Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 1771–1773, March 1984
Genetics

Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa

(sickle cell anemia/hemoglobinopathies/gene mapping/DNA polymorphism/β-globin gene cluster)

JOSÉE PAGNIER*, J. GREGORY MEARS†, OLGA DUNDA-BELKHODJA*, KIM E. SCHAEFER-REGO†, CHERIF BELDJORD*, RONALD L. NAGEL‡, AND DOMINQUE LABIE*

*Institut de Pathologie et Biologie Cellulaires et Moleculaires, Institut National de la Santé et de la Recherche Medicale U15, Paris, France; †Division of Hematology, College of Physicians and Surgeons, Columbia University, New York, NY 10032; and ‡Division of Hematology, Albert Einstein College of Medicine, Bronx, NY 10461

Communicated by Helen M. Ranney, December 1, 1983

ABSTRACT **Previous studies of the *Hpa* I cleavage site–sickle cell hemoglobin gene linkage in various African populations suggested that the sickle gene arose independently more than once. In the present study we have performed restriction endonuclease haplotype analysis for the β-globin-like gene cluster from four separate geographic areas in Africa, all of which possess the sickle gene. In Benin (Central West Africa) and Algeria (Arab North Africa) all chromosomes carrying the sickle gene possess an identical haplotype as defined by 11 different polymorphic restriction endonuclease sites within the 60-kilobase region of the β-globin-like gene cluster. In the Central African Republic (Bantu-speaking Africa) and in Senegal (Atlantic West Africa) a very large proportion of the sickle gene chromosomes were associated with a haplotype specific for each country. Thus, three different haplotypes are shown to be associated with the sickle gene in Africa, and each is present at a very high frequency in geographically separate regions. Since the three haplotypes differ from each other by at least three sites residing both 5′ and 3′ to a putative hot spot for recombination, it is most likely that the sickle gene arose at least three times on separate preexisting chromosomal haplotypes. This may have implications for a better understanding of the variable nature of the expression of sickle cell anemia, because clinically relevant sequences (for example, γ-globin gene regulatory sequences responsive to anemia) might be linked polymorphically to these haplotypes.**

Previous work by Kan and Dozy (1) and Mears *et al.* (2), using a single restriction endonuclease site polymorphism, have established that the origin of the sickle gene in Africa involved more than one mutation and that its subsequent expansion was in response to selective pressures. The observation of Kan and Dozy (3) suggested a dual origin, but the findings of Mears *et al.* suggested further diversity, with the possibility that the gene for sickle cell hemoglobin (Hb S) could have arisen independently in Atlantic Africa (Senegal), West Africa, and Bantu-speaking Africa (4). Nevertheless, single-base-pair polymorphisms are intrinsically incapable of providing conclusive data in this area. For this reason, attention is now centered on the use of multiple DNA polymorphisms in and surrounding the β-globin gene cluster for establishing the origin of the Hb S gene.

The conceptual basis for the use of DNA polymorphisms located in the β-globin gene cluster stems from the discovery of at least 11 sites accessible to detection by restriction endonucleases and several other sites detectable, at this point, only by sequence analysis (5–8). The 11 restriction endonuclease sites, spanning from just 5′ to the ε-globin gene to 8 kilobases (kb) 3′ to the β-globin gene, should be expected to be closely linked to the β-globin gene, and if the mutational event is reasonably recent a β-globin gene mutation should be accompanied by a defined haplotype formed by the set of polymorphic DNA sequences found in the β-globin-like gene cluster. This set of sequences would be those preexisting in the chromosome before the mutational event.

