Available online at www.aexpbio.com

ISSN : 2348-1935

RESEARCH ARTICLE

Annals of Experimental Biology
2014, 2 (1):53-57

Ava II-*Taq* I-*Hind* III represents a novel informative haplotype at the β-globin gene cluster: Application in carrier detection and prenatal diagnosis of beta Thalassemia in the Iranian population

Sedigheh Givi and Sadeq Vallian

Division of Genetics, Department of Biology, Faculty of Science, University of Isfahan, Isfahan, IR Iran

ABSTRACT

Thalassemia is one of the most common monogenic disorders, with a high demand for carrier detection and prenatal diagnosis in the Iranian population. In view of the presence of a large number of mutations associated with the disease, the polymorphic markers present in the β-globin gene cluster region were commonly used in linkage analysis of the disease. Markers usually show a population-based dependent haplotype frequency. Among the polymorphic markers, five markers including AvaII, RsaI, HinfI, TaqI and Hind III were genotyped in 150 unrelated healthy individuals from the Iranian population. The haplotype frequency was estimated using PHASE program and linkage disequilibrium (LD) was analyzed by MIDAS program. Among the eight possible haplotypes, five haplotypes showed relatively high frequencies (≥ 5%), of which the haplotype AvaII-TaqI-HindIII with the highest frequency could be suggested as an informative haplotype for possible carrier detection and prenatal diagnosis of beta thalassemia in the Iranian population. Moreover, the LD results showed that RsaI and HinfI (located in the hotspot region) were not associated with the 5' sub-haplotypes or 3' sub-haplotypes, and therefore, these markers might be excluded as strong molecular diagnostic markers in beta globin gene region in the Iranian population.

Key words: Beta-globin gene; haplotype; polymorphic markers; prenatal diagnosis; Iranian population.

INTRODUCTION

Beta thalassemia seems to be one of the most common hemoglobinopathy in the Middle East countries such as Iran [1].The beta-globin gene (HBB), part of the beta-globin cluster, is located at the short arm of chromosome 11 (11p15). More than 200 different mutations in the beta-globin gene have been reported to be associated with beta thalassemia [2]. Direct mutation analysis of the beta-globin gene mutations is a time-consuming and cost-effective procedure. Therefore, molecular markers which are linked to the beta globin gene region are usually used in carrier detection and prenatal diagnosis of the disease [3]. Prenatal diagnosis of beta-globin gene disorders using molecular markers usually relies on linkage disequilibrium (LD) of the beta-globin gene mutations and the adjacent markers [4]. This procedure may carry some degree of risk of recombination between mutant alleles and markers, due to meiotic recombination hotspots in the human genome [5]. The beta-globin gene cluster contains a hotspot region which lies immediately 5' to the beta-globin gene. The beta-globin gene cluster contains several well defined polymorphic markers, including twenty restriction fragment length polymorphisms (RFLPs) known as the 5'- and 3'-subhaplotypes located at the 5' and 3' of the hotspot region, respectively [6].

A limited number of RFLP-haplotypes of the beta-globin gene cluster have been demonstrated in some ethnic groups [7]. Orkin *et al* reported the presence of linkage between specific-thalassemia disease causing mutations with particular RFLP-haplotypes. The importance of beta-globin gene cluster haplotypes in carrier detection and prenatal diagnosis of beta-thalassemia have been documented [8]. Therefore, in the instances that direct mutation analysis fails, or to confirm the transmission of mutated alleles, haplotype analysis could be considered as an alternative approach with relatively high accuracy [9].

RFLP markers have been commonly applied in molecular diagnosis of beta-thalassemia in the Iranian population [10]. However, the frequency of beta-globin haplotypes and LD between RFLP markers in this population has not been clearly defined. In this study, five markers located in the beta globin gene region, including *Ava*II(in the β globin gene), *Rsa*I and *Hinf*I(in the hotspot region) and *Taq*I and *Hind*III (in the 5' sub-haplotypes of β) were genotyped to estimate their haplotype frequency in the Iranian population.

