

# Genetics and linguistics in sub-Saharan Africa

Presented at SAfA 2004

Bergen 27th June 2004

Roger Blench  
Mallam Dendo  
8, Guest Road  
Cambridge CB1 2AL  
United Kingdom  
Voice/Answerphone/Fax. 0044-(0)1223-560687  
E-mail R.Blench@odi.org.uk  
[http://homepage.ntlworld.com/roger\\_blench/RBOP.htm](http://homepage.ntlworld.com/roger_blench/RBOP.htm)

## TABLE OF CONTENTS

|   |           |
|---|-----------|
| <b>1. INTRODUCTION</b>                          | <b>1</b>  |
| <b>2. ACHIEVEMENTS OF GENETICS</b>              | <b>2</b>  |
| 2.1 Eve   | 2         |
| 2.2 Leaving home                                | 3         |
| 2.3 Back to Africa?                             | 3         |
| <b>3. METHODOLOGICAL ISSUES</b>                 | <b>5</b>  |
| <b>4. CASE STUDIES</b>                          | <b>7</b>  |
| 4.1 The Bantu expansion                         | 7         |
| 4.2 Afroasiatic, the Berber and the Guanche     | 10        |
| 4.3 The origin of the African pygmies           | 11        |
| <b>5. CONCLUSION: WHERE DO WE GO FROM HERE?</b> | <b>13</b> |
| <b>REFERENCES [DRAFT]</b>                       | <b>14</b> |

## TABLES

|                         |    |
|-------------------------|----|
| Table 1. The U6 lineage | 10 |
|-------------------------|----|

## 1. Introduction

Africa has never been so important to our broader understanding of what it means to be human. It is now widely accepted that hominids not ancestral to modern humans diffused out of Africa at least 1.8 million years ago (Swisher *et al.* 1994), that modern humans evolved in Africa (Allsworth-Jones 1993; Horai *et al.* 1995; Thomson *et al.* 2000; Ingman *et al.* 2000; Ke *et al.* 2001) and that they spread out of Africa more than 100,000 years ago (Stringer & McKie 1996; Mitchell 2002). In June 2003, fossils of the earliest modern human, *Homo sapiens*, were uncovered at Herto village in the Middle Awash area of Ethiopia, about 140 miles northeast of Addis Ababa, and were dated with radioisotopes at 154-160,000 years old (White *et al.* 2003). The findings provide strong evidence that *Homo sapiens* and pre-*Sapiens* hominids co-existed, rather than the former simply replacing the latter. Expanding modern *sapiens* displaced the existing hominids who populated the Old World so effectively that by ca. 30,000 BP these had been eliminated (Trinkaus 1983; Stringer & Gamble 1993). It is unlikely there was any genetic interchange between modern *Sapiens* populations and the resident *Homo erectus* (Krings *et al.* 1997). The exact dates and routes by which modern humans spread remain controversial, but early dates for Australia indicate that modern humans reached there between 60-50,000 BP (Connell & Allen 1998). The peopling of the New World remains controversial and dates earlier than the 'Clovis Horizon' (little more than 13,000 BP) were not accepted. This is now changing and many scholars accept dates of up to 16,000. Given the dates for modern humans in Siberia (e.g. the D'uktai culture central northeastern Asia and the Yana site, now dated to 30,000 BP (Pitulko *et al.* 2004)), there seems to be no reason in principle why such early dates are not acceptable.

Genetics can be divided into phenotypic and molecular. Phenotypic genetics began with the nineteenth century practice of skull measurements and well into the twentieth century African 'races' were ascribed particular head shapes (e.g. Seligman 1930). In the middle of the twentieth century there was a burst of interest in serology, the study of the distribution of blood groups in Africa. Roberts (1962) sampled the blood of a significant number of Nilotic populations (Dinka, Shilluk) with a view to finding an algorithm that would indicate how long ago these populations split apart. This has a certain structural resemblance to glottochronology and suffers from some of the same defects, namely the assumption that populations split apart in simple tree-like fashion. The key question, however, is whether the distribution of blood groups can be interpreted in a useful way by any other discipline. Whatever its scientific value, this thread of biological anthropology has virtually disappeared with the rise of molecular techniques. The reputations of traditional biological anthropologists have recently stood at an all-time low following analyses such as that of Gould (1981), who accurately skewered the underlying racial preoccupations of the supposedly scientific physical anthropologists of the nineteenth and early twentieth centuries, by showing just how problematic these procedures were in the case of Amerindian populations. There is every reason to think this type of work has very limited value in determining the pattern of the African past, although osteometrics remain acceptable in many European traditions, especially in France, as witness a standard text on human remains in the Sahara (Dutour 1989). Unless the patterns revealed by a biological or genetic parameter can be linked to the results of another historical result, this information is of limited value except within its own narrowly defined field.

Enter molecular biology. From the early 1990s, the development of modern techniques of DNA analysis constituted a major break with traditional biological anthropology; the introduction of molecular techniques in the early 1990s has largely revolutionised the study of human populations in Africa. DNA could potentially be recovered from archaeological material and analysis of DNA seemed to offer a way of relating present human populations both to one another and to past materials. The earliest work concentrated on mitochondrial DNA (mtDNA) but the analysis of nuclear or paternal DNA is now regarded as of equal significance.

DNA offered new insights into the development of modern humans, but its claims have gone further in recent times, to encompass the interpretation of archaeological and linguistic data. Despite great hopes and even greater claims, there has been deep scepticism from other disciplines about genetics. To judge by some of its exponents, the links between language, demographic movement and genetics in prehistory are well-established. These were enthusiastically promoted at the end of the 1980s and into the early 1990s as the 'New Synthesis' or 'Archaeogenetics' (see, for example, Cavalli-Sforza 1987; Cavalli-Sforza *et al.* 1988;

Renfrew 1992; Renfrew and Boyle 2000). The *opus magnus* of this trend was the appearance of ‘The History and Geography of Human Genes’ (Cavalli-Sforza *et al.* 1994), which essays a major revision of the methodology for exploring human history. Linguistic classifications of human populations purport to offer a tool for outflanking simple racial models; more abstract, they appear to provide an ideal analogue to the classificatory trees output from DNA analyses. If DNA phylogenies and language trees *were* to correspond, this would indeed be striking independent confirmation for models of human prehistory. Although this continues to play well in the pages of the journal *Nature*, most archaeologists and linguists remain deeply sceptical<sup>1</sup> (McEachern 2000). Some archaeologists are among those disturbed by the implications of the ‘New synthesis’ for encouraging narrow nationalistic readings of history, and restoring the discredited view of race, language and culture as generally coterminous (Pluciennik 1996). Part of this is innate conservatism and the fact that no academic career points are to be made in being interdisciplinary where established disciplines have developed internal formalisms. But it is also because DNA studies often do not deliver credible results; linguists are faced with trees that show linkages and dates quite contrary to established results and contradicting one another from one paper to the next. Many scholars feel that its claims have come unstuck at this point, that there is a fundamental incomparability between results in the different disciplines.

This paper will not take such a negative view; molecular biology has brought some real insights to the table. Nonetheless, it is often not its own best friend by taking a cavalier approach sampling frames and linguistic terminology. The paper looks at the methodological issues involved and explores some particular case studies from the literature, beginning with an assessment of some of the undoubted successes of genetics.

## 2. Achievements of genetics

### 2.1 Eve

Until the 1980s, the exact relationship between modern humans and the diversity of hominids previously inhabiting the planet was uncertain, with strong body of opinion favouring multi-regionalism, the view that modern human populations arose from the interaction of a new type of hominid with resident *in situ* populations. A significant early result of mtDNA was the ‘Eve’ hypothesis (Templeton 1993; Horai *et al.* 1995), which posited a single origin for humanity in the African continent in the relatively recent past. This result has come together with osteometric and archaeological data to form the now generally accepted ‘out of Africa’ hypothesis (Ke *et al.* 2001). During the 1990s, debates focused on whether *sapiens* were interfertile with the resident *Erectus* populations as an explanation for the diversity of modern humans. More recent work suggests that there is no Neanderthal or *Homo erectus* genetic component in modern humans and that as *sapiens* expanded out of Africa they simply outcompeted all other existing hominids. More recently, DNA work has focussed on the diversity of DNA lineages. A recent overview of African mtDNA (Salas *et al.* 2002) observes ‘Africa presents the most complex genetic picture of any continent, with a time depth for mitochondrial DNA (mtDNA) lineages >100,000 years.’ This pattern is certainly what we would expect from the ‘out of Africa’ hypothesis, although it is of little help in dating phylic dispersals within Africa.

