Hum Genet (1992) 90:443–449

human genetics

Analysis of β-globin gene haplotypes in Asian Indians: origin and spread of β-thalassaemia on the Indian subcontinent

Nermeen Y. Varawalla, Alison C. Fitches, John M. Old

Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

Received: 6 February 1992 / Revised: 29 April 1992

Abstract. β-globin gene haplotypes were determined for 196 normal (β-A) and 419 thalassaemia (β-Th) chromosomes of individuals from four different regions of the Indian subcontinent; North-west Pakistan, Gujarat, Punjab and Sindh. Analysis of β-A and β-Th haplotypes and haplotype-mutation associations in each regional group along with a consideration of Indian history provided information about the origin and spread of β-thalassaemia mutations on the Indian subcontinent. The data are consistent with relatively recent and local origins for most β-thalassaemia mutations. The frequencies of particular alleles differ markedly in various regions and these may be useful population markers. Of the high frequency alleles, intervening sequence 1 (IVS-1) nucleotide 5 (G-C) and codons 41/42 (-CTTT) appear to be older as suggested by multiple haplotype associations and a widespread geographical distribution. The microepidemiology of β-thalassaemia in this region reflects considerable ethnic diversity, gene flow from population migration and natural selection by malaria infection.

Introduction

The ethnic composition of the Asian Indian population is varied and complex with six main races described in prehistoric India, before the arrival of the Aryan-speaking tribes from Iran into the north-western region of the Indian subcontinent in 1500 B.C. (Thapar 1966). This is refelcted in the ethnographic complexity of the population that today inhabits the Indian subcontinent. The strict endogamy required by the caste system has perpetuated local genetic diversity. This extreme and long-standing genetic heterogeneity is a distinctive feature of the Asian Indian population and to a certain extent accounts for the uneven and variable distribution of various genetic markers such as blood group antigens (Mourant et al. 1983) as well as the haemoglobinopathies in this population (Brittenham 1983). β-Thalassaemia, which is widely prevalent in the malaria infested areas of the world, appears to be the most common monogenic disease on the Indian subcontinent. The incidence of the β-thalassaemia trait is not uniform in India and varies from 1%–15% in different regions (Sukumaran 1974). There are few published gene-frequency data for Pakistan, Bangladesh and Sri Lanka but anecdotal reports suggest a high incidence. At the molecular level the disease is heterogeneous with 16 different β-thalassaemia mutations identified and 5 common ones accounting for 93% of β-thalassaemia alleles (Varawalla et al. 1991a, b).

Analysis of restriction fragment length polymorphisms (RFLPs) in the globin gene cluster has been useful for determining the origin and spread of the thalassaemia and for defining the relationships of several human populations (Wainscoat et al. 1986; Hill et al. 1989). Since β-thalassaemia is widely prevalent and heterogeneous in the Asian Indian population we have carried out a large survey of β-globin gene haplotypes in disease carriers to study the origin and spread of this disease on the Indian subcontinent. Analysis of 615 β-globin gene haplotypes reveals clear evidence of local genetic differentiation, allows inferences to be made about the origin of particular β-thalassaemia mutations, and suggests that certain globin gene variants may be useful markers for anthropological studies in this region.

Subjects and methods

Subjects

Population sample. The population sample consisted of 708 unrelated carriers of β-thalassaemia originating from different parts of the Indian subcontinent. Twenty millilitres of whole blood was collected from each of these subjects along with their family members. In all 470 individuals were interviewed and bled in India and the blood samples were transported to Oxford. The remaining 238 were immigrants living in the U.K. DNA was extracted from the EDTA-anticoagulated blood samples by phenol extraction (Old and Higgs 1983). The β-thalassaemia mutation of each of these heterozygotes was identified using allele specific primers in the polymerase chain reaction (PCR) and direct DNA sequencing (Varawalla et al. 1991a, b).