MATERIALS AND METHODS

DNA samples and genotyping

Blood samples were collected from 150 healthy individuals from the Iranian population. Total genomic DNA was extracted using the standard salting out method as described [11]. DNA samples were genotyped for *Ava*II (+506β), *Rsa*I (-500β), *Hinf*I (-990β), *Taq*I (-9kb) and *Hind*III (located in the intron II of G_γ gene) using PCR with primers as described earlier [7, 12]. The PCRs were carried out in reactions of 50 µL total volume containing 100-200 ng genomic DNA, 50mM KCl, 10mM Tris-HCl (pH 8.4), 2.5-4mM $MgCl_2$, 0.4mM dNTP, 0.4U/µL Taq DNA polymerase and 0.2 µM of each primer. Initial denaturation was performed at 94℃ for 4 minutes, followed by 30 cycles of 94℃ denaturation (1 minute), different annealing temperatures depending on the primer (1 minute), 72℃ extention (1 minute) with the final extention of 72℃ for 5 minutes. The optimum annealing temperature for amplification of *Ava*II, *Rsa*I, *Hinf*I, *Taq*I and *Hind*III was 65℃, 62℃, 67℃, 58℃ and 65℃, respectively. The PCR products were digested with appropriate restriction enzymes, and separated by electrophoresis on 2% agarose gel and visualized by UV transillumination following staining with ethidium bromide using a UV gel documentation (Syngene, Cambridge, UK).

Statistical analysis

The allele frequency, observed heterozygosity (Ho) and expected heterozygosity (He) were estimated using GENEPOP software [13]. The haplotype frequency was estimated using the PHASE [14] computer program for unrelated individuals. D' and r^2, as the most important measures of LD, were estimated using inter-allelic disequilibrium analysis software (MIDAS) [15]. When D' is equal to 1, the LD is completed and two SNPs were not separated by recombination. In the condition that only two SNPs had not been separated by recombination, but had the same allele frequency, LD was considered as perfect (r^2 =1). Because D' alone could not provide good predictor of the potential usefulness for multiallelic markers, Yamazaki's standardized Chi-square, χ^2, another multiallelic LD statistics, was used [16]. Chi-square provides a direct evaluation of the strength of the association between alleles at the two loci. Also, the estimation of χ^2 was performed using MIDAS computer program.

RESULTS

The allelic frequencies and heterozygosity rates of *Ava*II, *Rsa*I, *Hinf*I, *Taq*I and *Hind*III loci for 150 healthy Iranian individuals were estimated using GENEPOP. As presented in Table 1, the degree of allelic frequency for the markers examined varied form 0.3 -0.9. *Hinf*I showed the lowest (0.9/0.1), and *Taq*I (0.44/0.56) and *Rsa*I (0.48/0.52) showed a higher degree of frequency. Moreover, all the markers showed a relatively high degree of allelic heterozygosity (Table 1).

The estimation of haplotype frequency was performed using PHASE computer program. The data revealed the presence of five haplotypes with frequencies above 0.05 (≥0.05) in Iranian population (Table 2).

Given the presence of relatively high proportion of informative haplotypes in the population, D', r^2 and χ^2 for the pairing of the examined five RFLP markers were calculated by MIDAS program (Table 3). The null hypothesis of random association between pairs of alleles at the two loci (D'= 0) was tested by χ^2value [17]. The χ^2 value for five paired markers was obtained by means of the MIDAS programs and was compared with χ^2obtained from the chi-square chart (P B 0.05). The data indicated that *Taq*I-*Hind*III haplotype had a strong LD with D'=0.783 and r^2=0.52. In contrast, a weak LD across the other haplotypes was observed. As shown in Table 3, D' values ranged from 0.0122 -0.783, whereas the r^2 values ranged from 0-0.52. As indicated, D' and r^2 for all haplotypes were almost positively correlated, except for *Rsa*I-*Hinf*I (D' = 0.547, r^2 = 0.034). As indicated in Table 3, χ^2 results confirmed the D' findings for all haplotype groups. These data suggested the presence of a weak LD for the *Rsa*I and *Hinf*I makers at the hotspot region.

DISCUSSION

Linkage analysis using polymorphic markers present in the beta-globin gene cluster region could facilitate carrier detection and prenatal diagnosis of beta thalassemia in families with affected individual. The degree of heterozygosity of the markers present at the beta-globin gene cluster and their informative haplotypes have been

reported to be population specific [3]. Moreover, the presence of hotspots in different genomic regions show that appropriate density of markers could be screened on either side of the hot spot for association studies. This implies that other areas of the genome experience low recombination, which may exhibit elevated level of linkage disequilibrium, and could require low marker density for successful association studies, with the assumption that levels of gene conversion are not too high. [18]

The nonrandom association of polymorphic restriction sites in the beta-globin gene cluster was first reported by Antonarakis *et al* (1982) [19]. Their report indicated that these polymorphic sites form a 5' sub-haplotype and a 3' sub-haplotype in the beta-globin. The 34.6kb 5' cluster consist of seven sites including *Hin*dII, *Hin*dIII-Gγ, *Hin*dIII-Aγ, *Hin*dII-ψβ, *Hin*dII-3'ψβ, *Taq*I-5'σ and the19.4kb 3'cluster consist of three sites from the 5' end of the beta-globin gene including *Hgi*AI-β ,*Ava*II-β ,*Bam*HI-β [6]. These findings have been fundamental to beta-globin linkage analysis, and were applied in beta thalassemia carrier detection and prenatal diagnosis [19].

