

## Report

# Ethiopians and Khoisan Share the Deepest Clades of the Human Y-Chromosome Phylogeny

Ornella Semino,<sup>1</sup> A. Silvana Santachiara-Benerecetti,<sup>1</sup> Francesco Falaschi,<sup>2</sup>  
L. Luca Cavalli-Sforza,<sup>3</sup> and Peter A. Underhill<sup>3</sup>

<sup>1</sup>Dipartimento di Genetica e Microbiologia, Università di Pavia, and <sup>2</sup>Servizio Pronto Soccorso, Accettazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Policlinico "San Matteo," Pavia, Italy; and <sup>3</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA

The genetic structure of 126 Ethiopian and 139 Senegalese Y chromosomes was investigated by a hierarchical analysis of 30 diagnostic biallelic markers selected from the worldwide Y-chromosome genealogy. The present study reveals that (1) only the Ethiopians share with the Khoisan the deepest human Y-chromosome clades (the African-specific Groups I and II) but with a repertoire of very different haplotypes; (2) most of the Ethiopians and virtually all the Senegalese belong to Group III, whose precursor is believed to be involved in the first migration out of Africa; and (3) the Ethiopian Y chromosomes that fall into Groups VI, VIII, and IX may be explained by back migrations from Asia. The first observation confirms the ancestral affinity between the Ethiopians and the Khoisan, which has previously been suggested by both archaeological and genetic findings.

Within extant African populations, both linguistic (Greenberg 1963) and genetic (Hiernaux 1975; Excoffier et al. 1987; Cavalli-Sforza et al. 1994, pp. 169–171) evidence indicates that most sub-Saharan populations are more closely related to each other, whereas Pygmy, Khoisan, and eastern African populations are the most differentiated. Paradoxically, genetic comparisons of Khoisan and Ethiopian populations show both polarity and affinity with respect to one another. This has been shown by the principal-components (PC) analysis of 79 classical protein polymorphisms (Cavalli-Sforza et al. 1993, 1994, p. 191). Although the second PC indicates that the Ethiopian and Khoisan populations are the most divergent, the third PC shows a close relationship. Although intermediary Bantu-speaking populations currently separate these two groups geographically, archaeological findings suggest that the Khoisan territory once extended above the equator, to present-day southern Ethiopia and Sudan (Nurse et al. 1985, p. 105).

In a previous study (Passarino et al. 1998), the genetic

structure of the Ethiopian population was investigated using mtDNA and some nonrecombinant Y-chromosome (NRY) markers previously studied in the Khoisan (Soodyall and Jenkins 1992; Spurdle and Jenkins 1992). These markers, because of their uniparental inheritance and lack of recombination, are particularly useful for inferring the history of populations through female and male lineages separately. Although the mtDNA did not reveal a particular relationship between Ethiopians and the Khoisan, affinities were suggested by Y-chromosome analyses. The YAP<sup>-</sup>/49a,f haplotype 26 (A<sub>2</sub>C<sub>0</sub>D<sub>0</sub>F<sub>0</sub>I<sub>1</sub>) combination appeared to be typical of these two groups (with a frequency of ~7% in the Ethiopians and 10%–15% in the Khoisan; for the Khoisan frequency, see the discussion by Passarino et al. [1998]). With the exception of some Jewish subjects, particularly Ethiopian Jews (Ritte et al. 1993; Santachiara-Benerecetti et al. 1993), the 49a,f haplotype 26 is absent or extremely rare in all surveyed populations (found only by Torroni et al. [1990] and Persichetti et al. [1992]). However, because of the variability of the complex 49a,f system, a polyphyletic origin for haplotype 26 could not be excluded. A later combined study of seven biallelic markers and four microsatellites showed that the Ethiopians and Khoisan shared the “archaic” haplotype 1A (Hammer et al. 1998), defined by the marker SRY10831 A→G (Whitfield et al. 1995), but that they did not share the

Received September 27, 2001; accepted for publication October 23, 2001; electronically published November 20, 2001.

