YEARBOOK OF PHYSICAL ANTHROPOLOGY 30151–194 (1987)

Genetics and History of Sub-Saharan Africa

LAURENT EXCOFFIER, BEATRICE PELLEGRINI, ALICIA SANCHEZ-MAZAS, CHRISTIAN SIMON, AND ANDRELANGANEY
Laboratoire de Génétique et Biométrie and Laboratoire de Préhistoire et Paléoanthropologie, Université de Genève, 1227 Carouge Genève, Switzerland

KEY WORDS Polymorphism, Prehistory, Africa

ABSTRACT This paper aims to review the contribution of genetic data to the prehistory and history of sub-Saharan African peoples. The authors review briefly paleontologic data, which give limited information about modem Homo *sapiens sapiens* origins and isolation of present African gene pools. Most linguistic and archaeological theories about *African* peoples' prehistory are then confronted with the most informative genetic data available. Rhesus, Gm, HLA, and DNA data are analyzed. Their frequent haplotypes are compared between populations by means of genetic distances and average linkage clustering. Despite heterogeneities between the quality and the quantity of data provided by different genetic systems, some clear conclusions can be drawn. Genetic differentiation clearly parallels the clustering of major linguistic families. These families of populations seem genetically homogeneous, suggesting either relatively recent origins or long-term important and continuous intragroup migrations. The well-known divergence between the historical theories suggested by immunological and DNA data about the relationship between Africa and other gene pools is discussed. Decisive conclusions about African origins of modem humans either from fossil or from DNA data Seem very premature. An alternative hypothesis issued from overall genetic variation is proposed.

History is clearly the major factor in the genetic differentiation of populations. A good interpretation of a continental set of gene frequency data would require information about chronology, effective population sizes, and migration throughout the process. Unfortunately, we are usually provided with scarce and biased fossil records, doubtful migration models, and purely hypothetical historical demography. Moreover, far from research centers, African population genetics has developed in an uncoordinated and random manner. Major areas and numerous populations remain untested from a genetic point of view, while some residual and atypic societies have practically become field laboratories for immunologists. It is not necessary to emphasize how hazardous and questionable any attempt at a synthesis can be. Nevertheless, it can be useful, at least as a first step toward better planning for future fieldwork, to gather complementary though not homogeneous data. In order to discern whether present populations may be linked to early African occupants, we shall consider first Pleistocene paleontological data. Second, information concerning the known history and linguistic affinities of present African populations will be presented. Third, genetic data coming from classical genetic markers and from DNA polymorphisms will be examined. Finally, these data will be analyzed together in an attempt to find any cultural link that may help to explain genetic interpopulation relationship.

For many people, Africa is the continent of human origins. We will not try to link its present inhabitants to remote origins, although some have made such attempts.

© 1987 Alan R. Liss, Inc.

The hypothesis of a continuous local human evolution for 2 million years is highly questionable and, up to now, cannot be disproved. But most authors admit that modern human populations diverged between 500,000 B.P. and 10,000 B.P. We will therefore have to take into account information about the last *Homo erectus* populations and about the subsequent Neanderthalian period. This time scale has seen the "emergence" of modern humans in a way which is still highly controversial, depending on what is considered as *H. sapiens sapiens* or *H. "presapiens."* The specific status of isolated fossils is, of course, arbitrary. Species clustering can be questioned even between H. *erectus* and *H. sapiens,* which are typological categories and not biologically different species. These considerations are important in our topic since a fairly well-known theory about the origin of modern man is that our ancestor first appeared in Africa to spread later on in the Old World (Brauer, 1984; Protsch, 1975; Trinkaus, 1986). This would explain why Africans cluster separately from others in some genetic systems (Cann et al., 1987; Wainscoat et al., 1986). Such a theory contradicts previous studies (Cavalli-Sforza and Edwards, 1965; Langaney, 1979, 1984; Piazza et al., 1981) and deserves discussion, even if the question is still open.

Another question of interest is the link between populations like those attested by archaeology to have existed in the Sahara as early as 40,000 years ago and present populations of sub-Saharan Africa. Ancient theories of various origins considered Black Africans as recent settlers following others who could be Pygmy or Khoisan ancestors. But the tendency now is to be more cautious because there is not much evidence supporting this point of view.

In such matters, we will try to describe the limited facts and mention the common hypotheses. But we must keep in mind that alternative models, very often, have not been developed. As an example, a relatively recent Middle East or Asian origin of Black Africans, which fits some immunological data, has not been discussed extensively as an alternative to the "African origin" of H. *sapiens sapiens.* Furthermore, it is also obvious that for the last 500,000 years, the choice of a gradual intraspecific model or of a punctuated equilibria theory (Langaney, 1984), which is still arbitrary, will produce very different interpretations of today's genetic diversity.

PALEONTOLOGICAL DATA

Theories concerning modern human "emergence"

The location and date of modem man's origin are still obscure today. Being sure that it is meaningful to determine the exact location where evolution toward a modem form of horninids might have taken place, many specialists have proposed a European origin for H. *sapiens sapiens.* The discovery of the first Neanderthals in Europe and more numerous European excavation sites have certainly influenced them to neglect other theories of origin.

Chronologically, H. *sapiens sapiens* succeeds *H. sapiens neanderthalensis,* which can be either "classical" in Europe or "progressive" in the Middle East or "Neanderthaloid" elsewhere. In Africa, for example, several fossils found from Morocco to Ethiopia and South Africa share a patchwork of anatomic features which prevent them from being attributed clearly to either H. *erectus* or *H. sapiens neanderthalensis* or some primitive H. *sapiens sapiens.*

Presently available datings situate Neanderthal disappearance in Europe around 34,000 and the *H. sapiens sapiens* apparition around 32,000 B.P. Over the past 20 years, African paleontology has revealed the presence of "archaic Rhodesoid" populations (African Neanderthaloids) around 40,000–30,000 B.P. Some authors (Brace, 1964; Tobias, 1961) have concluded that modern humans originated directly from the African *H. rhodesiensis,* but far later than in Europe. Others have proposed a Mediterranean origin of modern Africans (Coon, 1962; Thoma, 1973) or the existence of two independent lineages in Africa, one leading to H. sapiens *sapiens,* the other to a "Rhodesoid" population (Leakey, 1953).

Protsch (1975) has enunciated a radically different model. He postulates a unique origin for modem humans in sub-Saharan Africa and calls this ancestor *H. sapiens capensis,* a form of H. *sapiens.* This form would have been present in East Africa as early as 60.000–50,000 B.P. and would have migrated later further north out of Africa. Brauer (1984) also came to the conclusion that modern man appeared only once in Africa, but much earlier, between 100,000 and 70,000 B.P. He postulates a relatively gradual evolution of H. *sapiens* in Africa and distinguishes three main "grades" between H. *erectus* and modem *H. sapiens sapiens.*

Table 1 presents the main human fossils that were available for analysis to Protsch (1975) and Brauer (1984) before they proposed an African origin of modern man. Figure 1 shows the fossils' geographical location. One can see that the material is very sparse and limited, and thus should be interpreted very cautiously. Other problems prevent definitive conclusions to be drawn from fossil records: available samples are too small for any serious statistical study; Brauer's classification is based only on morphological characters, whose interpretation is considerably dependent on bone preservation. As far as dating is concerned, very few estimates are directly correlated with the fossils. Many of them have been produced on the basis of other chronological elements that were imprecisely associated with human remains (dating of the corresponding, superior, or inferior layer, or even the associated fauna).

These incertitudes added to sampling problems oblige us to remain doubtful about the factual basis of the most common models and hypotheses on modem human origins in Africa.

Theories concerning the apparition of modern human "types"

The origin of present-day morphologies is quite obscure. Human remains which could illustrate it are again rather sparse (Table 2) and allow only vague hypotheses. Before enunciating the fossils used for racial interpretations, it is necessary to recall

TABLE 1. The middle and upper Pleistocene horninids from Africa (after Brauer, 1984)

Specimen	Date of discovery	Dating (ky B.P.)
North		
1 Mugharet el'Aliya	1939	40–30
2 Jebel Irhoud 1,2,3,	1961	60–40
3 Temara	1975	30
4 Dares-Soltan	1975	30
5 Haua Fteah	1952	47[1]
6 Rabat	1933	200[1]
7 Sale	1971	400
8 Ternifine 1,4	1954	500–400
9 Sidi Abderraharn		300–200
10 Thomas Quarries 2	1972	
East		
11 Kanjera 1,5	1932	150–70
12 Singa	1924	120–80
13 Omo (Kibish 1)	1967	130
13 Omo (Kibish 31	1967	37
14 Murnba Rock Shelter	1977	130–110'
15 Laetoli (Ngaloba)	1976	120'
16 Diré Dawa	1933	80–40
17 Ndutu	1973	400–200
18 Bodo	1976	400–300
19 Eyasi 1,3	1935	120–40?
Olduvai H11[2]	1962	300
South		
20 Border Cave 5	1974	115–90'
21 Klasies River Mouth	1968	125–95'
22 Florisbad	1932	125–80
23 Cave of Hearths	1947	175–100
24 Hopefield	1953	600–300
25 Broken Hill	1921	200–125

[1]Absolute dating
[2]Same location as Olduvai specimen No 13 of Table 2

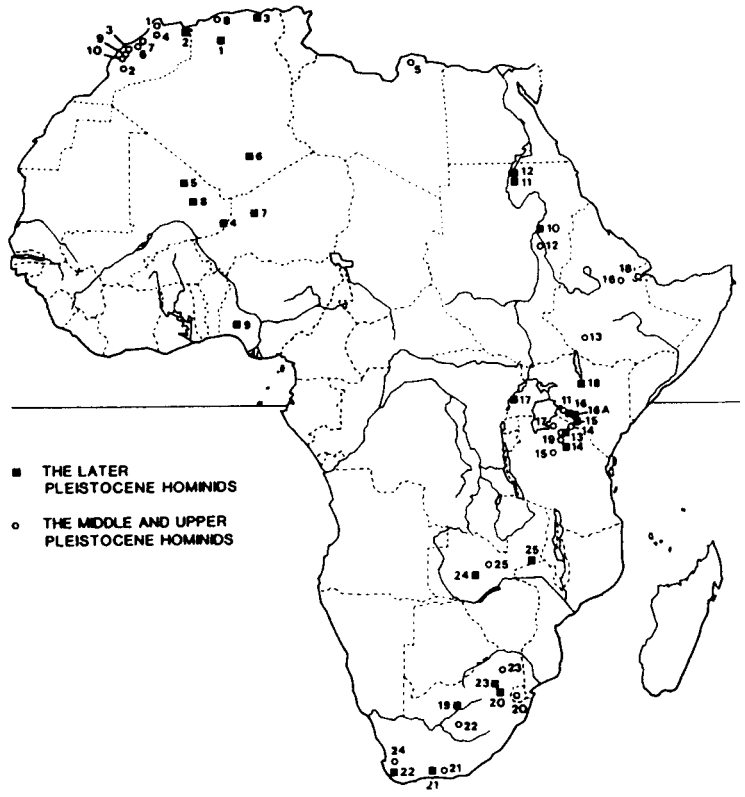

Fig. 1 Geographical location of Pleistocene hominid fossils found in *Africa*. Specimen numbers refer to Tables 1 and 2, where specimen names, dates of discovery. and datings are indicated.

two prevalent theories relative to the emergence of modern "races." The first one is the polytypic and polycentric hypothesis first described by Coon (1962), and also defended by Thoma (1973). According to Coon, modem man was "born" several times in different parts of the world. He postulates parallel but different origins of the races of modem man. The second hypothesis is monophyletic and monocentric. It sees the evolutionary process leading to modem man as unique. Present "races" would have diverged after the modem form evolved. Two subhypotheses can be advanced. Either a present-day race is older and gave birth to the others, or present-day races *are* issued from a vanished ancestral and primitive (*sensus stricto*) population.

From typological diagnosis of the few fossils presented in Table 2, a lot of authors have tried to establish a scenario for human differentiation in Africa. Among them, Brauer (1978) sums up the latest interpretations. According to him, main African morphotypes would have arisen during the late Pleistocene. Black Africans and Khoisan groups would have emerged from a common genetic pool around 15,000–10,000 B.P.

From the few fossils found, Brauer (1978) has hypothesized the following distribution of human groups by 11,000 B.P.:

TABLE 2. The later Pleistocene hominids from Africa (after Brauer. 1978)

Specimen	Date of discovery	Dating (ky B.P.)
North		
1 Columnata	1937	8–5[1]
2 Taforalt	1951	11[1]
3 Afalou-Bou-Rhummel	1928	12'
Sahara		
4 Ibalaghen	1958	
5 Tin Lalou	1960	
6 Tamanrasset	1964	7–4
7 Tamaya Mellet	1934	
8 Asselar	1927	Mesolithic
West		
9 Iwo Eleru	1965	11[1]
Northeast		
10 Khartoum	1949	7
11 Jebel Sahaba	1962	14–12
12 Wadi-Halfa	1972	Mesolithic
East		
13 Olduvai I	1914	17[1]
14 Lukenya Hill	1970	18[1]
15 Naivasha	1940	11[1]
16 Gamble's Cave II	1927	8[1]
16A Bromhead	1926	7[1]
17 Ishango	1935	7
18 Kangatotha	1965	5[1]
South		
19 Boskop	1913	34
20 Bushman Rock Shelter	1969	30[1]
21 Matjes River Cave	1930	10–9[1]
22 Fish Hwk	1927	36'
23 Tuinplaas	1929	Late Pleistocene
24 Mumbwa Cave	1923	20–18[1]
25 Kalemba 2	1971	8–7[1]
25 Kalemba 5	1971	5–4.5

'Absolute dating

- An Afro-Mediterranean stock in North Africa (Columnata, Taforalt, Afalou) and East Africa (Gamble's Cave, Bromhead, Olduvai 1, Naivasha) extending to the Great Lakes region.
- A Negroid stock in western Africa (Iwo Eleru), in the Sahara (Asselar and some Neolithic sites), in the northeast (Khartum, Jebel Sahaba, Wadi-Halfa), in the east (Kangatotha, Lukenya Hill), in Zaire (Ishango), and in South Africa (Bushman Rock Shelter).
- A Khoisan stock in southern Africa (Fish Hoek, Matjes River Cave, Mumbwa Cave) and eastern Africa covering also Tanzania and Kenya.

Brauer also observes mixed-type fossils. In Khartum, Jebel Sahaba, and Wadi-Alfa, for example, "Europoid" and "Negroid" traits have been noted. A mixture of "Europoid" and "Proto-Khoisanoid" is found in Bromhead, and finally, "Khoisanoid" and "Negroid" traits are present in Tuinplaas and Kalemba.

The location of the fossils seems to duplicate roughly the classic distribution of the current major populations. Nevertheless it seems to us that the use of morphological features established subjectively on present-day populations to interpret ancient human remains is not judicious. Morphological and metric data are indeed capable of relatively rapid change. Such traits represent mostly local adaptations to particular environments. Consequently, according to the available data, any schematic representation of an ancient population distribution would be highly hypothetical and premature.

PEOPLING HISTORY OF AFRICA

A striking feature of studies on present population affinities in Africa is the number of different peopling hypotheses that have been formulated on the basis of few and unreliable data. The human peopling history of Africa has always initiated endless debates. The first theses, appearing in books of the 19th century, tended to show that African people had no proper history. In this view, cultural evolution would have been dependent on Asian migrations into Africa. This interpretation has constituted the frame of the "Hamitic" theory, which has influenced "migrationist" theses for half a century and which is still found in recent studies.

Without going into the details of the debates which stirred the Africanist world in the 19th and 20th centuries, we must mention the very important role that historical linguistics played in the elaboration of most archaeological hypotheses. One cannot deny that archaeologists have sometimes preferred to interpret their data with a linguistic model rather than admit the lack of sufficient archaeological research. The aim of historical linguistics is to reconstruct the different stages of language evolution; so it is tempting to correlate it with cultural evolution.

The first task of historical linguistics is to establish classifications that link languages supposed to have a common origin (genetic classification). Greenberg (1963) constructed such a classification for the entire African continent.[1] He grouped all the languages into four phyla: Niger-Kordofanian, Afroasiatic, Nilo-Saharan, and Khoisan (Figs. 2, 3). The main innovations of this genetic classification are the following:

- Bantu is not considered to be a distinct language family but is related to Western Sudanic. The latter belongs to a greater group called "Niger-Congo" which includes also Adamawa-Eastern. Niger-Congo is part of the large Niger-Kordofanian family (Fig. 2A).
- The Nilo-Saharan group is formed of six branches: Songhai, Saharan, Maban, Fourian, Coman, and Chari-Nile (Eastern and Central Sudanic as well as Kunama and Berta) (Fig. 2B).

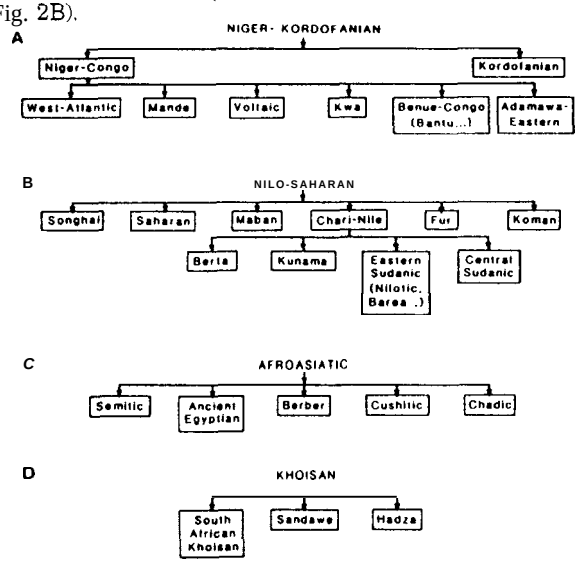

Fig. 2. The four African language phyla and their subclassifications (after Greenberg, 1963)

[1] This methodology has been described by Greenberg (1957).

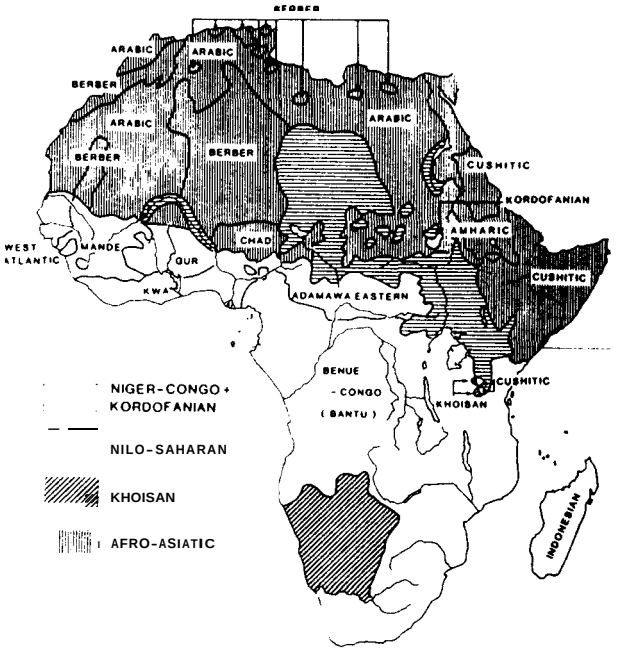

Fig 3 Linguistic map now prevailing in Africa. Modified from Greenberg (1963).

- The "Hamito-Semitic" family is reclassified as the Afroasiatic family, which includes five branches: Berber, Ancient Egyptian, Semitic, Cushitic, and Chadic (Fig. 2C).
- The Khoisan family is divided into three branches: Hadza and Sandawe in Tanzania, Khoisan in southern Africa (Fig. 2D).

Greenberg's proposals, although sometimes questioned, are generally accepted and we shall use his classification in this paper.

Beyond classification, linguistics has designed a set of methods to reconstruct the steps of language differentiation (see Gregersen, 1977, pp. 75–80, for a List of the methods). The underlying principle of these reconstructions is that the alterations affecting the languages during their evolution are universal. The task of the linguist is to identify the laws governing these modifications, to establish their relative chronology, and finally to reconstruct a hypothetical ancestral language (protolanguage). The main problem of historical linguistics is to provide "absolute" datings. This constitutes the aim of glottochronology, a method developed by Swadesh (1952). He postulated that the basic vocabulary is replaced at a constant rate. From the study of a dozen languages (most of them being Indo-European), he deduced that 20% of the basic vocabulary, whatever the language, is replaced every 1,000 years. This glottochronological method has been the subject of great controversy (see Bergland and Vogt, 1962; and Hymes, 1960, for the pros and cons), causing its datings to be more generally considered as rough estimates.

