Am J Hum Genet 36:212–217, 1984

Brief Communication

Hemoglobin E in Europeans: Further Evidence for Multiple Origins of the β^E-Globin Gene

HAIG H. KAZAZIAN, JR.,[1] PAMELA G. WABER,[1] CORINNE D. BOEHM,[1] JOSEPH I. LEE,[1] STYLIANOS E. ANTONARAKIS,[1] AND VIRGIL F. FAIRBANKS[2]

SUMMARY

We have determined haplotypes for the known restriction site polymorphisms in the β-globin gene cluster in two families of European ancestry containing individuals who are heterozygous for hemoglobin E. In both families, the β^E mutation is associated with a haplotype not previously found among the haplotypes of β^E chromosomes in Southeast Asia. Moreover, in one family, the mutation is present in a β-gene framework not found in β^E chromosomes of Southeast Asia. These data provide further evidence of multiple independent origins of the β^E mutation in human populations.

INTRODUCTION

Hemoglobin E ($\alpha_2\beta_2^{26\ glu\rightarrow lys}$) is a common hemoglobin variant in Southeast Asia and the second most prevalent abnormal hemoglobin in the world [1]. The β^E-globin gene is also a mild β-thalassemia gene such that β^E homozygotes have a clinical picture resembling classical β-thalassemia trait [2–5]. The β^E mutation affects β-gene expression by improving the efficiency of a normally inactive donor site for RNA splicing at codons 25–27 of the β-globin gene [6]. Through this mechanism, the mutation leads to a mild deficiency in normal β mRNA and production of small amounts of structurally abnormal β mRNA. The high frequency of the β^E gene in Southeast Asia is thought to be secondary to a selective advantage

Received May 19, 1983; revised August 2, 1983.

This work was supported by grant 2R01-AM-13983-14 from the National Institutes of Health, grant 6-194 from the March of Dimes, and the Cooley's Anemia Foundation of Maryland.

[1] Department of Pediatrics, Genetics Unit, Johns Hopkins University, School of Medicine, Baltimore, MD 21205.

[2] Department of Laboratory Medicine, Mayo Clinic, Rochester, MN 55091.

of the gene. β^E heterozygotes and homozygotes may have an increased resistance to malaria over that of homozygotes for hemoglobin A [1].

Recently, we showed that normal β-globin genes exist in four forms in various population groups. These forms have been called frameworks [7, 8]. Frameworks 1 and 2, which differ by a single nucleotide at IVS-2, position 74, are both found in Caucasians and Asians. Framework 3 Asian has the IVS-2, position 74 substitution of framework 2 and three additional nucleotide substitutions that are located at codon 2 of exon 1 and IVS-2, positions 16 and 666. Framework 3 of Caucasians is identical to framework 3 Asian except that it contains an additional nucleotide substitution at IVS-2, position 81. Moreover, there exists a strict correlation between the pattern at three polymorphic DNA restriction sites (sites 7, 8, and 9 of fig. 1) (an HgiA I site that includes codon 2, an Ava II site that includes IVS-2 position 16, and a Bam HI site 8 kb 3′ to the β gene) and the type of β-gene framework. When all of these sites are present (a + + + pattern), a framework 1 β-gene sequence has been found in 22 of 22 cases studied ([9] and S. H. Orkin and H. H. Kazazian, Jr., unpublished observations, 1983). When the pattern at these sites is + + −, a framework 2 β-gene sequence has been present in seven of seven cases. Finally, when the pattern at sites 7, 8, and 9 is − − +, the β-gene sequence has been framework 3 in seven of seven cases when the subject was of Mediterranean origin and framework 3 Asian in five of five cases when the subject was of Asian origin. Thus, when the pattern at these three polymorphic restriction sites is known for any β-globin gene cluster, the type of β-gene framework contained in that cluster can be predicted accurately.

In Southeast Asians, the β^E mutation is present in two β-gene frameworks, 2 and 3 Asian. Since these two frameworks differ at positions 70 nucleotides to the 5′ side of the β^E mutation and 382 nucleotides to the 3′ side of it, our data have suggested the existence of at least two independent origins of the β^E mutation in Southeast Asia [8].