Address for correspondence and reprints: Dr. Ornella Semino, Dipartimento di Genetica e Microbiologia, Università di Pavia, Via Ferrata, 1, 27100 Pavia, Italy. E-mail: semino@ipvgen.unipv.it

© 2002 by The American Society of Human Genetics. All rights reserved.  
0002-9297/2002/7001-0026\$15.00

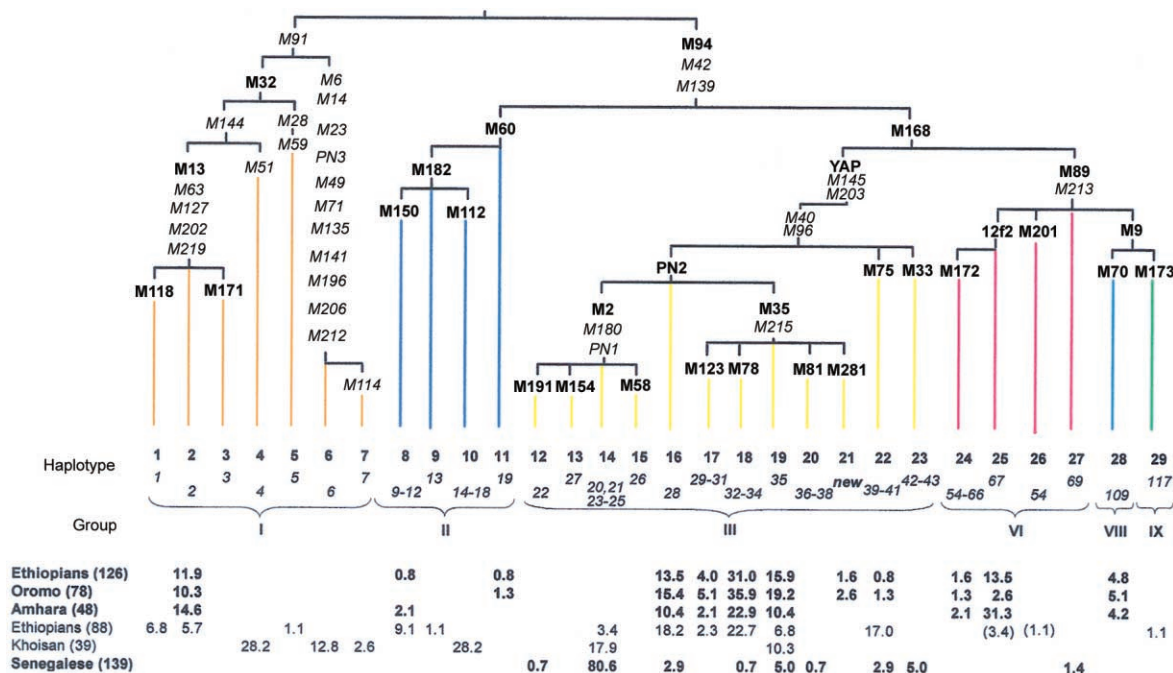
microsatellite variants (Scozzari et al. 1999). Most recently, a great number of Y-chromosome biallelic markers have become available (Underhill et al. 1997, 2000, 2001; Hammer et al. 2001). These markers, because of their very low mutability, have most likely arisen only once during human evolution, thus allowing a clear-cut definition of the worldwide Y-chromosome genealogy (Underhill et al. 2001).

To better understand the relationships between Ethiopians and the other African populations, we have now screened 126 Ethiopian (78 Oromo and 48 Amhara) and 139 Senegalese DNAs of our collection, for the diagnostic markers of the major haplogroups of the Y-chromosome genealogy (Underhill et al. 2001). The results obtained using a hierarchical approach are illustrated in figure 1, in which the Ethiopian and Khoisan samples examined by Underhill et al. (2000) are also included.

Groups I and II are essentially restricted to Africans and appear to be the most divergent clades within the tree. They show a patchy distribution, with high frequencies among isolated hunter-gatherer groups and in some peoples of Ethiopia and Sudan. Such a distribution was interpreted as the survival of some ancient lineages through more recent population events (Underhill et al.

2001). In particular, Group I, observed in 43.6% of the Khoisan (usually considered to be descendants of an early African population), is present in all of the Ethiopian samples: its frequency is 10.3% in the Oromo sample and 14.6% in the Amhara sample of the present study, and is 13.6% in the ethnically undefined sample reported by Underhill et al. (2000). In contrast, it was not found in the Senegalese. It is worth noting that the Ethiopian YAP<sup>-</sup>/49a,f haplotype 26 lineage, which is common within the Khoisan (Spurdle and Jenkins 1992; Passarino et al. 1998), corresponds to Group I and possibly reflects the signal seen in the third PC of classical polymorphisms (Cavalli-Sforza et al. 1993, 1994). However, figure 1 shows that the Ethiopian and Khoisan samples within Group I fall into different haplotypes (haplotypes 1, 2, and 5 in Ethiopians vs. haplotypes 4, 6, and 7 in the Khoisan), in agreement with an ancient divergence from the same ancestral population, as has been suggested by microsatellite data (Scozzari et al. 1999).