Correspondence to: N. Y. Varawalla

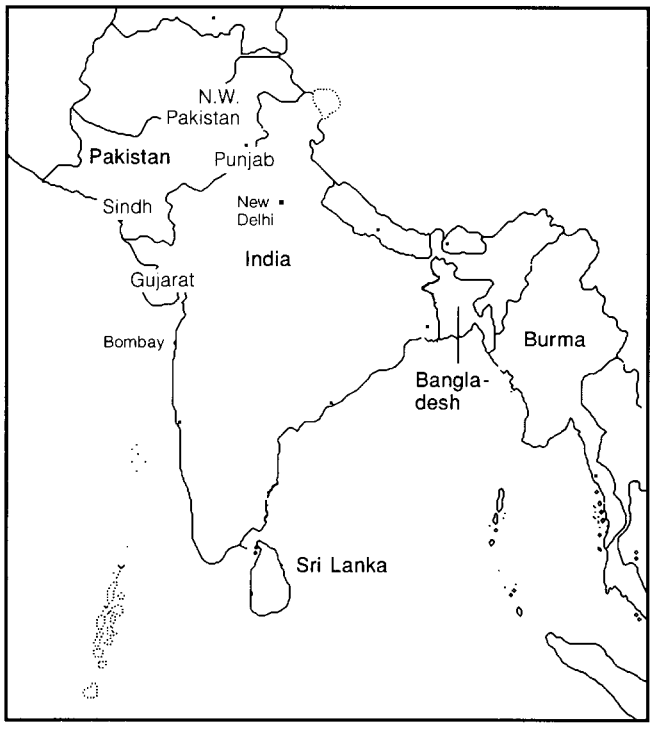

Fig. 1. Map of the Indian subcontinent showing the countries, regions and cities referred to in the text

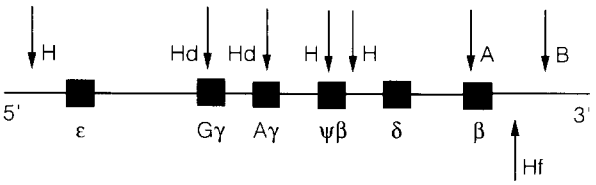

Fig. 2. Schematic representation of the β-globin gene cluster showing the restriction enzyme cutting sites used for haplotype construction. *H*, *Hin*dII; *Hd*, *Hin*dIII; *A*, *Ava*II; *B*, *Bam*HI; *Hf*, *Hin*fI. The *Hin*fI (Hf) site was examined instead of the *Bam*HI site, when the haplotypes were constructed by the polymerase chain reaction (PCR)

Regional origins. The regional and ethnic origins of each individual were ascertained by interview. Almost all of them originated from North-west Pakistan, Gujarat, Punjab and Sindh (Fig. 1), which are geographical regions not corresponding to political boundaries. The restriction of subjects studied to these four regions reflects in part the higher prevalence of β-thalassaemia in these regional groups compared with the subcontinent as a whole and in part our sampling procedure. The individuals living in India originated predominantly from the urban areas of Bombay and New Delhi: population movements in 1947 brought substantial numbers of β-thalassaemia carriers to Bombay from Sindh and into the area around New Delhi from areas of Punjab, which then became part of Pakistan. The immigrant Asian Indian population in the U.K. is comprised mainly of individuals originating from Pakistan and Punjab along with the Gujaratis who had initially settled in East Africa before migrating to the U.K.

Methods

Southern blot analysis. Seven RFLP sites in the β-globin gene cluster: *Hin*dII-ε, *Hin*dIII-Gγ, *Hin*dIII-Aγ, *Hin*dII-5′ψβ, *Hin*dII-3ψβ, *Ava*II-β and *Bam*HI-β (Fig. 2) were studied; the combination of these polymorphic restriction sites constituted the β-globin gene haplotype for that chromosome. In the earlier part of this study this was done by Southern blotting and hybridisation (Old and Higgs 1983) using genomic DNA probes; the probes used were 1.3-kb *Bam*HI/*Eco*RIε, 3.2-kb *Hin*dIIIγ, 3.9-kb *Pst*Iψβ and a 4.4 kb *Pst*Iβ.

Restriction enzyme digestion of PCR product. Later, the β-globin gene haplotypes were determined more simply and rapidly by PCR. Fragments of the β-globin gene cluster that contained the polymorphic restriction sites were amplified and subjected to restriction enzyme digestion. The RFLPs studied were those listed above with the exception of the *Bam*HI site. It is difficult to amplify the region around the *Bam*HI polymorphic site as it lies within a repetitive sequence. Hence the *Hin*fI polymorphism located 3′ to the β-globin gene, which is in complete linkage disequilibrium with the *Bam*HI polymorphism (Semenza et al. 1989), was studied instead. The oligonucleotide primers used for the amplification of the fragments containing the restriction sites are listed in Table 1, along with the fragment sizes following restriction enzyme digestion in the presence and absence of the restriction site. The primers used to amplify the *Hin*fI-β site were designed so that they avoided the region of the 619-bp deletion that is prevalent in Asian Indians.