East Africa

With regard to ethnic and linguistic contacts, East Africa has played a very important role. Representatives of Africa's four major language families can be found there: Niger-Congo, Khoisan, Nilo-Saharan, and Afroasiatic. The history of

contacts between these different populations is not yet well understood, no model being favored unanimously. Nevertheless, we are going to survey the main hypotheses concerning the peopling of East Africa, focusing here on Afroasiatic and Nilo-Saharan, the two other families being studied elsewhere in this paper.

Afroasiatic is the only family which is not exclusively African. Besides Berber, ancient Egyptian, Cushitic, and Chadic, it also includes Semitic, widely spread over the Middle East and the Arabic peninsula. Fleming (1976) and Ehret (1979) have suggested a time depth for the relationships within Afroasiatic of the order of 15,000 years, and Ehret (1979) locates their origin inland from the Red Sea, somewhere between Sudan and Ethiopia. The protovocabulary of subsistence seems to indicate that the ancient Afroasiatic communities had a way of life based on the intensive use of seeds and that this advantage would have enabled them to settle in the East African Horn, in the Sahara, in North Africa, and in the Middle East.

The Nilo-Saharan language family is widely spread in the High Nile Valley, in the oriental parts of Sahara and Sudan. The Songhaï, on the Niger River, is its most occidental representative. This wide repartition incited Sutton (1974) to correlate the expansion of the Nilo-Saharan to the aquatic "civilization" 8,000–10,000 years ago. At that time, the climate was much more humid and the lakes much more numerous. A mode of life based on fishing spread in a region extending from the Atlantic Ocean to the Nile basin. David (1982) deplores the use of the word "civilization" to designate what was only a variable technocomplex, and Phillipson (1977) suggests that the aquatic tradition would have been more likely to spread toward the already-established Nilo-Saharans.

In East Africa, available genetic data concern mainly speakers of Cushite (Afroasiatic) and Nilote (Nilo-Saharan) languages. The question of their establishment in the region of the Great Lakes is usually connected with the introduction of domestication. Linguistic history (Ehret, 1974) shows the entrance of two food-producing populations in the highlands of East Africa before the arrival of the Bantu. The relative sequence of the population movements would then be as follows (Ambrose, 1984; after Ehret, 1971, 1974):

> A Southern Cushitic-speaking population, with domestic stock and possibly a knowledge of agriculture, entered northern Kenya around the fifth millennium B.P. By the third or possibly as early as the fourth millennium B.P., Southern Cushites spread into central Kenya and northern Tanzania. At a later date, a population probably located near the wmmon border between Sudan, Uganda, Kenya and Ethiopia, and speaking a Southern Nilotic language, moved to the South into the western highlands and Rift Valley regions of Kenya. They arrived in western Kenya shortly before the introduction of iron into East Africa. The Southern Nilotes kept domestic stock and may have cultivated sorghum and finger millet.

The Eastern (Masaï, Turkana, Jie, etc.) and the Western (Luo, Nuer) Nilotes constitute the majority of the Sudanese, Kenyan, and North Tanzanian Nilotes. Their establishment in the southern regions occurred only recently, throughout the second millennium A.D., and they assimilated the former Nilotes (Sutton, 1980).

It would be futile to mention here all the attempts made by archaeologists to associate the arrival of linguistic groups with particular traditions (see Ambrose, 1982, 1984; David, 1982; Lynch and Robbins, 1979; Phillipson, 1977; Sutton, 1980; Vossen, 1982). The exact sequence of the cultures which have developed in East Africa is not well-enough documented and its interpretation differs from one author to the other. Meanwhile, it is worthwhile to point out a fact which seems accepted unanimously: the existence, before the coming of the Bantu, of close and extended contacts between some Cushitic and Nilotic populations in the Great Lakes region.

West Africa

West Africa, from a linguistic point of view, is quite homogeneous. The majority of its languages belong to the same family: the Niger-Congo (see Fig. 3). In this phylum, we can distinguish four West African language groups (Greenberg, 1980):

- Mande, very unlike the others. is divided into two branches: North Western Mande (Kpelle, Soninke. Loma, Bambara. etc.) and South Eastern Mande (Mano, Bisa, Samo, etc.).
- West Atlantic is clearly divided in two branches by a Mande incursion: North Western Atlantic (Fula, Serere, Woloff, Bassari, Bedik, etc.) and South Western Atlantic (Temne, Limba. etc).
- Gur or Voltaic (Dogon, Senoufo, etc.).
- Kwa (Kru, Ewe, Akan. Yoruba. Ebo, Ibo, etc.).

In West Africa, with the exception of classification, linguistic research is just beginning. This is particularly true in the area of historical reconstruction. Actually, there is no general model describing the origin and expansion of the different families. The only indications that we have concern the differentiation period of proto Niger-Congo. estimated around 8,000 B.P. (Ehret, 1984; Greenberg, 1964). Some authors express the idea that the Mande branch was the first to separate from proto-Niger-Congo (Welmers, 1973). The homeland of proto-Niger-Congo has not yet been located, but Ehret (1984) suggested that around the seventh or sixth millennium B.P., Niger-Congo-speaking people already lived in most of the regions of West Africa that are presently inhabited by speakers of this phylum. Painter (1966) suggested that Niger-Congo language differentiation may be correlated with desertification of the Sahara, which would have triggered migrations from the Sahel to the south.

Several authors agree in situating these migrations around the fifth millennium B.P. (Cornevin, 1982; McIntosh and McIntosh, 1983; Shaw, 1980). The most common model suggests that most of the nomadic herders followed the recession of the great rivers toward the south, while others searched for refuge in the Sahara central highlands. Even if it is generally admitted that there were migrations from the Sahel to the south, opinions diverge over the impact they had on older hunter-gatherer populations living in western Africa. We do not know the nature nor the size of the migrant populations. The Saharan rock paintings, showing light- and dark-skinned people, are the only indications we have.

Archaeologic studies have brought little information on the origin and the affinities of the present-day populations. Lack of West African data is due to relative delay in investigation in that region and also to the poor preservation conditions encountered in the tropical forest.

Central and southern Africa

The less numerous, but historically very important, people of southern Africa are the Khoisan. Greenberg (1963) classifies the Khoisan language family with three branches. The first one covers language with clicks now spoken in southern Africa by around 100,000 San (Bushmen). Khoi (Hottentots), and Dama. The two others, Hadza and Sandawe, are located further north, in Tanzania. Some authors (Köhler, 1975; Tucker, 1967) do not completely agree with Greenberg's affiliation of Hadza, but generally these two groups of hunter-gatherers are considered as being the remnants of an ancient eastern African Khoisan peopling. Rock paintings having several similarities in style and technique with those of southern Africa have been discovered on the territory of these groups. Moreover, the tools that have also been found seem to have been cut following the same methods as those attributed to the Khoisan of southern Africa.

On the cultural level, two groups of Khoisan may be distinguished: the hunter-gatherer San and the pastoral Khoi. Recent discoveries in southern Africa have overthrown old theories of the Khoisan peopling. Indeed, several excavations have revealed the existence of domestic animals and pottery associated with stone tools dating from the beginning of the first millennium A.D. (Klein, 1984; Sandelowsky et al., 1979), or perhaps even from 2,100 B.P. (Walker, 1983). Study of the fauna indicates that the ancestral wild forms of domesticated stocks did not previously

exist in southern Africa. Domestication in this part of the continent was known by Late Stone Age (LSA) populations before the arrival of Early Iron Age (EIA) Bantu and not after, as has been presumed. There are still several problems to be solved—for example, the source of the domestic strains, the people involved, and the routes of probable migrations. Many possible scenarios ensue. One of them is that hunter-gatherer populations of East Africa could have learned domestication from regional pastoral populations (Nilotes, Cushites?) 2,000 or 3,000 years ago. Then, they would have moved to southern Africa, meeting the local hunter-gatherers (Denbow, 1986; Denbow and Campbell, 1986; Denbow and Wilmsen, 1986).

Meanwhile, Ehret (1973, 1982a, 1984) suggests another model. It is based on the fact that words referring to herding have been borrowed from the Central Sudanic linguistic group, now located essentially in southern Sudan, the Central African Republic, and northern Zaire. A few Khoisan communities (the ancestors of Cape Khoi and Kwadi) could have acquired knowledge of domestication in a region located around northern Botswana before moving to the Atlantic Ocean and the Cape. Ehret (1982a) suggests that Central Sudanic populations extended as far as Zambia 2,000 years ago. But, as he emphasizes himself, the Central Sudanic vocabulary used in Khoisan languages is very limited. This could suggest either a short contact period or an indirect transmission by other populations. None of the different hypotheses can be favored over any other to date.

Bantu (Benue-Congo)-speaking people, with more than 100 million locutors, constitute the most important demographic group of Africa. In spite of many Bantu languages, one observes a quite marked linguistic homogeneity among this group, suggesting a recent differentiation. This homogeneity is found neither in the physical nor in the cultural traits. However, some authors did speak of a Bantu society and of a Bantu race, falsifying the debate on language differentiation. Hypotheses have been built in order to explain the very large repartition of these quite similar Bantu languages. They have often used the concept of a single Bantu people occupying progressively all the southern part of the African continent from a single region. Models describing Bantu migration routes are as numerous as they are different. In the frame of this article, it is not possible to enumerate them all (see, for linguistic studies, Bastin et al., 1983; Ehret, 1973, 1982b; Greenberg, 1972; Guthrie, 1962; Heine, 1973; Nurse, 1982; and for archaeological studies, David, 1980; de Maret, 1985; Huffman, 1970; Inskeep, 1976; Phillipson, 1975, 1977; Soper, 1971, 1982; Vansina, 1984). Nevertheless, one can distinguish some general lines accepted by the majority of linguists and archaeologists:

- Proto-Bantu speakers would have lived in the Middle-Benue Valley between Nigeria and Cameroon (between 5,000 and 3,000 B.P., depending on the authors) and, from there, would have spread in sub-Saharan Africa.
- The classification of Bantu languages differentiates the Western from the Eastern Bantu.
- The close similarity of Eastern languages to each other suggests a relatively recent and rapid expansion (in comparison with Western languages).
- The people the Bantu encountered in the equatorial forest were the Pygmies and in southern Africa, the Khoisan.
- In East Africa, the Bantu were probably in competition with Nilotes, and probably Cushites, but the nature and the exact location of the contacts are not known.

For dating the different steps, hopes have turned to archaeology in order to find, buried in the soil, the trace of Bantu expansion. The most generally admitted hypothesis was that metallurgy in eastern, central, and southern Africa came after Bantu expansion. However, facts connecting metallurgy diffusion to Bantu expansion are now disputed. The previous theory was built on the discovery of a 25th-century B.P. iron-working site (Nok culture) in Taruga (Nigeria), situated very near the presumed original Bantu site. This is contrast with an absence of metallurgy in the subequatorial region before our era. Thus, it was deduced that diffusion originated in Nigeria and spread toward eastern and southern Africa. More recent findings have shown that iron working was known at the same period, or even earlier, in the interlacustrine region in East Africa (Van Grunderbeek et al., 1983). Comparative linguistic studies (de Maret and Nsuka, 1977) also challenge the idea that proto-Bantu knew metallurgy before their expansion. In consequence, newer hypotheses suggest that in their first expansion the early Bantu did not possess iron technology. The currently favored model is (Vansina, 1984) that as early as the fourth millenium B.P., the Bantu penetrated the rainforest as yam-growers with Neolithic tools.

One can see that attempts to connect Bantu language expansion to the diffusion of some cultural traits, such as metallurgy, have not succeeded. The failure comes from the lack of data but also from the fact that the hypothesis stating that all subequatorial African colonization was done by a single population may be erroneous. Few authors have assumed the existence of populations other than Pygmies and Khoisan before the advance of Bantu speakers. Lwanga-Lunyiigo (1976) suggests "that the Negroes or their forebears were widespread from early times and did not come from one small region." This hypothesis has still to be proved like others, because the poor quality of fossils in the area does not allow arguments for or against this thesis (see theories concerning the emergence of different human types in Africa above). On the other hand, the notion of a single population expansion cannot be abandoned because no other language is spoken in subequatorial Africa. The Pygmies, for example, surely once lived in the whole equatorial forest, although their original language has not persisted.

GENETIC POLYMORPHISMS IN *AFRICA*

Genetic markers and available data

Population geneticists interested in population settlement history have few data at their disposal. These were not collected for such purposes and are poorly suited for genetic affinity calculations between populations. Most of the presently known genetic markers have been studied for medical reasons. Thus they are rarely obtained in sampling designs that fit perfectly with population genetics constraints.

A population history study, using genetic markers, supposes that the populations involved have had a relatively stable history since their common hypothetical origin. The resulting differences found should mainly reflect mechanisms based on migrations, mutations, and genetic drift. It is worth noting that one of the important components of evolution such as selection presents obstacles to historicogeneticists.

What are the characteristics of genetic markers that render them efficient in telling population history? If we are interested in a global study of possible links between a great number of populations, it is important to use regular phenotype systems, typed in most populations. Markers should not be under strong selection, in order to avoid adaptive convergence, or divergence from differential selective pressures on populations living in dissimilar environments. The markers must also be informative in allowing the characterization of each population in contrast with the others. One would prefer genetic systems having a great number of different alleles or haplotypes rather than less polymorphic systems, but, on the other hand, low-haplotype frequencies cannot be estimated with sufficient precision. These contradictory constraints make the systems having intermediate numbers of common haplotypes the most informative. There must exist no ambiguity in the definition of the marker, its characterization, and its typing. One must be sure that a clearly defined marker is the same in every population under study. It is imperative to differentiate between identity and isoaction. Unfortunately, this criterion may not be realistic presently as most allelic identification techniques are indirect. Furthermore, current DNA polymorphisms are ambiguous when not based on direct sequencing techniques. Finally. the typing should not be too costly if one hopes to carry out a valuable study on a great scale.

One can easily realize that these criteria are not satisfied and that the ideal marker does not really exist. Actually, every marker at our disposal has some disadvantages; that's why it is wise to use not only one but several.

Making a good choice of population also raises some practical problems. In fact, it is completely determined by the amount and the quality of available data. First of all, it is worth noting that many reports concern populations which are not representative of the largest African groups. Khoisan, for example, have been studied extensively (Botha et al., 1972; Jenkins, 1972; Jenkins et al., 1970; Nurse et al., 1975; Steinberg et al., 1975; Weiner and Zoutendyk, 1959; Zoutendyk et al., 1953, 1955), because their particular way of life engenders great interest. Nevertheless, their genetic characteristics may permit us to learn a lot about their likely isolation from other Black African populations or their different origin.

Extensive data, which could give us a deep insight into African genetics and furthermore into the origins and migrations of African peoples, are often missing. DNA polymorphism and HLA data overviews show how limited our investigation can be. For example, only a single study has reported HLA gene frequencies in West Africa (Woimant et al., 1980); only one has done so for East Africa (Spees et al., 1975). The numerous Bantu nations are represented by only a few small or heterogeneous samples (Festenstein et al., 1972; Govaerts et al., 1972), in contrast to six studies on Khoisanoid populations (Botha et al., 1972; Nurse et al., 1975). Such limitations prevent us not only from realizing a complete overview of the genic frequency variations inside Africa for a single system, like HLA, but also from combining the existing data from many systems to get a unique genetic distance picture of the whole continent. Each system will thus be studied separately, and the set of results, together with linguistic information, will then be compared.

Data of good quality are all the more precious because large populations are poorly represented. Unfortunately, we must add that samples are often not representative for a given population, as they are sometimes defined with a linguistic connotation that is too general—for example, "Bantu" (Festenstein et al., 1972). Gene frequencies obtained from such samples are not proper to any of the studied populations, whose linguistic affinities do not necessarily imply strong genetic similarities. If even a single one of these ethnic groups were genetically particular, the calculated frequencies would lose any meaning. Logically, each of them should be taken separately, but sample sizes would become far too small to be statistically significant. Fusing populations, which do not belong to the same linguistic family, is evidently not of great use for an historicogeneticist as they likely do not share the same history.

Other limitations which frequently occur concern sample sizes. In samples that are very small (all too common), only the highest gene frequencies are correctly estimated, while rare genes are often not even detected. On the other hand, large samples may be inconsistently defined. "Mixed population," "donors," "patients," "students," "aged 50," "males," "controls," or "recruits," as unique characterizations of the samples, render them, a priori, ethnically heterogeneous. They cannot be integrated in a study whose aim is to discover the possible origin of each population.

Finally, a lot of data are statistically refutable when recomputed by standard maximum likelihood programs. Antigen typing errors seem to be responsible for underestimated phenotypic frequencies as revealed by the usual chi-square tests, especially for the Rhesus system. Such data cannot be used unless careful correction, based on precise typing hypotheses, are made when no other equivalent sample is available (Sanchez-Mazas, 1985).

All these problems considerably restrict the amount of data actually computable. The samples we have finally chosen (listed in the appendix and geographically localized on Fig. 4) satisfy conditions of sue, ethnic homogeneity, and antigen typing. For each one, we have recomputed gene frequencies from published phenotypic distributions by means of a standard maximum likelihood program. Samples for which chi-square values were significant have been rejected. We have also defined the linguistic family of the ethnic groups, as well as their size, when some estimations were available. Focusing on historical genetics, we decided to exclude systems that were not sufficiently informative, as well as those in which frequency distributions seem to be selected, and those for which too few or too small samples were

Fig 4 Geographical location of all sub Saharan populations studied in this paper The most important North African samples are also shown

available. The beta^A-globin gene cluster and mitochondria1 DNA polymorphisms, for example, have not been studied with a sufficient number of samples to estimate African diversity. The sample sizes in these studies were also too small to obtain reliable estimates of gene frequencies. These two systems may give only limited indications of the relations of some African population with some other populations in the world.

Genetic distance and data representation

To preserve some biological meaning of similarity measurements and avoid random fluctuations of distance estimations, caused by the lowest genic frequencies, we will compare frequency distributions by means of the percentage of isoactive genes, or P, shared by two populations, according to

$$_2P = \sum_{k=1}^{m} \min(f_{ik}, f_{jk}) \cdot 100 \qquad ..$$

where f_{ik} and f_{jk} are the k^{th} frequencies observed in populations i and j, respectively, in a gene distribution defined by m elements (Sanchez-Mazas et al., 1986). The formula can be generalized to a set of n populations, according to

$$_nP = \sum_{k=1}^{m} \min(f_{ik}, \ldots, f_{nk})$$

The complement to 1 of $_2P/_{100}$ is a mathematical distance (Gregorius, 1978) whose equivalent formula

$$D_{ij} = 1/2 \sum_{k=1}^{m} |f_{ik} - f_{jk}|$$

is close to the *mean* character *difference* distance coefficient first described by Cain and Harrison (1958). This metric distance can be used for principal coordinate analysis (Gower, 1966) and dendrogram construction. The average linkage method seems *to* be the most appropriate clustering procedure (Wilmink and Uytterschaut, 1984), as it is not dependent (like minimum and maximum linkage methods) on extreme frequency distributions found in some populations. Moreover, it does not require average gene frequency calculations (as in the centroid linkage procedure), which would have no real biological meaning. Principal coordinate analysis has been shown to be equivalent to principalcomponent analysis of product-moment correlation coefficients (Gower, 1967). The main difference resides in the fact that principal coordinate analysis is referred to as a "Q" technique, which is here a study of similarity between pairs of populations, whereas principal component analysis is an "R" technique, which would study similarities between pairs of population characteristics.

Rhesus system

System definition

The Rhesus system is mainly known for its clinical importance, as it is responsible for haemolytic disease in newborns. The antigen D implied in this foetomaternal incompatibility is located on the erythrocyte surface, as are other numerous Rh specificities. Antigens C, c, E, and e are the best known, but more than 15 have been defined, which render the Rhesus system very complex. Rhesus antigens can be detected by means of hemoagglutination tests, specific Sera being isolated from the blood of multiparous women or polytransfused individuals. Such tests lead to phenotypic panels which, at a population level, allow the estimation of the corresponding Rhesus gene, or haplotype, frequencies.

