

Who Gave Whom Hemoglobin S: The Use of Restriction Site Haplotype Variation for the Interpretation of the Evolution of the β^S -Globin Gene

FRANK B. LIVINGSTONE

Department of Anthropology, University of Michigan, Ann Arbor, Michigan 48109

ABSTRACT The hypothesis that three separate mutations to the β^S -globin gene have occurred in Africa is challenged. The distributions of the β^S and other β -globin haplotypes and a simulation of the diffusion of the β^S gene are presented and argued to be more in accord with the rapid diffusion of a single mutant that by recombination and gene conversion now occurs on several different haplotypes.

There has always been the implicit belief that sickle cell anemia is a disease that occurs only in Blacks. This is undoubtedly due to a large extent to the fact that it was first discovered in a Black American and that the overwhelming number of cases in the United States occur in Blacks. When cases are found in whites, they are assumed to be due to racial admixture; and since these cases occur primarily in Mediterranean populations this seems plausible. When Lehmann and Cutbush (1952) first found high frequencies in some populations in Southern India, they immediately examined them for the so-called African blood group markers and found none in these populations. Nevertheless, the assumption persists, and the presence of African blood group markers in Middle Eastern and Sicilian populations with high frequencies of hemoglobin S is considered evidence for an African origin of hemoglobin S (Levene et al., 1976; Sandler et al., 1978).

More recently, the accumulation of data on restriction site polymorphisms (RSPs or RFLPs) and the distributions of the haplotypes defined by them has introduced a new kind of evidence to interpret the evolution of the hemoglobin variants and most especially that of the β^S -globin gene. There are now at least 17 known polymorphic restriction sites along the approximately 68 kilobases of the β^S -globin gene cluster, but most of the hypotheses about the origin and diffusion of the hemoglobin variants have been based on the 7 to 12 restriction sites that were first used in a significant number of population studies.

One of the first uses of restriction site data was a proposed association between the β^S -globin gene and one restriction site (HpaI) in the Black American population that would lead to the prenatal identification of sickle cell anemia (Kan and Dozy, 1978). However, hemoglobin S has since been found to be associated with many different restriction haplotypes, with HpaI both present and absent, in the Black American population (Antonarakis et al., 1984) and in most of the other populations in which this gene has relatively high frequencies (Kuzolik et al., 1986). In many areas of Africa the β^S gene is found predominantly on one haplotype, but the specific haplotype varies from area to area. These β^S haplotypes have been called *Senegal*, *Benin*, and *Bantu* after the area or people where they are found. They are defined by five restriction sites 5' to the β -globin gene and to three or four sites within it or 3' to it. According to the presence (+) or absence (-) of each restriction site, these β^S haplotypes have the patterns: - + - + + + +, Senegal; - - - - + + - +, Benin; and - + - - - + + +, Bantu (for the restriction enzymes involved and the haplotype frequencies, see Kuzolik et al., 1986).

There has been the general assumption that each of these haplotypes associated with the β^S gene is due to a separate mutation and thus that the β^S gene mutated in three different areas in Africa and also in the Middle East and/or India (Pagnier et al.,

Received September 13, 1988; accepted January 18, 1989.

1984; Kuzolik et al., 1986) and then diffused out from these four origins. This interpretation is now approaching the status of accepted fact (Nagel and Labie, 1985; Hill and Wainscoat, 1986; Sharon et al., 1988; Harvey et al., 1988; Chebloune et al., 1988).

From its 5' to its 3' end the β -globin cluster includes an embryonic ϵ gene, the two fetal globin genes G_γ and A_γ , a $\psi\eta$ pseudogene, the δ gene, and, finally, the β -globin gene itself. There is a region with an increased frequency of recombination immediately 5' to the β gene and 3' to the δ gene. Sites on different sides of this region are in linkage equilibrium, while those on the same side have varying degrees of linkage disequilibrium (Antonarakis et al., 1982a); and 75% of the crossing-overs occur in this region (Chakravarti et al., 1984). The frequencies of the restriction site haplotypes of the β -globin gene 5' to this region vary significantly among human populations, each major area of the world being characterized by a very few of the possible haplotypes. These frequency differences have been considered to be evidence for the origin of *Homo sapiens* in Africa (Wainscoat et al., 1986); but these conclusions have been questioned (Giles and Ambrose, 1986; Van Valen, 1986). The haplotypes of the 5' part of the β -globin cluster consist of the first five restriction sites of the Senegal, Benin, and Bantu haplotypes outlined above, and their variation is a major part of the evidence supporting the multiple mutation hypothesis.

Although these interpretations of the evolution of β -globin haplotypes seem plausible, they do not consider the effects of two other factors that now seem to be crucial to haplotype evolution. First, the enormous selective advantage of hemoglobin S heterozygotes in an environment with endemic falciparum malaria has enabled this gene to diffuse more rapidly through human populations and increase in frequency more rapidly than any other known human gene. Second, the recent evidence indicating high rates of recombination (Chakravarti et al., 1984) and gene conversion (Powers and Smithies, 1986) for the β -globin gene cluster could result in much more haplotype diversification than would be expected from mutation alone. Other interpretations are thus possible. This article will argue for a Middle Eastern origin of hemoglobin S and for the possibility that its worldwide distribution is due mostly to the diffusion of a single mu-

tant, which is postulated as having been diffused throughout Africa and the Middle East by Arab or Muslim expansion. A consideration of the worldwide distributions of the haplotypes of the β^S -globin and other β -globin variants and a simulation of haplotype distributions during the rapid advance of an advantageous gene are presented to support this alternate view.

DISTRIBUTIONS OF β -GLOBIN MUTANTS OTHER THAN HEMOGLOBIN S

It has been known for some time that the β -thalassemias are a very diverse group of molecular defects. When each one appeared to be found on one or a very few haplotypes (Hill and Wainscoat, 1986), it supported the obvious view that the association of one gene—one haplotype would indicate the occurrence of a single mutation, especially since most of the haplotypes associated with any particular β -thalassemia could be derived by simple cross-overs (Hill and Wainscoat, 1986). Antonarakis et al. (1982b) were the first to investigate a hemoglobin variant with several restriction enzymes and to find its association in a single population with different haplotypes. In a mixed group of Cambodians, Laotians, and Thais, they found two haplotypes associated with hemoglobin E in all three populations, and these could be derived from each other by a simple crossing-over at the approximate site of the known hot spot. In Cambodians, an additional haplotype was associated with hemoglobin E, and this haplotype was associated with another framework. A framework is defined by Antonarakis et al. (1982b) as the nucleotide sequence of the β gene. There are five nucleotides in the β gene, one in exon 1 and four in intron 2, that exhibit variation in all human populations. Four major frameworks based on this variation occur with varying frequencies in all human populations, and one of these major frameworks is also found on chimpanzees (Savatier et al., 1987). The finding of the same hemoglobin E mutation in two different frameworks (called frameworks 2 and 3), which differ at position 6 in exon 1 very close to the hemoglobin E mutation, was calculated by Antonarakis et al. (1982b) to be more probably due to two separate point mutations than to a series of single or double cross-overs. However, they did not include the possibility of gene conversion in their calculations.

Hemoglobin E is widely distributed in Southeast Asia, with its highest frequencies in a central area of Cambodia and Thailand and another smaller focus in the Assam valley of India (these and all subsequent frequencies cited can be found in Livingstone, 1985). This distribution suggests that the hemoglobin E mutation and/or intense endemic malaria originated in this central area of Southeast Asia and diffused out from there. Hundreiser et al. (1988a) have found that the major haplotype and the framework (No.2) associated with the β^E -globin gene in Assam is the same as those associated with it in Thailand. The β^E genes in Cambodia are found on both frameworks 2 and 3 (Hundreiser et al., 1988b), and there is a cline with an increase in framework 3 haplotypes eastward away from Thailand. The Vietnamese also have framework 3 haplotypes (Nakatsuji et al., 1986). These β^E genes on framework 3 may be due to a separate mutation, but the possibility of gene conversion is now being considered (Wainscoat, 1987; Hundreiser et al., 1988b).