The PCR was performed using approximately 0.5 μg of genomic DNA, in a 25-μl volume containing 20 pmol of each primer, 800 μ*M* total dNTPs, 10 m*M* TRIS-HCl, ph 8.4, 50 m*M* KCl, 1.5 m*M* $MgCl_2$ and 0.5 units *Taq* polymerase (AmpliTaq, Cetus). The thermal cycling regimen consisted of 30 cycles, with denaturation at 93°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1.5 min. In the final cycle extension was prolonged to 3 min. Amplification of the DNA fragment encompassing the *Hin*dIII-Gγ site, was achieved using an annealing temperature of 65°C. The amplified products were subjected to restriction enzyme digestion according to the manufacturer's recommendations. Following digestion, a 10-μl aliquot of the PCR product was examined by electrophoresis in a 1.5% agarose and 1.5% NuSieve (I.C.N. Biomedicals) gel, which was then stained with ethidium bromide and visualised by ultraviolet transillumination. The presence or absence of the polymorphic restriction site was determined from the size of the digested PCR product.

Haplotype construction. The seven polymorphic restriction enzyme cutting sites analysed were clustered in two groups: the 34-kb region 5′ to the δ gene contained five restriction sites and two sites were present within the β-globin gene and 3′ to it (Fig. 2), consituting 5′ and 3′ β-globin gene haplotypes. From these results the β-globin gene haplotypes were constructed for both the β-A and β-Th chromosomes (Antonarakis et al. 1982). In order to do this it was necessary to analyse some RFLPs in affected offspring or other family members. In 289 subjects a definite assignment of haplotypes was not possible owing to heterozygosity for a particular RFLP and nonavailability of family members. These subjects were excluded from this study.

Results

Haplotype analysis of the whole population sample

Amongst 196 normal Asian Indian chromosomes, 19 different β-globin gene haplotypes were identified (Table 2). Of the 32 (2^5) possible 5′ haplotypes, 8 were present and 3 5′ haplotypes, +−−−− (51%), −+−++ (20%) and −++−+ (19%), accounted for 90% of β-A chromosomes. Of the 3′ haplotypes, both the *Ava*II-β and *Bam*HI-β sites were present (framework 1) (Orkin et al. 1982) in 97 (50%) chromosomes, in 31 (16%) the *Bam*HI-

Table 1. Oligonucleotide primers used for analysis of restriction fragment length polymorphisms (RFLPs) by PCR. IVS-2, Intervening sequence 2

Restriction site	Primer sequence	Position of 5′ nucleotide with respect to the gene	Product size	Absence of site (−)	Presence of site (+)
ε-*Hin*dII	5′-TCTCTGTTTGATGACAAATTC	890 bp upstream from the first codon	760 bp	760 bp	446 and 314 bp
	5′-AGTCATTGGTCAAGGCTGACC	130 bp upstream from the first codon			
Gγ-*Hin*dIII	5′-AGTGCTGCAAGAAGAACAACTACC	IVS-2 position 709	328 bp	328 bp	91 and 237 bp
	5′-CTCTGCATCATGGGCAGTGAGCTC	22 bp downstream from termination site			
Aγ-*Hin*dIII	5′ATGCTGCTAATGCTTCATTAC	IVS-2 position 453	635 bp	635 bp	327 and 308 bp
	5′-TCATGTGTGATCTCTCTCAGCAG	93 bp downstream from termination site			
5′ψβ-*Hin*dII	5′-TCCTATCCATTACTGTTCCTTGAA	IVS-2 position 541	794 bp	794 bp	104 and 690 bp
	5′-ATTGTCTTATTCTAGAGACGATTT	357 bp downstream from termination site			
3′ψβ-*Hin*dII	5′-GTACTCATACTTTAAGTCCTAACT	2,431 bp downstream from termination site	914 bp	914 bp	435 and 479 bp
	5′-TAAGCAAGATTATTTCTGGTCTCT	3,345 bp downstream from termination site			
β-*Ava*II	5′-GTGGTCTACCCTTGGACCCAGAGG	Codon 33	328 bp	328 bp	228 and 100 bp
	5′-TTCGTCTGTTTCCCATTCTAAACT	IVS-2 position 112			
β-*Hin*fI	5′-AGTTAGAGGCTTGATTTGGAGG 5′-GTTAAGGTGGTTGATGGTAAC	368 bp downstream termination site 1,006 bp downstream from termination site	638 bp	336 bp constant fragment = 302 bp	123 and 213 bp

β site was absent (framework 2) and in 68 (35%) the *Ava*II-β site was absent (framework 3).

