Gene 521 (2013) 287–292

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Distribution of beta-globin haplotypes among the tribes of southern Gujarat, India

Aastha Aggarwal [a], Priyanka Khurana [a], Siuli Mitra [a], Bhavesh Raicha [b], K.N. Saraswathy [a], Yazdi M. Italia [b], Gautam K. Kshatriya [a,*]

[a] Department of Anthropology, University of Delhi, Delhi 110007, India
[b] Valsad Raktdan Kendra, R.N.C. Free Eye Hospital Complex, Kacheri Road, Valsad, Gujarat 396001, India

ARTICLE INFO

Article history:
Accepted 27 February 2013
Available online 14 March 2013

Keywords:
Indo-European
Heterozygosity
Linkage disequilibrium
Genetic affinity

ABSTRACT

The present study was carried out in Indo-European speaking tribal population groups of southern Gujarat (India) to elucidate the allelic and haplotypic content of β-globin system in individuals with HbAA genotypes. 6 neutral restriction sites of the β-globin system were analysed and various statistical parameters were estimated to draw meaningful interpretations. All the 6 sites were found to be polymorphic and most were in Hardy–Weinberg Equilibrium in the studied group. Haplotypes were constructed using two different combinations of the 6 restriction sites analysed. Analysis of the 5 sites revealed a set of three predominant haplotypes, '+−−−−', '−++−+' and '−+−++'; and haplotypes '+−−', '++−' and '+++' were found to be the most frequent when the 3 sites were used to construct the haplotypes. Haplotypic heterozygosity levels (>83%) observed in the present study group were comparable to those observed in African and Afro-American populations and greater than other world populations. All the ancestral haplotypes, +−−−−−, −++−+, −+−++ and −−−−+ were found in the study group. The distribution pattern of various haplotypes was consistent with the global pattern. The paucity of comparable data from other Indian populations restricted one from making interpretations about the study group's relationships with other Indian populations but the results were indicative of older population histories or experience of gene flow by the study group and their affinities with populations of southern India.

1. Introduction

β-Globin gene cluster is located on chromosome 11 in humans (11p15.5; Grzeschik and Kazazian, 1985) and consists of developmentally regulated genes involved in haemoglobin synthesis. The cluster is comprised 5 expressed genes and one pseudogene. It also consists of a 9.1 kb long recombination "hotspot" 5′ to the β-globin gene (Antonarakis et al., 1982). Mutations in β-globin gene located in this region are known to cause several diseases and hence the gene is under selection pressure. Numerous neutral polymorphic restriction sites have been looked into and investigated for constructing haplotypes in this region of the genome. These haplotypes have been explored in several populations both with clinical and evolutionary perspectives in mind. Many studies have focussed on searching for clinical relationship of these haplotypes with sickle cell mutation or thalassaemias (Adorno et al., 2004; Kulozik et al., 1987; Majumder et al., 1999; Mukherjee et al., 2004; Niranjan et al., 1999; Powars and Hiti, 1993) and hence have limited implications for evolutionary studies. Other studies with anthropological orientation have made use of these haplotypes to capture human variation and to study genetic relationships of populations across the globe (Chan et al., 1986; Chen et al., 1990; Latini et al., 2003; Long et al., 1990; Luz et al., 2010; Magaña et al., 2010; Ramsay and Jenkins, 1987; Wainscoat et al., 1986). Molecular diversity and gene frequencies at a neutral locus would be affected by selection at a linked locus (Fay and Wu, 2000). Keeping this in mind, haplotype analysis was carried out only for β^A chromosomes so that comparison could be made between populations in which selection is operating (because of presence of Hb*S) and populations in which it is not.

Different investigations have made use of variable number of polymorphic sequence variants located in non-coding regions to construct the haplotypes. Because of the presence of recombination "hotspot" in the region the haplotypes can be divided into 5′ and 3′ sub-haplotypes corresponding to the regions 5′ and 3′ to the "hotspot". The 5′ haplotypes have been widely used to investigate evolutionary relationships between human populations (Chan et al., 1986; Chen et al., 1990; Latini et al., 2003; Long et al., 1990; Luz et al., 2010; Magaña et al., 2010; Ramsay and Jenkins, 1987; Wainscoat et al., 1986) but this is not true for 3′ haplotypes. Such studies have mainly been carried out in populations of African, European or Oceanic origins or in admixed population groups from Europe and America. But only few studies have reported such data on Indian groups, and these few studies have made use of only three restriction sites (Saraswathy et al., 2008, 2009;

Abbreviations: LD, linkage disequilibrium; GIS, Gini–Simpson index; DNA, deoxyribonucleic acid.