Rhesus genes have been localized on the distal portion of the short arm of chromosome 1 (Marsh et al., 1974; Ruddle et al., 1972). Two theories have been proposed for its molecular structure. According to Fisher (1944), the Rhesus system includes three loci, D, C, and E, having alleles D and *d* (the latter being recessive), *C* and c (codominants), E and *e* (also codominants), respectively. Wiener (1943), on the contrary, considered one single locus having pleiotropic effects. These two theories gave the two usual nomenclatures. Both models can be considered as equivalents in our work, since they do not influence the interpretation of phenotypic observations. We will thus use the two nomenclatures simultaneously. Actually, Fisher's three loci are closely linked, and the resulting haplotypes follow a one-locus Mendelian transmission. Only one crossing-over event has been observed, though with some reservations (Steinberg, 1965). These rare recombination events, however, may be responsible for different haplotype origins in the human species, as suggested by Fisher (1947, 1953) and Fisher and Race (1946). The three most common European haplotypes, $R^1(CDe)$, $R^2(cDE)$, and $r(cde)$, could have given the less frequent $r'(Cde)$, $r''(cdE)$, $R^0(cDe)$, and $R^z(CDE)$ by single crossing-over. The order of loci would then be DCE. The $r^y(CdE)$ haplotype would have required two recombination events from the original ones *to* appear. This would explain its extreme rarity in actual populations. A complementary molecular model involving an operon regulative structure has since been proposed (Rosenfield et al., 1973).

Methodological constraints

Rhesus gene frequency estimations face precise antigen typing problems, added *to* the usual difficulties such as blood transport and its preservation from haemolysis. Antigen e, for example, often Seems difficult to detect, as attested by statistical incoherencies in phenotypic distribution analyses, when some {E} phenotypes appear *to* be overrepresented in comparison with their corresponding combination {Ee}. Many population samples have to be rejected when evidence for typing error arises from usual test procedures (Sanchez-Mazas, 1985). Incorrect antigen identifications (like D^u, see below), poor estimations, and the limited number of studies of rare variant frequencies also oblige us to consider only the seven well-known haplotypes $R^1(CDe)$, $R^2(cDE)$, $R^0(cDe)$, $R^z(CDE)$, $r(cde)$, $r''(Cde)$, and $r''(cdE)$.

D" problem

Unusual features found in the reactivity of some samples, when immunological tests were performed with anti-D sera, resulted in an official specification D" for those samples which reacted with some anti-D but not all (Stratton, 1946). It first seemed that there was a new specificity or allele at the D locus, but now D" is considered to be a result of some position effect (Ceppellini et al., 1955). In other words, the problem is far from being solved, and the degrees of reactions between different D" can be so different from one sample *to* another that it is wiser not to attempt interpopulation D^u frequency comparisons. Phenotypes including D" (found in a lot of Black African populations) are either neglected or added to their equivalents including only D or d. Neglecting them would inevitably introduce a bias in the statistical estimations. As some anti-D sera react with D" blood samples, we cannot consider that there is a recessive allele at this locus as long as the position effects are not completely elucidated. It is for these reasons that we have pooled D^u phenotypes with their D-positive equivalents.

Rhesus genes in the world and African features

A common feature of all Black African populations is the very high frequency of the $R^0(cDe)$ haplotype, which is not at all or only weakly present in other parts of the world. $R^1(CDe)$ frequencies are generally low compared to the other populations of the world, and even fall to zero in some groups. The $r(cde)$ gene is common in North Africans, Europeans, Near-Easterners, and Indians, whereas it is generally absent in Asiatics, Oceanians, and Amerindians. Low frequencies are found for $R^2(cDE)$ and $r'(Cde)$. Finally, $r''(cdE)$ and $R^Z(CDE)$ appear only exceptionally.

North Africans and Berbers are similar to Europeans and Near-Easterners according to most genetic systems (particularly Rhesus, Gm, and HLA). We did not include them in our study, as their past has to be related to "Caucasoid"[2] settlement history.

Rhesus genes in sub-Saharan Africa: ethnic differences

Figure 5 shows, on the left side, a dendrogram with 30 Black African populations, constructed according to the Rhesus frequencies represented on the right side. Figure 6 represents most of the good samples found in the literature concerning the Rhesus system in Africa. It must be regarded as complementary to the dendropam, which is less precise for detailed clusterings.

---

[2] "Caucasoid" (like "Negroid" and "Mongoloid") is an obselete term which used to refer to the traditional "races" formerly baaed on morphological characters and skin color studies. Because no other word can designate at the same time North African. European, Near-Eastern. and Indian populations. which are genetically close, we will use "Caucasoid" throughout this study to name the whole of them. But this has to be interpreted carefully. By opposition, "Black Africans" means all African populations not belonging to the Caucasoid group. i e , all but North Africans. "African Blacks" will not be used. as it emphasizes the skin color similarity with other black populations of the world (for example. Melanesians), which is not suitable in *our* genetic study.

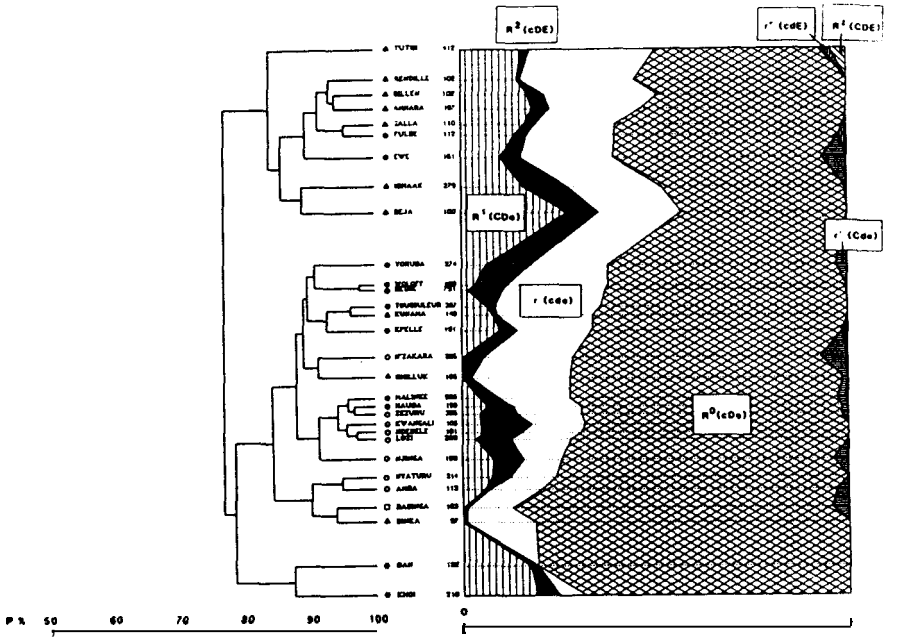

Fig. 5. Rhesus haplotype frequencies (cumulated) in 30 sub-Saharan African populations (right side). Sample sizes are indicated in front of ethnic names. General linguistic families of the samples are mentioned with the following symbols: A: Afroasiatic; A: Nilo-Saharan; O: Bantu and Adamawa-Eastern; ●: "West African" (Kwa + West Atlantic + Mandé); ∗: Khoisanoid; O: "Pygmy." The dendrogram is constructed according to the *average* linkage procedure and uses $1 - {}_2P/100$ distances (left side). A permutation procedure lowers the sum of distances between neighboring populations without modifying the dendrogram structure (Sanchez-Mazas et al., 1986). The proportionality of these distances is also conserved on the vertical axis.

Genic frequency differences inside *Africa can* be evaluated according to three genes or haplotypes: $R^0(cDe)$, $R^1(CDe)$, and $r(cde)$. The other genes always show low frequencies throughout the continent, which neither vary enough to raise particular features nor are sufficiently precisely estimated, due to small sample sizes.

Two main sets of populations *can* first be distinguished in Africa. Eastern ethnics belonging mainly to the Cushite linguistic family possess Rhesus frequencies close to Caucasoids, which make them appear as an intermediate group between the latter and Black Africans (top cluster of Fig. 5). Most other Black African populations, except the San, Khoi, and to a certain degree Pygmies, share more homogeneous distributions. Low $R^1(CDe)$ frequencies never exceeding 0.15, very high $R^0(cDe)$ frequencies between 0.55 and 0.80, and middle-ranged $r(cde)$ frequencies characterize a West-South cluster consisting of all Bantu subfamilies and West Atlantic, Kwa, Mande, Nilo-Saharan, and Chadic populations.

A rather clear distinction may further be made between Bantu and West Africans. As a matter of fact, they seem to form two overlapping but different clusters. This is especially evident in the principal coordinate analysis (Fig. 6). Turning to the gene frequency schema, one *can* discern higher $r(cde)$ frequencies in most West populations and lower ones in most southern ethnics composing Bantu tribes.[3] At first

[3]Haplotype frequencies used for the Karanga, Manyika, Ndebele, and Zezuru are those computed by Lowe (1969). Frequencies computed from published phenotypic distribution give such unusual values for Black Africans that we preferred not to use them.

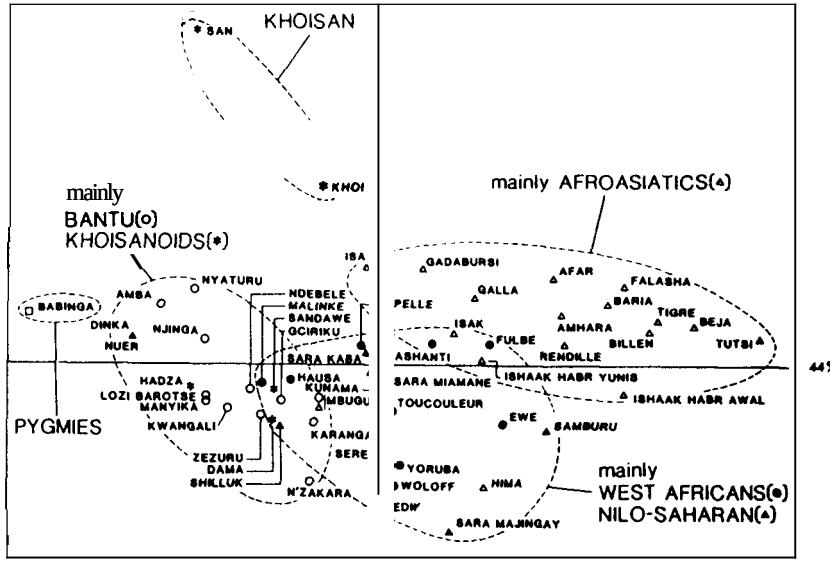

Fig. 6. Principal coordinate analysis for 53 sub-Saharan African populations. using $1 - {}_2P/100$ distances computed from Rhesus haplotype frequencies. General linguistic families of the samples are mentioned with the following symbols: A: Afroasiatic; A: Nilo-Saharan; O Bantu and Adamawa-Eastern; ●: "West African" (Kwa + West Atlantic + Mandé); ∗: Khoisanoid: 3:"Pygmy." Mbugu linguistic affinity is not clear, as two identically named populations exist.

sight, these genetic differences seem to be related to the geographical location of the populations (see Fig. 4). Whether they are more likely to be linguistic divisions will be discussed when more detailed comparisons are done (see below).

Concerning East Africans, the genetic data support quite clearly the main ethnic groupings defined by the linguistic approach. All genetically well-differentiated East Africans belong to the Cushite and Ethiosemite linguistic families. Nilotes (Dinka, Nuer, and Shilluk) as well as Central Sudanic Sara and Kunama (all belonging to Nilo-Saharan families) are quite distinct. The former are more closely related to the Bantu and the latter two are integrated in the western cluster.

Genetic results also completely concur with two other specific ethnic distinctions inside Africa, commonly known as Pygmies and Khoisan. In Figures 5 and 6, Pygmies (Babinga) are situated opposite the Cushite and Ethiosemite cluster and emerge from the Bantu cluster, as they have the highest $R^0(cDe)$ frequency (0.83) and the lowest $R^1(CDe)$ values (as an absence of this gene is not certain). The Nilotic Dinka and Nuer (genetically similar) come unexpectedly close to Babinga.

Khoisan are genetically linked to West and South African populations, due to their high R^0 frequencies. They are nevertheless somewhat particular, showing high R^1 frequencies like Afroasiatics and strikingly low $r(cde)$ frequencies. Their frequency distribution is consequently quite uncommon and makes them constitute a very differentiated cluster. Other "Khoisanoid" populations like Dama, Sandawe, and Hadza, on the contrary, possess common Bantu frequencies.

In some cases, genetic and linguistic subdivisions (and also geographical location) seem not to coincide if we refer to the main groups that we have just distinguished. However, they are interesting enough to be emphasized, as it will be possible to make some hypotheses about their history considering both their linguistic and genetic characteristics (see Discussion). These peculiar populations are Fulbe (Fulani or Peul), whose frequencies are similar to Cushite and Ethiosemite. and also to the

Ewe and Ashanti from Ghana; Tutsi, who are also connected to this eastern cluster in spite of their southern localization and Bantu language; and Hima, genetically far from their linguistic Bantu family and from the Nilo-Saharan populations, although these latter show an important genetic heterogeneity.

Gm *system*

System definition

The Gm system describes a polymorphism of gamma-immunoglobulins (IgG). IgGs are constituted of four polypeptide chains, two heavy and two light chains, linked by disulfide bridges. The variable parts of these seric proteins play a role in recognition reactions of foreign molecules. One may distinguish many specificities of IgG thanks to the antigenic properties of the constant parts of their heavy chains. These antigenic structures or allotypes are revealed with sera possessing specific antibodies directed against them. More than 20 different Gm allotypes have already been discovered.

IgG may be further divided into four subclasses or isotypes [IgG(1), IgG(2), IgG(3), and IgG(4)], each individual having all of them. The allotypes are not carried indifferently by any isotypes. Clear associations have been found between allotypes and isotypes. Table 3 shows this association and the correspondence between the numeric and alphanumeric nomenclatures commonly used.

The serum of an individual is tested for a series of allotypes. The combination of positive allotypes defines the Gm phenotype of this individual. From the phenotypic distribution found in a population assumed to be in Hardy-Weinberg equilibrium, one can estimate the presence of Gm haplotypes and their frequency.

At the DNA level, the different classes and subclasses of the constant parts of heavy-chain (IgCH) genes are clustered on chromosome 14 (Ellison and Hood, 1982; Flanagan and Rabbits, 1982; Kirsh et al., 1982). IgGCH genes all have three exons (CH1, CH2, CH3) and a smaller hinge region between CH1 and CH2. Allotypic Gm markers are located on CH1, CH2, or CH3 exons of gamma 1, gamma 2, and gamma 3 genes. No allotype has been yet recognized on gamma 4. Knowledge of the molecular structure of this cluster has provided an explanation of some Gm haplotype occurrences obtained either by single-point mutation or recombination between frequent haplotypes (Helal et al., 1981; Lefranc et al., 1982; Migone et al., 1985). Because Gm allotype characterization is not yet possible from DNA, typing still has to be done by immunological reactions. Consequently, direct knowledge of haplotypes present in a population is still impossible. Typing errors (faulty sera) and restricted number of tested allotypes cause the haplotype list of a population to be

TABLE 3. **Numeric** *and* **alphanumeric** *nomenclatures* **for** *the* **Gm** *system (allotypes mentioned in this study only)*

Isotype	Numeric	Alphanumeric
G1m	(1)	(a)
	(2)	(x)
	(3)	(f)
	(17)	(z)
G2m	(23)	(n)
G3m	(11)	(b0)
	(5)	(b1)
	(13)	(b3)
	(14)	(b4)
	(10)	(b5)
	(6)	(c3)
	(24)	(c5)
	(21)	(g)
	(15)	(s)
	(16)	(t)
	(28)	(Ll)
	(5,10,11,13,14) (5*)	(b*)

insufficiently reliable and often oblige us to make hypotheses about the actual presence of some haplotypes.

Gm haplotypes in the world and African features

Table 4 summarizes the Gm haplotype occurrences in the main worldwide groupings. Several Gm haplotypes seem to be proper to Black African populations. The most frequent one is $Gm^{1,17;;5,10,11,13,14}$, whose frequencies reach more than 0.80 in some ethnic groups like the Sara Majingay of Chad. Though it first seemed to be present in other parts of the world, and especially in Oceania, with very high frequencies in Australian Aborigines and Papuans, further studies using new antiallotypes suggested that non-African $Gm^{1,17;;5,10,11,13,14}$ is different. In Oceania, it appears to include the allotype G2m(23), but probably not in Africa, as it is shown at least in a study by van Loghem et al. (1978) of 214 Nigerian individuals. Unfortunately, allotype G2m(23) is rarely typed[1] and should be considered from now on as an important specificity. What seems also to differentiate the Black African haplotype from the non-African is its likely close association with the A2m(2) allele, whereas the non-African one (also found in Europe and Asia with low frequencies, generally less than 0.5) seems to be linked to A2m(1). It is to be hoped that more work will be done in order to have this problem elucidated, but enough information is nevertheless available to let one consider $Gm^{1,17;;5,10,11,13,14}$ as a Black African specificity.

Almost all Black African populations possess the two haplotypes $Gm^{1,17;;5,6,10,11,14}$ and $Gm^{1,17;;5,6,11,24}$, generally not found in other populations which appear to lack allotype G3m(6). This also characterizes Black Africans.

Finally, the haplotype $Gm^{1,17;;13,15}$ (or probably $Gm^{1,17;;10,11,13,15}$) has always been considered as typically Khoisan, but it is found in many other Black African populations with significant frequencies, as we shall discuss below. Outside of Africa, this haplotype is always associated with G3m(16) and thus cannot be considered as similar.

Thus, at least four haplotypes characterize Black African populations. The important discovery of the allotype G3m(28) by Blanc et al. (1976) and its typing in African populations (Rivat et al., 1978) reveal the striking feature that it is not systematically associated with G3m(21) as seems to be the case in Caucasoids and Asians. Its typing in Black Africans leads to the identification of two distinct

TABLE 4 Most common Gm **haplotypes** *found in the* **main worldwide groupings** *(occasional and* **rare haplotypes** *not* **mentioned)**

Groups	Very frequent	Common
Black Africans (including Khoisan)	$Gm^{1,17;;5*}$	$Gm^{1,17;;5,6,10,11,24}$ $Gm^{1,17;;5,6,11,24}$ $Gm^{1,17;;10,11,13,15}$
Europeans, Near Easterns Indians. North Africans	$Gm^{3;;5*}$	$Gm^{1,17;;21}$ $Gm^{1,2,17;;21}$
North East Asiatics	$Gm^{1,17;;21}$	$Gm^{1,2,17;;21}$ $Gm^{1,17;;10,11,13,15,16}$ $Gm^{1,3;;5*}$
Amerindians	$Gm^{1,17;;21}$	$Gm^{1,2,17;;21}$
South-East Asiatics	$Gm^{1,3;;5*}$	$Gm^{1,17;;21}$ $Gm^{1,2,17;;21}$
Melanesians	$Gm^{1,3;;5*}$ or $Gm^{1,17;;21}$ or $Gm^{1,17;23,5*}$	$Gm^{1,2,17;;21}$
Australian Aborigines	$Gm^{1,17;;21}$ or $Gm^{1,17;23,5*}$	$Gm^{1,2,17;;21}$

[1] As G2m(23) is not tested for the majority of the samples we have omitted mention of it in the complete haplotype descriptions. ";;" standing for this omission

$Gm^{1,17;;5,10,11,13,14,-+28}$ haplotypes having different proportions of G3m(28) according to the ethnic groups considered (Rivat et al., 1978). Other haplotypes like $Gm^{1,17;;5,6,10,11,14}$ or $Gm^{1,17;;5,6,11,24}$ are also split into two subspecificities G3m(+28) and G3m(−28). But not enough populations have been tested at the moment to use these more detailed characteristics in a population genetics study.

Gm haplotypes in sub-Saharan Africa: ethnic differences

North African people and the Tuareg present some Caucasoid specificities such as the presence of the haplotype $Gm^{3;;5,13,14}$ and a low frequency of $Gm^{1,17;;5,10,11,13,15}$ which make them appear again as intermediate populations.