If early modern humans indeed originated in Eastern and Southern Africa, then there may be some connection between the Khoisan languages still spoken there and this early speech. Regrettably, the grim history of European impacts in the region has caused the untimely disappearance of many Khoisan languages, and many others are threatened. Most Khoisan languages are too closely related to be reliable witnesses (Güldemann & Voßen 2000), but the isolates, so remote from any other language for the relationship to now be invisible, may indeed reflect these early speech-forms. Key languages in this respect are Kwadi (formerly in Southern Angola) and ꞤHōã in Botswana (Güldemann & Voßen 2000), Hadza and

---

<sup>1</sup> It would be unfair to say that there are no archaeologists who have taken an interest in ‘Archaeogenetics’, the publications of the McDonald Institute constituting a major focus of these ideas (e.g. Renfrew *et al.* 2000). But publications in this area seem to have taken on a momentum of their own; rather than influencing mainstream practitioners, a group of researchers spend their time going to conferences with one another.

Sandawe (Sands 1998) both in Tanzania. These languages may be the last descendants of the speech of early modern humans.

## 2.2 Leaving home

Earlier models assumed that modern humans left Africa via the Near East, crossing Sinai, but this assumption was based on the sparse remains recorded in Israel and the notion that Pleistocene humans could not make voyages across open sea. We now know this is false; at least by the time they reached insular SE Asia early humans had some kind of voyaging capacity. We have no idea what sort of water-craft were around in 30,000 BP, but we know that they existed, because Papuans reached islands that are only accessible by boat (Kirch 2000:68). Manus, in the Admiralty Islands, is 60-90 km. across open sea, not places that could be reached by individuals clinging to a drifting log (Spriggs 1997:29).

In the light of this, an exit via the Horn of Africa, to the southern coast of Arabia and thence to India and SE Asia is not only feasible, but seems to have occurred. It is now fairly widely accepted that there were two routes out of Africa, through the Sinai peninsula and across the Bab el Mandeb, from the Horn of Africa to Yemen (Quintana-Murci *et al.* 1999; Stringer 2000). When this second route opened is debated, but presumably prior to 70,000 BP, to account for the peopling of Australia. It is as yet unconfirmed by any incontrovertible archaeological evidence, but this is probably a reflection of the type of archaeology conducted around the shores of the Indian Ocean.

## 2.3 Back to Africa?

One of the recent observations of genetics may contribute to the explanation of one of the more puzzling features of the ethnolinguistic scene in Africa. The great majority of the African land mass is occupied by speakers of languages that are assigned to clearly defined phyla; isolates form a small and uncertain list. This is very much in contrast with Papua and the New World, where linguistic differentiation is at levels such that many groupings are disputed and many isolates have been identified. To illustrate the point, there are more language isolates in Colombia than the entirety of Africa (AILV 1994). This seems entirely counter to our present understanding of the relationship between time-depth and linguistic diversity; if modern humans did indeed come out of Africa, and they already had some form of language, then the languages of Africa ought to be considerably *more* diverse than those in Papua or South America.

There is an additional contrast that is equally surprising, the phonological and morphological diversity of African languages compared with Papuan. Both Papuan and Australian languages are distinguished by lexical diversity combined with surprisingly similar phonologies and morphologies (Dixon 1980; Foley 1986). In other words, despite the gradual diversification of words, the framework in which they are set has remained remarkably stable over a very long period. African languages, on the other hand, are strikingly diverse with very large and small consonant inventories often abutting one another and constant variability in tonal, morphological and syntactic systems.

Whatever the present situation, there must have been a stage in African prehistory when the continent was characterised by extreme linguistic and biological diversity. As modern humans diffused from southern and eastern Africa, they would have spread over the continent at extremely low population densities, either assimilating or out-competing existing *in situ* hominid populations. Whether modern humans would have been interfertile with resident African hominids is unclear, but it seems likely. Hominids within Africa a 100 kya would probably have been considerably closer genetically to *H. sapiens* than the hominids who left Africa in the first great outpouring several million years ago. The consequence of modern humans expanding within Africa would have been to create immense biological, social and linguistic complexity. The resultant populations would not necessarily have resembled Khoesanooids and we should look for their modern physical analogues among other residual groups such as the Hadza, the Ongota, the Kwadi and the Damara

(Blench 1999). Genetic studies indicates quite clearly that the Hadza, at least, are no closer to the Khoesan speakers than to any other African population with which they have been compared (Knight *et al.* 2003).

However, today this diversity has virtually disappeared, both linguistically and phenotypically. The most likely explanation for the present-day language situation is the expansion of the present-day language phyla in a relatively recent era and the assimilation of resident diverse groups. Watson *et al.* (1997) present some genetic evidence that seems to support this. We can calibrate this diversity in a simplistic manner by comparing Africa with other regions of the world. It is generally considered that Australia was populated by 55,000 bp and Papua must have been occupied at a similar era. Both Papua and Australia present a situation where one phylum is dominant (Trans-New Guinea in the case of Papuan and Pama-Nyungan in the case of Australia) and there are many isolates or small phyla on the periphery (Wurm 1982; Koch 1997). Given the lower level of language diversity in Africa, its phyla must have become established in the last 30-20,000 years and effectively assimilated the residual diverse languages. The loss of genetic diversity is more apparent than real; various studies have shown that Africa is the most diverse continent and retains the most ancient human lineages (Chen *et al.* 2000; Ke *et al.* 2001).

What could be the cause of this loss of ethnolinguistic diversity? Either a climatic or techno-environmental shift within Africa allowed some resident groups to become dominant, or there was a new influx of population from outside the continent which overwhelmed the *in situ* groups. This latter view would have previously seemed difficult to believe but recent observations in genetics are giving it increasing credibility. The source of such an influx relates to an old observation that Papuan populations are surprisingly similar, phenotypically, to African populations. An earlier generation of scholars had wondered if the two were not somehow related, but without archaeological or genetic results, this remained in the wild speculation box. The older explanation for phenotypical similarities, that melanin pigmentation is simply a result of an adaptation to exposure to ultra-violet, is still maintained in some quarters (Blum 1961, but also Jablonski & Chaplin 2000 for a more recent version).

However, once the 'out-of-Africa' hypothesis became well established in the scientific literature, it became reasonable to imagine that Papuans *were* related to Africans at the level of original migrations of *H. sapiens* out of Africa, which has strong support from genetics (Koda *et al.* 2003). In other words, when modern humans left Africa, one of their phenotypes was dark-skinned, curly-haired etc. and this physical type remained in Africa and in Papua but was driven out or assimilated in the intervening spaces. Relict groups such as the Andamanese, the Orang Asli of peninsular Malaysia, the Agta in the Philippines, perhaps the Vedda of Sri Lanka and the vanished Vazimba of Madagascar would then have been remnants of this movement.

However, this did not satisfactorily explain why the best candidates for the descendants of the original *Homo sapiens*, the Khoesoids of southern Africa, had a distinctly different phenotype. Strangely, there seems to be little trace of their physical type outside Africa. One possible explanation for this situation is that most present-day Africans resemble Papuans because their ancestors migrated from the eastern side of the Indian Ocean back westwards, re-entering Africa, with skills, technology and perhaps social/ritual systems, spread out across Africa, and gradually displaced or assimilated many of the resident populations. This argument seems to have been first put forward in its modern form by Kingdon (1993) although in the absence of modern genetic evidence it was little more than speculation. If this argument is accepted, these early immigrants would be at the origin of Nilo-Saharan, as this is the oldest of the resident phyla apart from Khoesan.