There is evidence of separate hemoglobin E mutations in other populations. A study of hemoglobin E in two German families found that it was on two very different haplotypes that are also very different from the haplotypes found with hemoglobin E in Southeast Asia (Kazazian et al., 1984b). These would thus seem to be separate mutations, since in addition there is no evidence of gene flow from Southeast Asia. However, Hundreiser et al. (1988a) have found the same haplotype as one of these European families in Assam and suggest that there may have been gene flow from India. Two other populations have high frequencies of hemoglobin E: The Eti-Turks in Southeast Turkey and the Veddas of Sri Lanka. These cases have been shown to be the same amino acid substitution, and it will be of considerable interest to determine their haplotypes. These populations are rather distant from populations in Southeast Asia with high frequencies of hemoglobin E, so that it is still a question as to whether their hemoglobin E genes are due to admixture or are separate mutations. In any case, all these occurrences of hemoglobin E do support other evidence that the hemoglobin E mutation appears to be relatively common as compared with other individual hemoglobins. Hemoglobin E is a glu-lys amino acid substitution, of which there are 21 known variants for the globin chain, while

there are nine glu-val variants including hemoglobin S (International Hemoglobin Information Center, 1988).

If the hemoglobin E mutation did originate in central Southeast Asia and even specifically among the Cambodians, then the hemoglobin E gene has been in these populations longer and in high frequencies longer than elsewhere, and thus there has been more opportunity for recombination and gene conversion to occur. Thus, its occurrence on more haplotypes in these populations is expected. But the same cannot be said for a mutation that should be equally likely to occur anywhere. Since hemoglobin E has such a high selective advantage and may be a more common mutant than hemoglobin S, it would most likely have survived in other parts of the world with endemic malaria.

Hemoglobin C, another β -globin variant found in high frequencies in West Africa, has also been found to associated predominantly with one haplotype, with the few known cases on other haplotypes being derivable from it by simple cross-overs (Boehm et al., 1985). Boehm et al. (1985) studied American Blacks; thus their cases of hemoglobin C originated in West Africa. It is noteworthy that hemoglobin C is found on a haplotype that is relatively rare in African populations. More recently, hemoglobin C in Sicily has been found on the same haplotype (Cataldo et al., 1987). Hemoglobin C has also been found in a small population of Palestinian Arabs. The distribution of hemoglobin C, with its peak frequencies in the interior of West Africa and a more or less gradual decrease in frequency in all directions, remarkably approximates the expected distribution of the diffusion of an advantageous gene (Cavalli-Sforza and Bodmer, 1971). It seems to have diffused outward from West Africa, with its diffusion spread caused in large part by its selective advantage in a malarious environment. However, it has diffused to the west and north much further than it has to the east. To the east in Nigeria there is a sharp border at the Niger River, with very few cases of hemoglobin C east of there (Okpara et al., 1986). This situation seems to be best interpreted by the replacement of hemoglobin C by hemoglobin S (Livingstone, 1976); since much higher frequencies of the β^S gene are found in Nigeria both east and west of the Niger River, and wherever very high frequencies of β^S are found in Central

and East Africa, other hemoglobin variants, including thalassemia, are rare.

In contrast to Africa and Asia, there is no hemoglobin variant with a widespread distribution in Europe. Instead, endemic malaria is usually associated with β -thalassemia. Sardinia has probably had the worst endemic malaria in Europe. Although adjacent to Sicily, it has had a much lower population density and relatively little admixture from outside. Brown (1981) has contested this view and cited evidence for considerable immigration during Carthaginian and Roman times. Nevertheless, the distribution of β -thalassemia and the predominance of a single mutant would indicate that diffusion within the Sardinian population and natural selection by malaria were the main forces determining the frequencies of β -thalassemia in Sardinia. Hemoglobins S and C are both present on Sicily but not on Sardinia, but Sardinia does have higher frequencies of β -thalassemia most of which are due to a single mutation, a β -39 nonsense mutation (Pirastu et al., 1987). This mutant is found on several different haplotypes, most of which can be derived by crossing-over from the predominant haplotype, but others require gene conversion or mutation. Since one would not expect the same thalassemia mutation to occur twice in the same population, this is strong circumstantial evidence for the occurrence of gene conversion in this region of the β -globin gene. β -Thalassemia frequencies in Sardinia are also closely correlated with the endemicity of malaria, with high frequencies in the lowlands and low frequencies in the highlands. This distribution can best be interpreted as the diffusion of one adaptive mutant throughout the whole population that has attained equilibrium frequencies almost everywhere (Livingstone, 1973). This distribution of β -thalassemia in Sardinia contrasts markedly with its distribution in Sicily, where there is a lower frequency of β -thalassemia but many different mutations are present (Maggio et al., 1986).

THE DISTRIBUTION OF THE β^S GENE IN AFRICA

The previous examples of the origins and diffusions of hemoglobin mutants that are responses to malarial selection seem to indicate the comparative recency of malarial selection in human evolution since the distributions are all somewhat localized and hence have not had time to diffuse through a

large part of the human species. The β^S -globin gene also seems to be a recent mutant that has diffused rapidly throughout the central areas of the Old World because of its enormous selective advantage in environments with endemic falciparum malaria. With the most reasonable estimates of the fitness values of the various genotypes, the $\beta^A\beta^S$ -gene heterozygotes have the highest fitness of any β -globin heterozygotes, and thus the β^S gene has a much faster rate of diffusion than other mutants; I have elsewhere called β^S a *predatory* gene (Livingstone, 1976). The fact that β^S is associated with different haplotypes in different areas is not evidence of separate mutations if, within the given time period, it could have diffused throughout the populations in which it is found and if other forces (recombination and interallelic gene conversion) could have changed the haplotype associations of the β^S gene during its diffusion.

Although hemoglobin S is found in high frequencies in populations in Europe, Africa, and Asia, it has not diffused to the limits of endemic falciparum malaria. This is to be expected, since malarial selection is the cause of these high frequencies and thus endemic malaria should precede them; gene diffusion is also a much slower process than is the epidemic diffusion of an infectious disease. The discordance between the distributions of endemic falciparum malaria and high frequencies of the β^S gene, however, does indicate nonequilibrium conditions in many populations that would seem to be due to recent diffusions of the disease and the gene. In Africa the β^S gene has allele frequencies that are close to equilibrium in much of the Congo Basin. In East Africa, there are extreme differences in altitude and rainfall that result in great differences in malaria endemicity, but β^S is still close to equilibrium in most populations. On the other hand, in Southern Africa there is endemic malaria but very low frequencies of the β^S gene. Wilson et al. (1950), in their review of hyperendemic malaria, cited a population in Zululand as one of the classic examples of hyperendemic malaria, but the Zulus and all the other Bantu tribes in Southern Africa have extremely low frequencies of the β^S gene. It is also possible that falciparum malaria has spread more recently to these populations, which seems likely since the Southern Bantu do not have high frequencies of β -thalassemia.