A total of 12 different β-globin gene haplotypes were identified in the 419 β-Th chromosome studied. With the exception of 5 (1%) chromosomes their 5′ haplotype was either +−−−− (78%), −+−++ (13%) or −++−+ (8%). The distribution of the 3′ haplotypes was different from that of the β-A chromosomes: 148 (35%) β-Th chromosomes had framework 1, 46 (11%) framework 2 and 225 (54%) framework 3. Three common haplotypes +−−−− −+, +−−−− ++ and −+−++ ++ accounted for 80% of the β-Th chromosomes.

The β-Th haplotypes were found to be less diverse than those of β-A. With the single exception of +−+−+ −+, all the β-Th chromosomes were represented amongst the β-A haplotypes. Further, the frequency of each β-Th haplotype was similar to that of its β-A counterpart. However, the commonest haplotype, +−−−− −+, occurred at twice the frequency amongst β-thalassaemia chromosomes (46%) as compared with normal ones (22%).

Strong linkage disequilibrium between the 15 β-thalassaemia mutations present and the β-globin gene haplotypes was observed (Table 3). However, the association between the mutations and haplotypes was not absolute. With the exception of some of the rare mutations, all were found to be associated with more than one haplotype. The association between the five most common mutations and their major haplotype ranged from 77% to 99%. Table 3 also shows the extent of association between each of the five most common β-Th haplotypes and their predominant mutation, thus by haplotype analysis it would be possible to predict the β-thalassaemia mutation with 60% to 86% accuracy. Six β-thalassaemia mutations were found to be present on two different frameworks.

The −+−++ and the ++−++ 5′ haplotypes have been found to be strongly linked to an *Xmn*I restriction site at −158 Gγ, which is associated with increased HbF production (Thein et al. 1987). The −+−++ 5′ haplotype was found in 53 out of 419 (12.7%) β-Th chromosomes, including 33 out of 36 (92%) chromosomes carrying the intervening sequence 1 (IVS-1) nucleotide 1 (G-T) mutation and the ++−++ 5′ haplotype was not found in a single β-Th chromosome. In contrast 47 out of the 196 (24.6%) β-A chromosomes studied had these haplotypes: 39 with −+−++ and 8 with ++−++.

Haplotype analysis in different regions

Of the 419 β-Th chromosomes haplotyped, 404 originated from either North-west Pakistan, Gujarat, Punjab or Sindh; correspondingly 187 of the 196 β-A chromosomes

Table 2. β-A and β-Th haplotypes in Asian Indians

Haplotype	β-A (%)	β-Th (%)
+−−−− ++	41 (21)	95 (23)
−+−++ ++	25 (13)	44 (10)
−++−+ ++	21 (11)	9 (2)
−++−− ++	3 (2)	–
−+−−+ ++	1 (0.5)	–
++−++ ++	4 (2)	–
−−−−− ++	1 (0.5)	1 (0.5)
+++−+ ++	1 (0.5)	–
+−−−− +−	15 (8)	38 (9)
−+−++ +−	4 (2)	5 (1)
−++−+ +−	5 (3)	3 (1)
−++−− +−	2 (1)	–
−+−−+ +−	1 (0.5)	–
++−++ +−	4 (2)	–
+−−−− −+	44 (22)	194 (46)
−+−++ −+	10 (5)	4 (1)
−++−+ −+	12 (6)	22 (5)
−++−− −+	1 (0.5)	–
−+−−+ −+	1 (0.5)	3 (1)
+−+−+ −+	–	1 (0.5)
Total	196	419

originated from one of these regions as well. The distribution of common β-A and β-Th haplotypes, i.e. those present in the population at a frequency of over 3%, are shown in Table 4a and the regional distribution of the five common β-thalassaemia mutations in this population sample is shown in Table 4b.

