Am. J. Hum. Genet. 50:394–398, 1992

The Origin of Glucose-6-Phosphate-Dehydrogenase (G6PD) Polymorphisms in African-Americans

Andrea C. Kay,* Wanda Kuhl,* Josef Prchal,† and Ernest Beutler*

*Scripps Clinic and Research Foundation, The Scripps Research Institute, Department of Molecular and Experimental Medicine, La Jolla, CA; and †University of Alabama, Birmingham

Summary

DNA samples from 54 male Afro-Americans were examined for glucose-6-phosphate dehydrogenase (G6PD) genotypes G6PD A(+)^{376G}, G6PD A – $^{202A/376G}$, and G6PD B and for polymorphisms in intron 5 (*Pvu*II), at nucleotide 1311, and at nucleotide 1116 (*Pst*I). In the G6PD B subjects, the nucleotide 1311 mutation and the *Pst*I site appeared to be in linkage equilibrium. No *Pvu*II + G6PD men were encountered. The G6PD A(+) mutation was in disequilibrium with respect to both the nucleotide 1311 mutation and the *Pst*I site. The G6PD A – nucleotide 202 mutation was in disequilibrium with all three polymorphic sites. No conclusion could be drawn with respect to the *Pvu*II site, except that it preceded the nucleotide 202 (A –) mutation. We conclude from these and our previous studies that G6PD B is the most ancient genotype. The nucleotide 1311 mutation, with its worldwide distribution, probably occurred next. The *Pst*I mutation, limited to Africans, probably arose next and is more ancient than the A(+) mutation, which occurred in a gene without either the *Pst*I or the 1311 mutation. G6PD A – $^{202A/376G}$ is the most recent of these mutations and is still in linkage disequilibrium with all of the sites. Presumably it occurred in an individual with both the A(+) and *Pvu*II mutations.

Introduction

Three glucose-6-phosphate dehydrogenase (G6PD) phenotypes are common in Africa—G6PD B, G6PD A(+), and G6PD A – . G6PD B is the most prevalent. G6PD A(+) is characterized by a mutation at nucleotide 376, resulting in an Asn→Asp mutation that increases electrophoretic mobility of the enzyme (Yoshida 1967; Takizawa et al. 1987). G6PD A – shares the mutation at nucleotide 376 with G6PD A(+) and has a second mutation that produces enzyme instability and therefore deficiency. Four different such second mutations have been found, three of them producing the G6PD A – phenotype (Beutler et al. 1989, 1991). Although G6PD A – is thus genetically heterogeneous, the vast majority of subjects with G6PD A – contain a second mutation at nucleotide 202 and are therefore designated G6PD A – $^{202A/376G}$ (Beutler 1989).

Only a small number of additional polymorphic sites have been found in G6PD. These include two in the coding region—one at nucleotide 1116 and the other at nucleotide 1311—and one in intron 5. All three of these sites are known to be polymorphic in African populations (Yoshida et al. 1988; Beutler and Kuhl 1990*a*, 1990*b;* Fey et al. 1990). One can therefore establish haplotypes with respect both to these polymorphic sites and to the sites that produce the G6PD A(+) and the G6PD A – $^{202A/376G}$ mutation.

Our previous studies showed that the G6PD A – $^{202/376G}$'s from both African and non-African populations were in complete linkage disequilibrium with the *Pvu*II and *Pst*I sites, suggesting a single origin for the G6PD A – mutation (Beutler and Kuhl 1990*a*). Moreover, we found that the DNA sequence of four chim-

Received April 30, 1991; final revision received October 14, 1991.

Address for correspondence and reprints: Ernest Beutler, M.D., The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037.

© 1992 by The American Society of Human Genetics. All rights reserved.