Figure 7 presents Gm frequencies and the corresponding dendrogram structure for 27 Black African populations. One can immediately distinguish three main clusters. East African populations, as with the Rhesus system, show intermediate characteristics between Caucasoids and Black Africans (top cluster). They possess the $Gm^{3;;5,13,14}$ and $Gm^{1,17;;21}$ haplotypes as Caucasoid populations do. Black African specificities like $Gm^{1,17;;5,10,11,13,14}$ are in any case very frequent. $Gm^{1,17;;5,6,10,11,14}$ and $Gm^{1,17;;5,6,11,24}$ are also present. Surprisingly, Khoisan populations show some genetic affinities with East Africans (second cluster). Even if they are not extremely close to each other, the dendrogram shows them as a cluster. These similarities are due to comparable $Gm^{1,17;;5,10,11,13,14}$ frequencies and to the presence of the haplotype $Gm^{1,17;;21}$ in the Khoisan. The other Black African populations are rather homogeneous. However, some differences emerge between a southern African group consisting of the Nguni (Zulu, Xhosa, Pondo) and Sotho (Pedi, Bechwana) populations (third cluster), and the other western, southern, and southwestern African populations (fourth cluster). This distinction is principally due to the rather high incidence of the $Gm^{1,17;;10,11,13,15}$ haplotype in Nguni and Sotho. This haplotype appears to be very frequent in Khoisanoid populations. Pedi and Xhosa also have the $Gm^{1,17;;21}$ haplotype in common with the Khoisan.

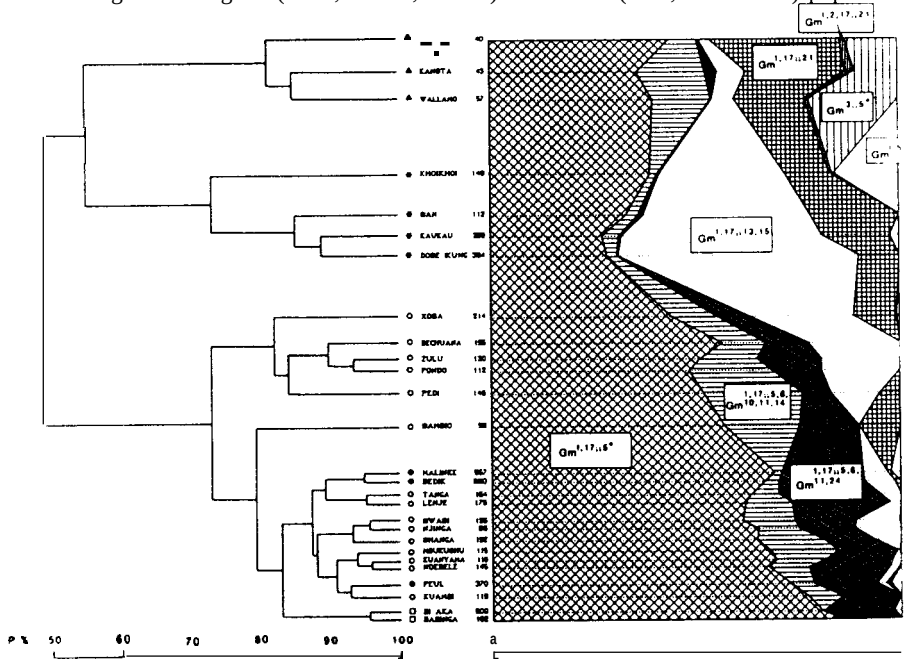

Fig. 7. Gm haplotype frequencies (cumulated) in 30 sub-Saharan African populations (right side). Sample sizes are indicated in front of ethnic names. General linguistic families of the samples are mentioned with the following symbols: A: Afroasiatic; A: Nilo-Saharan; O Bantu and Adamawa-Eastern; ●: "West African" (Kwa + West Atlantic + Mandé); ∗: Khoisanoid; O "Pygmy." The dendrogram is constructed according to the average linkage procedure and uses $1-{}_2P/100$ distances (left side). A permutation procedure lowers the sum of distances between neighboring populations without modifying the dendrogram structure (Sanchez-Mazas et al., 1986). The proportionality of these distances is also conserved on the vertical axis.

Thus, in addition, with quite clear distinctions of the main linguistic groups already seen with the Rhesus system, we observe a link between Khoisan, East Africans, and some southern Bantu tribes, whereas central and western Bantu seem quite different. These latter do not differ greatly from western Africans like the

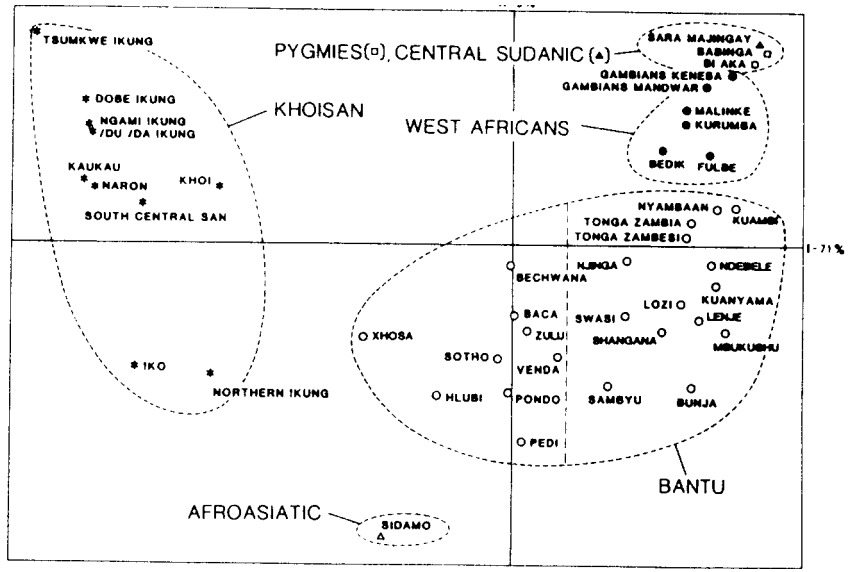

Fig 8 Principal coordinate analysis for 43 sub Saharan African populations. using $1-{}_2P/100$ distances computed from Gm haplotype frequencies General linguistic families of the samples are mentioned with the following symbols △ Afroasiatic; A Nilo Saharan, ○ Bantu and Adamawa Eaetern. ● West African" (Kwa + West Atlantic + Mandé); ∗ Khoisanoid Pygmy" The Bantu and Adamawa Eastern cluster may be clearly subdivided into two parts the left part includes all Nguni Sotho Venda populations except Ndebele

[5]As the three Sidamo samples have small sizes (40, 43, and 57 individuals) and belong to the same Sidamo family their phenotypic distributions have been grouped before computing their total haplotype frequencies This new sample 140 individuals. has been used for the principal coordinate analysis

Malinke or Bedik. Pygmies do not show great differences in contrast to West Africans, except the absence of $Gm^{1,17;;5,6,10,11,14}$ and a higher $Gm^{1,17;;5,10,11,13,14}$ frequency, which detach them slightly from the latter group. The Ndebele and the Swasi, two northern Nguni tribes, have a lower frequency of the $Gm^{1,17;;13,15}$ Khoisanoid haplotype. This makes them cluster with the non-Nguni Bantu populations.

HLA system

System definition

The HLA system is responsible for organ transplant rejection in humans. HLA antigens are subdivided into two main classes. Class I antigens are located on every one of our cells and are encoded by three allelic series, A, B, and C. Class II antigens are lymphocyte-specific and correspond to three series, DR, DQ, and DP (the latter two constituting the D specificities). A third class has been defined for the complement complex formed by seric proteins (see Dausset and Pla, 1985, for a review).

The HLA linkage group has been localized on the short arm of chromosome 6 (McKusick, 1978; Raum et al., 1976; van Rood et al., 1976). Each identified locus includes a considerable amount of alleles (20, 40, 8, 12, and 10 specificities have been defined for loci A, B, C, D, and DR, respectively), and it is presumed that more will be discovered. Because HLA loci are closely linked, each individual transmits his genes as haplotypes. Allelic variation is considerable; hence the probability for two unrelated individuals to be genetically similar is extremely low. However, some haplotypes appear to be more common in human populations. In spite of the HLA system's complexity, population specificities can be observed, thus promoting this system as very informative in population genetics. Unfortunately, population phenotypic panels obtained by immunological reactions do not yet permit a rigorous estimation of haplotype frequencies, unless elaborate family studies are performed. For the moment, this obliges us to consider single allele frequencies in large studies by taking the loci one by one. Maximum likelihood procedures, as well as other methods, can be used to estimate gene frequencies (Mattiuz et al., 1970; Williams, 1982), but phenotypic distributions are almost never published. This may prove to be a source of mistaken estimations, and it also prevents us from controlling for possible typing errors.

Methodological problems

HLA data from the African continent are unfortunately far less numerous than Rhesus or Gm data. As previously mentioned, western and eastern African populations are underrepresented. Moreover, the smallness of sample sizes has dramatic effects on gene frequency estimations for the HLA system, since most of the alleles have frequencies ranging below 0.1 or even 0.01. It is not surprising to find an absence of some specificities in many samples (like A11 or B22, for example), but we cannot conclude that they are actually lacking in these populations. HLA data raise another important problem, which is the occurrence of blank alleles, or in other words, of undefined specificities in the sample. Their frequency may reach as much as 0.22 for the HLA-A locus (in the Nigerian sample) and 0.47 for the HLA-B locus (in the Xhosa sample). Some important *African* allelic types seem not to be discovered yet, certainly due to the non-African origin of most antisera used for HLA antigen typings. The consequence is that distance measures should be corrected to take into account these meaningless frequencies. An acceptable hypothesis is to consider that two populations share a similar proportion of undefined alleles as they do of defined ones. Thus, the definition of P becomes:

$$_2P_{HLA} = [\sum_{k=1}^{m-1} \min(f_{ik},f_{jk})] + \min(f_{im},f_{jm}) \sum_{k=1}^{m-1} \min(f_{ik},f_{jk}),$$

where f_{im} and f_{jm} correspond to the blank frequencies of populations i and j. This can be considered as a good estimate of the common gene percentage shared by the two populations.

One of the reasons for the smallness of sample sizes is related to the needs and interests of medical research for the HLA system. Previous reports have shown high correlations between some HLA specificities and particular pathologies, like the well-known relationship between B27 and ankylosing spondylitis (Brewerton et al., 1973). Thus, a lot of studies are performed in order to find new associations with diseases. This is the case for studies carried out in eastern (Bodner et al., 1975; Hall et al., 1982), western (Famuyiwa et al., 1982), and southern Africa (Briggs et al., 1980; Hammond et al., 1977, 1980). In spite of the different aims pursued in such studies, they are interesting for us in that they contribute to the elucidation of the question of whether the HLA system is selected or not. For the moment, however, no conclusive answers have been advanced for excluding HLA data for such reasons in population studies.

Another problem resides in the fact that control samples in such studies cannot be used, because they are ethnically heterogeneous. We hope that future investigations. even if they are done for medical interest, will provide usable data for African ethnic comparisons. Finally, we have also been obliged to reject samples for which some interesting specificities were not tested. This is the case for the Shi sample not tested for B16, B18, and B21.

HLA genes in the world and African features

Some alleles seem to have specific frequencies in Black African populations compared with other ethnic groups. At the A locus, A11 is very rare or absent, whereas it is always present elsewhere and may even reach very high frequencies in Orientals (about 0.30 in Chinese and Vietnamese) or in New Guinea (0.59 in Kar Kar). On the other hand, A30 has a rather high frequency in Black Africans, reaching 0.22 in Zulu and 0.25 in San, compared to 0–0.05 in other parts of the world. Aw23 is generally more frequent than Aw24, thus contrasting with other world populations. Some recently defined alleles, though not tested in most populations, could also be typically Black African: Aw36 (or MO) found in Sarakole, Zulu, Xhosa, and especially in Nigerians (0.082), or Aw43 (or BK) found in Xhosa, Central !Kung, and Khoi. It would be necessary to test these alleles in other parts of the world in order to confirm or refute their specificity to African populations.

At the B locus, allele B17 is particularly frequent, compared to other populations. Its frequency ranges from 0.15 to 0.25, reaching even 0.37 in Central !Kung, whereas it generally does not exceed 0.10 in other populations except in Indians (0.14 in Hindus, 0.25 in Koya). B22 is often absent, though it is almost always present elsewhere except in many Amerindian populations. Finally, Bw42 (or MWA) could be typically Black African, as it has been found with appreciable frequencies in every Black African population so far tested, excluding Tigrinya. Its frequency is extremely low elsewhere. However, as it is again a recent discovery (like Aw36 and Aw43), it needs to be more extensively tested to confirm its peculiar distribution in Black Africans.

HLA in sub-Saharan Africa: ethnic differences

Within Africa, the Black African features we have just described are less apparent in the Erithrean Tigrinya sample. A11 reaches a frequency of 0.03 instead of being absent. A30 and Aw23 frequencies are really low (0.007). Some other alleles have an atypical frequency in this population: A1 frequency (0.12) is close to those found in Europe (0.10–0.25) and in North Africa (Egyptians have a 0.22 A1 frequency); A9 frequency is much lower than in other Black Africans (0.02 compared to 0.09–0.17); B12 frequency is also lower; B21 is particularly high (0.21), closer again to North African values (0.10–0.15 in Egyptians, Kabyls, and Tunisians). The Sarakole tribe is another population which shows an uncommon African frequency: B17 is much rarer than in other populations (0.05 compared to 0.15–0.25).

Figure 9 presents a principal coordinate analysis of distances among ten populations for the HLA system. The principal axes do not account for much of the total squared distance from the centroid, reflecting the fact that many factors must be considered in order to interpret HLA data. Nevertheless, it appears that the Tigrinya are extremely differentiated in one direction, and are situated opposite the Bantu populations, to which Pygmies are related. Khoisan populations are grouped alone in a single quadrant and show a relative closeness of both San samples, while the Khoi Nama stand closer to Bantu. West Africans seem somehow intermediate between East and southern African populations.

Although this topology is to be taken very carefully, it seems that the populations belonging to the main linguistic groups are clustered according to these two principal axes. This is in rough concordance with the Rhesus and Gm systems. Paradoxically the HLA system is not very informative, at the African continental level, for genetic population history, despite its numerous different alleles which should point out precise differentiations between populations.

DNA polymorphisms

Sampling problems in DNA polymorphism studies

Due to the complexity of DNA sequence analysis, actual data are often drawn from samples very different in size and representativeness compared with those constituted for more classical genetic markers (see Table 5). Moreover, the number of alleles, haplotypes, or DNA types found in such samples may be considerably fewer than those effectively present in the whole population. Ewens (1972) has shown that the number of alleles present in a sample may permit, at the level of a population, the estimation of a mutation parameter referred as teta (teta = $4Nu$ for nuclear genes and teta = Nu for haploid mtDNA types (Birky et al., 1983)), N standing for the effective population size and u being the mutation rate). Teta values vary greatly according to the genetic system under consideration and become espe-

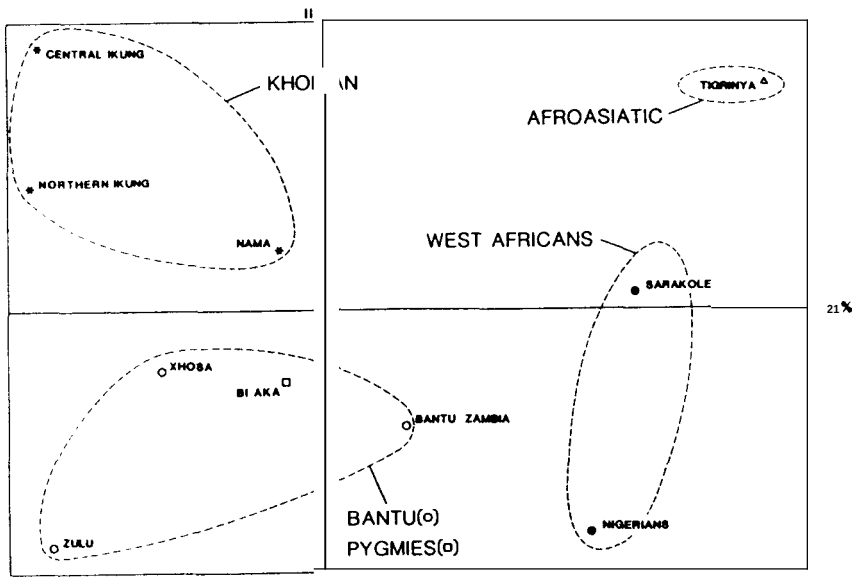

Fig 9 Principal coordinate analysis of ten sub-Saharan African populations, using 1 – $_2$P/100 distances computed from HLA (·A and ·B) gene frequencies. Samples have been chosen according to the number of tested specificities 11 for HLA·A (A1, A2, A3. A9, A10, All, A28, A29, A30+31, A32, A33) and 15 for HLA B (B5, B7, B8, B12, B13, B14, B15, B16, B17, B18, B21, B22, B27, B35. B40)

TABLE 5 Different sampling characteristics of the genetic systems used in this study

Genetic system[1]	No of samples studied	Mean No. of genes in samples	Mean No. of haplotypes or DNA types in sample	Mean teta[3] value (estimated)
Rhesus (7 hapl.)	30	439.4 (322.6)[2]	4.5 (0.86)	0.6 (0.13)
Gm (10 hapl.)	27	439.6 (447.6)	5 (1.41)	0.8 (0.38)
HLA·A (15 spec. + blank)	10	204.4 (57.3)	12 (0.82)	2.7 (0.28)
HLA·B (15 spec + blank)	10	204.4 (57.3)	13.3 (1.25)	3.2 (0.56)
Beta^A-globin (5 sites)	16	53.5 (27.96)	4.5 (1.21)	1.2 (0.55)
mtDNA (68 sites)	10	64.2 (33.2)	11.2 (3.16)	4.3 (1.72)

[1]Each system has been screened lor a partial set of possible specificities. haplotypes. or DNA types. Thus, teta values may constitute an underestimation ol the actual mutation parameters It is worth noting that the average teta values on a set of populations have no clear meaning but help to evaluate the required size of a sample for collecting the majority of the alleles (see Fig 10)
[2]Numbers in parentheses stand lor standard deviation
[3]For definition. see text

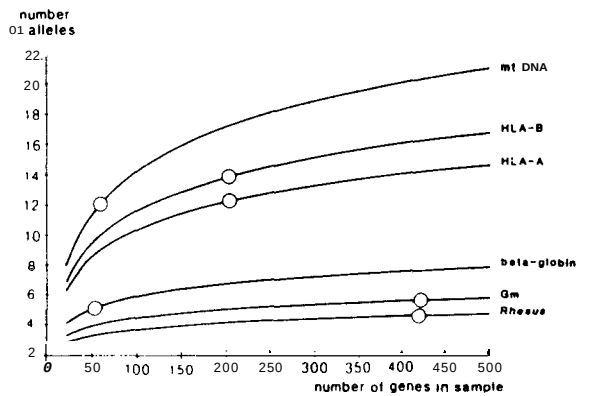

Fig 10 Expected number of alleles, haplotypes. or DNA types as a function of the number of genes in a random sample drawn from a stationary population. The curves have been computed on the basis of mean teta values tabulated in Table 5 and Ewens's (1972) equation 11 Circles are located on the vertical of the mean sample sizes in the different systems.

cially large for very polymorphic systems such as **HLA,** or DNA segments studied with several endonucleases. Thus, large samples are required in order to get the majority of the alleles present in the population and to allow reliable estimates of allele frequencies.

From Figure 10 and Table 5, one can see that Rhesus and Gm systems are generally studied from numerically good samples as the doubling of their size would not significantly increase the number of detectable alleles. The beta^A-globin gene system, when only five recognition sites are considered, would be better studied with samples of more than 200 genes per population instead of 55. As for the HLA system, samples of 200 genes are revealed as insufficient for the detection of all tested specificities. The case of mitochondria1 **DNA** (mtDNA) is critical from a sampling point of view. As the mean teta value is particularly high when 68 cutting sites are

considered, a mean sample size of 64 genes appears as inadequate for a complete sampling of the mtDNA types, when we see that a collection of 500 genes would almost double the number of detectable types. Future studies of DNA segments will probably be done on an increasing number of endonuclease recognition sites, or even on direct DNA nucleotide sequences. As a consequence, teta values will also increase and sample sizes will necessarily be much larger than those commonly found today. This material problem may limit the usefulness of too precise genotype characterization in population genetics studies and lead to studies of interindividual divergences which cannot give information on population differentiation.

