768 irregular heterozygous +/-
Lane 7: UH787 normal homozygous +/+
Lane 8: UH811 irregular heterozygous +/-
Lane 9: UH903 irregular homozygous +/+
Lane 10: UH956 irregular homozygous +/+
Lane 11: UH1264 irregular heterozygous +/-
Lane 12: UH461 normal heterozygous +/- (positive control)
a) approximately 205 bp uncut product
b) approximately 182 bp cut product
c)slightly less than 182 bp cut product

Numbers: Sample Reference Number.

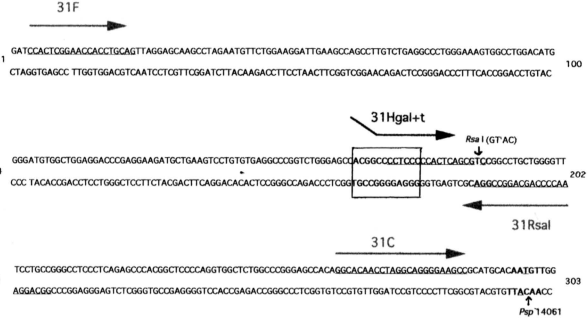

Fig.15: This figure shows the position of the 12 bp deletion (boxed, i.e., ACGGCCCCTCCC) that resulted in the irregular bands observed during the *Hga*I (*Rsa*I) assay of individuals UH542, UH560, UH768, UH811, UH903, UH956 and UH1264.

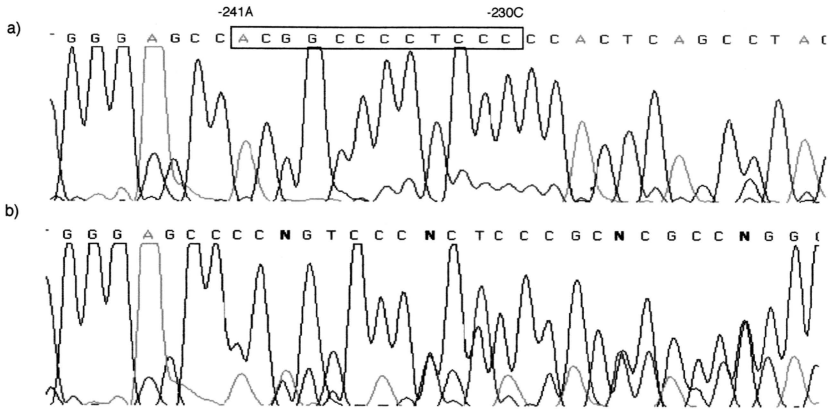

Figure 16: Comparison of base sequences between a normal heterozygous sample and an irregular heterozygous sample from -218A to -248G, showing the position of deletion (base positions are according to Neil and Jeffreys, 1993).
a) An example of automated sequencing chromatogram of normal heterozygous sample, UH461
b) An example of automated sequencing chromatogram of irregular heterozygous sample, UH560. The unclear sequence was due to overlapping of normal and irregular sequences.

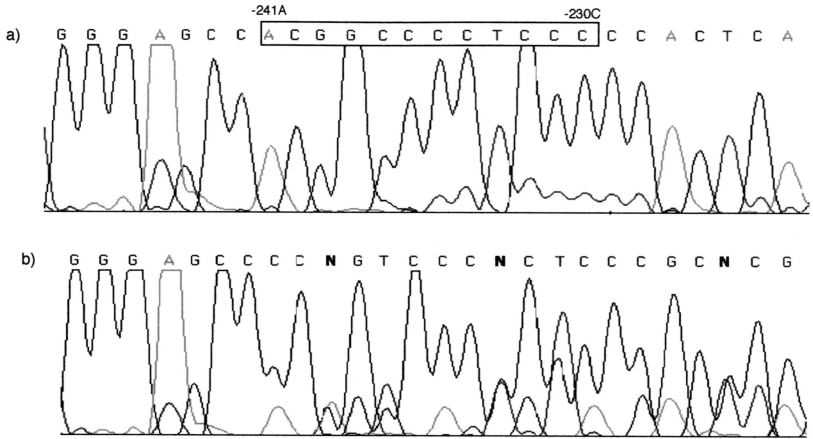

Figure 17: Comparison of base sequences between a normal heterozygous sample and an irregular homozygous sample from -223A to -248G, showing the position of deletion (base positions are according to Neil and Jeffreys, 1993).
a) An example of automated sequencing chromatogram of normal heterozygous sample, UH461
b) An example of automated sequencing chromatogram of irregular homozygous sample, UH956. The unclear sequence was due to overlapping of normal and irregular sequences.