In West Africa, malaria is also more widespread than is the β^S gene. Hemoglobin S is almost absent in the Kru peoples of Liberia and the Ivory Coast. Since these people are the indigenous inhabitants of West Africa, β^S seems most likely to have been introduced later. The diffusion of the β^S gene in West Africa is complicated by the presence there of the β^C -globin gene and a mild β -thalassemia. The latter has its highest frequencies in the Kru peoples, while the β^C gene is found in highest frequencies in the Gur-speaking peoples of the Ivory Coast, Ghana, and Burkina Faso. The β^S gene is found in a ring from Senegal to Ghana and Nigeria around these populations, and this distribution of β^S seems to be best explained as the diffusion into West Africa from elsewhere.

Further evidence for the recency of the β^S gene in West Africa is the number of populations in Guinea-Bissau with very low β^S frequencies. These populations are small and speak many different West Atlantic languages. They are surely the original inhabitants of this area and are surrounded by larger groups such as the Mandingo, Fulani, Saracole, and Wolof, who have been expanding their territories and numbers. The Senegal haplotype is the most common one in this area and is not found in high frequencies anywhere else, as far as is known from the meager samples of haplotypes studied thus far. This haplotype could have mutated here, but it is more probably a recombinant that occurred here or during the diffusion of the β^S gene to West Africa. Other haplotypes are present, and the Benin haplotype is the second most common (Pagnier et al., 1984).

Even in the central areas of Africa there are populations with low frequencies of the β^S gene but with endemic malaria. The Nilotic and Nuba peoples of the Sudan are two such groups. The Nilotes were probably more widespread in the past, since archaeological material and rock paintings throughout much of the Sahara indicate a similar pastoral way of life. Although malaria is endemic among the Nilotes, there is evidence that falciparum malaria may be comparatively recent among them (Henderson, 1932). Further north in the Sudan, the Arabic tribes have much higher frequencies of the β^S gene, and it seems likely that both falciparum malaria and the β^S gene were introduced into the Nilotes from these neighboring groups.

There are other possible reasons for the

low β^S gene frequencies in the Nilotes. They have not been influenced by the advance of either Arab or European cultures and have always had a distrust of "foreigners" (Evans-Pritchard, 1940), which should result in less gene flow into these groups. Their pastoral way of life also leads to a more nomadic existence, which can decrease the rate of transmission of malaria. Their close association with their cattle can also in some cases divert the mosquito from biting humans, and this can also decrease malaria transmission. Finally, there is some evidence that the Nilotic tribes have high frequencies of a mild β -thalassemia. These frequencies do show that these populations have had malaria selection for a considerable number of generations, but according to Henderson (1932) it was benign tertian malaria, probably caused by *Plamodium vivax*. Since the β^S gene has its major effect against falciparum malaria, it may well be that β -thalassemia was selected for because it has a greater effect against vivax malaria. But like the tribes of Liberia with β -thalassemia, the β^S gene has yet to be established among the Nilotes. Although the presence of high frequencies of β -thalassemia may retard the rate of increase of the β^S gene, the effect is minimal (Livingstone, 1983); and β^S can replace β -thalassemia or the β^C gene in the very short time of about 20 to 30 generations.

The Nuba people are an isolated population in the Nuba Mountains who also have resisted outside influences (Seligman and Seligman, 1932) and are most closely related to Central Africans. They speak languages of the Kordofanian branch of Congo-Kordofanian, while most of the peoples of Black Subsaharan Africa speak Niger-Congo languages, which constitute the other branch of this family (Greenberg, 1963). The Nuba did have malaria, but their isolation from Arabic influences seems to have prevented the β^S gene from diffusing to them. They cluster with the Nilotes and the populations of East Africa in genetic affinity, and their major difference from the Bantu is their low β^S frequency (Tay and Saha, 1988).

THE DISTRIBUTION OF THE β^S GENE IN EURASIA

The highest frequencies of the β^S gene in Eurasia are found in widely dispersed populations, and there is in all cases an association with severe hyperendemic falciparum

malaria. In Europe the gene has attained polymorphic frequencies in some populations in Sicily, Greece, Portugal, and the Soviet Union, and there are occasional cases in other countries such as Spain, France, Italy, Yugoslavia, Turkey, Bulgaria, and Albania. In the future a polymorphic frequency may well be found in some isolated population in these countries. β^S has probably reached Portugal from Africa, and haplotyping of the Portuguese β^S genes could decide whether they came primarily from Algeria or the Senegal region.

In Sicily the β^S gene is found primarily in the southeast. There is a very high frequency in Butera, with a gradual decrease in all directions from this town. This distribution has the appearance of the recent introduction of a single mutant, and its location seems to indicate an introduction from Greece (Barrari et al., 1987). β -Thalassemia is more widespread on the island and indicates that malaria has been endemic for a much longer time than the β^S gene has been there. However, there is disagreement about the origin of the β^S genes. Ragusa et al. (1988) have found that all the β^S haplotypes were of the Benin type and thus concluded that the gene originated in West Africa. The fact that the hemoglobin C haplotypes on Sicily were also identical to the major West African haplotype would seem to support this view, but this is not conclusive. If the β^S genes in Sicily came from West Africa, they would as likely be the Senegal as the Benin haplotype. The Sicilians could just as likely have obtained the β^C gene from Algeria or Tunisia and the β^S from some other population to the east, where the Benin haplotype is found. Another recent study from western Sicily (Sammarco et al., 1988) has also found that the Benin haplotype was most common and was associated with six β^S genes, but seven other β^S genes were found on other haplotypes. Most of these could be derived by simple cross-overs from the Benin type, but such haplotype diversity indicates either that the β^S gene has been present for some time if it was introduced once or that there were multiple introductions.

The highest β^S gene frequencies in Europe are found in Greece. These high frequencies are not as widespread as are those for β -thalassemia, which indicates that the latter has been there longer. β -Thalassemia in Greeks is known to include many different mutants (Antonarakis et al., 1985). In Mace-

donia, there is also a reciprocal relationship between the high frequencies of these two alleles, for which the most reasonable interpretation is that the β^S gene is in the process of eliminating β -thalassemia (Barnicot et al., 1963). Since high frequencies of β^S have now been found all along the southern coast of Turkey (Yuregir et al., 1987), this would appear to be the origin of the gene in Greece. In Turkey, the β^S gene is associated with the Benin haplotype (Bakioglu et al., 1985; Aluoch et al., 1986), so that the presence of this haplotype in Greece and Sicily could be due to gene flow from Turkey. However, the sporadic and erratic nature of this gene flow is shown by the absence of the β^S gene on Rhodes, which has high frequencies of thalassemia and endemic malaria.

In the Middle East and India, the β^S gene seems to have diffused to more of the populations with endemic malaria than is true for Africa and Europe. It has diffused through Iran to Azerbaijan in the Soviet Union and is found in most populations of the Indian peninsula with severe endemic falciparum malaria. It is also found in frequencies that seem close to equilibrium for the amount of malaria present in Saudi Arabia, southern Iran, and among Palestinian Arabs. Most of these populations have the β^S gene haplotype No. 31 (+ + - + + + -), which is now called the *Arab-Indian* haplotype (Ragusa et al., 1988). This haplotype is associated with a milder course of sickle cell anemia (Wainscoat et al., 1985; Kutlar et al., 1985). If it increases the fitness of the homozygote while maintaining the very high fitness of the heterozygote, then one would expect it to replace the other haplotypes with more severe effects.

Although Egypt is at the crossroads of the Old World, it is remarkable that there are practically no β^S genes in the Nile River Valley. The populations of the oases, however, have high frequencies, which are commensurate with the intense malaria. But this does seem to indicate that the β^S gene flow has not been via the Egyptians. The dividing line between the Benin and Arab-Indian haplotypes is through the middle of Saudi Arabia, with the Benin type found along the western mountain chain, where there is endemic malaria, and even in Riyadh, while the Arab-Indian haplotype is found in Eastern Saudi Arabia (Kuzolik et al., 1986; Ganeshaguru et al., 1987). Both the Benin and Arab-Indian haplotypes are

also found in South Turkey (Alouch et al., 1986) along with many other β^S haplotypes.