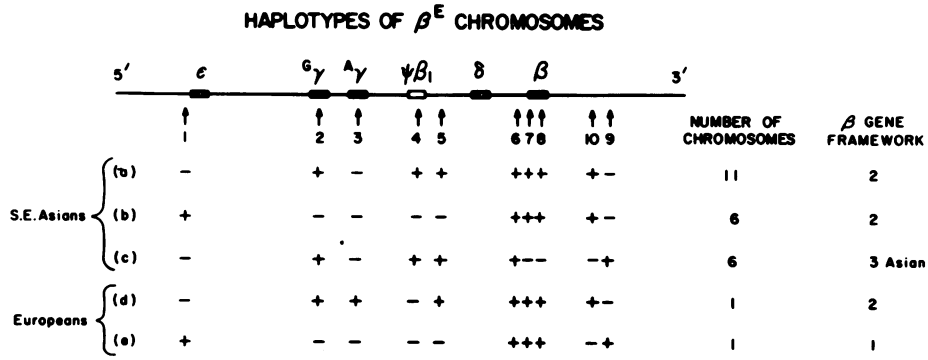

FIG. 1.—Polymorphism haplotypes of β^E-bearing chromosomes in Southeast Asians and Europeans. A 50-kb region of chromosomes 11p containing the β-globin gene cluster is *shown above*. Polymorphic sites designated *1–10* are discovered by the following restriction endonucleases: (1) Hinc II, (2) Hind III, (3) Hind III, (4) Hinc II, (5) Hinc II, (6) Hinf I, (7) HgiA I, (8) Ava II, (9) Bam HI, and (10) Hind III. The presence of a polymorphic site is designated + and the absence of a site is designated −. *Haplotypes a, b,* and *c* were reported [8].

Rarely are hemoglobin E heterozygotes observed in Europe, and they have the same clinical phenotype as heterozygotes in Southeast Asia. Fairbanks et al. described two such families with typical hemoglobin E trait: one of Czech extraction and the other of Northern European ancestry [2]. In both of these families, the β^E gene was documented in heterozygotes by amino acid analysis of the abnormal tryptic peptide, β 18-26 [2]. We have now determined haplotypes for the known polymorphic restriction sites in the β-globin gene cluster in these families [8, 10–13]. In both families, the β^E mutation is associated with a haplotype not previously found among the haplotypes of β^E chromosomes in Southeast Asia, while in one family the mutation is located in β-gene framework 1. These data provide further evidence of multiple independent origins of the β^E mutation in human populations.

METHODS

Subjects

Our subjects were members of the C. family (subjects 7 and 10 of [2]) and the W. family (subjects 19–21 of [2]). Certain family members had typical hemoglobin E trait characterized by MCVs of 68–75 fl, MCHs of 23–26 pg, MCHCs of 32%–35%, and hemoglobin E accounting for 30%–32% of their total hemoglobin. A detailed hematological description is provided in [2].

Analysis of DNA Polymorphisms

Polymorphic DNA restriction sites were analyzed as described [14–16]. The polymorphic sites analyzed were: (1) the Hinc II site 5′ to the ε gene; (2) the Hind III site in IVS-2 of the $^G\gamma$ gene; (3) the Hind III site in IVS-2 of the $^A\gamma$ gene; (4) the Hinc II site in the $\psi\beta_1$ gene; (5) the Hinc II site 3′ to the $\psi\beta_1$ gene; (6) the Hinf I site 5′ to the β gene; (7) the HgiA I site in the β gene; (8) the Ava II site in IVS-2 of the β gene; (9) the Bam HI site 3′ to the β gene; and (10) the Hind III site 3′ to the β gene. This latter site was analyzed using a 0.9-kb Bgl II-Eco RI fragment located 18 kb 3′ to the β gene [7]. For each family, the haplotype of the β-globin gene cluster containing the B^E gene was determined by analysis of these polymorphic restriction sites in a number of family members [12]. β-gene frameworks were identified by DNA polymorphisms and not by DNA sequence analysis.

RESULTS

The haplotypes of β-globin gene clusters containing β^E mutations that have been observed to date are shown in figure 1. Haplotypes a, b, and c were previously found in Southeast Asians, while d and e are reported here in Europeans. In the C. family of Czech extraction (haplotype d), the β^E mutation was present in a β-gene framework 2. This framework is associated with a + + − pattern at sites 7, 8, and 9, respectively, and it is often associated with β^E genes in Southeast Asia (see haplotypes a and b). On the other hand, the portion of the haplotype to the 5′ side of the β^E gene has not been previously observed in association with β^E genes in Southeast Asia. This new haplotype could have originated by a crossing-over event 5′ to a β^E gene of Southeast Asian origin. We previously suggested that the region 5′ to the β gene is relatively susceptible to crossing-over events [12].