Among the markers present in the beta-globin gene cluster region, the intra-genic marker, *Ava*II, and four extra-genic markers, *Rsa*I, *Hinf*I, *Taq*I and *Hind*III were investigated in the present study. The allele and haplotype frequency of the markers were estimated in the Iranian population.. *Taq*I and *Rsa*I were in much higher frequency than *Ava*II and *Hind*III. However, the minor allele frequency of *Hinf*I was 0.1 and the observed heterozygosity was 31%. Therefore, this marker was less informative than four other markers. Moreover, it was reported that the minor allele frequency of *Hinf*I in US black population, Greeks, Indian and Italian populations were 0.22, 0.04, 0.05 and 0, respectively [6]. This indicated that this marker was not very informative in carrier diagnosis applications. No significant difference between the observed and expected heterozygosity was evident except the *Hind*III. In view of the highest level of observed heterozygosity of *Hind*III, this marker could be more informative than other markers. Therefore, this marker could be considered as the first marker to be introduced in carrier detection and prenatal diagnosis of the beta thalassemia disease in the Iranian population.

The PHASE program was used to estimate haplotype frequency on unrelated individuals. Estimation of haplotype frequency of the markers showed five haplotypes to be common in the Iranian population (see Table 2) with frequencies above 5% (≥5%). The remaining 8.5% of subjects inherited rare or incomplete haplotypes. Furthermore, the estimation of D', r^2 and χ^2 for the above markers showed that they were not in linkage disequilibrium (except *Taq*I and *Hind*III). Therefore, a large percentage of the chromosomes in this population may display rare haplotypes. *Taq*I and *Hind*III markers, which are located within the 5' sub-haplotype, had a strong LD to each other. However, these markers showed a weak LD with *Ava*II (located at the 3' sub-haplotype) in the beta-globin gene. Moreover, *Rsa*I and *Hinf*I (located in the hotspot region) were not associated with the 5' sub-haplotypes (*Hind*III and *Taq*I) or 3' sub-haplotypes (*Ava*II) (Table 3). Kazazian *et al* (1984) have previously reported a beta thalassemia mutation in Asian Indians carrying identical haplotypes but differ at the polymorphic *Hinf*I site. This unusual finding led them to exclude *Hinf*I from haplotype designation. Furthermore, it was shown *that Hinf*I did not show significant association with the markers within 700bp both at 5' and 3' end. This could suggest that this marker lies within a limited region of sequence randomization [20]. There are several reports confirming this finding from populations such as Mediterraneans and Blacks [21, 22]. In another study in an Albanian family, two chromosomes carrying identical beta-globin haplotypes were found to differ at *Rsa*I site. They reported that because *Rsa*I was within an area of randomization, it did not seem to be associated with a specific haplotype or a specific beta-globin allele [23]. Furthermore, Valaei et al. (2009) demonstrated that a novel polymorphism of 330 base pair downstream of *Rsa*I/ β was not linked to a specific mutation in the HBB gene, and should be common in the Iranian population. Their findings suggested that this nucleotide change could interfere with precise interpretation of the *Rsa*I marker for linkage analysis and prenatal diagnosis [24].

The presence of two reported single nucleotide polymorphism (SNPs) in the recognition site of *Hind*III could decrease the accuracy of RFLP data using *Hind*III (Figure 2). Therefore, we prompted to sequence the PCR products with recognition site of *Hind*III (the Gγ-IVSII region) in several DNA samples. The data resulted in the identification of only one SNP in the samples with -/- genotype. This could support the application of *Hind*III in linkage and carriership analysis in the Iranian population. Several reports indicated that the presence of multiple SNPs within the *Hind*III restriction site located in the Gγ-globin intervening sequence II could produce the same RFLP pattern [25]. In the present study, eight samples from HbAA subjects were sequenced. The data showed that people with -/- genotype had only one SNP with this sequence AAGCTG.