THE ORIGIN(S) AND DIFFUSION OF THE β^S GENE

Despite the complicated distributions of the β^S gene and other hemoglobin variants outlined in this brief review, some general conclusions seem plausible. First, given the high frequencies of the β^S gene of close to 0.2 in some populations in Africa and India, the fitness of the heterozygote for the β^S gene is between 1.25 and 1.33 in these populations, assuming that the fitness of the β^S homozygote is close to 0.0. With this fitness advantage, this gene will diffuse more rapidly than any other known human gene. It could, in the last 3,000 to 5,000 years, have diffused throughout the geographical areas where it is found. With a small amount of long distance gene flow, a gene with a significant selective advantage can diffuse at a very high rate. With 0.2% long-range gene flow it could diffuse through a linear array of 40 populations in just 60 generations, or about 1,500 years (Livingstone, 1969).

The more widespread distribution of β -thalassemia and the fact that it comprises many different mutations is strong evidence that it has been the first response to malarial selection, but is now either being replaced or has been replaced by the β^S gene. There is a disagreement about the population dynamics of hemoglobin C in West Africa, but I think the most reasonable interpretation of all the evidence is that it is also being replaced by the β^S gene (Livingstone, 1976).

Since it seems plausible that the β^S gene could have diffused from a single source throughout the areas where it is now found in a relatively short time, from the sheer population dynamics aspect of the problem there is no need to postulate several different mutations to the gene. If the mutation rate to the β^S gene is estimated as 10^{-8} , or even 10^{-7} , then the rates of recombination and gene conversion would seem to be orders of magnitude higher and thus much more likely. The most extensive study of β^S haplotypes has been that of Antonarakis et al. (1984) on American and Jamaican Blacks. They examined 170 β^S and 47 β^A chromosomes. The majority (151) of the β^S haplotypes were either the Senegal, Benin, or Bantu types, but there were 19 other haplotypes. This is more variation than has been found in African populations, but this in-

creased haplotype variation could be a function of sample size and the fact that Black Americans are a recent mixture of many different African populations. In addition, Antonarakis et al. (1984) found only seven β^A haplotypes to match one of the three common β^S haplotypes, so that the β^S gene is not associated with the common haplotypes in this population. This study indicates that there has been considerable haplotype rearrangement since the establishment of these populations from African and European origins in the last 350 years. The short time span of this population implies a very considerable amount of such change.

The fact that African β^S genes are frequently associated primarily with one haplotype in any specific population can be considered to be evidence that β^S is a rather recent arrival by gene flow, although the other possibility of separate mutation of course cannot be ruled out. Whenever the β^S gene is introduced into a population with endemic falciparum malaria, it will be rapidly increased by selection, and thus at first most of the copies of the gene will be on the same haplotype as the introduced gene. Since the diffusion of such a highly adaptive gene is due in large part to its introduction into a population by a few migrants, then only one haplotype will usually be present at the front of the wave of advance of this advantageous gene. The fact that the African β^S haplotypes have been named for the populations in which they are the predominant one could thus have resulted in their being named for populations to which they have been very recently introduced.

The Benin haplotype is thus named because it was found in Benin, where it is the predominant haplotype. However, immediately to the west of Benin, in Togo and in Ghana, the β^S gene is found at lower frequencies and is in the process of replacing the β^C gene (Livingstone, 1976). The β^S gene would seem to be a recent introduction to this part of West Africa, and it has certainly arrived after the original dispersal of the Kwa speakers throughout the West African tropical forest and of the differentiation of the West Atlantic speakers to the north of the Kwa speakers. The Benin haplotype is also the predominant one in many Mediterranean populations, but here there are more β^S haplotypes. This may be because the Eastern Mediterranean is the approximate center of the distribution of β^S and it may

have originated here. Or it may be due to more admixture in these populations, somewhat comparable to the Black Americans. The Benin haplotype is also very common in the Middle East and in Western Saudi Arabia, where there are also more β^S haplotypes.

The Bantu haplotype is found as the predominant one in many Bantu populations, hence its name. But since it has not reached all Bantu populations—even many with endemic malaria—it has probably been introduced since the original Bantu expansion, which occurred about 4,000 years ago (Ehret and Posnansky, 1982). The name *Bantu* is thus another misnomer, and in addition this haplotype was first discovered in the Central African Republic (Pagnier et al., 1984). Most of these populations are not Bantu, but speak languages of the Adamawa-Eastern branch of Congo-Kordofanian (Greenberg, 1963). In Kenya, the Bantu haplotype is also the predominant one, but the highest frequencies of the β^S gene and the Bantu haplotype are found in the Luo tribe, who do not even speak a Congo-Kordofanian language but a Nilotic one that is very close to the norther Nilotic languages whose populations have very low β^S allele frequencies. This distribution of the β^S gene would seem to be more in accordance with the hypothesis of the recency of the β^S gene in central and eastern Africa.

The distributions of the β^S gene in Africa and of the three major haplotypes in Africa reflect to a remarkable extent the major diffusions of Arabic influence into Africa (Lewis, 1966). In Senegal most of the contact was with Morocco and Algeria, while in Nigeria Arabic influence and even Arabic tribes came from the Sudan to Bornu, and in East and Central Africa it came from along the East African coast. The populations with low β^S frequencies are those with the least Arab contact or influence. The situation in India looks different in that many of the most primitive tribal peoples have the highest β^S frequencies. However, these populations have the most severe malaria, and Muslim populations are found throughout the areas of India where the β^S gene is found. If the Arab expansion has carried the β^S gene throughout these areas, then this diffusion has occurred in the last approximately 1,300 years, or about 50–60 generations. This is faster than anyone has imagined or estimated, but it does seem possible given the very high fitness of the heterozygote and the

great amount of gene flow that has occurred among these populations. The simulations outlined below indicate that this fast diffusion is possible and the present distribution of haplotypes in Africa could be explained by this diffusion.

One further implication of this analysis is the effect of selection by the β^S gene on β^A haplotype evolution. Wainscoat et al. (1986) have used the 5' part of the β -globin haplotype to infer the population history of the human species. The haplotype - - - - +, which is the 5' part of the Benin haplotype, is the one that distinguishes the Africans in their sample, which is mostly Nigerians. The rapid increase in this haplotype by selection when it is associated with the β^S gene and subsequent crossing-over could account for some of the high frequency of this haplotype in Africa. Thus, selection could have considerable effect on their phylogenetic tree and subsequent conclusions.

SIMULATION OF HAPLOTYPE DIFFUSION

To determine the haplotype distributions during the diffusion of a selectively advantageous gene, the process was simulated through a linear array of 30 populations. The simulation was run for 100 generations, but the hemoglobin S allele was introduced into a single population (No. 3 along the cline) in generation 30. Since it is a simulation with a finite population size, the single introduced is likely to be lost; thus the gene was continually seeded in successive generations until one became established. In all cases this occurred within 15 generations. Gene flow was set at 0.05 between adjacent populations, 0.02 between those two away, and 0.01 random migration among the five on either side of any population. The fitnesses of the three genotypes for the β^S gene were set at: AA = 0.8, AS = 1.0, and SS = 0. These fitnesses result in an equilibrium frequency for β^S of 0.167, which is close to the average for much of Sub-Saharan Africa.

The population was stored as gametes or as haplotypes, and the next generation was determined by randomly selecting two gametes and then selecting, mutating, converting, crossing-over, and migrating the resultant zygote. Pairwise linkage disequilibria were calculated for selected loci within and between the 5' and 3' parts of the haplotype and with the β^S locus. Runs were made with population sizes of 100, 200, 400, 500,

1,000, and 2,000, although the major simulation through all 30 populations reported here was with 500 gametes or 250 individuals in each population, but there were many runs through 20 populations with a population of 200.