In each region the β-A haplotypes were found to be more diverse, i.e. the common haplotypes accounted for a smaller proportion of β-A haplotypes, as compared with β-Th chromosomes. With the exception of the −+−++ −+ β-Th haplotype in North-west Pakistan, each of the common β-Th haplotypes were represented amongst their normal counterparts in that particular region. The predominant β-Th haplotypes varied in each region. Further, the predominant β-Th haplotype in North-west Pakistan (+−−−− ++), Gujarat and Sindh (+−−−− −+) was represented at a substantially greater frequency amongst β-Th chromosomes as compared with the β-A ones in that region. The predominant β-Th haplotype in each region was the one that was most closely associated with the regions predominant mutations. In North-west Pakistan, the predominant haplotype, +−−−− ++, is associated with the codons 8/9 (+G) mutation, which account for 48% of β-thalassaemia mutations in this region. The predominant haplotype in Gujarat and Sindh, +−−−− −+, is associated with the IVS-1 nucleotide 5 (G-C) mutation and the 619 bp deletion, which are the two commonest mutations present in these neighbouring regions. In Punjab, the observations were slightly different, the +−−−− +− and −++−+ −+ β-Th haplotypes, which are associated with the codons 41/42 (-CTTT) and IVS-1 nucleotide 5 (G-C) mutations, respectively, occur at almost three times the frequency amongst β-Th chromosomes as compared with β-A ones. Amongst the four regions analysed, the codons 41/42 (-CTTT) mutation was most frequent in Punjab and 10 of the 19 −++−+ −+ β-Th haplotypes that were associated with the IVS-1

Table 3. Linkage disequilibrium between β-thalassaemia mutations and β-globin gene haplotypes. n, Nucleotide; c, codon; bp, base pair; del, deletion; IVS-1–1, IVS-1 minus 1

Haplotype	Mutation															
	IVS-1n5 (G−C)	C 8/9 (+G)	619 bp del	IVS-1n1 (G−T)	c 41/42 (−CTTT)	c 15 (G−A)	c 5 (−CT)	IVS-1−1 (G−C)	c 16 (−C)	IVS-2n-837 (T−G)	Cap site +1 (A−C)	IVS-1−1 (G−A)	IVS-2n1 (G−A)	−88 (C−T)	C 88 (+T)	Linkage with main mutation (%)
+−−−− ++	4	81		1	2	2		3				2				85
−+−++ ++	1	9		33				1								75
−++−+ ++					1	6		2								
+−−−− +−					27		4		3	3				1		71
−+−++ +−					3		1						1			
−++−+ +−					2										1	
+−−−− −+	117	2	71	2		1					1					60
−+−++ −+	3					1										
−++−+ −+	19					1	1				1					86
−+−−+ −+	2		1													
+−+−+ −+	1															
−−−−− −+	1															
Total	148	92	72	36	35	11	6	6	3	3	2	2	1	1	1	
Linakge with major haplotype (%)	78	88	99	92	77											

Table 4. Regional distribution of the common β-A and β-Th haplotypes on the Indian subcontinent, as well as the distribution of β-thalassaemia mutations

a Common β-A and β-Th haplotypes on the Indian subcontinent

Haplotype	North-West Pakistan		Gujarat		Punjab		Sindh	
	β-A (%)	β-Th (%)	β-A (%)	β-Th (%)	β-A (%)	β-Th (%)	β-A (%)	β-Th (%)
+−−−− ++	18 (25)	58 (45)	8 (17)	11 (9)	7 (19)	14 (17)	8 (36)	10 (15)
+−−−− +−	8 (10)	14 (11)	2 (4)	8 (6)	2 (5)	12 (14)	2 (9)	1 (1)
+−−−− −+	15 (19)	36 (28)	11 (23)	77 (62)	13 (36)	32 (37)	2 (9)	40 (60)
−+−++ ++	9 (10)	10 (8)	6 (13)	13 (11)	5 (14)	8 (9)	4 (18)	13 (19)
−+−++ −+	−	2 (2)	6 (13)	−	1 (3)	1 (1)	3 (14)	−
−++−+ ++	9 (12)	4 (3)	6 (13)	4 (3)	3 (6)	1 (1)	1 (5)	−
−++−+ −+	6 (7)	4 (3)	4 (8)	6 (5)	2 (5)	12 (14)	−	−
Total no. studied	81	128	48	124	36	85	22	67

b β-Thalassaemia mutations[a]