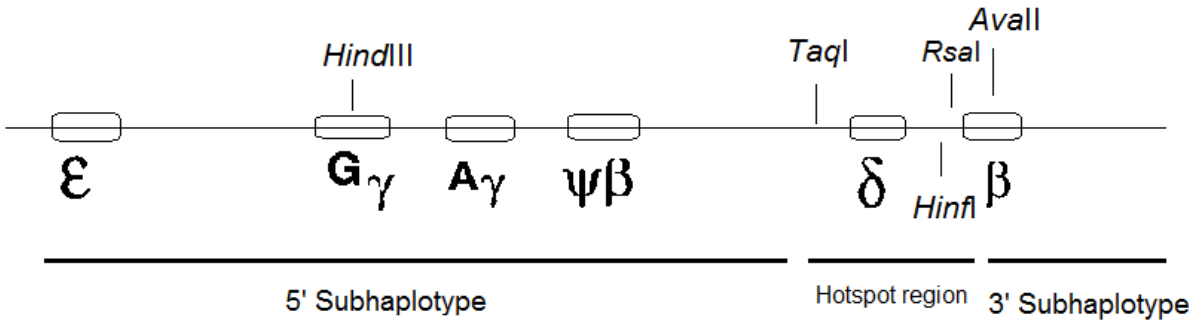

Figure 1. Map of the β-globin gene cluster with the 5' and 3' subhaplotypes and the 9-kb 'hotspot' region. RFLP sites are indicated by the lines.

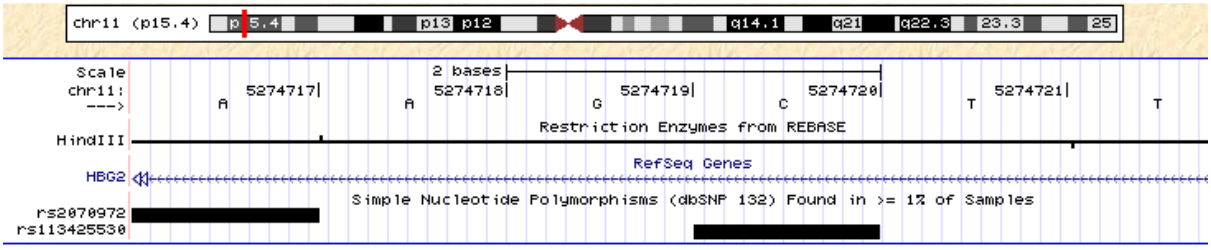

Figure 2: *Hind*III site with two SNP, rs2070972 and rs113425530. The location of the restriction enzyme on chromosome 11 and its nucleotide sequence were presented. *Hind*III site located in the intronic region of Gγ is shown in blue. Two SNPs that are located in the recognition site of *Hind*III are represented by thick black lines.

Table 1.Allele frequency, observed heterozygosity (H_o) and expected heterzygosity(H_e) of 5' beta-globin gene markers in the Iranian population.

RFLP marker	Frequencyof +allele	Frequencyof-allele	H_o	H_e
*Ava*II	0.63	0.36	45%	46%
*Rsa*I	0.48	0.52	46%	50%
*Hinf*I	0.903	0.096	31%	35%
*Taq*I	0.44	0.56	51%	49%
*Hind*III	0.40	0.60	62%	48%

Table 2.Haplotypes frequency determined by PHASE program

Index	*Ava*II-*Taq*I-*Hind*III Haplotype	Frequency
1	++−	0.056462
2	+++	0.212630
3	+−−	0.348056
4	+−+	0.025709
5	−+−	0.033964
6	−++	0.137939
7	−−−	0.160897
8	−−+	0.024343

Table 3.Analysis of D', r^2 and $\chi^2$for of 5' beta globin gene markers in the Iranian population.

two-locus haplotypes	D'	r^2	χ2	~Distance (kb)
*Ava*II-*Rsa*I	0.145	0.011	1.71	1
*Ava*II-*Hinf*I	0.228	0.003	0.51	1.5
*Ava*II-*Taq*I	0.052	0.002	0.32	9.5
*Ava*II-*Hind*III	0.106	0.010	1.54	25.5
*Rsa*I-*Hinf*I	0.547	0.034	5.2	0.44
*Rsa*I-*Taq*I	0.055	0.003	0.4	8.4
*Rsa*I-*Hind*III	0.0122	0.001	0.01	25.4
*Hinf*I- *Taq*I	0.214	0.006	1.03	8
*Hinf*I-*Hind*III	0.273	0.012	1.91	25
*Taq*I-*Hind*III	0.783	0.520	84.3	17

CONCLUSION

Together, the result from the present study demonstrated that *Ava*II-*Taq*I-*Hind*III could be used as an informative haplotype in performing carrier and prenatal diagnosis of beta thalassemia mutations in the Iranian population. These markers could be used in families with at least one affected individual in prenatal diagnosis and carrier detection testing. However, according to the previous reports, markers located in the hotspot region of the beta-globin gene (such as *Hinf*I and *Rsa*I) could be considered with great caution in prenatal diagnosis of beta thalassemia and haplotype association studies.