What would be the motivation for this extraordinary reverse movement? If we accept the route of the initial expansion it is most logical that these were coastal movements, strandlopers gathering shellfish, crustacean and small fish in rock-pools but without open sea capability. It is now fairly widely accepted that there were two routes out of Africa, through the Sinai peninsula and across the Bab el Mandeb, from the Horn of Africa to Yemen (Quintana-Murci *et al.* 1999; Stringer 2000). When this second route opened is debated, but presumably prior to 70,000 BP, to account for the dates of the peopling of Australia.

Does this have any support in the archaeological record or from genetics? In the case of the latter, the surprising answer is yes. Harding *et al.* (1997) have shown that ‘Asian’ lineages play an important role in human ancestry. Cruciani *et al.* (2002) undertook a major analysis of Y chromosome patterns in sub-Saharan Africa and find intriguing evidence for a ‘back-migration’ from Asia, based on Haplogroup IX chromosomes. They also refer to other studies that have reached similar conclusions;

Interestingly, phylogenetic analysis of primate T-cell lymphotropic viruses type 1 indicate a putative Asian origin (Vandamme *et al.* 1998) followed by a simian- or human-mediated introduction to Africa 20,000 years ago (Van Dooren *et al.* 2001). An ancient human back migration from Asia to Africa had already been proposed by Altheide and Hammer (1997) and Hammer *et al.* (1998, 2001), on the basis of nested cladistic analysis of Y-chromosome data.

Cruciani *et al.* (2002:1210)

The emphasis on Y-chromosomes in these data may well be significant; these ancient boat people could have had a pattern of exploration of new fishing and gathering grounds that involved groups of males setting off on survey journeys unaccompanied by their women. Landing back in Africa and encountering resident human populations, they could have intermarried extensively with these groups. The spectacular morphological and phonological diversity of African languages could perhaps be a consequence of the interaction between the speaker’s incoming languages and those already in situ.

Are there any possible linguistic correlates of this? To introduce wild speculation into the mix, it has long been noticed that there are certain similarities between the noun-class systems of African languages and those of Papua. Noun-classes are evidently a typological feature of world languages by the pattern shown by African and some Papuan languages is unusual (Aikhenvald 2000). Abu, an East Papuan language, has fifteen noun classes and concord, for all the world like a Bantu language. Typically, Papuan languages have many fewer classes. It is not yet clear whether Abu represents a survival of a more complex system or a ramping up of the typical simpler systems. Noun-classes are typical of Niger-Congo and most branches of Nilo-Saharan<sup>2</sup>. Is it possible there is a connection between the two, forged at a deep level of history?

### 3. Methodological issues

Much has been made of the ‘new synthesis’ in recent years, the integration of archaeology, linguistics and genetics. A key assumption of this type of trans-disciplinary enterprise is that results can be matched, that patterns of language distribution are in principle congruent with archaeology and genetics. The argument is that since both archaeology and linguistics are direct reflections of human activities, they must in some way be congruent. One good reason for thinking this is there is a clear congruence in the present; culture and language *are* clearly linked and divergences can be explained by relatively well-established sociolinguistic processes<sup>3</sup>.

This potential for congruence is *not* necessarily the case with genetics; genes are not people, and they can have a distributional logic quite different from languages and cultures. They reflect extensive and complex patterns of human interactions with each other and the environment at a one-to-one and one-to-many level. It seems to be no quirk of the analytic process that maps of different marker systems and haplogroups seem to reflect geography more than ethnicity and often do not map against each other. One consequence of this is ‘Cavalli-Sforza syndrome’ where the geneticist sorts frantically through a series of highly diverse maps and eventually finds one that approximately corresponds to a known linguistic or archaeological grouping. By extracting different statistical components the match is improved and a triumph for congruence is

---

<sup>2</sup> This might be taken as a controversial statement; Nilo-Saharan is not usually seen as a noun-class phylum. This is because the systems are everywhere eroding or have been recently rebuilt. But more and more examples have been reported of suffixing systems with semantic content, suggesting that this is typical of the phylum (see Dimmendaal 2000 on Nilo-Saharan number systems).

<sup>3</sup> English is the most intensively studied language in the world, and recent explorations of its varieties make it perfectly possible to account for both variation and the congruence or otherwise of the cultures of those who speak it.

announced. But this is frankly improbable; all other types of data suggest that on any large contiguous land mass, populations interact in such an intensive and complex fashion as to make congruence an unlikely outcome. Only on islands or other isolated locales where interaction with external populations is more constrained is there a conservation of genetic traits; hence there has been more success with integrated accounts in Polynesia than in Eurasia or Africa. There may also be an issue of scale; the conclusions of genetics seem better drawn over very large areas. Much of the recent successes of genetics have been in modelling of the expansion of modern humans out of Africa, the demonstration that there was no genetic interchange with *in situ* hominids and the demarcation of different migration routes. This is reflected in typical 'geographical' results, where patterns underlie the broad processes of human expansion, rather than the micro-movements associated with local cultural processes.

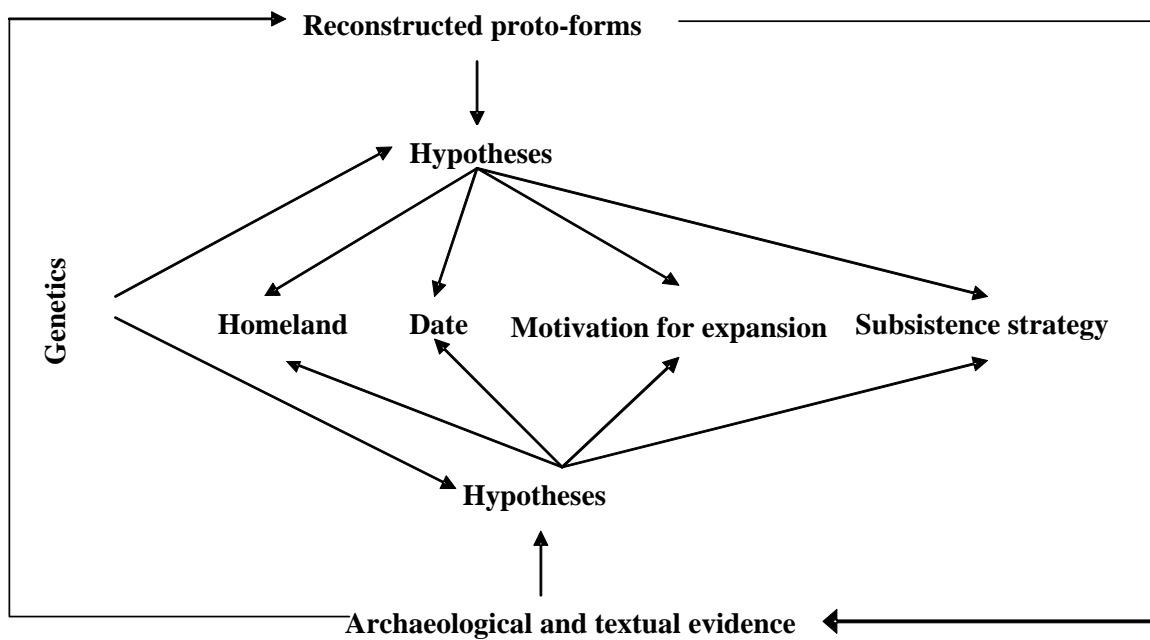
Another output of genetics where congruence is difficult to interpret are dates. Geneticists often believe in a 'clock' that can measure the speed of genetic divergence and thus they assign dates to human dispersals. It seems difficult to know how such a clock can be calibrated against anything except archaeology. Apart from the remaining adherents of glottochronology, few linguists now believe that language dispersals can be dated except by correspondences with archaeology; the same should hold for genetics.

Related to this is an issue often raised by geneticists, that of language diversity. Genetics can often put a quantitative measure on diversity and wonder whether this can be mapped against linguistic diversity. This seems as if it ought to work, but it doesn't, because languages diversify in different ways. Australian and Papuan are well-known for being highly diverse lexically and extremely uniform phonologically. Daic languages are quite uniform lexically but extremely diverse tonally. Khoesan and Nilo-Saharan languages are diverse in almost every conceivable way. Mountain *et al.* (1992) report on measures of diversity within Sinitic, but show that different categories of linguistic feature show different levels of diversity. This is not to say that diversity carries no information at all. The diversity within Australian and Papuan clearly indicate the long-term settlement of these regions; but whether anything more precise can be extracted from their variety is open to question.