The haplotypes were represented by a 10-tuple vector with positions 1–5 representing the 5' restriction sites; position 6, the β^S gene; and positions 7–10, the 3' restriction sites. The β^S gene was the only mutant allowed at position 6, and no other β -globin variants were programmed to occur. The N haplotypes were initially begun by randomly assigning a 0 (–) or a 1 (+) for the restriction sites, 1–5, and 7–10. At position 6, β^A was programmed as 0 and β^S as 1. With 10 sites, there are 2^{10} or 1,024 possible haplotypes, 512 with β^A . Very few of the 512 β^S haplotypes occur, since only one is begun in these populations and any others are generated by mutation, recombination, and gene conversion. The same haplotype distribution was begun in all 30 populations, and it generally took about 50 generations for the number of haplotypes to approach its equilibrium, although the number of haplotypes continued to decrease throughout 400 generations.

Mutation was programmed to occur at the same rate for all 10 sites. Estimates of the

mutation rate per nucleotide average about 10^{-8} (Neel, 1983). Since the restriction sites are several nucleotides long, the mutation rate was set at 10^{-7} . Recombination and gene conversion are the final and the most critical parameters to be set, but unfortunately there are little data on these rates for humans. However, given that both are now known to occur within the β -globin gene cluster (Gerhard et al., 1984; Smithies and Powers, 1986), any reasonable estimate of their frequency will have marked effects on the distribution of haplotypes in populations. The common average estimate for recombination is that 1 centimorgan is equivalent to 1,000 kb. Assuming a constant rate along the entire length of 68 kb of the β -globin cluster, the rate for the entire cluster would be 0.00068 (Chakravarti et al., 1984). However, Chakravarti et al. (1984) estimate the rate to be far higher, or 0.00386, and this is mostly due to the presence of the hot spot 5' to the β gene. Both estimates from Chakravarti et al. (1984) have been used to bracket possible values, and 90% of the crossing-overs have been programmed to be at the hot spot.

The rate of gene conversion is more difficult to estimate. It does seem to be frequent in the β -globin cluster (Powers and Smithies,

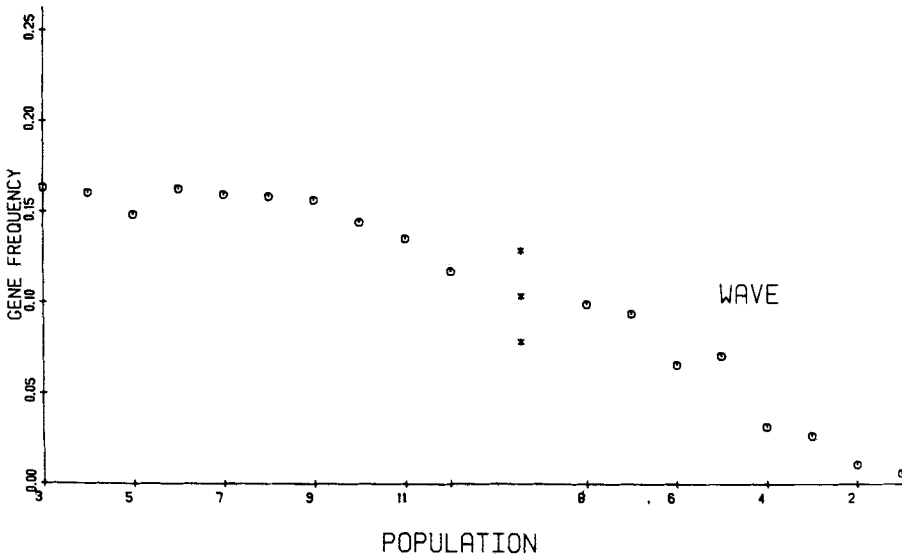


Fig. 1. The diffusion of the β^S gene through 30 populations after an average of 65 generations with a recombination rate of 0.00386, a population size of 500 gametes, and migration rates of 0.05 between adjacent populations, 0.02 between populations two away, and 0.01 randomly determined between five populations on either side. The wave of advance is separated to the right since its position varied among the runs.

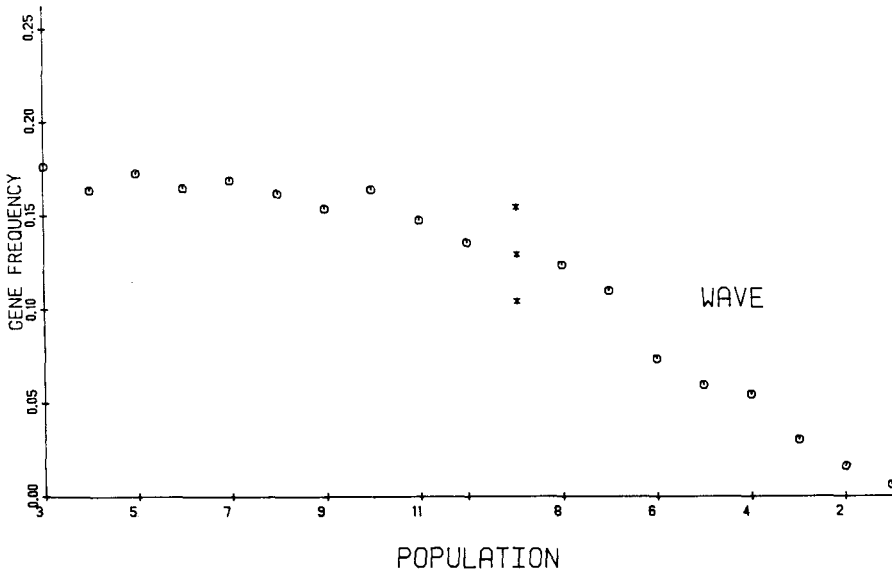


Fig. 2. The same diffusion as in Figure 1 but with a recombination rate of 0.00068.

1986). It can occur independently of crossing-over, but most conversion seems to be associated with crossing-over and to occur immediately 3' to it. This means for our model that the β -globin gene that is immediately 3' to the hot spot will have more conversion events than will the other nine restriction sites. In this model gene conversion will be proportional to the recombination rate and has been set to occur at 10% of the recombinations. It is also considered to occur independently, and this frequency has been set at 10^{-5} , although other values were also used.

RESULTS

The diffusion of the β^S gene in the 100 generations that the simulation was run is shown on Figures 1 and 2 for the two different recombination rates. It should be noted that the β^S gene has only been present for an average of about 65 generations. The wave is shown separately, since its specific position along the cline varied among runs. The average of the farthest population to which the gene had diffused was between populations 21 and 22 for both recombination rates, but varied among runs from population 15 to 30. This variation is due primarily to long distance random migration, which also was the major determinant of the rate of advance. If

the random rate of migration is doubled to 0.02, the gene advances through all 30 populations most of the time. This random migration combined with intense selection also led to a ragged wave front.

At the wave front only one haplotype was present as would be expected with these small frequencies. In 18 of the 20 runs the originally seeded S haplotype was the one present in the wave front, and it was the predominant haplotype in almost all of the populations in every run. With a recombination rate of 0.00068, the predominant haplotype was an average of 98% of the S haplotypes in populations 3 to 12, which were close to equilibrium; however, with a rate of 0.00386 it was 84% of the S haplotypes in populations 3 to 5 and 94% in populations 10 to 12. For the 0.00068 rate there was an average of 1.5 S haplotypes in populations 3 to 12, with a maximum of three S haplotypes ever occurring in any population; for the 0.00386 rate there was an average of 3.6 and a maximum of 7 S haplotypes. The average for the runs of the total number of S haplotypes present at generation 100 was 5.3 for the 0.00068 rate and 15.2 for 0.00386. Thus, these results do show that a considerable amount of haplotype diversity can occur in this short time within the total population of 7,500 individuals (30×250). Further-

more, in the 20 runs reported here all of this S haplotype variation was generated by recombination and gene conversion from the originally introduced S haplotype; no mutation to the S gene occurred.