Mutation	North-West Pakistan, no. (%)	Gujarat, no. (%)	Punjab, no. (%)	Sindh, no. (%)	Whole population, no. (%)
IVS-1n5	40 (31)	53 (43)	33 (40)	12 (18)	148 (35)
c 8/9	62 (48)	7 (6)	15 (18)	7 (10)	92 (22)
619 bp del	−	29 (23)	11 (13)	32 (48)	72 (17)
IVS-1n1	2 (2)	16 (13)	7 (8)	11 (16)	36 (9)
c 41/42	11 (9)	11 (9)	10 (12)	3 (4)	35 (8)

[a] Some of these results have been published previously

nucleotide 5 (G-C) mutation were from Punjabi carriers.

There were no differences in mutation-haplotype associations amonst the four regions studied. A difference in the regional distribution of mutations was present, but these mutations remained linked with the same haplotypes irrespective of the regional origin.

Discussion

The similarity of β-Th and β-A haplotypes on the Indian subcontinent as a whole and in the regional groups suggests that the β-thalassaemia mutations have arisen relatively recently on chromosomal backgrounds already existing in this population. The predominant β-Th haplotype occurred at a substantially greater frequency than its normal counterpart on the Indian subcontinent as a whole and within each of the regions studied. On the Indian subcontinent, the +−−−− −+ haplotype, associated with the commonest and most widespread β-thalassaemia mutation, IVS-1 nucleotide 5 (G-C), occurred in 46% of β-Th chromosomes and in only 22% of normal ones. The same phenomenon was observed in each genetically isolated regional "micro-population", an outstanding example being the presence of the +−−−− −+ haplotype in 60% of β-Th chromosomes in Sindh and in only 9% of its β-A chromosomes. This greater frequency of the predominant β-Th haplotype as compared with its normal counterpart occurred irrespective of its associated β-thalassaemia mutation; the codons 8/9 (+G) mutation in North-west Pakistan associated with the +−−−− ++ haplotype, the codons 41/42 (-CTTT) mutation in Punjab associated with the +−−−− +− haplotype and the IVS-1 nucleotide 5 (G-C) mutation in Gujarat and Sindh associated with the +−−−− −+ haplotype all having the same effect. These observations are consistent with the hypothesis that chromosomes carrying β-thalassaemia mutations experienced positive selection pressure, probably because of the protection against malaria experienced by β-thalassaemia carriers. Although endogamy is common, genetic drift is unlikely to contribute substantially in populations of this size (e.g. Gujarat has a population of about 60 million).

Considerable variation in the regional distribution of β-thalassaemia mutations is present on the Indian subcontinent (Varawalla et al. 1991a). The codons 8/9 (+G) mutation predominates in North-west Pakistan, the 619 bp deletion in Sindh and the IVS-1 nucleotide 1 (G-T) mutation is restricted to the neighbouring regions of Gujarat, Sindh and Punjab. These mutations could serve as useful anthropological markers for local populations. For example, the strong linkage disequilibrium between the 619 bp deletion and +−−−− −+ haplotype and the geographical distribution of this mutation suggests that the 619 bp deletion arose in Sindh and subsequent spread has been by gene flow through population migration.

The strong linkage disequilibrium between the 619 bp deletion, IVS-1 nucleotide 1 (G-T) and codons 8/9 (+G) mutations and their β-globin gene haplotypes as well as their localised distribution suggests a relatively recent origin of these mutations. In contrast, the IVS-1 nucleotide 5 (G-C) and the codons 41/42 (-CTTT) mutations have a more widespread distribution and are less closely

associated with their major haplotype. The great diversity of haplotypes associated with the IVS-1 nucleotide 5 (G-C) mutation, its high frequency and widespread distribution suggest that it may be the oldest β-thalassaemia mutation on the Indian subcontinent.