0002-9297/92/5002-0015$02.00

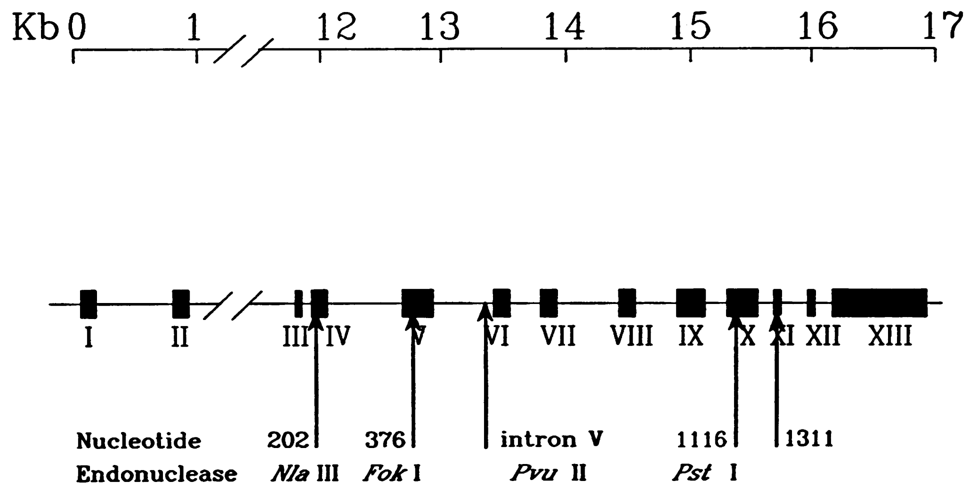

Figure 1 Polymorphic sites in G6PD gene, modified from Beutler and Kuhl (1990*a*). Gene structure is according to Martini et al. (1986).

panzees was that of G6PD B, indicating either that the common precursor of man and chimpanzee G6PD was G6PD B or that the polymorphism between G6PD A and G6PD B predated the divergence of man and chimpanzee. Accordingly, we regard G6PD B as the ancestral human genotype, i.e., the "wild type" G6PD (Beutler et al. 1989). Using all of the available markers we now examine linkage equilibrium with G6PD A(+) and reexamine the origin of G6PD $A-^{202A/376G}$ by using a third polymorphism unknown at the time of our original investigation: that at nucleotide 1311.

Material and Methods

Fifty-four blood samples were collected from American black males selected after enzyme assay and electrophoresis to provide increased numbers of those with G6PD A and G6PD A – . Only males were studied, so that X-linkage phase of the polymorphisms was certain. Samples were initially examined by electrophoretic mobility on cellulose acetate. DNA was then extracted from the whole-blood specimens. The polymorphic sites in the G6PD gene examined in the present study are illustrated in figure 1. Each site was examined by amplifying approximately 1 μg of genomic DNA by using PCR (Saiki et al. 1985, 1988). The sequences of the sense and antisense oligomers used have been published elsewhere (Hirono and Beutler 1988; Beutler and Kuhl 1990*a*). The *Fok*I, *Nla*III, *Pst*I, and *Pvu*II sites were detected using the appropriate endonucleases. In each case the site was scored positive (+) if it was cleaved and negative (–) if it was not. The polymorphic site at nucleotide 1311 was determined by sequence-specific oligonucleotide hybridization (Beutler and Kuhl 1990*b*). All samples were examined for the nucleotide 376 A→G mutation, which produces a new *Fok*I cleavage site. Samples without this site were classified as G6PD B. Those containing the nucleotide 376 mutation were further examined for the G6PD-deficient A – mutation, G→A, at nucleotide 202 by restriction analysis with *Nla*III. The samples containing the *Nla*III site were designated A – , and those without it were designated A(+). *Taq* polymerase and *Pst*I were obtained from Stratagene (La Jolla, CA), and *Pvu*II, *Nla*III, and *Fok*I were from New England Biolabs (Beverly, MA).

Results

Two of the 15 samples initially classified as G6PD A(+) by cellulose acetate electrophoresis were G6PD $A-^{202A/376G}$ by molecular analysis. Otherwise, all samples classified as G6PD A(+), A – , or B by electrophoresis were confirmed by DNA analysis. Nineteen of the samples proved to be G6PD B; 13 were G6PD A(+); and 22 were G6PD A – .

If only the *Pvu*II, *Pst*I and nucleotide 1311 polymorphisms are taken into account, four different haplotypes were found. The data are summarized in table 1. None of the G6PD B patient samples contained the *Pvu*II site; 75% contained the *Pst*I site; and 25% contained the C→T mutation at nucleotide 1311. DNA from only 1 of the 13 G6PD A(+) subjects contained the *Pvu*II site, but all had the *Pst*I site, and none

Table 1

Distribution of Males with G6PD B,A(+) and A(−) among Four Haplotypes

	No. with Haplotype[a]			
Genotype	−/−/−	−/+/−	−/+/+	+/+/−
B	4	11	4	0
A+	0	12	0	1
A−	0	0	0	22

[a] *Pvu*II/*Pst*I/1311 C→T.

carried mutations at the 1311 site. All of the samples from G6PD A− subjects contained both the *Pvu*II site and the *Pst*I site, and none contained the C→T mutation at nucleotide 1311.