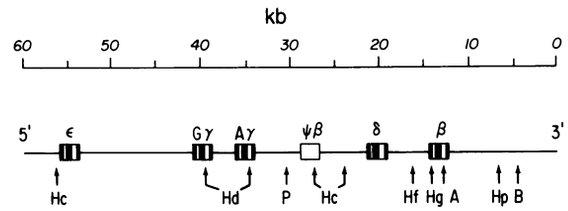

FIG. 1. Human β-globin gene cluster and 11 polymorphic sites revealed by 8 different endonuclease restriction enzymes. Hc, *Hinc*II; Hd, *Hin*dIII; P, *Pvu* II; Hf, *Hin*fI; *Hg*, *Hgi*AI; A, *Ava* II; Hp, *Hpa* I; B, *Bam*HI.

We have defined the haplotype for 11 polymorphic restriction endonuclease sites in the β-globin-like gene cluster (Fig. 1) in DNA harvested from sickle cell anemia patients in four geographic locations in Africa: Senegal (Atlantic West Africa), Benin (Central West Africa), the Central African Republic (Bantu-speaking Africa), and Algeria (Arab North Africa). We find not only extraordinary homogeneity of the haplotype found in each location but also specifically different ones for Senegal, the Central African Republic, and Benin. These data, together with the available data on the distribution of the *Hpa* I polymorphism in Africa, allow us to state that the Hb S gene arose in Africa in at least three independent historical instances and then proceeded to expand through selection.

MATERIALS AND METHODS

Subjects. Patients with sickle cell anemia were identified by hemoglobin electrophoresis and hemoglobin solubility testing. White cell pellets from 29 *SS* homozygotes from Senegal, 10 *SS* homozygotes from Benin, 14 *SS* homozygotes from the Central African Republic, and 10 from Algeria were obtained from sickle cell anemia clinics in the African countries. Ethnic origin of the subjects was assessed by identifying the tribal origin of the parents. A few normal subjects from each country were also studied.

DNA Analysis. DNA was prepared from the subjects' leukocytes by previously described techniques (2).

Haplotype analysis of the β-globin-like gene cluster was performed by using the following restriction endonucleases and DNA probes (Fig. 1). The numbers in parentheses represent the expected DNA fragment in kb if the polymorphic

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Hb S, sickle cell hemoglobin; kb, kilobase(s).

site is absent (−) or present (+): *Hin*cII, 1.3-kb ε-globin genomic probe (−, 8.0; +, 3.6); *Hin*dIII, 0.9-kb (γ-globin IVS2 genomic probe, which identifies one site in each gene ($^{G}\gamma$ = −, 7.9; +, 7.2; $^{A}\gamma$ = −, 3.4; +, 2.7); *Pvu* II, 1.7-kb *Bgl* II/*Xba* I fragment of pseudo-β-globin gene (−, 11.5; +, 14); *Hin*cII, probe that identifies two sites (−−, 7.6; −+, 6; ++, 3/3; +−, 3/4.6); *Hin*fI, 1.8-kb 5′ β-globin genomic probe (−, 1.0; +, 0.7); *Ava* II β-globin IVS2 genomic probe (−, 2.2; +, 2.0); *Hpa* I 4.4-kb *Pst* I β-globin genomic probe (−, 13; +, 7.6 or 7.0); and *Bam*HI, 4.4-kb *Pst* I β-globin genomic probe (−, 22; +, 9.3).

Restriction endonucleases were obtained from Amersham, Boehringer Ingelheim (Ingelheim, FRG), and New England BioLabs. Restriction endonuclease digestion employed the conditions recommended by the suppliers, with sufficient enzyme to ensure complete digestion.

Agarose gel electrophoresis was performed using 0.8–1.2% agarose, depending upon the size of the DNA fragments being studied, and an electric field to satisfactorily separate the fragments of interest as judged by concomitant separation of λ phage DNA size markers.

DNA was transferred to nitrocellulose by the method of Southern (9) and subsequently hybridized to the appropriate radiolabeled probe.

All probes were plasmid-purified genomic fragments (outlined above), which were nick-translated, as described (10), to a specific activity of 10^8–10^9 cpm/μg of DNA.