Beta-globin gene cluster

Hemoglobin is surely the protein which has most predominantly contributed to the comprehension of general mechanisms in physiology, biochemistry, and genetics. Its study at the molecular level may also promote it as an important genetic marker in anthropology.

Hemoglobin protein is a tetramer composed of two alpha chains and two beta chains. The alpha-globin gene cluster is coded on chromosome 16 whereas the beta-globin gene cluster is on chromosome 11. Different duplicated genes are expressed during the ontogeny of the individual (Honig and Adams, 1986).

DNA mutations in these complexes tend to provoke the appearance of more or less serious disorders in oxygen transport, resulting in different kinds of anemias. The persistence of some of these mutations in many populations of the world has led to the consideration of it as a real polymorphism. The observed correlation between the presence of the mutations leading to HbS, HbC, and HbE and the worldwide malaria repartition has led to the formulation of the hypothesis of a selective advantage for the heterozygotes carrying one of these mutations (for a review see Cavalli-Sforza and Bodmer, 1971).

Before beta-globin genes were studied, the system formed by HbA (adult hemoglobin), HbS, and HbC (its two more common variants in Africa) was seen as an example of polymorphism of three alleles at one locus (Cavalli-Sforza and Bodmer, 1971). The AS genotype is supposed to protect from *Plasmodium falciparum*, which is widely present in Africa, whereas AC genotype would protect from *P. malariae*, which is confined in West Africa (Livingstone, 1967). A unique origin for HbS supposes complex migratory movements in order to explain its vast distribution covering India, Saudi Arabia, and Africa (Lehmann, 1954; Livingstone, 1967). Geographic and climatic factors have been suggested as determinants for the understanding of the observed frequency variations (Charmot and Lefèvre-Witier, 1978; Lefèvre-Witier, 1985).

HbS. The first studies on the presence or absence of cleavage sites for the *Hpa* I restriction endonuclease in the beta-globin gene cluster have shown the existence of a polymorphism in American Blacks and a strong association of HbS with a particular haplotype (Kan and Dozy, 1978). In spite of the small size of the samples, the association of beta' with a 13.0-kb DNA fragment in West Africa (Ghana, Burkina Faso) would justify a possible origin in this geographic region. Another associated 7.6-kb DNA fragment present in Central Africa (Gabon), eastern Africa (Kenya), as well as in Saudi Arabia and India has constituted a base for the hypothesis of multiple mutation origin of $beta^S$ (Kan and Dozy, 1980). Small sample studies in Algeria and Morocco showed the same association of beta' as the one found in western Africa (Mears et al., 1981a). The West African group has also been shown to be heterogeneous for the *Hpa* I polymorphism, indicating that although the samples were polyethnic there is a difference between neighboring populations from Togo and Ivory Coast (Mears et al., 1981b).

The use of a larger number of restriction enzymes has made possible the definition of a greater amount of $beta^S$-associated haplotypes. Studies of Jamaican Blacks (Wainscoat et al., 1983) and American Blacks (Antonarakis et al., 1984) have respectively shown the existence of two and three frequently associated haplotypes, as well as many rare haplotypes supposed to be recombinants 5′ to the beta-globin gene.

The HbS multiple origin hypothesis prevails over the model of a common and very old origin (Solomon and Bodmer, 1979) when it is shown that the more frequent haplotypes cannot be deduced from the others by less than two crossing-over events. It is highly probable that these alleles have been the subject of independent mutations or genic conversions around the sixth codon of the beta-globin gene (Antonarakis et al., 1982, 1984). Moreover, the beta' mutation is found on three different frameworks of beta-globin which certainly appeared before the divergence of the great continental groups, as they occur in these groups in nonnegligible frequencies (Orkin et al., 1983).

Pagnier et al. (1984) have shown that these haplotypes could be attributed to three different regions of Africa. The first haplotype, called "Benin" after the location of its discovery, is also found with high frequencies among sicklemic patients in Algeria, Nigeria, Mediterranean European countries, and southern Saudi Arabia (Kulozik et al., 1986) and at lower frequencies in Senegal and the Central African Republic. The presence of a second haplotype called "Senegal" and, until now, apparently confined to this country, confirms West African heterogeneity in this system. Finally, a third haplotype called "Bantu." found in the Central African Republic, may characterize Bantu populations, as it is also found among sicklemic patients of southern African origin (Ramsay and Jenkins, 1986). We can note that these three hypothetical centers are located in regions possessing very high HbS frequencies (Cavalli-Sforza and Bodmer, 1971; Lefèvre-Witier, 1985; Livingstone, 1967, 1985). It would be interesting to determine which haplotypes are present in other HbS-rich regions like Gabon, Angola, Uganda, or Madagascar, in order either to find new haplotypes or to trace possible migration routes.

HbC. All associated haplotypes with $beta^C$-globin seem to be on the same DNA fragment concerning the *Hpa* I polymorphism, and thus to come from the same single origin (Boehm et al., 1985; Kan and Dozy, 1980). The more frequent haplotype is different from fragments commonly associated with beta', although the "Benin" haplotype also seems to carry the $beta^C$ mutation at lower frequencies (Boehm et al., 1985).

HbA. The frequenc of restriction sites in the beta-globin gene cluster differs greatly between beta' and its variants because of evident selection and "hitchhiking" effects (Orkin and Kazazian, 1984). Restriction fragment-length polymorphism (RFLP) studies of $beta^A$ genes are fewer than those of pathological genes and have been used for continental group comparisons (Long and Chakravarti, 1986; Wainscoat et al., 1986). This polymorphism could be very useful in anthropology provided some improvement in sampling design can be made and the proof of selective neutrality can be assured.

A difficulty stemming from such studies is the choice of representative populations for each continent. Unfortunately, this choice is rarely deliberate, but rather is dependent on the available data. A criterion for representativeness could be that the chosen population be close to all other populations of the considered group, as attested by other genetic systems. It seems to us that this criterion is not fulfilled for present studies on $beta^A$-globin DNA haplotypes. African continental diversity still has to be estimated on the basis of this polymorphism.

We have seen that the size of some samples is sometimes dramatically insufficient for the inference of precise haplotype frequencies, especially for African samples: 47 Bantu genes, 42 San genes (Ramsay and Jenkins, 1986), 26 genes of students in Nigeria, and 35 Nigerian genes (Wainscoat et al., 1986) constitute the only samples from which conclusions about African ancestry could be drawn.

Moreover, $beta^A$-globin genes do not constitute an independent genetic system from $beta^S$- or $beta^C$-globin genes in Africa and $beta^E$-globin genes in Asia. Frequency calculations of DNA haplotypes should not be made without care for the Hb variant genes, particularly if these ones are known to be selected. In these conditions of sampling deficiencies and nonattestation of selective uniformity, any attempt to trace human history in Africa or elsewhere from $beta^A$ haplotype frequencies would appear to us quite premature.

Mitochondrial DNA

The whole sequence of some 16,500 base pairs (bp) of the human mitochondrial DNA (mtDNA) is known (Anderson et al., 1981). Investigation of the variability of this extranuclear maternally inherited DNA has begun at a small scale with RFLP studies which show possible gain or loss of endonuclease cleavage sites. These changes are assumed to be due either to substitutions or length mutations. The RFLP techniques may be used in larger-scale surveys than direct sequencing of DNA, but they are much less informative. Sequencing studies have indicated higher point mutation (Brown et al.. 1979, 1982) and length mutation (Cann and Wilson, 1983) rates in mtDNA than in nuclear DNA. This evolutionary acceleration, whose causes are yet not clearly established, is interesting for anthropology: short divergence times, such as those thought to have occurred in the human species during the last 500,000 years, could be estimated.

Though very attractive in revealing population affinities, mtDNA studies have not brought decisive information to human settlement history. One of the reasons resides in the quality of the samples. Extraction of a large amount of mtDNA from placentas (Cann et al., 1982, 1987; Horai and Matsunaga, 1986; Horai et al., 1984) allows the examination of a greater number of restriction sites than a small amount of mtDNA extracted from peripheral blood, but the samples obtained cannot pretend to represent a homogeneous population. Thus, interpopulation mtDNA polymorphism studies have often been undertaken on poorly representative samples (Aquadro and Greenberg, 1983; Brown, 1980; Cann et al., 1987; Denaro et al., 1981; Johnson et al., 1983) and they have mainly confirmed the enormous diversity of mtDNA types present in each population (Horai and Matsunaga, 1986; Horai et al., 1984). Another important sampling problem occurs from the small size of the samples, as has already been noted.

The time elapsed since a fictive original monomorphic mtDNA population gave rise to actual polymorphic populations varies greatly according to the authors: 10,000–50,000 years (Denaro et al., 1981), 104,000 years (Johnson et al., 1983), 143,000–285,000 years (Cann et al., 1987), 180,000–360,000 years (Brown, 1980), or even 550,000 years (Cann et al., 1982). This incertitude mainly comes from the lack of information on the exact mechanisms for the high evolutionary rate observed (Avise and Lansman, 1983; Monnat and Loeb, 1985; Vawter and Brown, 1986) in mtDNA.

Nevertheless, we may note that, at the great continental group level, the studies of Denaro et al. (1981), Johnson et al. (1983), and Cann et al. (1987) agree in asserting that the mean number of restriction site differences is higher for people of African descent. Denaro et al. (1981) also suggest a greater affinity between San and Aka pygmies than with a South African Bantu sample, which contradicts blood group studies.

These results, based on data which derive from samples that would have been rejected in any other classical genetic marker study, have to be interpreted very cautiously. It would be wise to accumulate data on larger and more representative samples before drawing definitive conclusions. Founder effects, similar to those observed in Amerindians (Wallace et al., 1985), may also have occurred in African history and thus have eliminated some mtDNA alleles.

Other issues still have to be resolved to allow use of mtDNA in reconstructing history. A constant molecular clock, the necessary condition for mtDNA divergence studies to be plausible, seems to be missing in one San population, as suggested by Johnson et al. (1983). In this case, the utility of mtDNA RFLP studies for short evolutionary periods should be reconsidered. Maternal inheritance of mtDNA leads researchers to ignore paternal lineages, which may not be identical to maternal lineages. So. different evolutionary paths may have been followed by nuclear DNA and mtDNA, and thus their study may not lead to the same conclusions (Poulton, 1987).

A fact of importance which has been considerably neglected is the possibility for the mtDNA to be selected, most of its coding genes being highly vital for the respiratory chain. Moreover, interaction of the mitochondrial genotype on nuclear DNA cannot be excluded in man, as it has been observed recently in yeast (Parikh et al., 1987). When one considers mtDNA type frequencies found in some recent comparable studies (Bonné-Tamir et al., 1986; Brega et al., 1986a,b; Johnson et al., 1983; Wallace et al., 1985), one is obliged to remark on the high incidence of type number 1 (according to Johnson and collaborators' nomenclature). This fact may be explained either by a very recent divergence of most of the human populations or by some selective advantage carried by this mtDNA type.

In conclusion, it appears that present data on mtDNA polymorphism brings very few indications concerning affinities between African populations. On a worldwide scale, small nonhomogeneous samples and large interindividual differences lead to problematic data interpretation. As an example, intrapopulation genetic distances may be found to be greater than interpopulation distances (Cann et al., 1987). Given the enumerated ambiguities hanging over this system, available data are still difficult to interpret at a population level. It would appear wise to wait for large-scale population genetics studies, as has been done in other genetic systems, before drawing drastic conclusions from mtDNA.

DISCUSSION

Africa in the world

Despite the fact that numerous Plio-Pleistocene hominid fossils were found in Africa, one cannot put forward any evidence of their contribution to the direct descent of today's sub-Saharan Africans. Such hypotheses cannot be tested with the scarce and ambiguous remains found from the last 100,000 years. The origin of modern *H. sapiens sapiens* gave birth to quite different theories:

(1) Polytypic and polycentric origin long ago.
(2) European origin from *H. sapiens neanderthalensis*.
(3) African origin and later colonization of Europe and Asia.
(4) Middle East or West Asiatic origin.

It seems obvious to us that available data do not warrant a definitive decision for any one of these proposed theories. Nevertheless, genetic data permit an attribution of quite different likelinesses for each.

Hypothesis 1 supposes very unlikely convergences from very old *H. erectus* to anatomically very similar *H. saprens sapiens* on four different continents. Population genetics has provided arguments against this point of view (Langaney, 1979; Nei and Roychoudhury, 1974, 1982). Genetic drift alone would have created much more differentiation than that observed between present humans, had they been separated for more than 200,000 years. The polycentric theory seems almost refuted today.

Hypothesis 2: Many fossils in the Middle East and in Africa suggest that *H. sapiens,* either *sapiens* or *"presaprens."* were living in these areas long before the "apparition" of modern *H. sapiens* in Europe *30,000* years ago. Consequently, European Neanderthals seem to be a local specialized race or species and should be rejected as a direct ancestor of *H. saprens saprens.* At most they could be contributors to modern gene pools. if not intersterile with our ancestors. A European origin of modern man is quite unlikely.

Hypothesis 3 recently received quite determined support from some molecular biologists working on DNA RFLP and especially the beta-globin gene cluster (Wainscoat et al., 1986). They did not hesitate to claim that molecular evidence supports an African origin for our species. Cavalli-Sforza et al. (1986) tempered this judgment by underlining the apparent contradiction between immunological and DNA information. It was shown long ago (Cavalli-Sforza and Edwards, 1965) that erythrocyte and seric blood groups suggest a primary divergence between Occidentals and Oriental ancestors, and then between Caucasoids and Black Africans (Langaney, 1979, 1984; Sanchez-Mazas et al., 1986) (Fig. 11). DNA information evokes the divergence first

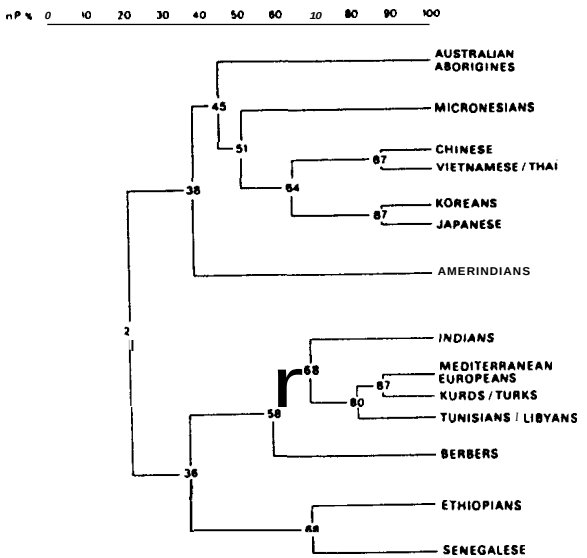

Fig. 11. Centroid cluster dendrogram for 14 world populations, using average $1 - {}_2P/100$ distances on five genetic systems: ABO. MNSs, Rhesus. Gm, and HLA (-A and -B). Branch lengths correspond to nP values (see text) computed at each level for a set of clustered populations and averaged for the five systems (Sanchez-Mazas et al., 1986).

between Africans and others. However, these findings must be questioned both for sampling reasons and because Wainscoat's arguments use principally hemoglobin cluster data which could have been shaped by selection rather than by history. Divergence time obviously cannot be inferred from hemoglobin data and it is questionable that phylogenic pattern could be, as hitchhiking effects prevent restriction polymorphisms from staying neutral in a highly selective genomic environment. Nevertheless, preliminary data on other DNA polymorphism also seem to support this hypothesis. The non-DNA polymorphism we reviewed give us an argument against this theory: if all the populations of the world, including sub-Saharan Africans, issued from the same ancestors, one could expect that these latter would still possess most of the alleles or haplotypes found elsewhere in the world. We have seen that West African, Bantu, and Khoisan were very different from Caucasoids or Orientals. Moreover, their gene pools could not be suspected to have produced, for example, a Middle East variation in a reasonable period of time, whatever the mechanisms. The only candidates for representativity of common ancestors would be the North-East Africans who are in genetic continuity with North Africans and Middle East peoples. Going so far obliges us to tend toward hypothesis 4, if one realizes that theories about gene pools can propose phylogenies but cannot localize ancestors geographically. Besides, some specific Oriental genes or haplotypes are present in the Middle East or India while none of them has been found, at least with Asian structure and noteworthy frequencies, in Black Africa. So, the only way to preserve this hypothesis in its original form would be to postulate that present Black Africans have lost their entire genetic heritage, except for some hemoglobin and other DNA features. This can still be held since the fourth hypothesis, too, supposes quite a lot of genetic drift between common ancestors and Black Africans.

Hypothesis 4 is supported by immunological data and challenged by limited DNA polymorphism data. As we previously pointed out (Langaney, 1979, 1984), a late origin and divergence of Black Africans is possible if both intensive selection and genetic drift occurred after a drastic founder effect and a somewhat long isolation of

very limited populations. Without being able to decide from available data, we will still favor this view, despite its complexity, as long as statistically acceptable molecular evidence will not confirm Wainscoat's proposals.

Nevertheless, some African peculiarities may have appeared only recently, concerning mainly selected characters. It is well established in Africa and in other intertropical regions that malaria is responsible for maintaining some advantageous alleles under the heterozygous form: G6PD deficiency, HbS, HbC, HbE (Livingstone, 1967; Cavalli-Sforza and Bodmer, 1971), beta-thalassemia (Livingstone, 1967), and alpha-thalassemia (Flint et al., 1986). The same disease under its hyperendemic form has also affected the Duffy system and the haptoglobin system. The genotype FyFy has been shown to be resistant to *P. vivax* (Miller et al., 1976), and the HpO phenotype appears to be due to malaria (Trape et al., 1985). Some systems such as ABO or HLA are known to be selected in some way, but the mechanisms involved have not yet been explained. Serious diseases exist in Africa that could have influenced genetic systems by hitchhiking or other interaction effects. Under these conditions, local singularities may be developed in few generations.

African subdiuiswns and their *interrelations*

From the genetic systems study, it seems that a quite clear distinction may be made between East Africans, West Africans, Pygmies, Bantu, and Khoisan. These genetic differentiations agree mostly with linguistic rather than with purely geographical divisions. Though distinct, these groups have certainly experienced some interactions during past periods. Migrations, sometimes very recent, have also contributed to a certain degree of genetic and cultural exchange. In the view of actual genetic and historical data, it seems possible to formulate certain hypotheses concerning these events.

East African Cushites and Ethiosemites, which both belong to the Afroasiatic linguistic family, appear to constitute a rather homogeneous and well-differentiated genetic group. In the Rhesus, Gm, and HLA systems, these populations show some Caucasoid characteristics as well as their undeniable African traits. This hypothetical Caucasoid contribution is also reflected in their physical traits and their language. Surprisingly, they appear to be closer to Khoisan than to Bantu, according to the Gm system. This affinity will be discussed below. On the other hand, Chari-Nile populations (Nilo-Saharan), mainly studied for the Rhesus system, may be differentiated genetically into at least two distinct groups. The first group consists of Nilotes (Eastern Sudanic) having a southern extension as far as Kenya, and the second incorporates Central Sudanic people extending westward to Chad and the Kunama situated in Erytrea. Samburu represent a recent southern Nilotic migration, and probably for that reason share genetic traits in the Rhesus system with the Cushite and Ethiosemite. The other Nilotic populations—Dinka, Nuer, and Shilluk, situated above the Great Lakes—seem to have been submitted to a strong founder effect as they apparently lack the R^1 haplotype. A likely random convergence might have made the former two resemble the Pygmies. The Kunama and Central Sudanic Sara are genetically quite close together and possess some West African traits. A Nilo-Saharan linguistic and technocultural continuity has been postulated for populations situated in the Sudanic zone 8,000–10,000 years ago (Sutton, 1974). These populations may have constituted a relatively homogeneous substrate for gene flows between East and West Africa, but further evidence for acknowledging them as part of this missing link has to be collected from other genetic systems. A genetic affinity exists between Fulbe and East Africans due to high R^1 frequencies. Ba and Dieterlen (1961) have suggested that the Fulbe could have had a Saharan origin, as they have recognized their initiation rituals in rock paintings of the "Bovidian" era (7,000–4,000 B.P.). Tauxier (1937) has supported a theory proposing migrations of Fulbe tribes from the eastern to the western coast, but no historical data exist to support it. A recent migration moved the Fulbe back again from Atlantic Ocean to eastern countries, crossing sub-Saharan regions. Another hypothesis which emerges from Rhesus data is a possible contact between Fulbe and Ghanaian tribes (Ewe, Ashanti).