Discussion

Frequency of Mutations in G6PD B African-Americans

Three different haplotypes were encountered in males of African origin who had G6PD B. The frequencies of some of these polymorphisms have been reported elsewhere (tables 2 and 3).

Among Algerians, Kenyans, and Zambians, the gene frequency of the *Pvu*II mutation is quite high, being .32–.40. In American blacks, as reported by Yoshida et al. (1988), only a .12 frequency was encountered. In the present study the *Pvu*II mutation was not found in any of the 19 G6PD B males. It is likely that the frequency of this mutation will differ among various African peoples. Moreover, among African-Americans the gene frequency is expected to be somewhat lower because of non-African admixture (Reed 1969).

D'Urso et al. (1988) found the *Pst*I site in all Europeans, but it was missing from 20% of Africans. Similarly, we found this site to be missing in 4 of 19 G6PD B African-American males.

Four of the 19 G6PD B males had the nucleotide 1311 mutation, but they were not the same individuals who lacked the *Pst*I site. This frequency of the nucleotide 1311 mutation is very similar to the 25% we have reported elsewhere (Beutler and Kuhl 1990*b*).

Table 2

Frequency of G6PD Polymorphisms in Various Populations Not Characterized by G6PD Status

Investigators and Ethnic Origin (no. of X chromosomes)	Proportion (frequency) of		
	*Pst*I[a]	*Pvu*II[b]	1311[c]
Fey et al. 1990:			
African (117).....................	...	41/117 (.35)	...
West Indian (22)	...	8/22 (.36)	...
British (74).......................	...	0/74	...
Saudi (20).........................	...	0/20	...
Icelandic (10)....................	...	0/10	...
Filipino (20).....................	...	0/20	...
D'Urso et al. 1988:			
African (64)	51/64 (.80)	...	
West Indian (24)	23/24 (.96)	...	
European (54)	54/54 (1.00)	...	
Beutler and Kuhl 1990[b]:			
African (20)	...	...	5/20 (.25)
Asian Indian (20)	...	...	9/20 (.45)
White non-Jewish (68)........	...	...	9/68 (.132)
White Jewish (41)..............	...	...	9/41 (.220)
Oriental (59)	...	...	3/59 (.051)
Kurdi-Haida et al. 1990:			
Middle Eastern (36)...........	...	...	5/36 (.14)

[a] The *Pst*I site is found if a G is present at NT 1116. In this event, cleavage of a 3.0-kb genomic fragment produces 2.1-kb and 0.9-kb fragments (Fey et al. 1990).

[b] A new *Pvu*II site is created in intron V if a mutation has occurred; it has been designated as *Pvu*II type 2 by Yoshida et al. (1988).

[c] 1311 C→T mutation present.

Table 3

Occurrence of *Pvu*II Site with Respect to G6PD Type in African-Americans

	PROPORTION (frequency) OF					
INVESTIGATORS	B	A (+)	A − (not genotyped)	A − $^{202A/376G}$	A$^{376G/680T}$	A$^{376G/968T}$
Yoshida and Takizawa 1988 ...	5/43 (.12)	7/17 (.41)	5/12 (.42)	...	...	...
Beutler and Kuhl 1990[a]	...	...	...	16/16 (1.00)	0/1	0/1
Authors of present study	0/19	1/13 (.08)	...	22/22 (1.00)	...	...

Haplotype Analysis

We find three of the four possible combinations with respect to the *Pst*I and nucleotide 1311 mutations—i.e., combinations − / −, + / −, and + / + —lacking only individuals without the *Pst* site and with the nucleotide 1311 mutation, i.e., individuals who are − / +. The absence of such individuals is not a surprising finding at equilibrium because of the low frequency of both. It appears, then, the *Pst*I and the nucleotide 1311 mutation have approached linkage equilibrium in the African population, suggesting that, like the nucleotide 1311 mutation, the *Pst*I is of very ancient origin.