An important but little-discussed aspect of the methodology of genetics is the targeting of sample collection. The hard science aspect of genetics has often blinded journal referees to the highly unscientific nature of the samples that are analysed. Thus we can find 'three West Africans' compared to 'Mbuti pygmies'. Even now, most of the articles cited above depend on 'out of the freezer' materials, often exchanged between laboratories. But if we are really to solve some of the major problems of ethnic and language correlations then what is required is targeted sampling; i.e. collecting samples that are statistically valid and reflect closely the particular groups that are the focus of the study. It is thus unacceptable to make claims about 'Nilo-Saharan' when in fact only one or two groups have been sampled, often from related populations. Ethnolinguistically targeted DNA collection is presently under way and some more coherent results may emerge within a few years.

Figure 1 shows the feedback loops involved in building a convincing model that relates historical reconstruction in linguistics with archaeology and genetics. Although ideally hypotheses generated by different disciplines should be developed independently, at least in the initial phase, this is rarely the case.

**Figure 1. Feedback relations in historical reconstruction**



In future, a diagram like this will certainly have a more integrated place for genetics, but the lack of present consensus on exactly what role genetics can play in the interpretation of phylic expansion means that it presently plays only a supporting role. In other words, archaeologists and linguists are pleased if genetics appears to support their hypotheses, but they are unlikely to change them because of a recent article in *Nature*.

#### 4. Case studies

##### 4.1 The Bantu expansion

The Bantu expansion is relatively recent compared in terms of the world’s language phyla and its early phase took place within sparsely inhabited tropical rain forest. Hence, it is almost certainly an example of demic expansion. The widely accepted model has the Bantu splitting into at least two groups, one heading east along the northern edge of the rainforest and the other staying in the west and moving south and southeast through the rainforest. The relatively recent date of these events has made it possible to link particular groupings with pottery styles in a manner that is so far not possible elsewhere in Africa (Phillipson 1977). Eggert (1992), while taking a critical approach to simplistic correspondences between pottery styles and Bantu subclassification, nonetheless makes it evident that the different ceramic traditions, notably the *Pikunda-Munda*, on the Sangha and neighbouring rivers in Congo-Brazzaville and which date to ca. 2200 BP represent an ‘aquatic settlement’ of this inhospitable region. Wotzka’s (1995) detailed study of archaeological pottery types in Central Africa led him to link the intrusion of the ‘Imbonga’ style of ceramic on the main waterways of the DRC, dated 400-100 BC, to the incoming Bantu populations. Denbow (1986, 1990) describes the ceramics of Tchissanga near the mouth of the Congo, which consistently date to around the 6<sup>th</sup> century BC, and are linked to the Okala traditions in Gabon and those of Ngovo in the DRC. He links these to a major movement of Western Bantu-speakers towards the Kalahari, where they encountered Khoisan speakers. Leakey *et al.* (1948) first defined the ‘dimple-based ware’ that is characteristic of much of the East African region. This was later-renamed Urewe ware and is essentially similar to Kwale ware and first occurs in sites near the coast as early as 200 AD (Forsslund 2003). There is every reason to link this with the expansion of the Bantu east from the Great Lakes region to the coast. Urewe has been found in Mozambique and at Nelspruit in South Africa (Huffman 1970, 1980, 1989a,b, 1998) and this is potentially linked to the coastal movement southwards of the Eastern Bantu. Herbert & Huffman (1993) proposed that the other major ceramic tradition south of the rainforest, the so-called ‘Kalundu’ tradition, is linked with the

Western Bantu. In their version, the bearers of the Kalundu tradition emerge from the rainforest and migrate both eastward and southeast, eventually interlocking with the Urewe tradition somewhere in Zambia.

The potential for correlations between the distribution of the Bantu languages and genetics would seem to be high, and unpublished evidence suggests that there *are* sequences that link geographically distant Bantu-speakers with each other and with their immediate relatives in Cameroun. Underhill *et al.* (2001) have suggested the haplotypes defined by M2/PN1/M180 polymorphisms as markers of that expansion. These are the analogues of haplogroup 8 defined by the YAP<sup>+</sup>/sY81G polymorphisms. They present evidence of strong founder effects in that sub-clade (40% of the members share the M191 mutation). This was independently supported by results from Y-STR haplotypes in a South African Bantu population (Thomas *et al.* 2000), where the proportion of YAP<sup>+</sup>/sY81G lineages was 80±5%, of which more than half shared the same 6 Y-STR based haplotype or its one-step neighbours. Pereira *et al.* (2002) who tested Bantu groups from around Iberophone Africa<sup>4</sup> (including Cabo Verdeans, who are not Bantu and have an extremely mixed ancestry and Central African Pygmies whose status is at best controversial) support the notion that the Y-STR haplotypes are associated with the Bantu expansion.

These are exciting possibilities, but paradoxically they are rather marginal to the main literature on the Bantu. The preoccupations of those who fund genetics and the potential for media exposure have entirely displaced what might be described as more central approaches to the question. Genetics, like much science driven by a continuous need for funding, researches the fashionable. The slave trade is certainly *à la mode*, and the diaspora certainly has more resources to commit than are available in Africa. Commercial interests in identifying the ethnic origin of modern individuals is driving research on routes and mechanisms of the slave trade<sup>5</sup>. This is an extremely well-researched area historically and it is hard to see how genetics can make a genuine contribution, when our knowledge of the base populations within Africa remains so weak.

As a consequence, analyses of former translocated populations tend to show what would be expected. Mateu *et al.* (1997) look at the island populations of Bioko and São Tome and show that the Bubi of Bioko are the result of the ancient migration of a small founder population with virtually no admixture, whereas the populations of São Tome are more mixed and result from multiple recent movements (the São Tomeans were transported by the Portuguese in the 17<sup>th</sup> century and have no language of their own). Clist (1998) notes that although there was apparently an LSA population on Bioko, the Bubi probably reflect the earliest Neolithic pottery, the Carboneras tradition, presently dated to 560 AD. Trovoada *et al.* (2001, 2004) show that the populations of São Tome have little or no European lineages and that they have very limited substructure, presumably being transported from the opposite coast from the fifteenth century onwards. Pereira *et al.* (2001) have looked at some Southern African populations, especially in Mozambique, with a perspective of tracking European incursions and the impact of the slave trade. They confirm the importance of the Y-STR haplotypes as a marker of the Bantu expansion, but also conclude that some 8.8% of lineages can be attributed to a pre-Bantu substrate (presumably Khoisan). At the same time, evidence from 11Y-binary markers suggests that European males made a contribution of 5.9% to the present population.

Bandelt *et al.* (2001) have a more puzzling account of a different haplogroup that shows up in both Bantu and translocated populations. They say, 'The mtDNA haplogroup L3e, which is identified by the restriction site +2349 *MboI* within the Afro-Eurasian superhaplogroup L3 (-3592 *HpaI*), is omnipresent in Africa but virtually absent in Eurasia (except for neighbouring areas with limited genetic exchange). L3e was hitherto poorly characterised in terms of HVS-I motifs, as the ancestral HVS-I type of L3e cannot be distinguished from the putative HVS-I ancestor of the entire L3 (differing from the CRS by a transition at np 16223). An *MboI* screening at np 2349 of a large number of Brazilian and Caribbean mtDNAs (encompassing numerous mtDNAs of African ancestry), now reveals that L3e is subdivided into four principal clades, each characterised by a single mutation in HVS-I, with additional support coming from HVS-II and partial RFLP analysis. The apparently oldest of these clades (transition at np 16327) occurs mainly in central Africa and

---

<sup>4</sup> Much as Bantu might be thought to be a transboundary, precolonial entity, it seems its modern representatives are selected according to the distribution of colonial languages as far as genetic sampling goes.