Table 1 shows the number of S haplotypes that have been found in human populations. Many of the samples are very small, and the numbers are certainly not realistic estimates of the number of S haplotypes in these large populations. Nevertheless, they do show that in its travels around the world the hemoglobin S gene has undergone a significant amount of recombination, gene conversion, and, perhaps, mutation.

The amount of haplotype variation in a population is proportional to the number of restriction sites being considered and to population size and is maintained by migration, recombination, and gene conversion. With the values used for these variables, recombination had the most effect on haplotype variation. With a recombination rate of 0.00386, the average numbers of haplotypes in a population after 200 generations are for a population of 100, 7.3 ± 1.5 ; of 200, 16.5 ± 2.8 ; 400, 26.6 ± 3.9 ; 1,000, 54.2 ± 5.9 ; and 2,000, 112.5 ± 3.5 .

Recombination also had a pronounced effect on haplotype number. With a population of 500, the number of haplotypes at generations 100, 300, and 400 were 46.4 ± 5.0 , 29.0 ± 2.4 , and 24.9 ± 1.5 , respectively, for a

recombination rate of 0.00068 and 65.3 ± 5.8 , 53.1 ± 7.6 , and 42.8 ± 4.8 with a rate of 0.00386. These figures are not strictly comparable to haplotype variation in actual populations because of the method of starting the haplotype distributions in the simulation. As was previously noted, the initial haplotype distribution was determined by randomly selecting — haplotypes from the 512 possible β^A haplotypes with nine restriction sites with replacement, where N is the population size. This distribution was the starting distribution in all 30 populations. It approximates a Poisson distribution, so that 512 $(1.0 - e^{-Np})$ haplotypes would be expected at the start. The numbers of haplotypes given above show how much genetic drift has reduced this number, but it does also seem to show that with the population size of the simulation, 500, reduction was still occurring 300–400 generations later.

Despite these qualifications, a comparison with the data on the number of haplotypes in human populations does seem to indicate a number that could be expected from these simulations. Table 2 shows the haplotype numbers found for the β^A gene in several populations. The number is obviously proportional to the number of sites examined and the size of the sample, which is very small in most cases.

Linkage disequilibrium values were quite puzzling. There seemed to be no relationship

TABLE 1. Number of S haplotypes found in human populations

Population	No. sampled	No. sites	No. S haplotypes	Reference
Black Americans and Jamaicans	170	11	16	Antonarakis et al. (1984)
Black Americans	98	8	8	Hattori et al. (1986)
Black Americans	118	6	12	Sharon et al. (1988)
Black Europeans	50	9	8	Aluoch et al. (1986)
Senegalese	56	11	4	Pagnier et al. (1984)
Beninians	20	11	1	Pagnier et al. (1984)
Nigerians	34	8	2	Kulozik et al. (1986)
Central Africans	28	11	4	Pagnier et al. (1984)
Bantu	23	11	4	Ramsay and Jenkins (1987)
Luo and Luhya	116	8	3	Ojwang et al. (1987)
Saudis, Riyadh	22	8	2	Kulozik et al. (1986)
Saudis, Eastern	50	8	4	Kulozik et al. (1986)
Saudis, Eastern	28	5	4	Wainscoat et al. (1985)
Mediterraneans	11	11	2	Antonarakis et al. (1984)
Algerians	20	11	1	Pagnier et al. (1984)
West Sicilians	13	10	8	Sammarco et al. (1988)
East Sicilians	38	10	1	Ragusa et al. (1988)
Turks	136	9	9	Aluoch et al. (1986)
Indians, Poona	8	8	4	Kulozik et al. (1986)
Indians, Orissa	13	8	1	Kulozik et al. (1986)
Tribal Indians, Orissa	25	8	4	Kulozik et al. (1986)

TABLE 2. Number of haplotypes found in human populations

Population	No. sampled	No. sites	No. haplotypes	Reference
Black Americans and Jamaicans	47	10	22	Antonarakis et al. (1984)
Africans	26	5	5	Wainscoat et al. (1986)
Nigerians	35	5	5	Wainscoat et al. (1986)
Bantu	34	9	17	Ramsay and Jenkins (1987)
San	27	9	13	Ramsay and Jenkins (1987)
Asian Indians	33	9	10	Kazazian et al. (1984a)
Asian Indians	111	5	9	Wainscoat et al. (1986)
Asians	84	7	9+	Old et al. (1984)
Mediterraneans	122	10	17	Antonarakis et al. (1984)
Cypriots	82	5	5	Wainscoat et al. (1986)
Cypriots	40	7	7+	Old et al. (1984)
Greeks	42	7	7	Athanassiadou et al. (1987)
Sardinians	100	7	13	Pirastu et al. (1987)
Italians	80	8	12	Carestia et al. (1987)
Italians	50	5	4	Wainscoat et al. (1986)
Germans	16	7	9	Oehme et al. (1985)
British	37	5	4	Wainscoat et al. (1986)
Cambodians	47	9	9	Antonarakis et al. (1982b)
Thais	32	5	3	Wainscoat et al. (1986)
Chinese	31	9	8	Cheng et al. (1984)
Chinese	96	11	25	Chan et al. (1986)
Japanese	56	7	9	Shimizu (1987)
Melanesians and Polynesians	228	5	8	Wainscoat et al. (1986)

+, Several haplotypes lumped as "rare."

between distance and disequilibrium, and frequently adjacent loci were not in significant disequilibrium when those more distant were. None of the data are shown here, but this absence of a relationship between distance and disequilibrium is now a frequent finding for many diverse loci occurring on this short length of chromosome (Barker et al., 1984; Borresen et al., 1988; Litt and Jorde, 1986; Thompson et al., 1988).

Another interesting finding of the simulations was the great divergence of haplotypes among these 30 populations. With an average number of 27.2 haplotypes in each population, there was only an average of 4.0 shared between population 1 and the forefront of the β^S gene advance. That such a great amount of differentiation could occur in the short time of 100 generations was quite unexpected. Since it does not seem to occur in human populations, other forces may be responsible for limiting the amount of variation found among human populations.

Although the simulations really do not provide new evidence for the view advocated here of the recent diffusion of the β^S gene, but they do show that with bracketing estimates of the most reasonable values of the major parameters of conversion and recombination, this is at least as plausible an

explanation of the world distribution of this remarkable gene as is the assumption of several separate mutations, especially in Africa where the limited amount of S haplotype variation is more likely indicative of the recent diffusion of the S gene to these populations. If the β^S gene has only been present in human populations for as long as postulated here, then the large number of haplotypes found among them seems to indicate a recombination rate closer to the larger value of 0.00386 used here and that was calculated on other data by Chakravarti et al. (1984).