The nucleotide 376 mutation that gives rise to G6PD A(+) is probably of more recent origin. All 13 subjects with G6PD A(+) also had the *Pst*I site and lacked the nucleotide 1311 mutation, i.e., were + / − with respect to these two polymorphic sites. Only 11 of 19 of the G6PD B individuals had this haplotype. A two-tailed analysis utilizing Fisher's exact test indicates that this difference is significant at the .01 level.

G6PD A − $^{202A/376G}$ is in complete linkage disequilibrium with respect to all three polymorphic sites. Indeed, the rare *Pvu*II site was present in none of the G6PD B subjects, in only one of 13 G6PD A(+) subjects, but in all 22 G6PD A − samples examined. The absence of the *Pvu*II sites from some of the G6PD A − subjects studied by Yoshida and Takizawa (1988) could be due to inclusion of genetically different forms of G6PD A −, such as the G6PD A − $^{376G/680T}$ and G6PD A − $^{376G/968T}$ mutations. Elsewhere we have shown that the *Pvu*II site is missing from subjects who have both of these mutations (Beutler and Kuhl 1990*a*).

Order in Which Mutations Have Occurred

Our findings allow us to consider the order in which various mutations may have occurred in the G6PD gene. The first of these mutations to arise must have been the one at nucleotide 1311. The C at this position may be regarded as the wild type, since it was present in the common precursor of man and chimpanzee (Beutler and Kuhl 1990*b*). Since the mutant type, with a T at nucleotide 1311, has a worldwide distribution, it must have preceded the dispersion of the human race. All of the other mutations appear to be confined to Africans and therefore probably came later. The *Pst*I mutation, although limited to Africa, appears to be in linkage equilibrium with the nucleotide 1311 mutation and therefore is the next oldest of the group. The nucleotide 376 mutation, giving rise to G6PD A(+), is more recent, since it is still in marked linkage disequilibrium with the *Pst*I/nucleotide 1311 haplotype. Presumably the nucleotide 376 mutation arose in someone who had not sustained the loss of the *Pst*I site and who had a C at nucleotide 1311. The G6PD A − mutation at nucleotide 202 is clearly the most recent. Not only did it arise in a gene that already had both the A(+) mutation at nucleotide 376 and the *Pvu*II mutation, but it is still in complete linkage disequilibrium with all of the other sites. Thus the data on this group of men with predominantly African ancestry, along with our earlier studies, suggest the following sequence in which mutations of the G6PD gene were introduced into the human species: nucleotide 1311, *Pst*I, *Fok*I [A(+)], and *Nla*III [A −]. The *Pvu*II mutation was too rare to be assigned an accurate place in this sequence, but it must have preceded the *Nla*III mutation, since the latter mutation apparently occurred in an individual with the *Pvu*II restriction site.

It is interesting to speculate how recent the origin of the nucleotide 202 mutation may be. It is located only slightly more than 1,500 genomic nucleotides from the *Pvu*II site, and no crossovers between these sites have been documented. In the present, previous (Beutler and Kuhl 1990*a*), and unpublished studies we have examined DNA from 42 G6PD A − $^{202A/376G}$ X chromosomes, without finding a single crossover with the *Pvu*II site. Crossovers at the more 3′ polymorphic sites at nucleotides 1116 and 1311 are less informative be-

cause most of the population carries the same sequence as do the G6PD A – $^{202A/376G}$ genes. Thus, most crossovers would be inapparent and therefore uninformative. If we use the rule of thumb that 1 cM equals 1 Mb of sequence (White et al. 1989), then, if we correct for the fact that only two-thirds of X chromosomes are at risk of recombining in each generation, the probability of recombination between these two polymorphic sites is 10^{-5} per generation. There is an 82% ($1 - .96^{42}$) binomial probability that the population frequency of crossover is less than 4% with the 0/42 recombinants being observed. If there were 10^{-5} crossovers per generation, then achieving 4% crossovers would require 4,000 generations, or 80,000 years. This places a very rough upper limit on the time that has elapsed since this mutation first occurred. On the other hand, the origin of the G6PD A – mutation could have been quite recent. Without any information about the survival value of this mutation, one can only speculate about the time involved. A 5% selective advantage of the mutation having occurred in 1 individual in 200,000 would result in a 10% frequency in the population in only 250 generations, or approximately 5,000 years. If such an advantage were expressed only in heterozygotes, as has been suggested (Bienzle et al. 1972), then nearly 500 generations, or 10,000 years, would have been required to achieve this level in the population. Thus, a much more recent origin of the G6PD A – mutation is possible if its selective advantage is sufficiently strong.