<sup>5</sup> See <http://www.manataka.org/page267.html> for a sceptical view from the perspective of American Indians.

was probably carried to southern Africa with the Bantu expansion(s). The most frequent clade (transition at np 16320) testifies to a pronounced expansion event in the mid-Holocene and seems to be prominent in many Bantu groups from all of Africa. In contrast, one clade (transition at np 16264) is essentially restricted to Atlantic western Africa (including Cabo Verde). We propose a tentative L3e phylogeny that is based on 197 HVS-I sequences. We conclude that haplogroup L3e originated in central or eastern Africa about 46,000 ( $\pm 14,000$ ) years ago, and was a hitchhiker of much later dispersal and local expansion events, with the rise of food production and iron smelting. Enforced migration of African slaves to the Americas translocated L3e mitochondria, the descendants of which in Brazil and the Caribbean still reflect their different regional African ancestries.' The argument is that the L3e haplogroup originated with the pre-Bantu populations of Central Africa, presumably relics of the early expansion of modern humans, to judge by the date, and entered the expanding Bantu gene pool through inter-marriage, a few thousand years ago. Its incidence was high enough to be reflected in translocated populations in the New World.

From high antiquity to recent history<sup>6</sup>, one of the more bizarre episodes in the attempt to link genetics and linguistics is the case of the Lemba in Southern Africa<sup>7</sup> (Hendrickx 1991; Spurdle & Jenkins 1996; Thomas *et al.* 2000). If you believe the many websites, the Lemba are a black Southern African Bantu speaking group who have Jewish ancestry. They purportedly observe customs such as not eating pork, male circumcision, and keeping one day a week holy. According to their oral history, they came to Africa from "Sena in the north by boat". The original group, which is said to have been almost entirely male, made its way to the coasts of Eastern Africa. If the Lemba do indeed have Jewish ancestry then one might expect to find a similarity between the Y chromosomes of Lemba men and those of Jewish men living in other parts of the world.

Needless, to say, this has stimulated the 'lost tribes of Israel' lobby. Tudor Parfitt, a lecturer in Jewish Studies has made a miniature media career through a book and television programme, *Journey to the Vanished City: Search for a Lost Tribe of Israel* (Parfitt 1997). In this bold essay into the unknown the brave hero ventures into South Africa and lo and behold uncovers the Lemba. This was linked to a study that compared the Y chromosomes of around 136 Lemba to those of Ashkenazic and Sephardic Jews, Yemeni and non-Lemba Bantu speakers (Thomas *et al.* 2000). Researchers found evidence of Semitic origin in the Lemba, although it was not clear whether this origin was Jewish or Arab, or a mixture of both. The study also found that the Lemba carry the Cohen modal haplotype (CMH) at a frequency similar to that found in Jewish populations. The CMH has been suggested as a signature for the ancient Hebrew population. Non-Lemba Bantu speakers in the study did not carry the CMH. The researchers concluded that the Lemba most likely have a mixture of Jewish, Arab and Bantu origins, although the CMH present in Lemba men could have an exclusively Jewish origin. It is therefore claimed that the genetic evidence is therefore consistent with the Lemba oral tradition of a Jewish origin. Wilson & Goldstein (2000) even go so far as to refer to them as a 'Bantu-Semitic Hybrid Population'.

**Figure 2. Rejudaisation of Lemba leaders**



---

<sup>6</sup> It is hard not to be reminded of Karl Marx' observation "Hegel remarks somewhere that all great, world-historical facts and personages occur, so to speak, twice. He has forgotten to add: the first time as tragedy, the second as farce." Opening sentences of 'The Eighteenth Brumaire of Louis Napoleon'.

<sup>7</sup> See an uncritical review at [http://www.bioethics.umn.edu/genetics\\_and\\_identity/case.html#lemba](http://www.bioethics.umn.edu/genetics_and_identity/case.html#lemba)

The whole story has more than a whiff of Wilbur Smith. These ‘traditions’ are almost entirely spurious and do not date from the earliest records of the Lemba, but are an example of reinvention spurred by the interests of outsiders. The Lemba only now claim to be of Jewish origin because they have told this is the case, just as they now wear skull-caps and shawls in conformity with this spurious tradition. It seems very likely that even the claim that there is ‘Semitic’ DNA would be difficult to support and the likelihood that the frequency of CMH is really similar to modern day Jewish populations unlikely in the extreme. Bizarrely, the Lemba are now also claiming to have built Great Zimbabwe. Assuming the genetics result reflect anything at all, it is probably intermarriage with Arab traders in the past few centuries. It is noticeable that the Lemba have no language of their own and indeed no linguistic trace remains of their supposed Jewish forbears. Nonetheless, as quoted in Thomas *et al.* (2000), the Lemba are now writing books about themselves, recounting traditions of apparent Jewish origin. Search websites on Bantu genetics and this is the main topic they want to discuss; Judaic websites have begun to elaborate an entire mythology of the lost Jewish populations of sub-Saharan Africa<sup>8</sup>. Ruwita (1997) has indeed pointed to this reinvention but to no good effect. By some irony, a series of studies of the Falasha, the ‘Black Jews’ of Ethiopia, who have always claimed to be Orthodox Jews and who were certainly practising Jewish religion when first encountered by outsiders, show no Jewish genetic traits at all (Lucotte & Smets 1999).

**Figure 3. Lemba claiming to have built Great Zimbabwe**



#### 4.2 Afroasiatic, the Berber and the Guanche

Maca-Meyer *et al.* (2003) is a study of the U6 lineage, usually thought to be autochthonous in North Africa. It seems to have originated in the Near East >30 kya and spread along the North African coast. It then divides into (Table 1);

| <b>Sublineages</b> | <b>? Date</b>   | <b>Possible correlate</b> |
|--------------------|-----------------|---------------------------|
| U6                 | >30kya          |                           |
| U6a                |                 |                           |
| U6a1               | 13,289 ± 5,470  | Afroasiatic?              |
| U6b                | 24,411 ± 15,200 | Capsian                   |
| U6b1               | 5,830 ± 4,551   | Guanche                   |
| U6c                | 17,658 ± 12,862 |                           |
| U6c1               |                 | Guanche                   |

Source: Maca-Meyer *et al.* (2003)

These dates can only be given limited credence, but what is interesting is the U6a1 lineage which spread from East Africa (unspecified) back towards the Maghrib and the Near East in the frame of 10-20,000 kya. This does rather suggest the early expansion of Afroasiatic, for those who accept its origin in the Ethiopian region.

The Berber languages still today represent an extremely far-flung group, spoken from a remote isolated group in Mauritania, the Zenaga, to Siwa Oasis in Egypt. It is generally accepted that the languages of the Canaries, collectively known as Guanche, were Berber, but these became extinct before they could be recorded by professional linguists (Wölfel 1965). Recent work on the genetics of former Guanche

<sup>8</sup> See <http://dickinsg.intrasun.tcnj.edu/diaspora/part1/pwrpnt/Judaism/sld013.htm>

populations suggests that the Guanche represent an early migration of Berber populations, while those remaining on the mainland have undergone substantial reshaping (Maca-Meyer *et al.* 2003, 2004).

### 4.3 The origin of the African pygmies

The literature on African pygmies is extremely voluminous, although coverage of the many groups remains extremely uneven. The common view, however, is that the pygmies are the ancient denizens of the forest zone, dating from at least the Middle Stone Age (MSA) (e.g. Cavalli-Sforza 1968a). They would have lived by hunting and gathering until they encountered expanding Central Sudanic, Adamawa-Ubangian and Bantu-speaking farmers ca. 4000 bp<sup>9</sup>. Since that date they have lived in a symbiosis with the farmers, often as a despised and marginalised group. If this is the case, then major MSA archaeological sites in the area of the present-day rain-forest are presumed to be the traces of these ancient pygmy groups. There is no doubt the Central African rainforest has been occupied for a very long time (Clist 1995; Mercader and Marti 1999), but there is no direct evidence as to the racial or genetic affiliations of the populations whose stone tools have been recovered. These sites have problems of dating, but it is usually assumed that the sites, 'Sangoan' or 'Lupemban' are >40,000 years old (the usual limit of radio-carbon dating).

In view of their physical distinctiveness, the Pygmies have long been the subject of biological and genetic studies, such as those collected in Cavalli-Sforza (1986b). None of these results are unambiguous but the tendency is to underline the distinctiveness of the Pygmy populations. Recent genetic trees published in studies of the genomes of African populations appear to confirm this point of view, especially those in Cavalli-Sforza *et al.* (1994) which place the Mbuti as a primary branching of sub-Saharan populations.