LITERATURE CITED

- Aluoch JR, Kilinc Y, Aksoy M, Yuregir GT, Bakioglu I, Kutlar A, Kutlar F, and Huisman THJ (1986) Sick cell anaemia among Eti-Turks: Haematological, clinical and genetic observations. *Br. J. Haematol.* 64:45-55.
- Antonarakis SE, Boehm CD, Giardina PJV, and Kazazian HH (1982a) Nonrandom association of polymorphic restriction sites in the β -globin gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* 79:137-141.
- Antonarakis SE, Boehm CD, Serjeant GR, Theisen CE, Dover GJ, and Kazazian HH (1984) Origin of the β^S -globin gene in Blacks: The contribution of recurrent mutation or gene conversion or both. *Proc. Natl. Acad. Sci. U.S.A.* 81:853-856.
- Antonarakis SE, Kazazian HH, and Orkin SH (1985) DNA polymorphism and molecular pathology of the human globin gene clusters. *Hum. Genet.* 69:1-14.
- Antonarakis SE, Orkin SH, Kazazian HH, Goff SC, Boehm CD, Waber PG, Sexton JP, Ostrer H, Fair-

- banks VF, and Chakravarti A (1982b) Evidence for multiple origins of the β^E -globin gene in Southeast Asia. *Proc. Natl. Acad. Sci. U.S.A.* 79:6608-6611.
- Athanassiadou A, Zarkadis I, Papahadjopoulou A, and Maniatis GM (1987) DNA haplotype heterogeneity of β -thalassemia in Greece: Feasibility of prenatal diagnosis. *Br. J. Haematol.* 66:379-384.
- Bakioglu I, Hattori Y, Mathew C, and Huisman THJ (1985) Five adults with mild sickle cell anemia share a β^S chromosome with the same haplotype. *Am. J. Hematol.* 20:297-300.
- Barker D, Holm T, and White R (1984) A locus on chromosome 11p with multiple restriction site polymorphisms. *Am. J. Hum. Genet.* 36:1159-1171.
- Barnicot NA, Allison AC, Blumberg BS, Belyannis G, Krimbas C, and Ballas A (1963) Haemoglobin types in Greek populations. *Ann. Hum. Genet.* 26:229-236.
- Barral I, Schiliro G, Beretta M, Mazzetti P, Russo A, and Mancuso GR (1987) Population structure of Sicily: Beta-thalassemia and HbS. *Hum. Genet.* 75:1-3.
- Boehm CD, Dowling CE, Antonarakis SE, Honig GR, and Kazazian HH (1985) Evidence supporting a single origin of the β^C -globin gene in Blacks. *Am. J. Hum. Genet.* 37:771-777.
- Borresen AL, Moller P, and Berg K (1988) Linkage disequilibrium analyses and restriction mapping of four RFLPs at the pro $\alpha 2(I)$ collagen locus: Lack of correlation between linkage disequilibrium and physical distance. *Hum. Genet.* 78:216-221.
- Brown PJ (1981) New considerations on the distribution of malaria, thalassemia, and glucose-6-phosphate dehydrogenase deficiency in Sardinia. *Hum. Biol.* 53:367-382.
- Carestia C, Pagano L, Fioretti G, and Mastrobuoni A (1987) β -Thalassemia in Campania: DNA polymorphism analysis in β^A and $\beta^{th\alpha 1}$ chromosomes and its usefulness in prenatal diagnosis. *Br. J. Haematol.* 67:231-234.
- Cataldo F, Sammarco P, Trippiedi MA, Lo Gioco P, Albeggiani A, and Maggio A (1987) Evidence of Negro origin of HbC in Sicily. *Haematologica* 72:95-96.
- Cavalli-Sforza LL, and Bodmer WF (1971) *The Genetics of Human Populations*. San Francisco: W.H. Freeman.
- Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, and Kazazian HH (1984) Nonuniform recombination within the human β -globin gene cluster. *Am. J. Hum. Genet.* 36:1239-1258.
- Chan V, Chan TK, Cheng MY, Leung NK, Kan YW, and Todd D (1986) Characteristics and distribution of β thalassemia haplotypes in South China. *Hum. Genet.* 73:23-26.
- Chebloune Y, Pagnier J, Trabuchet G, Faure C, Verdier G, Labie D, and Nigon V (1988) Structural analysis of the 5' flanking region of the β -globin gene in African sickle cell anemia patients: Further evidence for three origins of the sickle cell mutation in Africa. *Proc. Natl. Acad. Sci. U.S.A.* 85:4431-4435.
- Cheng T-C, Orkin SH, Antonarakis SE, Potter MJ, Sexton JP, Markham AF, Giardina PJV, Li A, and Kazazian HH (1984) β -Thalassemia in Chinese: Use of in vivo RNA analysis and oligonucleotide hybridization in systematic characterization of molecular defects. *Proc. Natl. Acad. Sci. U.S.A.* 81:2821-2825.
- Ehret C, and Posnansky M (1982) *The Archaeological and Linguistic Reconstruction of African History*. Los Angeles: University of California Press.
- Evans-Pritchard EE (1940) *The Nuer*. Oxford: Clarendon Press.
- Ganeshaguru K, Sejeny SA, Acquaye JK, and Omer J (1987) Sickle gene in Western Arabia is mostly associated with the 13.0 kb Hpa I fragment. *Scand. J. Haematol.* 35:249.
- Gerhard DS, Kidd KK, Egeland JA, and Housman DE (1984) Identification of a recent recombination event within the human β -globin gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* 81:7875-7879.
- Giles E, and Ambrose SH (1986) Are we all out of Africa? *Nature* 322:21-22.
- Greenberg JH (1963) *The Languages of Africa*. The Hague: Mouton.
- Harvey PH, Keymer AE, and May RM (1988) Evolving control of diseases. *Nature* 332:680-681.
- Hattori Y, Kutlar F, Kutlar A, Mckie VC, and Huisman THJ (1986) Haplotypes of β^S chromosomes among patients with sickle cell anemia from Georgia. *Hemoglobin* 10:623-642.
- Henderson LH (1932) Some observations on the incidence of malaria amongst the Nilotic tribes. *Trans. R. Soc. Trop. Med. Hyg.* 25:281-286.
- Hill AVS, and Wainscoat JS (1986) The evolution of the α - and β -globin gene clusters in human populations. *Hum. Genet.* 74:16-23.
- Hundreiser J, Deka R, Gogoi BC, Papp T, and Flatz G (1988a) DNA haplotypes and frameworks associated with the beta-globin gene in the Kachari population of Assam (India). *Hum. Hered.* 38:240-245.
- Hundreiser J, Sanguansermstri T, Papp T, Laig M, and Flatz G (1988b) β -Globin gene linked DNA haplotypes and frameworks in three South-East Asian populations. *Hum. Genet.* 80:90-94.
- International Hemoglobin Information Center (1988) IHC variants list. *Hemoglobin* 12:209-282.
- Kan YW, and Dozy AM (1978) Polymorphism of DNA sequence adjacent to the human β -globin structural gene: Relationship to sickle mutation. *Proc. Natl. Acad. Sci. U.S.A.* 75:5631-5635.
- Kazazian HH, Orkin SH, Antonarakis SE, Sexton JP, Boehm CD, Goff SC, and Waber PG (1984a) Molecular characterization of seven β -thalassemia mutations in Asian Indians. *EMBO J.* 3:593-596.
- Kazazian HH, Waber PG, Boehm CD, Lee JI, Antonarakis SE, and Fairbanks VF (1984b) Hemoglobin E in Europeans: Further evidence for multiple origins of the β^E -globin gene. *Am. J. Hum. Genet.* 36:212-217.
- Kutlar A, Hattori Y, Bakioglu I, Kutlar F, Kamel K, and Huisman THJ (1985) Hematological observations on Arabian SS patients with a homozygosity or heterozygosity for a β^S chromosome with haplotype #31. *Hemoglobin* 9:545-557.
- Kuzolik AE, Wainscoat JS, Serjeant GR, Kar BC, Al-Awamy B, Essan GJF, Falusi AG, Haque SK, Hilali AM, Kate S, Ranasinghe WAEP, and Weatherall DJ (1986) Geographical survey of β^S -globin gene haplotypes: Evidence for an independent Asian origin of the sickle-cell mutation. *Am. J. Hum. Genet.* 39:239-244.
- Lehmann H, and Cutbush M (1952) Sickle-cell trait in Southern India. *Br. Med. J.* 1:404-405.
- Levene C, Rachmilewitz EA, Ezekiel E, Freundlich E, and Sandler G (1976) Blood group phenotypes and hemoglobin S. *Acta Haematol.* 55:300-305.
- Lewis IM (ed.) (1966) *Islam in Tropical Africa*. London: Oxford University Press.
- Litt M, and Jorde LB (1986) Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2Q. *Am. J. Hum. Genet.* 39:166-178.
- Livingstone FB (1969) An analysis of the clines at the β hemoglobin locus and their simulation by models involving differential selection. *Hum. Biol.* 41:223-236.
- Livingstone FB (1973) Gene frequency differences in