The β-globin gene cluster containing the β^E mutation in the W. family of Northern European ancestry is even more interesting (fig. 1, haplotype e). The β^E mutation was present in a chromosome that is + + + at sites 7, 8, and 9 of figure 1. Of 22 β genes previously studied from chromosomes with this polymorphism pattern, all 22 have been framework 1 ([9] and S. H. Orkin and H. H. Kazazian, Jr., unpublished observations, 1983). Thus, in all likelihood, this β^E mutation is present in a framework 1 β gene that lacks all the nucleotide polymorphisms of frameworks 2, 3 Asian, and 3. Although only four common β-gene frameworks (1, 2, 3, and 3 Asian) have been recognized [8], the β^E gene has now been found in three of them (1, 2, and 3 Asian). Note that framework 3 Asian is seen in haplotype c and is associated with a − − + pattern at sites 7, 8, and 9, respectively.

The β^E mutation of the W. family was present in a β-gene cluster that was unusual for β^E genes not only in the β gene itself but also in the partial haplotype 3′ to the β gene (sites 7, 8, 9, and 10, fig. 1). In this family, the mutation was contained in a β-gene cluster that lacked a newly discovered polymorphic Hind III site (site 10, fig. 1). This Hind III restriction site polymorphism is located about 7 kb 3′ to the β gene (B. Forget, personal communication, 1982). We have studied this site in 85 β^A chromosomes representing all four β-gene frameworks and a comparable number of β-thalassemia and β^S chromosomes. The site is uniformly present adjacent to β genes of framework 2 (site 10 in haplotypes a, b, and d of fig. 1) and is absent next to all β genes of frameworks 3 and 3 Asian (site 10 in haplotype c of fig. 1). On the other hand, chromosomes containing β genes of framework 1 can be subdivided into those that contain the polymorphic Hind III site and those that lack this site. Among these chromosomes in Caucasians, Indians, and blacks, 20%–25% lack this site.

To place these data into perspective and to review, we previously reported three common patterns of polymorphic restriction sites involving the β gene and the 8 kb 3′ to it [12]. Each of these patterns correlates strictly with a type of β-gene framework. The Hind III site continues the strict linkage disequilibrium in this region, but subdivides one of the common patterns into two, so that we now recognize four patterns instead of three. However, the sequence of the β gene derived from a chromosome with the fourth pattern (+ + − + at sites 7–10 of fig. 1) is identical to that of a β gene derived from a chromosome with a more common pattern + + + + at sites 7–10 of figure 1, that is, both contain framework 1 (S. Orkin, personal communication, 1983). The β^E gene of the W. family is present in a chromosome with a + + − + pattern at sites 7–10 of figure 1.

DISCUSSION

Only one type of mutation can produce β^E globin, a G→A substitution at the first nucleotide of codon 26 of the β-globin gene. This mutation has now been observed in association with five different haplotypes in the β-gene cluster; three in Southeast Asians and two in Europeans. Among these five haplotypes, the mutation was previously found in two different β-gene frameworks in Southeast Asians, and now it has been found in a third β-gene framework in a family of European origin. Since the haplotypes involved are ancient and found commonly

in all ethnic groups studied, we previously suggested that the β^E mutations observed today arose after the divergence of human races and had at least two independent origins in Southeast Asia [8].

We now believe that recurrent mutation is but one of three possible mechanisms by which a mutation may appear on more than one haplotype in a given population group. The other two mechanisms are: (1) meiotic recombination 5′ to the β-globin gene [8] and (2) a specific interallelic gene conversion event [18]. Meiotic recombination could account for the presence of the β^E mutation in haplotypes a and b. Although it is still highly speculative, a specific interallelic gene conversion event could theoretically explain the β^E mutation in haplotype c in Southeast Asians. This event would transfer the sequence at codon 26 of a β^E gene of haplotype a to a β^A gene of haplotype c, converting the latter to a β^E gene. Data presented here strongly suggest at least one additional independent origin of the β^E mutation in Europe (haplotype e, fig. 1).

Evidence from studies of other mutations suggests that the nucleotide that is substituted to produce β^E is not necessarily a site susceptible to mutation. The β^S mutation has now been observed associated with 16 different haplotypes and three β-gene frameworks in blacks [19]. The β-thalassemia gene produced by a nonsense mutation at codon 39 of the β gene has been observed in four different haplotypes and two β-gene frameworks in Mediterraneans [8, 9]. Using an oligonucleotide probe, we recently obtained evidence for the existence of a common Indian β-thalassemia mutation, a nucleotide substitution at IVS-1, position 5, in a Chinese β-gene cluster containing a different β-gene framework from that associated with the mutation in Indians (S. H. Orkin and H. H. Kazazian, Jr., unpublished observations, 1983). Thus, even though meiotic recombination and possibly interallelic gene conversion may account for a sizable fraction of the diversity of haplotypes associated with these mutations, there is good evidence that these three mutations, β^S and the two thalassemia mutations, in addition to the β^E mutation, have occurred more than once in human populations. It follows that any nucleotide substitution may have arisen several times in human history.