Indeed, it is interesting to note that the only way through the tropical forest from the sub-Saharan savanna to the sea is located precisely in the Ghanaian region. Thus, Fulbe could have brought their pastoral customs to this region and mixed with local tribes. Other data in the Gm and HLA system need to confirm these possibilities. The Toucouleur are supposed to be partly from Fulbe origin (Murdock, 1959).Their affinity with Kunama in the Rhesus system perhaps reveals indirectly the relationship between Fulbe and East African populations. The Tutsi and Hima are quite close to Cushites and Ethiosemites. This affinity is in agreement with their supposed history, as they migrated probably from the North (Ethiopia?) at the end of the 13th and beginning of the 14th centuries (Liesegang et al., 1979). This may be confirmed by the fact that the sickle-cell trait, almost absent from East Africa, is very rare among them, too (Livingstone, 1967), although they are surrounded by Bantu possessing it with a frequency exceeding 0.30, one of the highest in Africa. As previously seen, West African populations often seem to have some genetical traits in common with East Africans, particularly with Cushites and Ethiosemites (higher *r* frequencies, lower $Gm^{1,17;;5,6,10,11,14}$ frequencies, presence of $Gm^{1,17;;21}$). Apart from the fact that they share a linguistic kinship with the Bantu, West Africans also present genetically clear relations with them. For the Rhesus system, for example, the Ndebele are very similar to the Bedik, Yoruba, and Toucouleur, which have quite high *r* frequencies. The Gm system also presents the same pattern concerning West Africans and Bantu affinities, but allows even more precision by showing that West Africans are closer to southern Central Bantu than to the Nguni, Sotho, and Venda (see Fig. 6). Among West Africans a distinction is allowed on the basis of beta^S-globin DNA haplotypes. A different mutation has appeared and has been maintained in Senegal and in the Benin-Nigeria regions. They may constitute specific markers for recent migrations between and from these areas.

Pygmies, although recent Bantu speakers, appear quite peculiar, for they share affinities with West Africans in the Gm, with Bantu in the HLA, and with Nilotes in the Rhesus systems. As residual populations of a much wider group, extending probably throughout the equatorial forest in a geographically central part of Africa, Pygmies, should have been in contact with most of the other great groups at one time or another.

Bantu-speaking populations, extending from sub-Saharan Central Africa to southern Africa, appear genetically quite homogeneous. This fact suggests that their expansion was quite recent and rapid, as is suggested by linguists. Bantu speakers are often found genetically close to western Africans. This is in concordance with their linguistic affiliation and the assumed Bantu homeland located on the Nigeria/Cameroon border (Greenberg, 1964). Every genetic system shows a clear separation between Cushite and Ethiosemite populations on one hand and Bantu on the other, although they have certainly been in contact during the expansion of the latter. Gm data suggest a quite fine distinction between Nguni, Sotho, and Venda populations, which share $Gm^{1,17;;13,15}$ with the Khoisan, and northern Bantu tribes which are closer to western Africans (Fig. 6). This haplotype has even been presented as a percentage estimator of San admixture (Jenkins et al., 1970). In our view, such a hypothesis supposes that gene frequency fluctuations have nothing to do with genetic drift, but it may be regarded as an indicator of some past contacts between Bantu and Khoisan populations. This distinction among Bantu is not perceptible with Rhesus data, for Nguni or Sotho tribes have not been tested with all antigens in this system. Further studies would be necessary in order to confirm this result.

Genetic data, without any doubt, always separate Khoisan from other African populations. Although presently forming a small, almost genetically uniform group, they are supposed to have had a wide northern extension (Nurse et al., 1985), prior to Bantu migrations. In comparing them with African populations other than the Bantu, the Khoisan show some affinities with Cushitic and Ethiosemitic populations. This is exemplified by the presence of the $Gm^{1,17;;21}$ haplotype specific to East African populations. East African-Khoisan affinities, as well as Nguni-Sotho-Khoisan

ones, may be explained by the following hypothesis: Khoisan were present early in southern and eastern countries of Africa, as suggested by linguistics and some archaeological discoveries. There they could have had contacts with some East African tribes from whom they aquired the $Gm^{1,17;;21}$ haplotype. Eventually they moved southward, pushed by Bantu or followed by them at a later date. The Nguni and Sotho acquired the $Gm^{1,17;;10,11,13,15}$ haplotype as well as residual frequencies of $Gm^{1,17;;21}$—which is of East African origin. Complete absorption of Khoisan tribes would explain their current absence in eastern coastal regions of southern Africa where the Nguni and Sotho are living. Consecutive Bantu migrations toward the south brought tribes like Kuambi, Kuanyama, and Sambyu to the Southwest, the Mbukushu, Tonga, and Lenje to Middle-Zambese regions, and the Shona eastward. All these populations, in all likelihood, did not mix with the Khoisan and thus did not integrate the $Gm^{1,17;;13,15}$ haplotype.

Although the presence of remnant Khoisanoid populations such as Hadza and Sandawe in Kenya would confirm a prior northern extension of Khoisan people, these populations seem nowadays genetically undistinguishable from other neighboring Bantu as far as the Rhesus system is concerned. HLA and Rhesus data, as well as other genetic systems (Nurse et al., 1985), permit differentiation between the Khoi and the San, the former being closer to East Africans, but also, in a general manner, to the whole Black African human type. The San, on the contrary, reveal by their gene frequencies a likely early isolation. This distinction slightly modifies our hypothesis concerning Khoisan migration: the Khoi could have learned herding from East Africans or Central Sudanic populations, as they must have had a contact with other Black Africans. They might have migrated later than the San to southern Africa, where further Bantu migrations would have allowed some transmission of genetical and cultural traits.

Differentiating factors

After having discussed the relations between major groups in Africa, we may wonder how the differences found could have emerged and what kind of mechanisms were involved. Though it is not clearly established whether all sub-Saharan African populations have one single origin or not, they clearly possess common genetic traits. Some populations may have been submitted to historical factors that influenced their genetic pools, and these have led to what we have called their peculiarities.

Thus, genetic drift added to a founder effect may account for the apparent loss of R' haplotype in Pygmies and Nilotes and of *r* in the San (or at least their drastically low frequency). Random genetic drift is also a candidate for explaining the local fluctuations among Cushites/Ethiosemites and Bantu populations, as well as in West Africans. Recent migrations may account for some affinities found between distant populations. The Tutsi and Hima. for example, as we have seen, surrounded by Bantu populations but closer genetically to Cushites and Ethiosemites, are known to have migrated from northern territories recently. Population fusion and admixture processes could be responsible for major genetic exchange, as hypothesized for the relations between the Khoisan and southern Bantu, as well as between the Khoisanoid Hadza and Sandawe populations and the Central Bantu.

Without searching for overly simplified genetic clines, most of the relations between populations may be explained satisfactorily by genetics when confronted with linguistic and historical data. More generally, small interpopulational migrations may have contributed to the maintaining of a relative genetic homogeneity among the groups described previously. In this respect, what is observed is in contradiction with what would have happened if populations had been differentiating independently from each other for many thousands of years. Although claimed (Hiernaux, 1968), traces of recent and strong founder effects, at least in most of the studied populations, seem uncommon in Africa.

Current Gm and Rhesus systems data suggest the following hypothetical scenario concerning a recent peopling of Africa. An ancestral population, still possessing

most of the specificities found nowadays in Africa, Europe. and Asia, moved to East Africa. From these migrants, two populations emerged. The first one colonized southern Africa. Khoisan populations descended from them and were submitted to a strong and probably ancient founder effect and to genetic drift. The second daughter population moved westward and differentiated progressively. More recently, Bantu populations, probably coming from Central West Africa, migrated toward the South and met the Khoisan in southern Africa. In this schema, the heterogeneous Nilo-Saharan populations may be the scattered remnants of an ancient East to West Sudanic peopling zone. That would explain their present affinities with East and West Africans as well as with Bantu. This scenario is compatible with HLA data and beta^S DNA haplotype polymorphism, as the latter is thought to have occurred only recently. This model seems verifiable, because of the current precision obtained from HLA typing and DNA polymorphism concerning population kinship measurements.

CONCLUSIONS

The presence of the genus *Homo* in Africa has been dated to 2 million years ago, but no evidence of a continuum exists between these early *Homo* and extant populations. Fossils on which observations may be drawn are rare, chronologically distant, and geographically dispersed. Physical characteristics are prone to rapid changes and adaptations to environmental factors. Thus, any attempts to determine the probable ethnicity of tens-of-thousands-of-years-old skeletal remains seem highly questionable.

Though Africans are genetically clearly differentiated from other populations of the world, they do not seem to constitute the latter's direct ancestors. If we accept a hypothetical initial divergence between human groups around 200,000 B.P., it is very difficult to imagine that all the known genetic differentiations could have been developed from an African gene pool in such a short period of time on other continents. That is why we favor the hypothesis of non-African populations colonizing Africa after having been subjected to a drastic founder effect and random genetic drift. The resulting loss of some alleles or haplotypes and a frequency increase of others are in agreement with our results based on blood group data. In this view, populations still possessing numerous different haplotypes may be regarded as more or less representative of ancestral populations. This could be the case of some East Africans.

Studies of DNA polymorphism, though showing that Africans have accumulated more DNA changes than others, cannot presume the direction of a primordial migration into or out of Africa. Distances based on mean number of codon differences do not take into account gene flow (Slatkin and Maruyama, 1975) or historical events. At any rate, as long as the questions of mutation rate and its mechanisms are not solved, any attempt in setting intraspecific divergence time seems premature.

Due to numerous studies of the Rhesus and Gm systems (53 and 43 samples, respectively), it is possible to compare many populations representative of different parts of sub-Saharan Africa. These systems, theoretically not as informative as HLA or DNA polymorphism, are, however, the only ones on which as many African samples have been gathered. They clearly separate four main groups of linguistically related populations. West Africans are often related to Bantu populations and sometimes to East Africans, being in some way intermediate between these latter two. The Khoisan group, possessing local specificities, is supposed to have influenced genetically southern Bantu and to have had ancient contacts with East Africans. Our study also reveals some finer details on certain population affinities that would sometimes Seem contradictory without the help of linguistics and historical sources: for example, Nilotes (Dinka, Nuer, Shilluk) have been shown to differ from other East Africans concerning the Rhesus system as they lack the R^1 haplotype; Chari-Nile populations (Kunama and Sara), also belonging to the Nilo-Saharan family but established, respectively, in Chad and Ethiopia, are quite similar and related to West Africans; the Bantu may be subdivided into two distinct groups on the basis of the presence of a Khoisan Gm specificity; the Tutsi and Hima, established around Victoria Lake, differ greatly from their Bantu neighbors, but it is well known that they came only recently from northern territories. In this regard, it appears that most of the genetic peculiarities concerning present populations can be explained, when confronted with historical and linguistic sources, rather than being considered as examples of founder effects, random genetic drift, or frequency convergence. This is not to say that these phenomena did not happen in African history, but their local importance has probably been either greatly overestimated or attenuated by numerous migrant exchanges between distant populations. Nowadays relative intragroup genetic homogeneity does not reflect long isolations and strong differentiation at this level. Most of the time, genetic specificities correspond to linguistic differentiations. Observation of this link between cultural and genetic traits, to be confirmed by further studies, should lead to close collaboration between these two scientific fields in the future.

ACKNOWLEDGMENTS

We are deeply indebted to Professor Alain Gallay, Dept. d'Anthropologie, Université de Geneve, for help, counselling, and documentation; to Dr. Madeleine Blanc of the Centre d'Hémotypologie du C.N.R.S.,Toulouse, France, for allowing the use of unpublished data in the overall analysis; to Dr. Marie-Paule Ferry of the UA 49 (CNRS Paris) for her comments in the linguistic field; and to Andrea Kindermann for her precious data collection on the Gm system. We are also grateful to Georges Puissant for programming; to Serge Aeschlimann, Jean-Gabriel Elia, and Yves Reymond for the illustrations; and to Robin Postel for her careful reading of the manuscript.

This work was supported by grant 3.954-0.84 of the Swiss National Foundation for Scientific Research.

LITERATURE CITED

Ambrose. SH (1982) Archaeology and linguistic reconstructions of history in East Africa. In C Ehret and M Posnansky (eds): The Archaeological and Linguistic Reconstruction of African History Berkeley: Univ. of California Press. pp. 104–157

Ambrose. SH (1984) The introduction of pastoral adaptations to the Highlands of East Africa In JD Clark and SA Brandt (eds): From Hunters to Farmers. The Causes and Consequences of Food Production in Africa. Berkeley Univ of California Press, pp. 212–239.

Anderson, S, Bankier, AT, Barrel, BG. De Bruijn, HL. Coulson, AR, Dronin. J, Eperon. IC, Nierlich. DP, Roe, BA. Sanger, F, Schreier, PH, Smith, AJH. Staden, R, and Young. IG (1981) Sequence and organization of the human mitochondria1 genome Nature *290*:457–465

Antonarakis, SE. Boehrn, CD. Giardina. PJV, and Kazazian, H. Jr (1982) Nonrandom association of polymorphic restriction sites in the beta globin gene cluster. Proc. Natl. Acad. Sci. USA *79*:137–141.

Antonarakis, SE. Boehm, CD. Serjeant, GR. Thei sen, CE, Dover, CJ. and Kazazian, H, Jr (1984) Origin of the beta S globin gene in Blacks: The contribution of recurrent mutation and/or gene conversion. Proc. Natl. Acad. Sci USA *81*:853–856.

Aquadro, CF, and Greenberg, BD (1983) Human mitochondria1 DNA variation and evolution. Analysis of nucleotide sequences from seven individuals. Genetics *103*:287–312.

Avise, JC. and Lansman, RA (1983) Polymorphism of mitochondrial DNA in populations of higher animals. In M Nei and RK Kwhn (eds.): Evolution of Genes and Proteins. Sunderland: Sinauer Associates Inc , pp. 147–164.

Ba. AH, and Dieterlen. G (1961) Koumen. Texte Initiatique des Pasteurs Peuls. Mouton: La Haye.

Bastin, Y, Coupez. A, and de Halleux, B (1983) Classification lexicostatistique des langues bantoues (214 relevés). Bull. Seanc. Acad. Sci. Outre. Mrr 27 173–199

Baumann, H (ed) (1979) Die Volker Afrikas und lhre Traditionellen Kulturen. Wiesbaden: Franz Steiner Verlag GmbH.

Bergland. K, and Vogt, H (1962) On the validity of glottochronology. Curr. Anthropol. *3*:115–153.

Birky, CW, Maruyama. T, and Fuerst. P (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts. and some results. Genetics *103*:513–527.

Blanc. M. Dugoujon, JM, and Ruffié, J (1976) Hemotypologie: un nouveau marqueur genetique des IgG humaines habituellement associé au facteur Gm(21) dans les populations caucasoïdes et mongoloïdes C.R Acad. Sci. (Paris) *283*:D1461–D1463

Blanc. M. and Langaney. A (unpublished).

Bodmer, JG. Bodmer, WF, Ziegler. J, and Magrath. IT (1975) HL-A and Burkitt's tumour A study in Uganda Tissue Antigens *5*:59–62.

Boehm. CD. Dowling, CE. Antonarakis. SE. Honig, GR. and Karazian. H, Jr (1985)Evidence supporting a single origin of the beta C globin gene in Blacks Am J Hum. Genet. *37*:771–777

Ronne-Tamir. B. Johnson, MJ. Natali, A. Wallace, AD, and Cavalli-Sforza, LL (1986) Human mitochondrial DNA types in two Israeli populations

A comparative study at the DNA level. Am. J. Hum. Genet. *38:*341-351.

Botha. MC. Campbell, E, Briggs. B, and du Toit, ED (1975) Differences in serological definition and frequency of some HLA antigens among southern African populations. In F Kissmeyer-Nielsen (ed): Histocompatibility Testing 1975. Copenhagen: Munkagaard. pp. 151–172.

Botha. MC. du 'bit. ED. Jenkins, T, Van Leeuwen. A. D'Amaro, J, Meera Khan, P, Van Der Steen, G, Van Rood. JJ. and Van Der Does. JA (1972) The HLA-A system in Bushmen (San) and Hottentot (Khoikoi) populations of South West Africa. In J Dausset and J Colombani (eds): Histocompatibility Testing 1972. Copenhagen: Munksgaard. pp. 421-432.

Brace, CL (1964) The fate of the "classic" neanderthals. A consideration of hominid catastrophism. Curr. Anthropol. 53-43.

Brauer, C (1978) The morphological differentiation of anatomically modern man in Africa. with some regard to recent finds from East Africa. Z. Morphol. Anthropol. *69:*266-292.

Brauer. G (1984) A craniological approach to the origin of anatomically modern *Homo sapiens* in Africa and implications for the appearance of modern Europeans. In FH Smith and F Spencer (eds): The Origins of Modern Humans: A World Survey of the Fossil Evidence. New York: Alan R. Liss. Inc, pp. 327–410.

Brega, A, Cardella. R, Semino, O, Morpurgo. G, Astaldi Ricotti, CB, Wallace, DC, and Santachiara Benerecetti. AS (1986a) Genetic studies on the Tharu population of Nepal: Restriction endonuclease polymorphism of mitochondrial DNA. Am. J. Hum. Genet. *39:*502-512.

Brega. A, Scozzari, R, Maccioni. L, Iodice. C, Wallace, DC, Bianco. I. Cao. A, and Santachiara Benerecetti. AS (1986b) Mitochondrial DNA polymorphism in Italy: I. Population data from Sardinia and Rome. Ann. Hum. Genet. *50:*327-338.

Brewerton, DA. Hart, FD. Nicholls. A. Caffrey. M. James, DCO, and Sturrock. RD (1973) Ankylosing spondylitis and HL-A27. Lancet *1:*904-907.

Briggs. BR. Jackson, WPU, du 'bit, ED, and Botha. MC (1980) The histocompatibility (HLA) antigen distribution in diabetes in southern African Blacks (Xhosa). Diabetes *29:*68-70.

Brown, WM (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc. Natl. Acad. Sci. USA *77:*3605-3609.

Brown, WM, George. M, Jr, and Wilson, AC (1979) Rapid evolution of animal mitochondrial DNA Proc. Natl. Acad. Sci. USA *76:*1967-1971.

Brown, WM, Prager. EM, Wang. A. and Wilson, AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. *18:*225-239.

Cain. AJ. and Harrison, CA (1958) An analysis of the taxonomist's judgement of affinity. Proc. Zool. Soc. Lond. *131:*85-98.

Cann. RL. Brown, WM, and Wilson. AC (1982) Evolution of human mitochondrial DNA: A preliminary report. In B. Bonné-Tamir (ed): Human Genetics, Part A: The Unfolding Genome. New York: Alan R. Liss. Inc., pp. 157-165.

Cann. RL. and Wilson. AC (1983) Length mutations in human mitochondrial DNA. Genetics *104:*699-711.

Cann. RL. Stoneking. M. and Wilson. AC (1987) Mitochondrial DNA and human evolution. Nature *325:*31-36.

Cavalli-Sfona. LL. and Bodmer, WF (1971) The Genetics of Human Populations. San Francisco: WH Freeman and Co.

Cavalli-Sforza, LL. and Edwards. AWF (1965) Analysis of human evolution. In SJ Geerts (ed): Genetics Today. Proceedings of the Eleventh International Congress of Genetics. The Hague, Sept. 1963, Vol 3. New York: Pergamon Press.

Cavalli-Sforza, LL, Kidd. JR. Kidd. KK. Bucci, C, Bowcock, AM, Hewlett. BS. and Friedlaender. JS (1986) DNA markers and genetic variation in the human species. Cold Spring Harbor Symp. (in press).

Cavalli-Sfona. LL. Zonta, LA, Nuzzo, F, Bernini. L, De Jong. WWW, Meera Khan, P, Ray, AK, Went, LN. Siniscalco. M. Nijenhuis, LE. van Loghem, E, and Modiano. G (1969) Studies on African Pygmies. I. A pilot investigation of Babinga Pygmies in the Central African Republic (with an analysis of genetic distances). Am. J. Hum. Genet. *21:*252-274.