Group III of Underhill et al. (2001) includes three main clusters (PN2 [Hammer et al. 1997], M75, and M33) that are uniquely characterized by M40 (corresponding to SRY4064 transition of Whitfield et al.



**Figure 1** Phylogenetic tree of the Y-chromosome haplotypes and their percent frequencies in the two Ethiopian groups (Oromo and Amhara) and in the Senegalese of the present study, compared with the frequencies in the Ethiopians and Khoisan previously reported by Underhill et al. (2000). Numbering of mutations is according to Underhill et al. (2001): those examined in the present study are shown in boldface type; those inferred or reported by Underhill et al. (2000) are shown in italics. Haplotype numbers in the present study are shown in boldface type; those in italics are from Underhill et al. (2001). Group numbers refer to groups reported by Underhill et al. (2001). The frequency values in parentheses correspond to the subclassification of the haplotype defined only by the M89 mutation in the report by Underhill et al. (2000).

[1995]). The majority of African Y chromosomes fall into either the M2 or M35 subclades of the PN2 cluster (Underhill et al. 2001).

Figure 1 shows that virtually all of our Ethiopian YAP<sup>+</sup> Y chromosomes fall into either haplotype 16, characterized only by the PN2 mutation (Hammer et al 1997), or the M35-related haplotypes 17, 18, and 19. A new M35 haplotype (haplotype 21) was observed in two Oromo. This is defined by the G→A transition (M281 in fig. 1) at position 280 within the sequence-tagged site containing the M67 and M68 mutations associated with Group VI, which were described by Underhill et al. (2000). Noteworthy is the particularly high frequency of haplotype 18, defined by M78, which also characterizes most of the European YAP<sup>+</sup> chromosomes (O.S., unpublished data), and the absence of the haplotype 20, identified by the M81 mutation, which is the most frequent M35 lineage in North Africa (Bosch et al. 2001). In a comparison of the different groups of Ethiopians, the Oromo show an incidence (62.8%) of the M35 cluster higher than that in the Amhara (35.4%,  $P < .005$ ); the Amhara value is similar to the frequency (31.8%) found in the Ethiopian sample of Underhill et al. (2000). A consistent proportion (17.0%) of Y chromosomes belonging to the M75 cluster (haplotype 22) is a distinctive feature of the latter sample. In contrast, almost all Senegalese (98.6%) are YAP<sup>+</sup>, and the majority of them (81.3%) fall into the M2 subclade, but only one of them shows the M191 mutation (haplotype 12) (Underhill et al. 2001). This mutation accounts for ~40% of the M2 members, who are mainly Pygmies (Underhill et al. 2000). Group III is less frequent in the Khoisan (28.2%), who share with Ethiopians only the M35 haplotype 19 (10.3%). Conversely, the M2 component, which occurs at a frequency of 17.9% in the Khoisan, is virtually absent in the Ethiopians.

Group VI was observed almost exclusively as the 12f2 subgroup in the Ethiopians. Among them, the Amhara are by far the most important component (33.4%, vs. 3.8% for the Oromo [ $P < .0001$ ] and 3.4% for the other Ethiopian data [ $P < .0001$ ]). This difference, not revealed in the study by Passarino et al. (1998), in which the Oromo were underrepresented, might reflect distinct population histories. It is reported (Levine 1974) that the Amhara experienced a strong influence from Middle Eastern populations, in which the 12f2 8-kb allele has a very high frequency and probably originated (Santachiara-Benerecetti et al. 1993; Semino et al. 1996; Quintana-Murci et al. 2001). This is further supported by the opposite distribution of the M35 subclade (35.4% for the Amhara, vs. 62.8% for the Oromo [ $P < .005$ ] and 31.8% for the other Ethiopian data). Group VI also includes two Senegalese who, however, are currently defined only by the M89 mutation (haplotype 27) and lack

any other known mutation characterizing the M89 subgroups.

Groups VIII and IX were also found in the Ethiopians as haplotypes characterized by the mutations M70 (haplotype 28) and M173 (haplotype 29), respectively. M70 was observed in few of our Ethiopians (~5%), and M173 was found in just one subject in the Ethiopian sample of Underhill et al. (2000). The finding of M70 is intriguing, since it has so far been observed to be widely scattered in several continents at a low frequency (Semino et al. 2000; Underhill et al. 2000). The M173 and related lineages are common and widespread in European and in western and central Asian populations (Semino et al. 2000; Underhill et al. 2000; Bosch et al. 2001; Wells et al. 2001); the observation of one M173 in Ethiopia could, therefore, represent a recent admixture event.