* Corresponding author. Tel.: +91 9899915374; fax: +91 11 27666614.
E-mail address: g26_51@yahoo.co.in (G.K. Kshatriya).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.02.039

Vishwanathan et al., 2004). The present study using 6 restriction sites in different combinations has been undertaken in tribal groups from Gujarat state of India. To the best knowledge of the authors, the present study is a first attempt of its kind to elucidate the allelic and haplotypic contents of β-globin system in individuals with HbAA genotypes from India.

2. Materials and methods

5 ml intravenous blood samples were collected from 174 normal AA individuals from Indo-European speaking tribal population groups from Valsad district of southern Gujarat, India. The samples belonged mainly to 2 tribal population groups namely Dhodia and Dubla and a few samples from other tribal population groups of the area. These groups have been previously demonstrated to have underlying genetic and socio-cultural homogeneity and probably of proto-Australoid racial affinity (Kshatriya et al., 2010). The blood samples were collected by a trained medical practitioner after taking informed written consent from the individuals. DNA was isolated from the collected samples using salting-out method (Miller et al., 1988). The samples were screened for six restriction sites in the β-globin gene cluster; *Hinc*II 5′ε, *Hind*III G_γ, *Hind*III A_γ, *Hinc*II 5′ψβ, *Hinc*II 3′ψβ, and *Hinf*I 5′β. The structure of the β-globin gene cluster and the location of the six restriction sites are shown in Fig. 1.

The concerned genomic regions were amplified through Polymerase Chain Reaction using appropriate primers and protocols (Lee et al., 2002; Majumder et al., 1999; Sutton et al., 1989). The amplified products were subsequently digested with appropriate enzymes followed by typing using agarose gel electrophoresis and the results were then documented.

Allele frequencies were computed for the different loci by gene counting method. The adaptation of genotype frequencies to Hardy–Weinberg Equilibrium was tested locus by locus using chi-square goodness of fit test. DISPAN software was used to estimate heterozygosity levels and gene diversity measures (Nei, 1973). The haplotype frequencies were estimated using an expectation maximisation (EM) algorithm, as implicated in Arlequin software package version 3.1 (Excoffier et al., 2005). Linkage disequilibrium (LD) between all pairs of loci was also computed using Arlequin. Gini–Simpson index was estimated to measure genetic variation (Rao, 1982).

5 of the six sites, namely *Hinc*II 5′ε, *Hind*III G_γ, *Hind*III A_γ, *Hinc*II 5′ψβ and *Hinc*II 3′ψβ occur 5′ to the recombination "hotspot" and site *Hinf*I 5′β lies within the highly recombining region. The six restriction sites were used in two combinations to construct haplotypes. Analysis of restriction polymorphisms at the sites *Hinc*II 5′ε, *Hind*III Gγ, *Hind*III Aγ, *Hinc*II 5′ψβ and *Hinc*II 3′ψβ spread across 31.3 kb (Orkin et al., 1982) yielded 5′ sub-haplotypes; and the 3 sites *Hinf*I 5′β, *Hinc*II 3′ψβ and *Hinc*II 5′ψβ were used to form another set of haplotypes. After excluding individuals with missing genotypes for one or more restriction sites, the data on 5 restriction sites from 162 individuals was used to construct the 5′ sub-haplotypes; and those from 163 individuals (3 sites) was used to construct the other set of haplotypes.

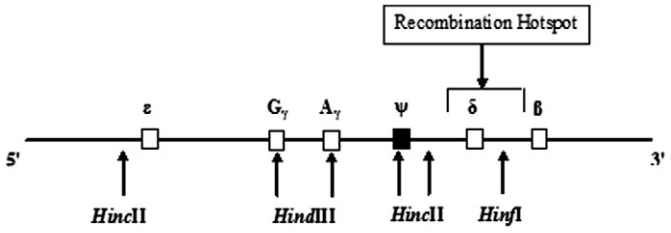

Fig. 1. Illustration of β-globin gene cluster showing the location of the different genes of the cluster, recombination hotspot and restriction sites analysed in the present study.