Ceppellini. R, Dunn. LC, and Turri, M (1955) An interaction between alleles at the Rh locus in man which weakens the reactivity of the Rh_0 factor (D^u). Proc. Natl. Acad. Sci. USA *41:*283-288.

Charmot, G, and Lefevre-Witier, P (1978) Géographie de la drépanocytose: Les causes de sa répartition dans l'Afrique au sud du Sahara. Med. Trop. (Mars.) *38:*167-174.

Coon. CS (1962) The Origin of Races. New York: Alfred A. Knopf.

Cornevin. M (1982) Les Neolithiques du Sahara central et l'histoire generale de l'Afrique. Bull. Soc. Prehist. Fr. *79*(10-12):439–450.

Dausset. J, and Pla, M (1985) HLA. Complexe Majeur d'Histocompatibilité de l'Homme. Paris: Flammarion Médecine-Sciences.

David, N (1980) Early Bantu expansion in the context of central African prehistory. 4000-1 B.C. In L Bouquiaux (ed): L'Expansion Bantoue. Actes du Colloque International du CNRS. Viviers. France. 4-16 avril 1977 Paris. CNRS-SELAF, pp. 609-647

David. N (1982) Prehistory and historical linguistics in Central Africa. Points of contact. In C Ehret and M Posnansky (eds): The Archaeological and Linguistic Reconstruction of African History. Berkeley: Univ. of California Press, pp. 78–95.

de Maret. P (1985) Recent archaeological research and dates from central Africa. J. Afr. Hist. *26:*129-148.

de Maret. P, and Nsuka. F (1977) History of Bantu metallurgy: Some linguistic aspects. Hist. Afr. *4:*43-65.

Denaro. M, Blanc. H, Johnson. MJ. Chen, KH, Wilmsen. E. Cavalli-Sfona. LL. and Wallace, DC (1981) Ethnic variation in Hpa I endonucleaae cleavage patterns of human mitochondrial DNA. Proc. Natl. Acad. Sci. USA *78:*5768-5772.

Denbow. JR (1986) A new look at the later prehistory of the Kalahari. J. Afr. Hist. *27:*3-28.

Denbow. JR. and Campbell. A (1986) The early stages of food production in southern Africa and some potential linguistic correlations. SUGIA *7:*83-103.

Denbow, JR, and Wilsem, EN (1986) Advent and course of pastoralism in the Kalahari. Science 234:1509-1515.

Ehret, C (1971) Southern Nilotic History: Linguistic Approaches to the Study of the Past Evanston Northwestern Univ. Press.

Ehret. C (1973) Patterns of Bantu and Central Sudanic settlement in central and southern Africa. Transafr. J. Hist. *3:*1-71.

Ehret. C (1974) Ethiopians and East Africans: The Problem of Contacts. Nairobi: East African Publishing House.

Ehret. C (1979) On the antiquity of agriculture in Ethiopia. J. Afr. Hist. 20.161-177

Ehret, C (1982a) The first spread of food production to Southern Africa. In C Ehret and M Posnansky (eds): The Archaeological and Linguistic Reconstruction of African History. Berkeley: Univ. of California Press, pp. 158-181.

Ehret, C (1982b) Linguistic inferences about early Bantu history. In C Ehret and M Posnansky (eds): The Archaeological and Linguistic Reconstruction of African History. Berkeley: Univ. of California Press, pp. 57-65.

Ehret, C (1984) Historical/linguistic evidence for early African food production. In JD Clark and SA Brandt (eds): From Hunters to Farmers. The Causes and Consequences of Food Production in Africa. Berkeley: Univ. of California Press. pp. 26–35.

Ellison, J, and Hood. L (1982) Linkage and sequence homology of two human immunoglobulin gamma heavy chain constant region genes. Proc Natl. Acad Sci. USA *79:*1984-1988

Ewens. WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 387-112.

Famuyiwa. OO, Nwabuebo. IE, and Abioye. AA (1982) Pattern of histocompatibility (HLA) antigen distribution among Nigerian (West African Black) diabetics. Diabetes *31:*1119-1122

Festenstein. H, Adams. E. Brown, J. Burke. J. Lincoln, P, Oliver, TD, Rondiak, G, Sachs. JA. Welch. SG, and Wolf, E (1972) The distribution of HL-A antigens and other polymorphisms in Bantu. speaking Negroids living in Zambia In J Dausset and J Colombani (eds) Histocompatibility Testing 1972. Copenhagen: Munksgaard. pp 397–407

Fisher. RA (1947) The Rhesus factor. A study in scientific method Am. Sci 35 95-103

Fisher. RA (1953) Population genetics Proc R Soc Lond. [Biol.] *141:*510-523

Fisher, RA. cited by Race. RR (1944) An 'incomplete' antibody in human serum Nature *153* 771-772.

Fisher, RA. and Race, RR (1946) Rh gene frequencies in Britain. Nature *157:*48

Flanagan. JG, and Rabbits. TH (1982) Arrangement of human immunoglobulin heavy chain constant region genes implies evolutionary duplication of a segment containing gamma, epsilon and alpha genes Nature *300:*709-713.

Fleming, HC (1976) Sociology, ethnology and history in Ethiopia. Int. J. Afr Hist. *9:*248-278

Flint, J, Hill, AVS. Bowden, DK. Openheimer. SJ, Sill. PR, Serjeantson, SW. Bana-Koiri, J, Bhatsu, K, Alpers, MP. Boyce. AJ. Weatherall. DJ. and Clegg. JB (1986) High frequencies of alpha thalassemia are the result of natural selection by malaria. Nature *321:*744-750.

Godber, M. Kopec, AC, Mourant. AE. Teesdale, P, Tills, D, Weiner. H, El-Niel. H, Wood. CH. and Barley, S (1976) The blood groups, serum groups and red cell isoenzymes and hemoglobins of the Sandawe and Nyaturu of Tanzania. Ann. Hum. Biol. *3:*463–473.

Govaerts, A. Massart, T, Rivat, L, and Brocteur. J (1972) Serological characteristics of a Bantu population. In J Dausset and J Colombani (&): Histocompatibility Testing 1972. Copenhagen: Munksgaard. pp 409–413.

Cower. JC (1966) Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika *53:*325-338.

Cower. JC (1967) Multivariate analysis and multidimensional geometry. Statistician 17: 13–28.

Greenberg. JH (1957) Essays in Linguistics. New York: Viking Fund Publications in Anthropology 22.

Greenberg. JH (1963) The Languages of Africa. The Hague: Mouton.

Greenberg. JH (1964) Historical inferences from linguistic research in sub-Saharan Africa. In J Butler (ed): Boston Univeraity Papers in African History I, pp 1–15.

Greenberg. JH (1972) Linguistic evidence regarding Bantu origins. J Afr. Hist. *13:*189-216

Greenberg, JH (1980) Classification des langues d'Afrique In J Ki-Zerbo (ed): Histoire Générale de I'Afrique. Vol I Paris: Jeune Afrique/Unesco, pp. 321-338

Gregersen. EA (1977) Language in Africa. An Introductory Survey. New York: Cordon and Breach.

Cregorius. H-R (1978) The concept of genetic diversity and its formal relationship to heterozygosity and genetic distance. Math. Biosci. *41:*253-271

Cuthric. M (1962) Some developments in the prehistory of the Bantu languages. J. Afr Hist. *3:*273-282

Ha[illegible]in, AG. Entwistle. CC. Knight, SC, Wasunna, A. Kung'u, A, and Brubaker. G (1982) HLA antigens in East African Black patients with Burkitt's lymphoma or nasopharyngeal carcinoma and in controls. A pilot study. Hum Immunol. *6:*91-106.

Hammond. MG. Appadoo. B, and Brain. P (1975) HLA antigens in Bantu and Indians. In F Kissmeyer Nielsen (ed) Histocompatibility Testing 1975 Copenhagen Munksgaard. pp. 173-178

Hurnrnond MG, Appadoo. B. and Brain, P (1977) HLA and cancer in South African Negroes Tissue Antigens. 9: [illegible]

Hammond. MG, Asmal. AC. and Omar, MAK (1980) HLA and insulindependent diabetes in South African Negroes Diabetologia *19:*101-102.

Heine. B (1973) Zur genetishe Gliederung der Bantu-Sprachen Afr Ubersee *56:*164-185.

Helal, AN, Lefranc. MP, Ellouze, F, Boukef. K, Sfar Gandoura. M. Sfar. M. and Lefranc. G (1981) Les allotypes des immunoglobulines humaines dans la population tunisienne De précieux marqueurs en genetique des populations et immunogénétique Rev Fac Sci Tunis *1:*121-136

Hiernaux, J (1968) Bantu expansion. The evidencr from physical anthropology confronted with linguistic and archaeological evidence. J Afr Hist 9 505-515

Hiernaux, J (1976) Blood polymorphisms frequencies in the Sara Majingay of Chad. Ann Hum Biol *3* 127-140.

Honig, GR, and Adams. JG III (1986) Human Hemoglobin Genetics New York: Springer Verlag. p. 35

Horai. S, Gojobori, T, and Matsunaga. E (1984) Mitochondrial DNA polymorphism in Japanese. I Analysis with restriction enzymes of six base pair recognition. Hum. Genet. 68.324–332.

Horai. S, and Matsunaga. E (1986) Mitochondrial

DNA polymorphism in Japanese. II. Analysis with restriction enzymes of four or five base pair recognition. Hum. Genet. *72*:105–117.
Huffman, TN (1970) The Early Iron Age and the spread of the Bantu. S. Afr. Arch. Bull. *25*:3–21.
Huizinga, J (1969) Human biological observations on some African populations of the Thorn Savanna Belt II. Proc. Ned. Akad. Wet. *C71*:373–390.
Hymes. DH (1960) Lexicostatistics so far. Curr. Anthropol. *1*:3–44.
Inskeep, R (1976) The problem of Bantu origins. In G de G Sieveking, IH Longworth. and KE Wilson (eds): Problems in Economic and Social Archaeology, Duckworth. pp. 25–37.
Jaeger. G, Pinerd, JM, Tomassone, R, Blanc. M, and van Loghem, E (unpublished).
Jenkins, T (1972) Genetic Polymorphisms of Man in Southern Africa. Thesis, London. XIV. 572pp.
Jenkins, T, and Steinberg, AG (1966) Some serum protein polymorphisms in Kalahari Bushmen and Bantu: Gammaglobulins. haptoglobins and transferrins. Am. J. Hum. Genet. *18*:399–407.
Jenkins, T, Zoutendyk, A. and Steinberg. G (1970) Gammaglobulin groups (Gm and Inv) of various southern African populations. Am. J. Phys. Anthropol. *32*:197–218.
Johnson. MJ. Wallace. DC, Rattazi, MC. and Cavalli-Sforza, LL (1983) Radiation of human mitochondrial DNA types analyaed by restriction endonuclease cleavage patterns. J. Mol. Evol. *19*:255–271.
Kan. YW, and Dozy, AM (1978) Polymorphism of DNA sequence adjacent to human beta globin structural gene. Relation to sickle mutation. Proc. Natl. Acad. Sci. USA *75*:5631–5635.
Kan. YW. and Dozy, AM (1980) Evolution of the hemoglobin S and C genes in world populations. Science *209*:388–391.
Kirsch, IR, Morton. CC. Nakahara. K, and Leder. P (1982) Human immunoglobulin heavy chain genes map to a region of translocation in malignant B lymphocytes. Science *216*:301–303.
Klein. R (1984) The prehistory of Stone Age herders in South Africa. In J Clark and S Brandt (eds): From Hunters to Farmers. The Causes and Consequences of Food Production in Africa. Berkeley: Univ. of California Press, pp. 281–289.
Kohler. O (1975) Geschichte und Probleme der Gliederung der Sprachen Afrikas. In H Baumann (ed): Die Volker Afrikas und Ihre Traditionellen Kulturen. Teil 1. Wiesbaden: Franz Steiner Verlag GmbH, pp. 137–373.
Kulkarni, AG. Ibazebo, RO, Dunn. DT, and Fleming. AF (1985) Some red cell antigens in the Hausa population of Northern Nigeria. Hum. Hered. *35*:283–287.
Kulozik, AE. Wainscoat, JS, Serjeant. GR, Kar. BC. Al-Awamy, B. Essan, 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 sicklecell mutation. Am. J. Hum. Genet. *39*:239–244.
Langaney. A (1979) Diversité et histoire humaines. Population *6*:985–1006.
Langaney. A (1984) La nouvelle démographie de l'evolution. Population *39*:587–606.
Langaney. A, Jaeger. G, Marti, Y, and Blanc. M (1978) Généalogies et groupes sanguins des Sara Kaba N'Dindjo de Miamane. L'Anthropologie *82*:95–101.
Leakey, LSB (1953) Kenya et Tanganyika. Catalogue des Hommes Fossiles Alger.
Lefèvre-Witier, P (1985) An anthropological perspective on the epidemiology of hemoglobin defects and G-6-PD deficiencies in the northern half of the african continent. In R Chakraborty and EJE Szathmary (eds): Diseases of Complex Etiology in Small Populations: Ethnic Differences and Research Approaches. New York: Alan R. Lisa, Inc.. pp. 367–400.
Lefranc. G, Lefranc. MP. Helel. AN, Boukef. K. Chaabani, H. Sfar-Gandoura. M, and Van Loghem. E (1982) Unusual heavy chains of human IgG immunoglobulins: Rearrangements of the CH domains exons. J. Immunogenet *9*:1–9
Lehmann. H (1954) Distribution of the sickle cell gene. A new light on the origin of East Africans. Eugen. Rev. *46*:101–121.
Liesegang, G, Seitz. S, and Winter, JC (1979) Das Aequatoriale Ostafrika. In H Baumann (ed): Die Volker Afrikas und Ihre Traditionellen Kulturen. Teil 2. Wiesbaden: Franz Steiner Verlag GmbH, pp. 1–67.
Livingstone. FB (1967) Abnormal Hemoglobins in Human Populations. Chicago: Aldine Publishing Company.
Livingatone. FB (1985) Frequencies of Hemoglobin Variants. New York: Oxford University Press.
Long, JC. and Chakravarti. A (1986) A phylogeny of beta-globin haplotypes and its implications for human evolution. Am. J. Hum. Genet. [Suppl.] *39*:A240.
Lowe, RF (1969) Rhodesian tribal blood groups. Centr. Afr. J. Med. *15*:151–164
Lwanga-Lunyiigo. S (1976) The Bantu problem reconsidered. Curr. Anthropol. *17*:282–286.
Lynch, RM. and Robbins. LH (1979) Cushitic and Nilotic prehistory: New archaeological evidence from North-West Kenya. J. Afr. Hist. *20*:319–328.
Marsh. WL, Chaganti. RSK. Gardner. FH. Mayer. K. Nowell, PC, and German, J (1974) Mapping human autosomes: Evidence supporting assignment of Rhesus to the short arm of chromosome No. 1. Science *183*:966–968.
Mattiuz. PL. Ihde, D. Piazza. R. Ceppellini. R. and Bodmer. WF (1970) New approaches to the population genetic and segregation analysis of the HLA system. In PI Terasaki (ed): Histocompatibility Testing 1970. Copenhagen: Munksgaard. pp. 193–205.
McIntosh. SK. and McIntosh, RJ (1983) Current directions in West African prehistory. Ann. Rev. Anthropol. *12*:215–258.
McKusick, VA (1978) Mendelian Inheritance in Man. Baltimore: Johns Hopkins Univ. Press.
Mears. JG, Boldjord, C, Benabadji, M, Belghiti, YA, Baddou. MA, Labie, D. and Nagel. RL (1981a) The sickle gene polymorphism in North Africa. Blood *58*:599–601.
Meara. JG, Lachman. HM, Cabannes. R, Anegnizin. KPE. Labie. D, and Nagel. RL (1981b) Sickle gene: Its origin and diffusion from West Africa. J. Clin. Invest. *68*:606–610.
Migone, N, De Lange. G, Piazza. A, and Cavalli-Sforza, LL (1985) Genetic analysis of eight linked polymorphisms within the human immunoglobulin heavy chain region. Am. J. Hum. Genet. *37*:1146–1163.
Miller. LH. Masin, SJ, Clyde, DF, and McGinnis, MH (1976) The resistance factor to Plasmodium vivax in Blacks. The Duffy-blood-group genotype FyFy N. Engl. J. Med. *295*:302–304.

Monnat. RJ, and Loeb, LA (1985) Nucleotide sequence preservation of human mitochondrial DNA. Proc Natl. Acad. Sci USA *82*:2895–2899.
Mourant. AE. Kopec, AC. and Domaniewska-Sobczak. K (1976) The Distribution of the Human Blood Groups and Other Polymorphisms. London: Oxford Univ. Press.
Mourant. AE. Kopec, AC. Ikin, EW. Lehmann. H. Bowen-Simpkins, P, Fergusson, ILC, Hellier, MD. Jones, RD. and Roberts. IAM (1974) The blood groups and hemoglobins of the Kunama and Baria of Erithrea, Ethiopia. Ann Hum. Biol. *1*:383–392.
Muller, JY. Halle, L. and Jaeger. G (1981) HLA-A and B antigens in Aka Pygmies Tissue Antigens *17*:372–375.
Murdock. GP (1959) Africa. Its Peoples and Their Culture History New York: McGraw-Hill Book Company.
Nei. M, and Roychoudhury. AK (1974) Genic variation within and between the three major races of man. Caucasoids. negroids and mongoloids. Am J. Hum. Genet. *26*:421–443.
Nei. M, and Roychoudhury, AK (1982) Genetic relationship and evolution of human races. Evol. Biol. *14*:1–59.
Nurse, D (1982) Bantu expansion into East Africa: Linguistic evidence. In C Ehret and M Posnansky (eds): The Archaeological and Linguistic Reconstruction of African History. Berkeley: Univ. of California Press, pp. 199–222.
Nurse. GT, Bodmer. JG, Bodmer. WF, Van Leeuwen, A, Van Rood. JJ, du Toit, ED, and Botha. MC (1975) A reassessment of the HL-A system in Khoisan populations of South West Africa. Tissue Antigens *5*:402–414.
Nurse, GT, and Jenkins, T (1977) Serogenetic studies of the Kavango peoples of South West Africa. Ann. Hum. Biol. *4*:465–478.
Nurse, GT, Jenkins, T, Santos David, JH, and Steinberg. AG (1979) The Njinga of Angola: A serogenetic study. Ann. Hum. Biol. *6*:337–348.
Nurse. GT, Lane, AB, and Jenkins, T (1976) Serogenetic studies on the Dama of Southwest Africa. Ann. Hum. Biol. *3*:33–50.
Nurse. GT, Weiner, JS, and Jenkins, T (1985) The Peoples of Southern Africa and Their Affinities. Oxford: Clarendon Press.
Okoye, RC, Williams, E, Alonso, A, Doyle, P, Awad, J. Navarrete. C, Jaraquemada. D, Ollier, WER, and Festenstein, H (1985) HLA polymorphisms in Nigerians. Tissue Antigens *25*:142–155.
Orkin, SH, Antonarakis. SE, and Kazazian. H, Jr (1983) Polymorphism and molecular pathology of the human beta globin gene. Prog. Hematol. *13*:49–73.
Orkin, SH. and Kazazian. H. Jr (1984) The mutation and polymorphism of the human beta globin gene and its surrounding DNA. Ann. Rev. Genet. *18*:131–171.
Pagnier, J, Mears, JG. Dunda-Belkodja. O, Schaefer-Rego, KE, Beldford, C. Nagel. RL. and Labie. D (1984) Evidence for the multicentric origin of the beta S globin gene in Africa. Proc. Natl. Acad. Sci. USA *81*:1771–1773.
Painter, C (1966) The Guang and West African historical reconstruction. Ghana Notes Queries *9*:58–66.
Parikh, VS, Morgan, MM, Scott, R, Scott Clements. L. and Butow, RA (1987) The mitochondrial genotype can influence nuclear gene expression in yeast. Science *235*:576–580.
Phillipson, DW (1975) The chronology of the Iron Age in Bantu Africa. J. Afr. Hist. *16*:312–342.
Phillipson. DW (1977) The Later Prehistory of Eastern and Southern Africa. London: Heinemann.
Piazza. A, Menozzi, P, and Cavalli-Sforza, LL (1981) Synthetic gene frequency maps of man and selective effects of climate. Proc. Natl. Acad. Sci. USA *78*:2638–2642
Poulton, J (1987) All about Eve. New Scientist *1560*:51–53.
Protsch, R (1975) The absolute dating of Upper Pleistocene sub-Saharan fossil hominids and their place in human evolution. J. Hum. Evol. *2*:529–536.
Ramsay, M. and Jenkins, T (1986) β^S- and B^A- globin gene cluster haplotypes in southern African populations. 7th Int. Congr. Hum. Genet. (Berl.) *2*:462 (Abstr.).
Raum, D. Glass. D, Carpenter, CB. Alper. CA. and Schur. PH (1976) The chromosomal order of genes controlling the major histocompatibility complex, properdin Factor B and deficiency of the second component of complement. J. Clin. Invest. 58: 1240–1248.
Rivat. L. Rivat. C, Cook. CE, and Steinberg. AG (1978) Gm(28), un nouvel allotype du systéme Gm present sur les IgG3: Intérêt particulier de son etude dans les populations négroïdes. Ann. Immunol. *129*C:33–45.
Rosenfield, RE. Allen, FH, Rubinstein. JR, and Rubinstein. P (1973) Genetic model for the Rh blood group system. Proc. Natl. Acad. Sci. USA *70*:1303–1307.
Ruddle. F. Ricciutti. F, McMorris, FA, Tischfield, J. Creagan, R. Darlington, G, and Chen. T (1972) Somatic cell genetic assignment of peptidase C and the Rh linkage group to chromosome A-1 in man. Science *176*:1429–1431.
Sanchez-Mazas, A (1985) Le Système Rhésus et l'Histoire du Peuplement. Diplôme, Université de Genève.
Sanchez-Mazas, A. Excoffier, L. and Langaney. A (1986) Measurement and representation of the genetic similarity between populations by the percentage of isoactive genes. Theoria 4:143–154.
Sandelowsky. B. Van Rooyen, J, and Vogel. J (1979) Early evidence for herders in the Namib. S. Afr. Archaeol. Bull. *34*:15–32.
Shaw. T (1980) Prehistoire de l'Afrique occidentale In J Ki-Zerbo (ed): Histoire Générale de l'Afrique, Vol 1. Paris: Jeune Afrique/Unesco, pp. 643–668.
Slatkin, M. and Maruyama. T (1975) The influence of gene flow on genetic distance. Am. Nat. *109*:667–676.
Solomon, E. and Bodmer. WF (1979) Evolution of sickle variant gene. Lancet April 28:923.
Soper. R (1971) A general review of the Early Iron Age of the southern half of Africa. Azania *6*:5–37
Soper, R (1982) Bantu expansion into eastern Africa: Archaeological evidence. In C Ehret and M Posnansky (eds): The Archaeological and Linguistic Reconstruction of African History Berkeley Univ. of California Press. pp. 223–238.
Spedini. G, Capucci. E. Rickards. O, Fuciarelli. M. Giaccaia. L. Aebischer. ML, Manella, E, and Loreti. O (1981) Some genetic erythrocyte polymorphisms in the Mbugu and other populations of the Central African Republic with an analysis of genetic distances. Anthropol. Anz. *39*:10–19.
Spees. EK. Pool, P, Sullinger. WO. Clark, GB. Passeretti. FA, Sperry. D, Woodbury, MA. and Amos. DB (1975) HL-genetic structure of an Eritrean semitic group In F Kissmever-Nielsen (ed)