The explanations for the association of a particular mutation with more than one haplotype are that the mutation has arisen independently on multiple occasions, as has been postulated for the codon 39 (C-T) mutation (Pirastu et al. 1987) or there have been recombination events. We have found six mutations associated with different β-globin gene frameworks, thus these mutations in particular, may have arisen more than once, as has been argued for the sickle cell mutation in Africa (Chebloune et al. 1988). However, in spite of significant differences in the regional distribution of β-thalassaemia mutations, no difference in haplotype-mutation associations amongst the different regions was found: this argues against multiple independent origins of these mutations in the different regions. Hence, the finding of multiple haplotype associations may have simply resulted from interallelic gene conversion events, which have been well documented in globin genes (Slightom et al. 1980).

Throughout its recorded history the Indian subcontinent has had political, military and commercial interactions with Central, Western and South-East Asia and later Europe (Thapar 1966), which are very likely to have resulted in gene flow resulting in the spread of some β-thalassaemia alleles. The IVS-2 nucleotide 1 (G-A) mutation was identified in an individual from Punjab on the same haplotype as has been described in Turkey and Northern Cyprus (Diaz-Chico et al. 1988), which could be explained by the Turkish invasion of Northern India in the 11th century (Thapar 1966). Similarly, the codon 5 (-CT) mutation has been described in Greeks on a haplotype identical to that found in this population (Kollia et al. 1989), also suggesting a common origin with subsequent gene flow. In Indonesia (Lie-Injo et al. 1989), Burma (Brown et al. 1992) and Malaysia (Yang et al. 1989) the IVS-1 nucleotide 5 (G-C), codons 41/42 (-CTTT) and IVS-1 nucleotide 1 (G-T) mutations have been described in association with haplotypes similar to those found in β-Th chromosomes carrying these mutations on the Indian subcontinent. Also in Thailand, the IVS-1 nucleotide 5 (G-C) mutation has been described in association with the +−−−− −+ haplotype (Laig et al. 1989). As early as the 6th century B.C., Indian sea-traders have been exploring the coasts of Burma, the Malay peninsular and western Indonesia offering possible routes for the spread of these alleles. However, the codons 41/42 (-CTTT) and the IVS-1 nucleotide 5 (G-C) mutations have also been described in other population groups associated with different haplotypes and frameworks. The codons 41/42 (-CTTT) mutation is present in the Chinese and Thais on two different frameworks, +−−−− ++ and +−−−− −+ (Kazazian et al. 1986; Laig et al. 1989) while the IVS-1 nucleotide 5 (G-C) mutation has been described in association with the +−−−− ++ haplotype in Chinese (Kazazian et al. 1986) and Melanesians (Hill et al. 1988) and with the −+−−+ −+ haplotype in Lebanon (Chehbab et al. 1987). These observations suggest that these two mutations may have had more than one independent origin.

Further detailed microepidemiological analysis of β-thalassaemia on the Indian subcontinent including the isolated tribal groups should be of value in providing evidence of historical gene flow in this region and in defining local population relationships.

Acknowledgements. We wish to thank Professor Sir D. J. Weatherall for his support. Dr. A. V. S. Hill for helpful advice with the manuscript, and Dr. P. G. Natrajan and Professor I. C. Verma for their assistance with the collection of blood samples for analysis. N. Y. V. is a Rhodes Research Fellow at Somerville College, Oxford.

References

Antonarakis SE, Boehm CD, Giardina PJ, Kazazian HH (1982) Nonrandom association of polymorphic restriction sites in the beta globin cluster. Proc Natl Acad Sci USA 79:137–141

Brittenham GM (1983) The geographic and ethnographic distribution of haemoglobinopathies in India. In: Bowman JE (ed) Distribution and evolution of haemoglobin and globin loci. Elsevier, Amsterdam, pp 169–178

Brown JM, Thein SL, Weatherall DJ, Mar KM (1992) The spectrum of β-thalassaemia in Burma. Br J Haematol 81:574–578

Chebloune Y, Pagnier J, Trabuchet G, Faure C, Verdier G, Labie D, Nignon V (1988) Structural analysis of the 5′ flanking region of the β-globin gene in African sickle cell anemia patients: further evidence for three origins of the sickle cell mutation in Africa. Proc Natl Acad Sci USA 85:4431–4435

Chehbab FF, Der Kaloustian V, Khouri FP, Deeb SS, Kan YW (1987) The molecular basis of β-thalassemia in Lebanon: application to prenatal diagnosis. Blood 69:1141–1145