3. Results

All the six sites were polymorphic in the study group (Table 1). All the markers except for *Hind*III A_γ and *Hinf*I 5′β were found to be in Hardy–Weinberg Equilibrium ($p > 0.05$) in the study group. Analysis of the 5 sites revealed average heterozygosity values of 0.4563.

3.1. Haplotype analysis using 5 sites

Haplotypes were constructed only for those individuals for whom data was available for all the 5 restriction sites. When 5 restriction sites were used, a total of 22 5′ haplotypes were observed in the study group (Table 2). The frequency of these haplotypes varied from as low as 0.32% (H22 haplotype, +++++−) to as high as 31.17% (H2 haplotype, +−−−−). Haplotypes with frequencies ≥2.0% revealed a subset of 8 haplotypes. Three haplotypes, '+−−−−' (H2), '−+++−' (H3) and '−+−++' (H4), constituted the predominant haplotypes in the group under study. These three haplotypes were associated with 64.6% of individuals with AA genotype.

Haplotypic analysis at individual's level revealed that in the study group 27.78% individuals ($n = 45$) were homozygous for some haplotype. Of these individuals, 44.44% were homozygous for haplotype H2 (+−−−−), 22.22% for H14 (+++−+) and 8.89% each for H3 (−+++−+) and H4 (−+−++). In other words, the haplotypes found in greater frequency in the study group were present in homozygous state. In the remaining heterozygous samples for which the haplotypic combinations could be decidedly determined, the predominant haplotypes were found to be also the most frequently present in heterozygous state. In the case of indistinguishable combinations, it was assumed that the combination of two common haplotypes was present or a common haplotype was present with a rare haplotype rather than two rare haplotypes (Romana et al., 2000).

Significant values were observed for all pairwise LD comparisons except between *Hinc*II 5′ε and *Hind*III A_γ sites and between *Hind*III A_γ and *Hinc*II 5′ψβ (Table 3).

Heterozygosity estimated using haplotype frequencies was found to be of the order of 0.8321 in the studied tribal groups of Gujarat (Table 4). Gini–Simpson index (GSI) was found to be 0.8296 and effective number of haplotypes was found to be 5.9.

3.2. Haplotype analysis using 3 sites

Haplotypes were constructed only for those individuals for whom data was available for all the 3 restriction sites. Analysis of the 3 sites namely; *Hinf*I 5′β, *Hinc*II 3′ψβ and *Hinc*II 5′ψβ; in the study population yielded 8 haplotypes (Table 5). Of these, 6 haplotypes were present in frequencies greater than 2%, ranging from 4.52% (H8 haplotype, −−+) to 32.54% (H3 haplotype, +−−). Haplotypes '+−−' (H3), '+++' (H1), '++−' (H2) and '−+−' (H7) were unambiguously determined.

Significant pairwise linkage disequilibrium was found to exist only between the sites *Hinc*II 3′ψβ and *Hinc*II 5′ψβ (Table 6). This observation is in concordance with the fact that *Hinf*I site lies within the "hotspot".

Table 1
Allele frequencies at the six loci of β-globin cluster in the study group of Gujarat.

Marker	Genotype frequencies (%)			Allele frequencies		2n[c]
	++	+−	−−	+[a]	−[b]	
*Hinc*II 5′ε	24.29	49.72	25.99	0.491	0.509	348
*Hind*III Gγ	26.95	50.3	22.75	0.521	0.479	334
*Hind*III Aγ	21.76	31.76	46.48	0.376	0.624	340
*Hinc*II 5′ψβ	6.98	36.04	56.98	0.250	0.750	344
*Hinc*II 3′ψβ	26.01	54.34	19.65	0.532	0.468	346
*Hinf*I 5′β	58.79	19.39	21.82	0.685	0.315	330

[a] Frequency of allele for presence of restriction site.
[b] Frequency of allele for absence of restriction site.
[c] Number of chromosomes analysed.