Histocompatibility Testing 1975. Copenhagen: Munksgaard, pp. 213–218.
Steinberg. AC (1965) Evidence for a mutation or crossing over at the Rh locus. Vox Sang. *10*:721–724.
Steinberg. AC (1973) Cm and Inv allotypes of some Sidamo Ethiopians. Am. J. Phys. Anthropol. *39*:403–408.
Steinberg. AC, and Cook. CE (1981) The Distribution of the Human Immunoglobulin Allotypes. Oxford: Oxford Univ. Press.
Steinberg, AC. Jenkins, T, Nurse, GT, and Harpending. HC (1975) Cammaglobulin groups of the Khoisan peoples of southern Africa: Evidence for polymorphism for a Gm1,5,13,14,21 haplotype among the San. Am. J. Hum. Genet. *27*:528–542.
Stratton. F (1946) A new Rh allelomorph. Nature *158*:25.
Sutton. JEC (1974) The Aquatic civilization of mid. dle Africa. J. Afr. Hist. *15*:527–546.
Sutton, JEC (1980) L'Afrique orientale avant le VIIe siécle. In C Mokhtar (ed): Histoire Générale de l'Afrique, Vol II. Paris: Jeune Afrique/Unesco, pp. 613–639.
Swadesh, M (1952) Lexicostatistic dating of prehistoric ethnic contacts. Proc. Am. Philos. Soc. *96*:452–463.
Tauxier, L (1937) Moeurs et Histoire des Peuls. Paris: Payot.
Thoma, A (1973) New evidence of the polycentric evolution of *Homo sapiens*. J. Hum. Evol. *2*:529–536.
Tills, D, Kopec, AC. and Tills, RE (1983) The Distribution of the Human Blood Croups and Other Polymorphism. Supplement 1. Oxford: Oxford Univ. Press.
Tills, D, Kopec, AC. Warlow, A, Barnicot, NA, Mourant. AE. Marin, A, Bennett, FJ. and Wood. burn. JC (1982) Blood group. protein, and red cell enzyme polymorphism of the Hadza of Tanzania. Hum. Genet. *61*:52–59.
Tobias, PV (1961) New evidence and new views on the evolution of man in Africa. S. Afr. J. Sci. *57*:25–38.
Trape, JF, Fribourg-Blanc, A, Bosseno, MF. Lallemant, M. Engler. R. and Mouchet. J (1985) Malaria, cause of ahaptoglobinaemia in Africans. Trans. R. Soc. Trop. Med. Hyg. *79*:430–434.
Trinkaus. E (1986) Les Néanderthaliens. La Recherche *180*:1040–1047.
Tücker, AN (1967) Fringe Cushitic. An experiment in typological comparison. Bull. Sch. Oriental Afr. Stud. *30*(3):655–680.
van Crunderbeek, MC. Roche, E, and Doutrelepont, H (1983) Le Remier Age du Fer au Rwanda et au Burundi: Archéologie et Environnement, Brussels: I.F.A.Q.
van Loghem, E. Salimonu, L, Williams, AIO. Osunkoya. BO. Boyd. AM, de Lange, G, and Nijenhuis, LE (1978) Immunoglobin allotypes in African populations. I. Cm-Am haplotypes in a Nigerian population. J. Immunogenet. *5*:143–147.
van Rood. JJ. van Leeuwen, A, Termytelen, A, and Kenning. JJ (1976) The HLA linkage group. Exp. Hematol. *4*:237–242.
Vansina. J (1984) Western Bantu expansion. J. Afr. Hist. 25:129–145.
Vawter, L, and Brown, WM (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science *234*:194–196.
Vossen, R (1982) The Eastern Nilotes: Linguistic and Historical Reconstructions. Berlin: Dietrich Reimer Verlag.
Wainscoat. JS, Bell, JI, Thein. SL, Higgs. DR, Serjeant. CR. Peto, TEA, and Weatherall. DJ (1983) Multiple origins of the sickle mutation: Evidence from beta S globin gene cluster polymorphism. Mol. Biol. Med. *1*:191–197.
Wainscoat. JS, Hill, AVS. Boyce, AL. Flint, J. Hernandez, M. Thein. SL, Old. JM, Lynch, JR, Falusi. AC. Weatherall, DJ, and Clegg, JB (1986) Evolutionary relationships of human populations from an analysis of nuclear DNA polymorphisms. Nature *319*:491–493.
Walker, N (1983) The significance of an early date for pottery and sheep in Zimbabwe. S. Afr. Archaeol. Bull. *38*:88–92.
Wallace, DC, Garrison, K. and Knowler. WC (1985) Dramatic founder effects in Amerindian mitochondrial DNAs. Am. J. Phys. Anthropol. *68*:149–155.
Weiner. JS. and Zoutendyk. A (1959) Blood group investigation on Central Kalahari Bushmen. Nature *183*:843–844.
Welmers. W (1973) African Language Structures. Berkeley: Univ. of California Press.
Wiener, AS (1943) Genetic theory of the Rh blood types. Proc. Soc. Exp. Biol. *NY 54*:316–319.
Williams, RC (1982) HLA. Yrbk. Phys. Anthropol. *25*:91–112.
Wilmink. FW. and Uytterschaut. HT (1984) Cluster analysis, history, theory and applications. In CN van Vark and WW Howells (eds): Multivariate Statistical Methods in Physical Anthropology. Dordrecht: D. Reider Publishing Company, pp. 135–175.
Woimant. F, Lesobre, B, Dehay, C. and Piéron, R (1980) HLA-A and B typing in the Sarakole population of West Africa. Tissue Antigens *15*:216–219.
Zoutendyk, A. Kopec. AC. and Mourant. AE (1953) The blood groups of the Bushmen. Am. J. Phys. Anthropol. *11*:361–368.
Zoutendyk. A. Kopec, AC. and Mourant, AE (1955) The blood groups of the Hottentots. Am. J. Phys. Anthropol. 13r691–697.

APPENDIX. Complete list of sub-Saharan African populations tested for Rhesus, Gm, and HLA, according to the criteria defined in text

Population	Location	Linguistic family	Sample size	Population size[6]	Reference
Rhesus					
Afar	Ethiopia	East Cushitic	198	110,000	Fourquet (1969)[1]
Amba	Uganda	Interlac. Bantu	113	30,000	Ikin et al. (1952)[1]
Amhara	Ethiopia	Ethiosemitic	107	11,000,000	Ikin and Mourant (1962)[1]
Amhara	Ethiopia	Ethiosemitic	104	11,000,000	Bat-Miriam (1962)[1]
Ashanti	Ghana	Kwa	113	700,000	Armattoe et al. (1953)[1]
Babinga	Zaire	"Pygmies"	163	27,000	Cavalli-Sforza et al. (1969)[1]
Barea	Eritrea	Eastern Sudanic	129	10,000	Mourant et al. (1974)
Bedik	Senegal	West Atlantic	791	2,000	Bouloux et al. (1972)[2]
Beja	Sudan	Northern Cushitic	100	20,000	el Hassan et al. (1968)[1]
Bergdama (Dama)	Namibia	Khoisan	119	30,000	Nurse et al. (1976)
Billen	Ethiopia	Central Cushitic	102	30,000	Ikin and Mourant (1962)[1]
Dinka	Sudan	West Nilotic	97	1,100,000	Roberts et al. (1955)[1]
Ewe	Ghana	Kwa	161	700,000	Armattoe et al. (1962)[1]
Falasha	Ethiopia	Ethiosemitic	152		Bat-Miriam (1962)[1]
Fulbe (Peul)	Senegal	West Atlantic	112	6,000,000	Moullec and Linhard (1956)
Gadabursi	Djibuti	Eastern Cushitic	137	125,000	Fourquet (1962)[1]
Galla	Ethiopia	Eastern Cushitic	110	12,000,000	Bat-Miriam (1962)[1]
Geiriku	Namibia	Bantu Kavango	110		Nurse and Jenkins (1977)[1]
Hadza	Tanzania	Khoisan	162	200	Tills et al. (1982)
Hausa	Nigeria	Chadic	190	20,000,000	Kulkarni et al. (1985)
Hima	Uganda	Bantu	117	4,000	Allison et al. (1954)[1]
Isa	Djibuti	Eastern Cushitic	187	125,000	Fourquet (1962)[1]
Isak (Ishaak?)			156		Fourquet (1962)[1]
Ishaak (Awal)	Somalia	Eastern Cushitic	279	420,000	Goldsmith (1959)[1]
Ishaak (Yunis)	Somalia	Eastern Cushitic	230	420,000	Goldsmith (1959)[1]
Karanga	Zimbabwe	Bantu Shona	159		Lowe (1969)[1]
Khoi	Namibia	Khoisan	210		Zoutendyk et al. (1955)[1]
Kpelle	Liberia	Mande	191	250,000	Livingstone et al. (1960)[1]
Kwangali	Angola Namibia	Bantu Kavango	106		Nurse and Jenkins (1977)
Kunama	Eritrea	Chari-Nil	148	40,000–70,000	Mourant et al. (1974)
Lozi 2 (Barotse)	Zambia	Bantu Sotho	209	180,000	Barclay et al. (1969)[1]

(continued)

APPENDIX. Complete list of sub-Saharan African populations tented for Rhesus, Gm, and HLA, according to the criteria *defined in text* (continued):

Population	Location	Linguistic family	Sample size	Population size[3]	Reference
Malinke	Senegal	Mande	586	3,000,000	Blanc and Langaney (unpublished)
Manyika	Zimbabwe	Bantu Shona	163		Lowe (1969)'
Mbugu	C.A.R.'	Adamawa-Eastern?	131	13,600	Spedini et al. (1981)
Ndebele	Zimbabwe	Bantu Nguni	101	410,000	Lowe (1969)'
Njinga	Angola	Bantu Ngola	109		Nurse et al. (1979)
Nuer	Sudan	Western Nilotic	100	3,000,000	Roberts et al. (1955)[1]
Nyaturu	Tanzania	Bantu	214	32,000	Godber et al. (1976)
N'Zakara	C.A.R.	Adamawa-Eastern	296		Creata (1964)[1]
Rendille	Kenya	Cushitic	102	15,000,000	Corrain (1973)[2]
Samburu	Kenya	Nilotic	118	12,000	Corrain (1972)[2]
San 1	Botswana	Khoisan	122		Tobias and Zoutendyk (1958)[1]
Sandawe	Tanzania	Khoisan	215	30,000	Godber et al. (1976)
Sara (Majingay)	Chad	Central Sudanic	257	750,000	Hiernaux (1976)
Sara (Kaba N'Dindjo)	C.A.R.	Central Sudanic	302	750,000	Langaney et al. (1978)
Sara (Kaba N'Dindjo)	C.A.R.	Central Sudanic	295	750.000	Jaeger (1974)[2]
Serere	Senegal	Western Atlantic	363	300,000	Moullec and Linhard (1956)[1]
Shilluk	Sudan	Northern Nilotic	106	110,000	Roberts et al. (1955)[1]
Tigre	Eritrea	Ethiosemitic	104	200,000	Ikin and Mourant (1962)'
Toucoueleur	Senegal	West Atlantic	307		Kane (1963)'
Tutsi	Rwanda		412		Jadin and Bruynoghe (1952)[1]
Wolof	Senegal	West Atlantic	408	1,500,000	Moullec and Linhard (1956)[1]
Yoruba	Nigeria	Kwa	274	8,000,000	Worlledge et al. (1966)'
Zezuru	Zimbabwe	Bantu	365		Lowe (1969)[1]
Gm					
Babinga	C.A.R.	"Pygmies"	162	27,000	Cavalli-Sforza et al. (1969)
Baca	R.S.A.[5]	Bantu Nguni	137		Jenkins et al. (1970)
Bechwana	Botswana	Bantu Sotho	155	1,000,000	Jenkins et al. (1970)
Bedik	Senegal	West Atlantic	880	2,000	Blanc and Langaney (unpublished)
Bi Aka	C.A.R.	"Pygmies"	900		Jaeger et al. (unpublished)
Bunja	Namibia	Bantu Kavango	111		Jenkins et al. (1970)
Fulbe (Peul)	Senegal	West Atlantic	370	4,000,000	Blanc and Langaney (unpublished)
Gambians (Keneba)[1]	Gambia	West Atlantic?	822		
Gambians (Manduar)[2]	Gambia	West Atlantic?	307		
Hlubi	R.S.A.	Bantu Nguni	145		Jenkins et al. (1970)
Khoi	Namibia	Khoisan	149		Steinberg et al. (1975)
Kuambi	Ovamboland	Bantu ovambo	119		Jenkins et al. (1970)
Kuanyama	Ovamboland	Bantu	118		Jenkins et al. (1970)
Kurumba	Burkina Faso	Voltaic	150	90,000	Huizinga (1969)[3]
Lenje	Zambia	Bantu Ila-Tonga	175	40,000	Jenkins et al. (1970)
Lozi 1 (Mlozi)	Zambia	Bantu Sotho	189	180,000	Jenkins et al. (1970)
Malinke	Senegal	Mande	557	3,000,000	Blanc and Langaney (unpublished)
Mbukushu	Angola	Bantu Kavango	115	10,000	Jenkins et al. (1970)
Ndebele	R.S.A.	Bantu Nguni	103		Jenkins and Steinberg (1966)
Njinga	Angola	Bantu Ngols	86		Nurse et al. (1979)
Nyambaan	Mozambic	Bantu	119		Jenkins et al. (1970)
Pedi	R.S.A.	Bantu Sotho	146	1,600,000	Jenkins et al. (1970)
Pondo	R.S.A.	Bantu Nguni	112	300,000	Jenkins et al. (1970)
Sambyu	Namibia	Bantu Kavango	98		Jenkins et al. (1970)
San 1					
!Kung					
Northern	Botswana	Khoisan	100		Steinberg et al. (1975)
Ngami	Botswana	Khoisan	152		Steinberg et al. (1975)
Dobe	Botswana	Khoisan	394		Steinberg et al. (1975)
/du/da	Botswana	Khoisan	100		Steinberg et al. (1975)
Kaukau	botswana	Khoisan	259		Steinberg et al. (1975)
Naron	Botswana	Khoisan	138		Steinberg et al. (1975)
South Central	Botswana	Khoisan	112		Jenkins and Steinberg (1966)
Ko	Botswana	Khoisan	72		Jenkins and Steinberg (1966)
San 2					
!Kung Tsumkwe	Namibia	Khoisan	197		Steinberg et al. (1975)
Sara (Majingay)	Chad	Central Sudanic	255	750,000	Hiernaux (1976)
Shangana	Mozambic	Bantu Tonga	153		Jenkins et al. (1970)
Sidamo (+Wallamo + Kambta)	Ethiopia	Eastern Cushitic	140	500,000	Steinberg (1973)

(continued)

APPENDIX. Complete list **of** *sub-Saharan* African *populations* tested for *Rhesus, Gm,* **and** *HLA, according* **to** the criteria *defined* in *text* (continued)

Population	Location	Linguistic family	Sample size	Population size[6]	Reference
Sotho (Basutu)	Lesotho	Bantu Sotho	149	3,000,000	Jenkins et al. (1970)
Swasi	R.S.A.	Bantu Nguni	126	490,000	Jenkins et al. (1970)
Tonga	Zambia	Bantu **Ila-Tonga**	120		Jenkins et al. (1970)
	Zambesi		164		Jenkins et al. (1970)
Venda	R.S.A.	Bantu	80	360,000	Jenkins et al. (1970)
Xhosa	R.S.A.	Bantu Nguni	214	3,900,000	Jenkins et al. (1970)
Zulu	R.S.A.	Bantu Nguni	130	3,900,000	Jenkins **et** al. (1970)
HLA					
Aka (Bi Aka)	C.A.R.	"Pygmies"	85		Muller et al. (1981)
"Bantu"	Zambia	Bantu	166		Festenstein et al. (1972)
Khoi					
Nama	Namibia	Khoisan	99		**Nurse** et al. (1975)
Hottentot	Namibia	Khoisan	110		Botha et al. (1972)
Nigerians			114		Okoye et al. (1985)
Sen 2					
!Kung Tsumkwe	Namibia	Khoisan	76		Botha et al. (1972)
North	Namibia	Khoisan	82		
Heikum	Namibia	Khoisan	65		
!Kung Central	Namibia	Khoisan	116		Nurse et al. (1975)
North	Nambia	Khoisan	79		
Sarakole	**West** Africa	Mande	119	360,000	Woimant **et** al. (1980)
Shi	Zaire	Eastern Bantu	93		Govaerts et al. (1972)
Tigrinya	Eritrea	Ethiosemitic	68	3,560,000	Spees et al. (1975)
Xhosa	R.S.A.	Bantu Nguni	76		Botha et al. (1975)
Zulu	R.S.A.		100	3,000,000	Hammond et al. (1975)

[1]In Mourant **et al. (1976).**
[2]**In Tills et al. (1983).**
[3]**In** Steinberg and **Cook (1981).**
[4]Central African Republic
[5]Republic of South Africa.
[6]**Most** demographic estimations have **been** found in Murdock (1959) and Baumann (1979)