In conclusion, the present study underscores the complexity and substructure of the Ethiopian Y-chromosome gene pool. First, the presence of different Y-chromosome haplotypes belonging to African-specific Group I in all groups of Ethiopians and in the Khoisan (at frequencies of ~13% and 44%, respectively) confirms that these populations share an ancestral paternity, as was previously suggested by the 49a,f data (Passarino et al. 1998), and it indicates that Group I was part of the proto-African Y-chromosome gene pool. The virtual absence of this clade in the other African ethnic groups suggests that they could derive from a more recent ancestral population that went through a long period of differentiation before expansion. In addition, Group II, the next closest to the NRY genealogy root and typically an African group, is shared by Ethiopians and the Khoisan but to a lesser degree. In the case of Group II, the split responsible for the differences observed between Ethiopian and Khoisan haplotypes is also old. Second, most of the Ethiopian Y chromosomes, the rest of the Khoisan Y chromosomes, and the majority of the Senegalese Y chromosomes belong to Group III, which is also mainly African but whose precursor is believed to be involved in the first migration out of Africa (Underhill et al. 2001). Third, the remainder of the Ethiopian Y chromosomes (Groups VI, VIII, and IX) may be explained by back migrations from Asia.

## Acknowledgments

We warmly thank L. Excoffier, who provided us with some Senegalese samples. This research was supported by the Italian Ministry of the University "Progetti Ricerca Interesse Nazionale" and by a "Fondo d'Ateneo per la Ricerca" dell'Università di Pavia grant to A.S.S.-B. P.A.U. was supported by National Institutes of Health grants GM 28428 and GM 55273 to L. L.C.-S.

## References

- Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* 68:1019–1029
- Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. *Science* 259:639–646
- (1994) *The history and geography of human genes*. Princeton University Press, Princeton
- Excoffier L, Pellegrini B, Sanchez-Mazas A, Simon C, Langaney A (1987) Genetics and history of sub-Saharan Africa. *Yearb Phys Anthropol* 30:151–194
- Greenberg JH (1963) *The languages of Africa*. Indiana University, Bloomington
- Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol Biol Evol* 15:427–441
- Hammer MF, Karafet MT, Redd AJ, Jarjanazi H, Santachiara-Benerecetti AS, Soodyall H, Zegura SL (2001) Hierarchical patterns of global human Y-chromosome diversity. *Mol Biol Evol* 18:1189–1203
- Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchel RJ, Horai S, Jenkins T, Zegura SL (1997) The geographic distribution of human Y chromosomes. *Genetics* 145:787–805
- Hiernaux J (1975) *The people of Africa*. Scribner, New York
- Levine DN (1974) *Greater Ethiopia*. University of Chicago Press, Chicago
- Nurse GT, Weiner JS, Jenkins T (1985) The San yesterday and today. In: Harrison GA (ed) *The peoples of southern Africa and their affinities*. Clarendon Press, Oxford
- Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, Santachiara-Benerecetti AS (1998) Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am J Hum Genet* 62:420–434
- Persichetti F, Blasi P, Hammer M, Malaspina P, Jodice C, Terrenato L, Novelletto A (1992) Disequilibrium of multiple DNA markers on the human Y chromosome. *Ann Hum Genet* 56:303–310
- Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K (2001) Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet* 68:537–542
- Ritte U, Neufeld E, Broit M, Shavit D, Motro U (1993) The differences among Jewish communities—maternal and paternal contributions. *J Mol Evol* 37:435–440
- Santachiara-Benerecetti AS, Semino O, Passarino G, Torroni A, Brdicka R, Fellous M, Modiano G (1993) The common, Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. *Ann Hum Genet* 57:55–64
- Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellitto D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A (1999) Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. *Am J Hum Genet* 65:829–846
- Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 59:964–968
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155–1159
- Soodyall H, Jenkins T (1992) Mitochondrial DNA polymorphisms in Khoisan populations from southern Africa. *Ann Hum Genet* 56:315–324
- Spurdle A, Jenkins T (1992) Y chromosome probe p49a detects complex *PvuII* haplotypes and many new *TaqI* haplotypes in southern African populations. *Am J Hum Genet* 50:107–125
- Torroni A, Semino O, Scozzari R, Sirugo G, Spedini G, Abbas N, Santachiara-Benerecetti AS (1990) Y chromosome DNA polymorphisms in human populations: differences between Caucasoids and Africans detected by 49a and 49f probes. *Ann Hum Genet* 54:287–296
- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996–1005
- Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 65:43–62
- Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet* 26:358–361
- Wells RS, Yuldashva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L et al. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA* 98:10244–10249
- Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation on the human Y chromosome. *Nature* 378:379–380