After hybridization, filters were washed free of specifically unbound probe as described before (11) and exposed to X-ray film at −70°C for 24–48 hr.

Analysis of the identified fragments was performed as outlined above. Homozygosity for a polymorphic site was identified by the presence of only one of the two possible DNA fragments for a site and scored as − if only the larger fragment was seen, + if only the smaller fragment was seen, and ± if both were seen.

RESULTS

In Benin, of the 10 sickle cell homozygotes investigated by these methods, all 20 chromosomes were found to have the same haplotype (Table 1).

Similarly, in Algeria all 20 sickle cell anemia chromosomes have the same haplotype as seen in Benin.

Of the 28 chromosomes in the Central African Republic from individuals with sickle cell anemia, 26 (84%) possessed a specific haplotype (Table 1). In addition, one individual was a homozygote for the chromosome most commonly found in Benin. Two additional chromosomes were found to possess an atypical haplotype, assuming the accompanying chromosome was in one case of the Benin type and in the other of the Central African Republic type.

In Senegal, 48 of 58 chromosomes (82%) studied from sickle cell anemia patients were found to have a unique haplotype. In addition, 8 chromosomes from this population carried the Benin haplotype, and 2 carried atypical haplotypes, provided their accompanying chromosomes were of the Benin and Senegalese types, respectively.

Of the 124 Hb S gene-bearing chromosomes studied in these few African countries, 110 haplotypes were geographically specific and 120 could be classified into three types (Fig. 1). In other words, only 4.4% of the sample had atypical haplotypes.

DISCUSSION

The study of the molecular history of the Hb S gene was started with the pioneering work of Kan and Dozy (1), who described the polymorphism of the *Hpa* I site located about 5 kb 3′ to the β-globin gene. This polymorphism allowed these investigators to link the Hb S gene with two types of chromosomes (one bearing the sequence recognized by *Hpa* I and the other not), each found in defined geographical locations. The limited initial data were interpreted to mean that one of the chromosomes was characteristic of West Africa and the other of Equatorial and East Africa (3). Mears *et al.* (4) extended this data to a larger and more diversified sample and determined that the 13-kb fragment generated by the absence of the *Hpa* I recognition sequence was characteristic of the central portion of West Africa but not of the populations found towards the Atlantic Coast of Africa. In addition, it was confirmed that Bantu-speaking populations (Central African Republic, Gabon, and portions of Cameroon) have a significant number of Hb S genes linked to the 7.6-kb fragment.

These findings have anthropologic importance. Available data (12) suggest that black populations were restricted originally to West Africa. About the time of the collapse of the Roman Empire, West African groups migrated east and south to the present areas of Equatorial and East Africa. Linguistic data suggest that the originators of this migration were inhabitants of the margins of the Benue River (eastern Nigeria). This migration explains the degree of homology between the Bantu languages spoken in this area, in contrast to

Table 1. β-Globin gene cluster haplotypes in four African populations

	*Hin*cII	*Hin*dIII		*Pvu* II	*Hin*cII		*Hin*fI	*Hgi*AI	*Ava* II	*Hpa* I	*Bam*HI	No. of chromosomes		
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 11	Obs.	Total	%
Benin	−	−	−	+	−	+	−	+	+	−	+	20	20	100
Algeria	−	−	−	+	−	+	−	+	+	−	+	20	20	100
Central	−	+	−	+	−	−	−	+	+	+	+	24	28	85.7
African	−	−	−	+	−	+	−	+	+	−	+	2	28	7.3*
Republic	+	−	−		−	−	−	+	+	+	+	1	28	3.6†
	−	+	+		−	+	−	+	+	+	−	1	28	3.6‡
Senegal	−	+	−	+	+	+	+	+	+	+	+	46	56	82.1
	−	−	−	+	−	+	−	+	+	−	+	8	56	14.3*
	+	−	−				−		+	+	+	1	56	1.8§
	−	+	−		−	+	+	+	+	+	+	1	56	1.8‡