Diaz-Chico JC, Yang KG, Stoming TA, Efremov DG, Kutlar A, Kutlar F, Aksoy M, Altay C, Gurgey A, Kininc Y, Huisman THJ (1988) Mild and severe β-thalassemia among homozygotes from Turkey: identification of the types by hybridisation of amplified DNA with synthetic probes. Blood 71:248–251

Hill AVS, Bowden DK, O'Shaughnessy DF, Weatherall DJ, Clegg JB (1988) β-Thalassemia in Melanesia: association with malaria and characterization of a common variant (IVS-1 nt 5 G-C). Blood 72:9–14

Hill AVS, O'Shaughnessy DF, Clegg JB (1989) Haemoglobin and globin gene variants in the Pacific. In: Hill AVS, Serjeantson SW (eds) The colonization of the Pacific: a genetic trail. Clarendon Press, Oxford, pp 246–285

Kazazian HH, Dowling CE, Waber PG, Huang S, Wilson HYL (1986) The spectrum of β-thalassemia genes in China and Southeast Asia. Blood 68:964–966

Kollia P, Gonzalez-Redondo JM, Stoming TA, Loukopoulos D, Politis C, Huisman THJ (1989) Frameshift codon 5 [FSC-5(-CT)] thalassemia: a novel mutation detected in a Greek patient. Hemoglobin 13:597–604

Laig M, Sanguansermsri T, Wiangnon S, Hundrieser J, Pape M, Flatz G (1989) The spectrum of β-thalassaemia mutations in northern and northeastern Thailand. Hum Genet 84:47–50

Lie-Injo LE, Cai SP, Wahidijat S, Moeslichan S, Lim ML, Evangalista L, Doherty M, Kan YW (1989) β-Thalassemia mutations in Indonesia and their linkage to β haplotypes. Am J Hum Genet 45:971–975

Mourant AE (1983) Asia. In: Mourant AE (ed) Blood relations, blood groups and anthropology. Oxford University Press, Oxford, pp 57–72

Old JM, Higgs DR (1983) Gene analysis. In: Weatherall DJ (ed) Methods in haematology: the thalassaemias. Churchill Livingstone, Edinburgh, pp 74–96

Orkin SH, Kazazian HH, Antonarakis SE, Goff SC, Boehm CD, Sexton JP, Waber PG, Giardina PJV (1982) Linkage of β-thal-

assaemia mutations and β-globin gene polymorphisms with DNA polymorphisms in the human β-globin gene cluster. Nature 296:627–631

Pirastu M, Galanello R, Doherty MA, Tuveri T, Cao A, Kan YW (1987) The same β-globin gene mutation is present on nine different β-thalassemia chromosomes in a Sardinian population. Proc Natl Acad Sci USA 84:2882–2885

Semenza GL, Dowling CE, Kazazian HH (1989) *Hin*fI polymorphism 3′ to the human β-globin gene detected by the polymerase chain reaction (PCR). Nucleic Acids Res 17:2376

Slightom JL, Blechl AE, Smithies O (1980) Human fetal Gγ- and Aγ- globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21:627–638

Sukumaran PK (1974) Abnormal haemoglobins in India. In: Sen NN (ed) Trends in haematology. Calcutta, pp 225–261

Thapar R (1966) A history of India, vol 1. Penguin, Harmondsworth

Thein SL, Wainscoat JS, Sampietro M, Old JM, Cappellini D, Florelli G, Modell B, Weatherall DJ (1987) Association of thalassaemia intermedia with a beta-globin gene haplotype. Br J Haematol 65:367–373

Varawalla NY, Old JM, Sarkar R, Venkatesan R, Weatherall DJ (1991a) The spectrum of β-thalassaemia mutations on the Indian subcontinent: the basis for prenatal diagnosis. Br J Haematol 78:242–247

Varawalla NY, Old JM, Weatherall DJ (1991b) Rare β-thalassaemia mutations in Asian Indians. Br J Haematol 79:640–644

Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JM, Lynch JR, Falusi AG, Weatherall DJ, Clegg JB (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature 319:491–493

Yang KG, Kutlar F, George E, Wilson JB, Kutlar A, Stoming TA, Gonzalez-Redondon JM, Huisman THJ (1989) Molecular characterization of β-globin gene mutations in Malay patients with HbE-β-thalassaemia and thalassaemia major. Br J Haematol 72:73–80