Acknowledgments
This study was supported by internal postgraduate research fellowship (PAGOOHANEH) from the department of research of University of Isfahan.

REFERENCES

[1] Z. Rahimi, A. Veysi Raigani, A. Merat, M. Haghshenas, N. Gerard, R. Nagel, R. Krishnamoorthy, I*ranian journal of medical sciences (IJMS)*, **2006**, 31, 70-73.
[2] S.L. Thein, *Baillière's clinical haematolo*gy, **1998**, 11, 91-126.
[3] J. Old, R. Petrou, F.K. Modell, D. Weatherall, *The Lancet*, **1982**, 320, 1413-1416.
[4] C.D. Boehm, S.E. Antonarakis, J.A. Phillips III, G. Stetten, H.H. Kazazian Jr, *New England Journal of Medicine*, **1983**, 308, 1054-1058.
[5] R.A. Smith, P.J. Ho, J.B. Clegg, J.R. Kidd, S.L. Thein, *Blood*, **1998**, 92, 4415-4421.
[6] A. Chakravarti, K. Buetow, S. Antonarakis, P. Waber, C. Boehm, H. Kazazian, *American journal of human genetics*, **1984**, 36, 1239-1258.
[7] Y.-J. Lee, S.S. Park, J.Y. Kim, H.-I. Cho, *Journal of Korean medical science*, **2002**, 17, 475.
[8] S.H. Orkin, H.H. Kazazian, S.E. Antonarakis, S.C. Goff, C.D. Boehm, J.P. Sexton, P.G. Waber, P.J. Giardina, *Nature*, **1982**, 296, 627-631.
[9] B. Forget, H. Pearson, Philadelphia. JB Lippincott Company, **1995**, 1525-1590.
[10] Z. Rahimi, A. Muniz, R. Akramipour, F. Tofieghzadeh, H. Mozafari, A. Vaisi-Raygani, A. Parsian, *Blood Cells, Molecules, and Diseases*, **2009**, 42, 140-143.
[11] S. Miller, D. Dykes, H. Polesky, *Nucleic acids research*, **1988**, 16, 1215.
[12] Z. Rahimi, M. Karimi, M. Haghshenass, A. Merat, *American journal of hematology*, **2003**, 74, 156-160.
[13] M. Raymond, F. Rousset, *Journal of heredity*, (**1995**), 86, 248-249.
[14] J. Marchini, D. Cutler, N. Patterson, M. Stephens, E. Eskin, E. Halperin, S. Lin, Z.S. Qin, H.M. Munro, G.R. Abecasis, *The American Journal of Human Genetics*, **2006**, 78, 437-450.
[15] T.R. Gaunt, S. Rodriguez, C. Zapata, I.N. Day, *BMC bioinformatics*, **2006**, 7, 227.
[16] H. Zhao, D. Nettleton, M. Soller, J. Dekkers, **2005**, 86, 77-87.
[17] Z. Fazeli, S. Vallian, *Int J Hum Genet*, **2009**, 9 115-121.
[18] J.A. Schneider, T.E. Peto, R.A. Boone, A.J. Boyce, J.B. Clegg, **2002**,11, 207-215.
[19] S.E. Antonarakis, C.D. Boehm, P.J. Giardina, H.H. Kazazian, **1982**, 79, 137-141.
[20] H. Kazazian Jr, S. Orkin, S. Antonarakis, J. Sexton, C. Boehm, S. Goff, P. Waber, *The EMBO journal*, **1984**, 3, 593.
[21] N. Maschonas, E. de Boer, R.A. Flavell, *Nucleic acids research*, **1982**, 10, 2109-2120.
[22] S.E. Antonarakis, C.D. Boehm, G.R. Serjeant, C.E. Theisen, G.J. Dover, H.H. Kazazian, P*roceedings of the National Academy of Sciences*, **1984**, 81, 853-856.
[23] G. Semenza, P. Malladi, S. Surrey, K. Delgrosso, M. Poncz, E. Schwartz, *Journal of Biological Chemistry*, **1984**, 259, 6045-6048.
[24] A. Valaei, F. Bayat, A. Kordafshari, S. Zeinali, M. Karimipoor, *Hemoglobin*, **2009**, 33, 417-421.
[25] L. Liu, S. Muralidhar, M. Singh, C. Sylvan, I.S. Kalra, C.T. Quinn, O.C. Onyekwere, B.S. Pace, *Blood Cells, Molecules, and Diseases*, **2009**, 42, 16-24.