*Benin type.
†Atypical haplotype, assuming accompanying chromosome is of the Central African Republic type.
‡Atypical haplotype, assuming accompanying chromosome is of the Benin type.
§Atypical haplotype, assuming accompanying chromosome is of the Senegal type. May be the same as one of the atypical Central African Republic haplotypes.

the enormous diversity and lack of formal connections between the many languages spoken in West Africa. With this background it was surprising that the Hb S gene was linked to the 7.6-kb fragment both in Atlantic Africa and in Bantu-speaking Africa, and the possibility arose that the Bantu expansion might have started in populations living in the most western portions of West Africa since the evidence for the Benue River origin is purely linguistic.

Single DNA polymorphism determination does not allow for the precise assignment of the mutated gene to a particular chromosome or definitively determine the number of mutations that might have occurred independently, since recombination events could in theory explain the previous results. The discovery of multiple polymorphic sites in the β-globin-like gene cluster (5–8) permits us to address the questions posed. Eleven sites, detectable by endonuclease mapping, can be determined, and the haplotype can be defined for each carrier of the Hb S gene.

We have selected four populations of Africa: one in North Africa (Algeria), one in Atlantic Africa (Senegal), one in Central West Africa (Benin), and one in Bantu-speaking Africa (Central African Republic). As we have more extended data on the distribution of the *Hpa* I polymorphism (a part of the haplotype), the data gathered in these four populations can be used to establish a general picture of the region.

The data presented here demonstrate that the haplotypes are almost entirely homogeneous in each population of Hb S homozygotes. The few exceptions (in which the individuals exhibit a heterozygous form for the haplotype or a homozygous form for a different haplotype) can be explained by migration from surrounding areas. Of interest, the most common haplotype that appears to have migrated has been the one associated with Benin (Central West Africa). In particular, appreciable admixture of the Benin haplotype with the local haplotype is seen in Senegal. The Senegalese study group was obtained from Dakar. Historically, Dakar has been a major commercial port for West Africa, and it is probable that the haplotype admixture present today reflects migration associated with commerce. The rare atypical haplotypes found in our sample could be the product of mutation in the DNA sequences recognized by the restriction endonuclease or the product of recombination.

Importantly, of the three different haplotypes identified by us to be associated with the Hb S gene, at least three different restriction site polymorphisms distinguish each from the others. The polymorphism differences between any two of the haplotypes occur both 5′ and 3′ to the putative "hot spot" for recombination that has been postulated to exist for some ethnic groups between the 3′ pseudo-β-globin *Hin*cII site and the *Hin*fI site 5′ to the β-globin gene (5, 7). Recombination at this point cannot explain the generation of the three haplotypes.

The data presented here provide strong evidence that the Hb S gene was generated in Africa by at least three separate mutational events involving three or more different chromosomes. In addition, the data suggest that the Hb S gene migrated from West Africa to North Africa through the well-documented trans-Saharan caravan routes (12). Also, the data are entirely compatible with the Bantu expansion having originated in an area close to the frontier of present day Nigeria and Cameroon.

It should be remembered, nevertheless, that in other populations and other genetic diseases, haplotype associations are not as easy to interpret. Y. W. Kan and his colleagues have found eight haplotypes associated with the β-39 nonsense mutation in β-thalassemia carriers in the island of Sardinia (13). This finding is not unequivocally interpretable. This situation could have originated by multiple origins of the gene, particularly if there is hypermutability of the β-39 site, but also by recombination (most of the haplotypes involved only changes 5′ to the δ-globin gene), gene conversion, or gene flow. None of these alternatives need to be envoked to explain our data.

Of anthropological interest is that these haplotypes could provide an objective tool capable of defining the places of origin of Blacks dispersed through the Americas by the slave trade. This approach will allow the calculation of the composition by African origin of many Caribbean and other New World Black communities, as long as they exhibit an appreciable Hb S gene frequency.