A feature of the pygmy/Twa complex that is seldom fully analysed is the widespread presence of 'Twa' populations well south of the rainforest proper (Blench 1999). Indeed many maps of Pygmies (e.g. Bahuchet 1993) seem to ignore these southern groups presumably on the grounds that they disrupt the image of the *civilisation forestière*. However, there are 'Twa' populations in semi-arid Angola, Namibia, Botswana and Zambia (Dornan 1925; De Almeida 1965, 1994; Estermann 1976 [1958], 1983). These are characteristically found in swamps, especially those in Zambia and Botswana. Documentation on the Twa groups of the Namibia/Angola region is very limited and tends to confuse Khoisan populations with Twa. Estermann (1976) provides useful material on the Twa of Angola. He says:

The southern Twa today live in close economic symbiosis with the tribes among which they are scattered — Ngambwe, Havakona, Zimba and Himba. None of the individuals I have observed differs physically from the neighboring Bantu.

Estermann trans. Gibson (1976:32)

The evidence is that Twa populations exist even in areas where there is no historical evidence for rainforest. Most of the non-forest Twa are taller than the Pygmies and in some cases physically indistinguishable from the Bantu. There are two possible explanations for this;

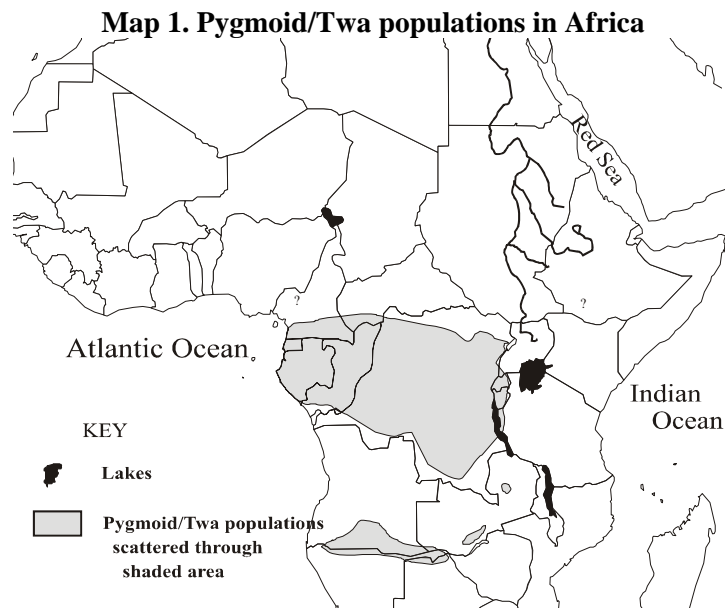
- a) either these are simply specialised hunter-gatherer groups who never were pygmies
- b) *or* they came along with the Bantu as they emerged from the forest and grew correspondingly taller in a savannah environment.

In most cases the savannah Twa seem to have the same despised 'shameful' status as the rainforest peoples, suggesting b) as the most likely scenario.

---

<sup>9</sup> Lower case dates indicate uncalibrated or estimated figures.

Map 1 shows the distribution of Pygmoid/Twa populations in Africa;



The biological literature has tended to argue that the pygmies are an ancient and separate race (Hiernaux 1968). Cavalli-Sforza *et al.* (1994:175 & 180) say ‘The San differ from other sub-Saharan Africans  $0.1082 \pm 0.0140$  that is, more than any sub-Saharan group differs from any other’ and ‘the San show no special association with Pygmies more than that with other sub-Saharan Africans’.

One of the most intractable problems in reconstructing African linguistic prehistory is whether the pygmies ever had their own language. Letouzey (1976) made a preliminary attempt to recover a substrate language through the use of plant names but without any very convincing results. Bahuchet (1992, 1993) presents a challenging view of the history of the pygmy populations, in particular the Aka and the Baka. Despite speaking Niger-Congo languages of quite different genetic affiliation, these groups prove to have common vocabulary, concerning especially with food-gathering in the rain-forest. If Bahuchet is right, then this vocabulary constitutes a trace of the lost language of the pygmies. Bahuchet further argues that the reduction in the rain-forest at the end of the Pleistocene isolated pygmoid groups in relict forest. These groups diffused outwards when the forest began to expand again, eventually encountering the incoming Bantu cultivators.

An alternative view is presented in Blench (1999), which argues that the absence of a true pygmy language is no accident: that the pygmies are to be identified genetically with their cultivator neighbours. Instead, the fragmentary hunter-gatherer peoples and isolated languages which today form a ring around the rain-forest represent the remaining traces of a lost complex of non-pygmoid hunter-gatherer populations speaking highly diverse language who inhabited Africa prior to the spread of the major language phyla. This diversity would then have been largely eliminated in the regions where the major language phyla expanded.

In the genetic analysis of the pygmy data the populations are divided, rather unsatisfactorily, into three groups;

- a) Mbuti- Eastern pygmies
- b) Aka - Western pygmies
- c) Pygmoids- all other pygmies including the Cameroun groups, the Rwandese Twa and those of NW Zaire (Baka?)

Of these groups only the Mbuti show any striking result; the others are so affected by ‘admixture’ as to be hardly distinct from other sub-Saharan groups. Cavalli-Sforza *et al.* (1994: Fig. 3.5.1) show a genetic tree mapping the genetic distance of sub-Saharan populations, and the Mbuti appear as one of the first branching

of sub-Saharan populations (although still closer to these than to NE Africans, which include most Afroasiatic speakers).

This might initially appear to be strong counter-evidence to the claim made in this paper. However the tree *also* groups together closely the following;

Sandawe (central Tanzania) with Fulfulde, Wolof and Serer (Senegambia)  
San (Southern Africa) with Somali (Horn of Africa)  
Kunama (northern Ethiopia) with SE Bantu  
Bantoid (central Cameroun) with Hausa (savannah West Africa)

These groupings are all geographically remote from one another and neither their cultures nor their languages have anything in common, being part of different phyla. Such conjunctions correspond to no known historical or archaeological data. It seems very difficult to know what meaning to attach to them or how to use them in any credible reconstruction of African prehistory. Another more disturbing aspect of this type of analysis is the way inconvenient conjunctions are removed when Cavalli-Sforza is writing directly of the links with language. Thus in Cavalli-Sforza (1991) where the standard genetic classifications of language phyla are mapped against the results from DNA, these inconvenient results have disappeared, appearing to make the match between disciplines more convincing than actually it is. As superficially attractive as DNA is for building models of African ethnohistory, all that cladograms show are mappings of conjunctions and frequencies of genes. There seems to be no particular reason why these *should* be codistributed with language phyla and apparently they are not. The results from genetics are so remote from the results derived from other disciplines that they cannot presently contribute to this debate.

## 5. Conclusion: where do we go from here?

At present the relationship between the configurations of African languages and genetics is more potential than actual. The achievements of genetics are more in the sphere of 'long ago and far away' than in the near present. We may find that there *are* deep-level relationships between the world's language phyla that correspond to the two routes out of Africa, and that the pattern of phyla such as Niger-Congo is related to the substrate languages that genetics suggests must once have existed. But in last few thousand years, the links are so far rather weak, and some of the claims in the literature need to be evaluated with scepticism.

One of the issues is that there is a mismatch between the categories geneticists use and the claims they make. 'Semitic' is a language grouping thrown around freely in the discussions over the affiliation of the Lemba. But Semitic is a large, complex linguistic grouping, many members of which have never been sampled for DNA and there is as yet no evidence for any polymorphism that characterises Semitic as a whole. Similarly, a few samples from two ill-defined ethnic groups in South Africa are made to stand for Bantu as a whole, encompassing some 1000 languages. Ethnolinguistic categories are being matched with polymorphisms in quite inappropriate ways.