A particular origin of a specific mutation may have a low frequency in a population if that origin was a recent event or the mutation is of no selective advantage in the environment in which it is placed. Either or both of these factors may account for the low frequency of the β^E mutation in European populations.

ACKNOWLEDGMENTS

We thank Dianne Andrews and Emily Pasterfield for help in manuscript preparation.

REFERENCES

1. FLATZ G: Hemoglobin E: distribution and population dynamics. *Humangenetik* 3:189–234, 1967
2. FAIRBANKS VF, GILCHRIST GS, BRIMHALL B, JEREB JA, GOLDSTON EC: Hemoglobin E trait reexamined: a cause of microcytosis and erythrocytosis. *Blood* 53:109–115, 1979
3. FAIRBANKS VF, OLIVEROS R, BRANDABUR JH, WILLIS RR, FIESTER RF: Homozygous hemoglobin E mimics beta-thalassemia minor without anemia or hemolysis: hema-

tologic, functional, and biosynthetic studies of first North American cases. *Am J Hematol* 8:109–121, 1980
4. TRAIGER J, WOOD WG, CLEGG JB, WEATHERALL DG: Defective synthesis of HbE is due to reduced levels of β^E mRNA. *Nature* 288:497–499, 1980
5. BENZ EJ, BERMAN BW, TONKONOW BL, ET AL.: Molecular analysis of the β-thalassemia phenotype associated with inheritance of hemoglobin E ($\alpha_2\beta_2^{26\,Glu\rightarrow Lys}$). *J Clin Invest* 68:118–126, 1981
6. ORKIN SH, KAZAZIAN HH JR, ANTONARAKIS SE, OSTRER H, GOFF SC, SEXTON JP: Abnormal RNA processing due to the exon mutation of the β^E-globin gene. *Nature* 300:768–769, 1982
7. ORKIN SH, KAZAZIAN HH JR, ANTONARAKIS SE, ET AL.: Linkage of β-thalassemia mutations and β-globin gene polymorphisms in the human β-globin gene cluster. *Nature* 296:627–631, 1982
8. ANTONARAKIS SE, ORKIN SH, KAZAZIAN HH JR, ET AL.: Evidence for multiple origins of the β^E-globin gene in Southeast Asia. *Proc Natl Acad Sci USA* 79:6608–6611, 1982
9. ORKIN SH, ANTONARAKIS SE, KAZAZIAN HH JR: Polymorphism and molecular pathology of the human β-globin gene, in *Progress in Hematology*, vol XIII, edited by BROWN GB, New York, Grune & Stratton. In press, 1983
10. JEFFREYS AJ: DNA sequence variants in the $^G\gamma$-, $^A\gamma$-, δ-, and β-globin genes of man. *Cell* 18:1–10, 1979
11. KAN YW, LEE KY, FURBETTA M, ANGUIS A, CAO A: Polymorphism of DNA sequence in the β-globin gene region. *N Engl J Med* 302:185–188, 1980
12. ANTONARAKIS SE, BOEHM CD, GIARDINA PVJ, KAZAZIAN HH JR: Non-random association of the polymorphic restriction sites in the β-globin gene cluster. *Proc Natl Acad Sci USA* 79:137–141, 1982
13. MOSCHONAS N, DEBOER E, FLAVELL RA: The DNA sequence of the 5′ flanking region of the human beta-globin gene: evolutionary conservation and polymorphic differences. *Nucleic Acids Res* 10:2109–2120, 1982
14. SOUTHERN EM: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503–517, 1975
15. MANIATIS T, KEE GS, EFSTRATIADIS A, KAFATOS FC: Amplification and characterization of the β-globin gene synthesized *in vitro*. *Cell* 8:163–182, 1976
16. GEORGE DL, PHILLIPS JA, FRANCKE U, SEEBURG PH: The genes for growth hormone and chorionic somatomammotropin are on the long arm of human chromosome 17 in region q21 to qter. *Hum Genet* 57:138–141, 1981
17. FEARON ER, KAZAZIAN HH JR, WABER PG, ET AL.: The entire β-globin gene cluster is deleted in a form of γδβ-thalassemia. *Blood* 61:1273–1278, 1983
18. DOVER G: Molecular drive: a cohesive mode of species evolution. *Nature* 299:111–116, 1982
19. ANTONARAKIS SE, BOEHM CD, SERJEANT GR, THEISEN CE, DOVER GJ, KAZAZIAN HH JR: DNA polymorphism analysis suggests multiple origins of the β^S mutation in blacks. *Proc Natl Acad Sci USA*. In press, 1984