Furthermore, an interesting possibility arises with respect to the clinical expression of sickle cell anemia. The expression of the γ-globin gene, in the presence of anemia, is probably under genetic control and possibly is modulated by specific DNA sequences within the β-globin-like gene cluster (14). The switch from Hb F (fetal) to Hb A (adult) is retarded in sickle cell anemia patients (15), and genetic diversity among individuals could exist with respect to this feature. The genetic basis of these phenotypic features may be linked to specific haplotypes. Of considerable interest would be to define the biochemical, cellular, and clinical features of sickle cell anemia patients belonging to the three haplotypes described here. Variability in the course of sickle cell anemia may in part relate to the multiple independent origins of the Hb S gene arising in different preexisting chromosomal haplotypes. Further studies are required to elucidate this possibility.

We are grateful for the assistance and cooperation of the following colleagues: Dr. M. Benabadji from Centre National de Transfusion, Alger, Algeria; Dr. J. Teyssier and Dr. H. Plassart from Hospital Principal, Dakar, Senegal; Dr. J. Roffi from Institut Pasteur, Dakar; Dr. I Zohoun from Hopital de Cotonou, Republique democratique du Benin; Dr. G. Jaeger C.E.A.B.H. from Université René Descartes, Paris, France; and Dr. H. Baya from Hopital de Bangui, Republique Centre Africaine. This work was supported by Grants of Institut National de la Sante et de la Recherche Medicale to Unite 15, National Institutes of Health Grants HL28381 and HL21016, and the Sinsheimer Foundation.

1. Kan, Y. W. & Dozy, A. M. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 5631–5635.
2. Mears, J. G., Lachman, H. M., Cabannes, R., Amegnizin, K. P. E., Labie, D. & Nagel, R. L. (1981) *J. Clin. Invest.* **68,** 606–610.
3. Kan, Y. W. & Dozy, A. M. (1980) *Science* **209,** 388–391.
4. Pagnier, J., Labie, D., Lachman, H. M., Dunda-Belkhodja, O., Kaptue-Noche, L., Zohoun, I., Nagel, R. L. & Mears, J. G. (1983) *Distribution and Evolution of Hemoglobin and Globin Loci* (Elsevier, New York), pp. 145–158.
5. Antonarakis, S. E., Boehm, C. D., Giardina, P. G. V. & Kazazian, H. H. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 137–141.
6. Orkin, S. H., Kazazian, H. H., Antonarakis, S. E., Goff, S. C., Boehm, C. D., Sexton, J. P., Waber, P. G. & Giardina, P. J. V. (1982) *Nature (London)* **296,** 627–631.
7. Antonarakis, S. E., Orkin, S. H., Kazazian, H. H., Goff, S. C., Boehm, C. D., Waber, P. G., Sexton, J. P., Ostrer, H., Fairbanks, V. F. & Chakravarti, A. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 6608–6611.
8. Old, J. M. & Wainscoat, J. S. (1983) *Br. J. Haematol.* **53,** 337–361.
9. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–577.
10. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72,** 1184–1188.
11. Mears, J. G., Beldjord, C., Benabadji, M., Belghiti, Y. A., Baddou, M. A., Labie, D. & Nagel, R. L. (1981) *Blood* **58,** 599–601.
12. Fage, J. D. (1979) *A History of Africa* (Knopf, New York).
13. Pirastu, M., Doherty, M., Galanello, R., Cao, A. & Kan, Y. W. (1983) *Blood* **62,** 75a (abstr.).
14. Dover, G. J., Boyer, S. H. & Pembrey, M. E. (1981) *Science* **211,** 1441–1444.
15. Schneider, R. G. (1973) *Sickle Cell Disease*, eds. Abramson, H., Bertles, J. F. & Wethers, D. J. (Mosby, St. Louis), p. 230–243.