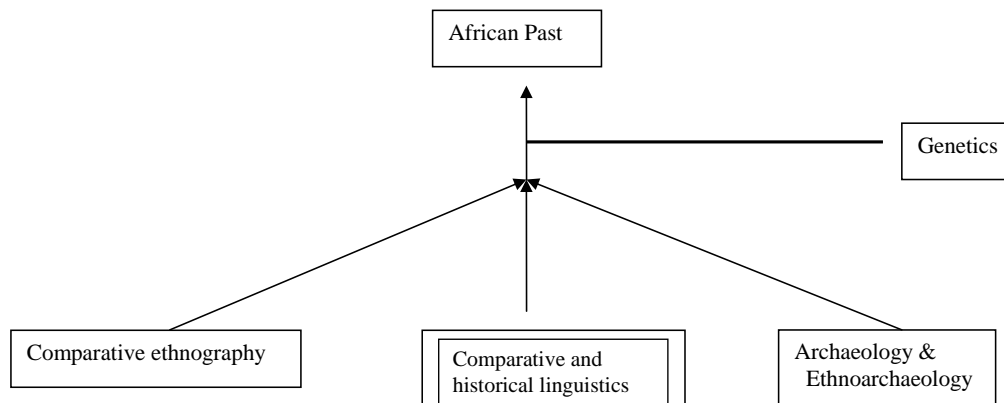
Genetic arguments are in danger of a certain circularity; to be congruent with other disciplines they will simply confirm their conclusions, as tangential observations will simply be ignored. In the case of the slave trade, we know that the ancestors of the African diaspora were carried to the offshore islands and the New World since 1500 and that there was a substantial genetic input from European males. Brakez et al. (2001) analysing Moroccan Berbers, concluded that the majority of their mtDNA was West Eurasian with a 26% contribution from sub-Saharan Africa, a conclusion which could have been fairly reached without recourse to a laboratory. Geneticists need to ask difficult questions.

For a more fruitful interchange, geneticists will need to develop a more ethnolinguistically informed procedure for obtaining their samples and a collaboration that asks interesting questions. Too often it is acceptable simply to use materials already 'in the freezer' and publish results from their analysis, regardless of whether any useful conclusion has emerged. Genetics can clearly generate illuminating results at great

levels of antiquity but we should be deeply suspicious of congruence with recent language expansions, especially where there has been substantial interaction with resident populations. Genetics *should* show different results from language and ethnographic studies if all we understand about the diversity of marital patterns and population shift is to be taken into account.

If these arguments are accepted, then a possible future configuration for reconstructing the African past can be imagined. Figure 4 imagines these varied disciplines converging on a single point;

**Figure 4. A possible future configuration for reconstructing the African past?**



## References [DRAFT]

- Aikhenvald, Alexandra Y. 2000. *Classifiers: a typology of noun categorization devices*. Oxford: Oxford University Press.
- Altheide, T.K., & M.F. Hammer 1997. Evidence for a possible Asian origin of YAP\_ Y chromosomes. *American Journal of Human Genetics*, 61:462–466.
- Arnaiz-Villena, Antonio, Jorge Martínez-Laso & Eduardo Gómez-Casado 2000. *Prehistoric Iberia : genetics, anthropology, and linguistics*. New York : Kluwer Academic/Plenum Publishers.
- Bahuchet, Serge 1993. *La rencontre des agriculteurs: les pygmées parmi le peuples d’Afrique centrale*. Histoire d’une civilisation forestière, II. Paris: Peeters-SELAF.
- Bandelt, H.-J. et al. 2001. Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Annals of Human Genetics*, 65,6: 549-563.
- Bastin Y., A. Coupeuz & M. Mann 1999. Continuity and divergence in the Bantu languages: perspectives from a lexicostatistic study. *Annales Sciences Humaines du Musée royal de l’Afrique Centrale de Tervuren*, vol. 162, Tervuren.
- Blench, Roger M. 1999. Are the African pygmies an ethnographic fiction? In: *Central African hunter-gatherers in a multi-disciplinary perspective: challenging elusiveness*. K. Biesbrouck, S. Elders & G. Rossel eds. 41-60. Leiden: Centre for Non-Western Studies.
- Blench, Roger M. 2004. Archaeology and Language: methods and issues. In: *A Companion To Archaeology*. J. Bintliff ed. 52-74. Oxford: Basil Blackwell.
- Bosch E, Calafell F, Pérez-Lezaun A, Comas D, Mateu E and Bertranpetit J. 1997. Population history of North Africa: Evidence from classical genetic markers. *Hum Biol*, 69:295-311.
- Brehm, A., Pereira, L., Bandelt, H.-J., Prata, M. J. & Amorim, A. 2002. Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of the Atlantic slave trade. *Ann. Hum. Genet.*, 66: 49-60.
- Brakez Z, Bosch E, Izaabel H, Akhayat O, Comas D, Bertranpetit J, Calafell F. 2001. Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol*. 28(3):295-307.
- Cavalli-Sforza, L.L. 1986a. *Pygmies as a model of the Upper Palaeolithic*. In Cavalli-Sforza, L.L. ed. *African Pygmies*. 422-426. Orlando, Florida: Academic Press.
- Cavalli-Sforza, L.L. ed. 1986b. *African Pygmies*. Orlando, Florida: Academic Press.
- Cavalli-Sforza, L. Luca 1997. Genes, peoples, and languages. *Proc. Natl. Acad. Sci. USA*. 94:7719–7724.

- Cavalli-Sforza, L. Luca *et al.* 1988. Reconstruction of human evolution: bringing together genetic, archaeological and linguistic data. *Proceedings of the National Academy of Science, USA*. 85, 6002-6006.
- Cavalli-Sforza, L.L., Piazza, A. and P. Menozzi 1994. *History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.
- Chen, J., R.R. Sokal, and M. Ruhlen, 1995. Worldwide analysis of genetic and linguistic relations of human populations. *Human Biology*, 67 (4), 595-612.
- Chen, Yu-Sheng *et al.* 2000. mtDNA Variation in the South African Kung and Khwe—and their genetic relationships to other African populations. *Am. J. Hum. Genet.* 66:1362–1383.
- Clist, Bernard. 1995. *Gabon: 100, 000 ans d'Histoire*. Libreville: Sepia.
- Clist Bernard 1998. Nouvelles données archéologiques sur l'histoire ancienne de la Guinée-équatoriale. *l'Anthropologie*, 102(2):213-217.
- Cruciani, Fulvio *et al.* 2002. A back migration from Asia to Sub-Saharan Africa is supported by high-resolution analysis of human Y-Chromosome haplotypes. *American Journal of Human Genetics*, 70:1197–1214.
- d'Errico, F., Henshilwood, C. S. & Nilssen, P. 2001. An engraved bone fragment from ca.70 ka year old Middle Stone Age levels at Blombos Cave, South Africa: implications for the origin of symbolism and language. *Antiquity*, 75:309-318.
- Dimmendaal, Gerrit J. 2000. Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics* 42(2): 214-261.
- Dina, Jeanne & Jean Michel Hoerner 1976. Étude sur les populations Mikea du Sud-Ouest de Madagascar. In : *Omalysy Anio, Hier et Aujourd'hui, Revue du Département d'Histoire, Antananarivo* : 269-286.
- Ehret C. 1998. *An African Classical Age: Eastern and Southern Africa in World History, 1000 BC to AD 400*, Charlottesville, University Press of Virginia.
- Ehret C. 1999. Subclassifying Bantu: the evidence of stem morpheme innovations. In J.-M. Hombert & L.M. Hyman (eds.), *Bantu Historical Linguistics : Theoretical and Empirical Perspectives*. 43-147. Stanford, CSLI Publications.
- Excoffier L. & al. 1987. Genetic and history of Sub-Saharan Africa, *Yearbook of Physical Anthropology*, 30, pp. 151-194.
- Excoffier L. & al. 1991. Spatial differentiation of RH and GM haplotype frequencies in sub-Saharan Africa and its relation to linguistic affinities, *Human Biology*, 63, 3, pp. 273-307.
- Hammer, M. 1994. A recent insertion of an Alu element on the Y chromosome is a useful marker for Human Population studies, *Mol. Biol. Evol.*, 11 (5), : 749-761.
- Hammer, M.F. *et al.* 1998. Out of Africa and Back Again: Nested Cladistic Analysis of Human Y Chromosome Variation. *Molecular and Biological Evolution*, 15(4):427–441.
- Hammer, M.F. *et al.* 2001. Hierarchical patterns of global human Y-chromosome diversity. *Molecular and Biological Evolution*, 18:1189–1203.
- Harding, R.M. *et al.* 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. *American Journal of Human Genetics*, 60:772–789.
- Hendrickx, B. 1991. The ancient origin of the Lemba (Mwenye): A critical overview of the existing theories. *Journal of Oriental and African Studies*, 3:172-193.
- Henshilwood, C.S. & Sealy, J. 1997. Bone artefacts from the Middle Stone Age at Blombos Cave, southern Cape, South Africa. *Current Anthropology* 38, 890-895.
- Henshilwood, C.S., d'Errico, F.E., Marean, C.W, Milo, R.G, Yates, R. 2001. An early bone tool industry from the Middle Stone Age at Blombos Cave, South Africa: implications for the origins of modern human behaviour, symbolism and language. *Journal of Human Evolution*, 41:631-678.
- Horai, S. *et al.* 1995. Recent African origin of modern human revealed by complete sequences of hominoid mitochondrial DNAs. *Proceedings of the National Academy of Science, USA*, 92:6720-22.
- Ingman, M, Kaessmann H., Pääbo S., Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature*, 408:708–713.
- Ke, Yuehai *et al.* 2001. African Origin of Modern Humans in East Asia: A Tale of 12,000 Y Chromosomes. *Science*, 292: 1151-1153.
- Kirch, Patrick V. 2000. *On the road of the winds*. Berkeley: University of California Press.