- human populations: Some problems of analysis and interpretation. In MH Crawford and PL Workman (eds.): *Methods and Theories of Anthropological Genetics*. Albuquerque: University of New Mexico Press, pp. 39-66.
- Livingstone FB (1976) Hemoglobin history in West Africa. *Hum. Biol.* 48:487-500.
- Livingstone FB (1983) The malaria hypothesis. In JE Bowman (ed.): *Distribution and Evolution of Hemoglobin and Globin Loci*. New York: Elsevier Science Publ. Co., pp. 15-44.
- Livingstone FB (1985) *Frequencies of Hemoglobin Variants*. New York: Oxford University Press.
- Maggio A, Acuto S, Lo Gioco P, Di Marzo R, Giambona A, Sammarco P, and Caronia F (1986) β^A and β^{thal1} DNA haplotypes in Sicily. *Hum. Genet.* 72:229-230.
- Nagel RL (1984) The origin of the hemoglobin S gene: Clinical, genetic, and anthropological consequences. *Einstein Q. J. Biol. Med.* 2:53-62.
- Nagel RL, and Labie D (1985) The consequences and implications of the multicentric origin of the Hb S gene. In G Stamatoyannopoulos and AW Nienhuis (eds.): *Experimental Approaches for the Study of Hemoglobin Switching*. New York: Alan R. Liss, Inc., pp. 93-103.
- Nakatsuji T, Kutlar A, Kutlar F, and Huisman THJ (1986) Haplotypes among Vietnamese hemoglobin E homozygotes including one with a γ -globin gene triPLICATION. *Am. J. Hum. Genet.* 38:981-983.
- Neel JV (1983) Frequency of spontaneous and induced "point" mutations in higher eukaryotes. *J. Hered.* 74:2-15.
- Oehme R, Kohne E, and Horst J (1985) DNA-polymorphic patterns linked to the β -globin gene in German families affected with hemoglobinopathies and thalassemias: A comparison to other ethnic groups. *Hum. Genet.* 71:219-222.
- Ojwang PJ, Ogada T, Beris P, Hattori Y, Lanclos, KD, Kutlar A, Kutlar F, and Huisman THJ (1987) Haplotypes and a globin gene analysis in sickle cell anaemia patients from Kenya. *Br. J. Haematol.* 65:211-215.
- Okpara RA, Nkan EU, and Essien EM (1986) Haemoglobin patterns in the Cross River State of Nigeria with particular reference to haemoglobin C. *East Afr. Med. J.* 63:417-420.
- Old JM, Petrou M, Modell B, and Weatherall DJ (1984) Feasibility of antenatal diagnosis of β thalassaemia by DNA polymorphisms in Asian Indians and Cypriot populations. *Br. J. Haematol.* 57:255-263.
- Pagnier J, Mears JG, Dunda-Belkhodja O, Schaefer-Rego KE, Beldjord C, Nagel RL, and Labie D (1984) Evidence for the multicentric origin of the sickle cell hemoglobin gene in Africa. *Proc. Natl. Acad. Sci. U.S.A.* 81:1771-1773.
- Pirastu M, Galanello R, Doherty MA, Tuveri T, Cao A, and Kan YW (1987) The same β -globin gene mutation is present on nine different β -thalassaemia chromosomes in a Sardinian population. *Proc. Natl. Acad. Sci. U.S.A.* 84:2882-2885.
- Powers PA, and Smithies O (1986) Short gene conversions in the human fetal globin gene region: A by-product of chromosome pairing during meiosis? *Genetics* 112:343-358.
- Ragusa A, Lombardo ML, Sortino G, Lombardo T, Nagel RL, and Labie D (1988) β^S Gene in Sicily in linkage disequilibrium with the Benin haplotype: Implications for gene flow. *Am. J. Hematol.* 27:139-141.
- Ramsay M, and Jenkins T (1987) Globin gene-associated restriction-fragment-length-polymorphisms in Southern African peoples. *Am. J. Hum. Genet.* 41:1132-1144.
- Sammarco P, Giambona A, Lo Gioco P, Di Marzo R, and Maggio A (1988) Evidence of the African origin of sickle cell hemoglobin in Western Sicily. *Hemoglobin* 12:193-196.
- Sandler SG, Schiliro G, Russo A, Musumeci S, and Rachmilewitz EA (1978) Blood group phenotypes and the origin of sickle cell hemoglobin in Sicilians. *Acta Haematol.* 60:350-357.
- Savatier P, Trabuchet G, Chebloune Y, Faure C, Verdier G, and Nigon VM (1987) Nucleotide sequence of the beta-globin genes in gorilla and macaque: The origin of nucleotide polymorphisms in human. *J. Mol. Evol.* 24:309-318.
- Seligman CG, and Seligman BZ (1932) *Pagan Tribes of the Nilotic Sudan*. London: Routledge and Kegan Paul.
- Sharon B, Poncz M, Surrey S, and Schwartz E (1988) Non-random association of the RSA I polymorphic site 5' to the β -globin gene with major sickle cell haplotypes. *Hemoglobin* 12:115-124.
- Shimizu K (1987) DNA polymorphism patterns in the β -globin gene cluster in the Japanese population. *Hemoglobin* 11:487-496.
- Smithies O, and Powers PA (1986) Gene conversions and their relation to homologous chromosome pairing. *Phil. Trans. R. Soc. Lond. [Biol.]* 312:291-302.
- Tay JSH, and Saha N (1988) Genetic heterogeneity among the Negroid and Arab tribes of the Sudan. *Am. J. Phys. Anthropol.* 76:211-216.
- Thompson EA, Deeb S, Walker D, and Motulsky AG (1988) The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. *Am. J. Hum. Genet.* 42:113-124.
- Van Valen LM (1986) Speciation and our own species. *Nature* 322:412.
- Wainscoat JS (1987) The origin of mutant β -globin genes in human populations. *Acta Haematol.* 78:154-158.
- Wainscoat JS, Hill AVS, Boyce AL, Flint J, Hernandez M, Thein SL, Old JN, Lynch JR, Falusi AG, Weatherall DJ, and Clegg JB (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. *Nature* 319:491-493.
- Wainscoat JS, Thein SL, Higgs DR, Bell JI, Weatherall DJ, Al-Awamy BH, and Serjeant GR (1985) A genetic marker for elevated levels of haemoglobin F in homozygous sickle cell disease. *Br. J. Haematol.* 60:261-268.
- Wilson DB, Garnham PCC, and Swellengrebel NH (1950) A review of hyperendemic malaria. *Trop. Dis. Bull.* 47:677-698.
- Yuregir GT, Donma O, Dikmen N, Isbir T, and Cinar M (1987) Population studies of hemoglobin S and other variants in Cukurova, the Southern part of Turkey. *Acta Haematol.* 50:757-765.