Acknowledgments

This work was supported by NIH grants HL25552 and RR00833 and by the Sam Stein and Rose Stein Charitable Trust Fund. This is publication 6825-MEM from the Scripps Clinic and Research Foundation, La Jolla, CA.

References

Beutler E (1989) Glucose-6-phosphate dehydrogenase: new perspectives. Blood 73:1397–1401

Beutler E, Kuhl W (1990*a*) Linkage between a PvuII restriction fragment length polymorphism and G6PD A – $^{202A/376G}$: evidence for a single origin of the common G6PD A – mutation. Hum Genet 85:9–11

——— (1990*b*) The NT 1311 polymorphism of G6PD: G6PD Mediterranean mutation may have originated independently in Europe and Asia. Am J Hum Genet 47:1008–1012

Beutler E, Kuhl W, Sáenz GF, Rodriguez W (1991) Mutation analysis of G6PD variants in Costa Rica. Hum Genet 87:462–464

Beutler E, Kuhl W, Vives-Corrons J-L, Prchal JT (1989) Molecular heterogeneity of G6PD A – . Blood 74:2550–2555

Bienzle U, Lucas AO, Ayeni O, Luzzatto L (1972) Glucose-6-phosphate dehydrogenase and malaria: greater resistance of females heterozygous for enzyme deficiency and of males with non-deficient variant. Lancet 1:107–110

D'Urso M, Luzzatto L, Perroni L, Ciccodicola A, Gentile G, Peluso I, Persico MG, et al (1988) An extensive search for RFLP in the human glucose-6-phosphate dehydrogenase locus has revealed a silent mutation in the coding sequence. Am J Hum Genet 42:735–741

Fey MF, Wainscoat JS, Mukwala EC, Falusi AG, Vulliamy TJ, Luzzatto L (1990) A *Pvu*II restriction fragment length polymorphism of the glucose-6-phosphate dehydrogenase gene is an African-specific marker. Hum Genet 84:471–472

Hirono A, Beutler E (1988) Molecular cloning and nucleotide sequence of cDNA for human glucose-6-phosphate dehydrogenase variant A(–). Proc Natl Acad Sci USA 85:3951–3954

Kurdi-Haidar B, Mason PJ, Berrebi A, Ankra-Badu G, Al-Ali A, Oppenheim A, Luzzatto L (1990) Origin and spread of the glucose-6-phosphate dehydrogenase variant (G6PD-Mediterranean) in the Middle East. Am J Hum Genet 47:1013–1019

Martini G, Toniolo D, Vulliamy T, Luzzatto L, Dono R, Viglietto G, Paonessa G, et al (1986) Structural analysis of the X-linked gene encoding human glucose-6-phosphate dehydrogenase. EMBO J 5:1849–1855

Reed TE (1969) Caucasian genes in American negroes. Science 165:762–768

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–494

Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354

Takizawa T, Yoneyama Y, Miwa S, Yoshida A (1987) A single nucleotide base transition is the basis of the common human glucose-6-phosphate dehydrogenase variant A(+). Genomics 1:228–231

White R, Lalouel JM, Leppert M, Lathrop M, Nakamura Y, O'Connell P (1989) Linkage maps of human chromosomes. Genome 31:1066–1072

Yoshida A (1967) A single amino acid substitution (asparagine to aspartic acid) between normal (B +) and the common negro variant (A +) of human glucose-6-phosphate dehydrogenase. Proc Natl Acad Sci USA 57:835–840

Yoshida A, Takizawa T (1988) The same extra *Fok*I cleavage site exists in glucose-6-phosphate dehydrogenase variants A(+) and A(–). Am J Hum Genet 43:131–133

Yoshida A, Takizawa T, Prchal JT (1988) RFLP of the X chromosome–linked glucose-6-phosphate dehydrogenase locus in blacks. Am J Hum Genet 42:872–876