- Knight, Alec *et al.* 2003. African Y Chromosome and mtDNA Divergence provides insight into the history of click Languages. *Current Biology*, 13, 464–473.
- Koda, Yoshiro, Takafumi Ishida, Hidenori Tachida, Baojie Wang, Hao Pang, Mikiko Soejima, Augustinus Soemantri & Hiroshi Kimura 2003. DNA sequence variation of the human ABO-secretor locus (FUT2) in New Guinean populations: possible early human migration from Africa. *Human Genetics*, 113:534–541.
- Liesenbang, G. 1977. New light on Venda traditions: Mahumane’s account of 1730. *History in Africa*, 4:162–181.
- Lucotte, Gerard and Pierre Smets 1999. Origins of Falasha Jews studied by Haplotypes of the Y chromosome. *Human Biology*,
- Maca-Meyer, Nicole *et al.* 2003. Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genetics*, 4:15. <http://www.biomedcentral.com/1471-2156/4/15>.
- Maca-Meyer, Nicole *et al.* 2004. Ancient mtDNA analysis and the origin of the Guanches. *European Journal of Human Genetics*, 12:155-162.
- Mateu, E., D. Comas, F. Calafell, A. Pérez-Lezaun, A. Abade & J. Bertranpetit. 1997. A tale of two islands: population history and mitochondrial DNA sequence variation of Bioko and São Tomé, Gulf of Guinea. *Ann. Hum. Genetics*, 61:507–518.
- Mercader, J. & R. Marti 1999. Middle Stone Age Sites in the tropical forests of Equatorial Guinea. *Nyame Akuma*, 51:14-24.
- Molet, Louis 1960. Aperçu sur un groupe nomade de la forêt épineuse des Mikea. In : *Bulletin de l’Académie Malgache*, t. XXXVI : 241-243.
- Parfitt, T. 1997. *Journey to the vanished city*. London: Phoenix.
- Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M. J. & Amorim, A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Annals of Human Genetics*, 65:439-458.
- Pereira, L. *et al.* 2002. Bantu and European Y-lineages in Sub-Saharan Africa. *Annals of Human Genetics*, 66:369-378.
- Pitulko, V.V. *et al.* 2004. The Yana RHS Site: humans in the Arctic Before the last Glacial Maximum *Science*, 303: 52-56.
- Poloni E. S. & al. 1997. Human genetic affinities for Y-chromosome p49a,f/TaqI haplotypes show strong correspondence with linguistics, *American Journal of Human Genetics*, 61, pp. 1015-1035.
- Ruwitah, A. 1997. Lost tribe, lost language? the invention of a false Remba identity. *Zimbabwea*, 5:53–71.
- Salas, Antonio *et al.* 2002. The Making of the African mtDNA Landscape. *Am. J. Hum. Genet.* 71:1082–1111.
- Sands, B. 1998. *Eastern and Southern African Khoisan: evaluating claims of a distant linguistic relationships*. Quellen zur Khoisan-Forschung 14. Köln: Rüdiger Köppe.
- Scozzari R. & al. 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African Populations, *American Society of Human Genetics*, 65, pp. 829-846.
- Semino, Ornella *et al.* 2002. Ethiopians and Khoisan Share the Deepest Clades of the Human Y-Chromosome Phylogeny. *Am. J. Hum. Genet.* 70:265–268.
- Soodyall H. & al. 1996. MtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans, *American Society of Human Genetics*, 58, pp. 595-608.
- Soodyall H. & T. Jenkins (1993), Mitochondrial DNA polymorphisms in Negroid populations from Namibia : new light on the origins of Dama, Herero and Ambo, *Ann. Hum. Biol.*, 20, pp. 477-485.
- Spriggs, Matthew 1997. *The Island Melanesians*. Cambridge: Blackwell Publishers.
- Spurdle, A.B. & T. Jenkins 1996. The origins of the Lemba “black Jews” of southern Africa: evidence from p12F2 and other Y-chromosome markers. *Am J Hum Genet*, 59: 1126–1133.
- Stringer, Chris & R. McKie 1996. *African exodus: the origins of modern humanity*. New York: Henry Holt.
- Stringer, Chris 2000. Palaeoanthropology: coasting out of Africa. *Nature*, 405:24–5, 27
- Templeton, A.R. 1993. The “Eve” hypothesis: a genetic critique and reanalysis. *American Anthropologist*, 95:51–72.

- Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB. 2000. Y Chromosomes Traveling South: The Cohen Modal Haplotype and the Origins of the Lemba - the 'Black Jews of Southern Africa'. *American Journal of Human Genetics*, 66(2):674-686.
- Thomson, R. *et al.* 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci USA*, 97:7360–7365.
- Trovoada, M. J., Alves, C., Gusmão, L., Abade, A., Amorim, A. & M. J. Prata 2001. Evidence for population sub-structuring in São Tomé e Príncipe as inferred from Y-chromosome STR analysis. *Ann. Hum. Genet.*, 65:271-283.
- Trovoada, M. J., L. Pereira, L. Gusmão, A. Abade, A. Amorim, M. J. Prata Pattern of mtDNA 2004. Variation in Three Populations from São Tomé e Príncipe. *Annals of Human Genetics*, 68,1:40.
- Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Lahr, M. M., Foley, R. A., Oefner, P. J. & Cavalli-Sforza, L. L. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* 65:43-62.
- Van der Veen, Lolke J. & Jean-Marie Hombert n.d. *On the origin and diffusion of Bantu: a multidisciplinary approach.* Available at:[http://www.ddl.ish-lyon.cnrs.fr/Publications/Pdf/Van%20Der%20Veen/Lolke\\_ACAL.pdf](http://www.ddl.ish-lyon.cnrs.fr/Publications/Pdf/Van%20Der%20Veen/Lolke_ACAL.pdf)
- Vandamme, A-M., Salemi, M. & J. Desmyter 1998. The simian origins of the pathogenic human T-cell lymphotropic virus type I. *Trends in Microbiology*, 6:477–483.
- Vansina J. (1995), New linguistic evidence and 'The Bantu Expansion', *Journal of African History*, 36, pp. 173-195.
- Watson E. & al. (1996), MtDNA sequence diversity in Africa, *American Journal of Human Genetics*, 59, pp. 437-444.
- Watson, E., Forster P., Richards M. & H-J. Bandelt 1997. Mitochondrial footprints of human expansions in Africa. *American Journal of Human Genetics*, 61:691–704.
- Westphal, E.O.J. 1963. The linguistic prehistory of Southern Africa: Bush, Kwadi, Hottentot and Bantu linguistic relationships. *Africa*, 33: 237-265.
- Wilson, James F. and David B. Goldstein 2000. Consistent Long-Range Linkage Disequilibrium Generated by Admixture in a Bantu-Semitic Hybrid Population. *Am. J. Hum. Genet.* 67:926–935.
- White, Tim D. *et al.* 2003. Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature*, 423:742–747.