Ann. Hum. Genet. (2001), **65**, 43–62
Printed in Great Britain

The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations

P. A. UNDERHILL[1]*, G. PASSARINO[1,5], A. A. LIN[1], P. SHEN[2], M. MIRAZÓN LAHR[3,4], R. A. FOLEY[3], P. J. OEFNER[2] AND L. L. CAVALLI-SFORZA[1]

[1] *Department of Genetics, Stanford University, 300 Pasteur Dr., Stanford, CA 94305–5120, USA*
[2] *Stanford DNA Sequencing and Technology Center, 855 California Ave, Palo Alto, CA 94304, USA*
[3] *Department of Biological Anthropology, University of Cambridge, Downing Street Cambridge CB2 3DZ, UK*
[4] *Departamento de Biologia, Inst. de Biociencas, Universidad de São Paulo, Rua do Matão, Travessa 14, No. 321, 05508–900 Cidade Universitária, São Paulo, Brasil*
[5] *Department of Cell Biology, Calabria University, Rende, Italy*

(Received 24.8.00. Accepted 16.11.00)

SUMMARY

Although molecular genetic evidence continues to accumulate that is consistent with a recent common African ancestry of modern humans, its ability to illuminate regional histories remains incomplete. A set of unique event polymorphisms associated with the non-recombining portion of the Y-chromosome (NRY) addresses this issue by providing evidence concerning successful migrations originating from Africa, which can be interpreted as subsequent colonizations, differentiations and migrations overlaid upon previous population ranges. A total of 205 markers identified by denaturing high performance liquid chromatography (DHPLC), together with 13 taken from the literature, were used to construct a parsimonious genealogy. Ancestral allelic states were deduced from orthologous great ape sequences. A total of 131 unique haplotypes were defined which trace the microevolutionary trajectory of global modern human genetic diversification. The genealogy provides a detailed phylogeographic portrait of contemporary global population structure that is emblematic of human origins, divergence and population history that is consistent with climatic, paleoanthropological and other genetic knowledge.

INTRODUCTION

A model for the origins of human diversity deduced from palaeontological evolutionary geography maintains that while the modern human species originates from a single evolutionary event, diversity is a result of subsequent multiple evolutionary events associated with various geographic range expansions, migrations, colonizations and differential survival of populations (Lahr & Foley, 1994). Overall, current paleoanthropological evidence would suggest an early set of dispersals across Africa and into Western Asia; an early southern dispersal into Asia and Melanesia; and a later one into Northern Eurasia. Overlain on these events are the contractions associated with the Last Glacial Maximum (LGM), and subsequent post-glacial expansions of both hunter-gatherers and agriculturists.

DNA sequences offer an evidentiary alternative to fossil-based pre-historical reconstructions (Jorde *et al.* 1998, Owens & King 1999). The uniparentally inherited non-recombining haploid mtDNA and the Y chromosome loci are particularly sensitive to the influences of drift, especially founder effect. Consequently these loci are ideal for assessing the origins of contemporary population diversity, and provide context for paleontological hypothesis testing (Foley, 1998). The combination of a recent

Correspondence: P. A. Underhill.
E-mail: under@stanford.edu

molecular age (Shen *et al*. 2000), and geographical structure, makes the NRY a sensitive genetic index capable of tracing the microevolutionary patterns of novel modern human diversity. Any and all population level forces and possible localized natural selection that reduces the effective male population size relative to females, will influence the genetic landscape.

We combine 205 PCR compatible binary NRY polymorphisms (Underhill *et al*. 2000; Shen *et al*. 2000) together with 13 additional markers from the literature to examine phylogeographical patterns that may record historical population migrations, mergers and divisions that account for the current spectrum of human variability. While extrapolating variation associated with a single gene to population history must be interpreted cautiously, the phylogeographic reconstruction presented here offers one such interpretation. It comprehensively integrates the prehistoric and Y-chromosome data, along with inferences from mtDNA and autosomal haplotypes, into a possible hypothesis for the evolution of human diversity. We attempt here to discuss the observed phylogeographic patterns of NRY variation in the context of global population diversification, and integrate it with paleoclimatological, paleoanthropological and other genetic knowledge. In developing our synthesis we have aimed at producing palaeodemographic hypotheses that are consistent with as many other lines of evidence as possible, and that are amenable to testing by further studies from a number of disciplines.

Materials and methods

Samples

DNA from 1062 men belonging to 21 populations was analysed. Further details on the ethnic affiliations of these samples are given in Underhill *et al*. (2000).

PCR

Primers designed for SMCY, DFFRY, UTY, and DBY covered all unique sequences and repeat elements other than LINE, yielding overlapping amplicons 300–500 bp in length. PCR conditions are given in Underhill *et al*. 2000 and Shen *et al*. 2000. All 218 polymorphisms are given in Appendix I (deposited at http://www.gene.ucl.ac.uk/anhumgen/) which lists primers, the primary reference for each marker, the specific DNA sequence variant and its location in the fragment. Two new markers (M223, M224) found while genotyping other markers are included.

DHPLC analysis

Unpurified PCR products were mixed at an equimolar ratio with a reference Y chromosome and subjected to a 3-min 95 °C denaturing step followed by gradual reannealing from 95 °C to 65 °C over 30 min. Ten μl of each mixture were loaded onto a DNASep® column (Transgenomic, San Jose, CA), and the amplicons were eluted in 0.1 M triethylammonium acetate, pH 7, with a linear acetonitrile gradient at a flow rate of 0.9 ml/min (14). Using appropriate temperature conditions, which were optimized by computer simulation (http://insertion.stanford.edu/melt.html), mismatches were recognized by the appearance of two or more peaks in the elution profiles.

DNA sequencing

Polymorphic and reference PCR samples were purified with QIAGEN (Valencia, CA) QIAquick spin columns, cycle sequenced with ABI Dye-terminator cycle sequencing reagents and analysed on a PE Biosystems 373A sequencer. Chimpanzee, gorilla and orangutan samples were also sequenced for each human polymorphic locus.

Results and discussion

The 218 NRY polymorphisms were used to deduce a phylogenetic tree based on the principle of maximum parsimony, in which a network of branches is drawn that minimizes the number of mutational events required to relate the lineages (Fig. 1). The ancestral alleles were deduced using

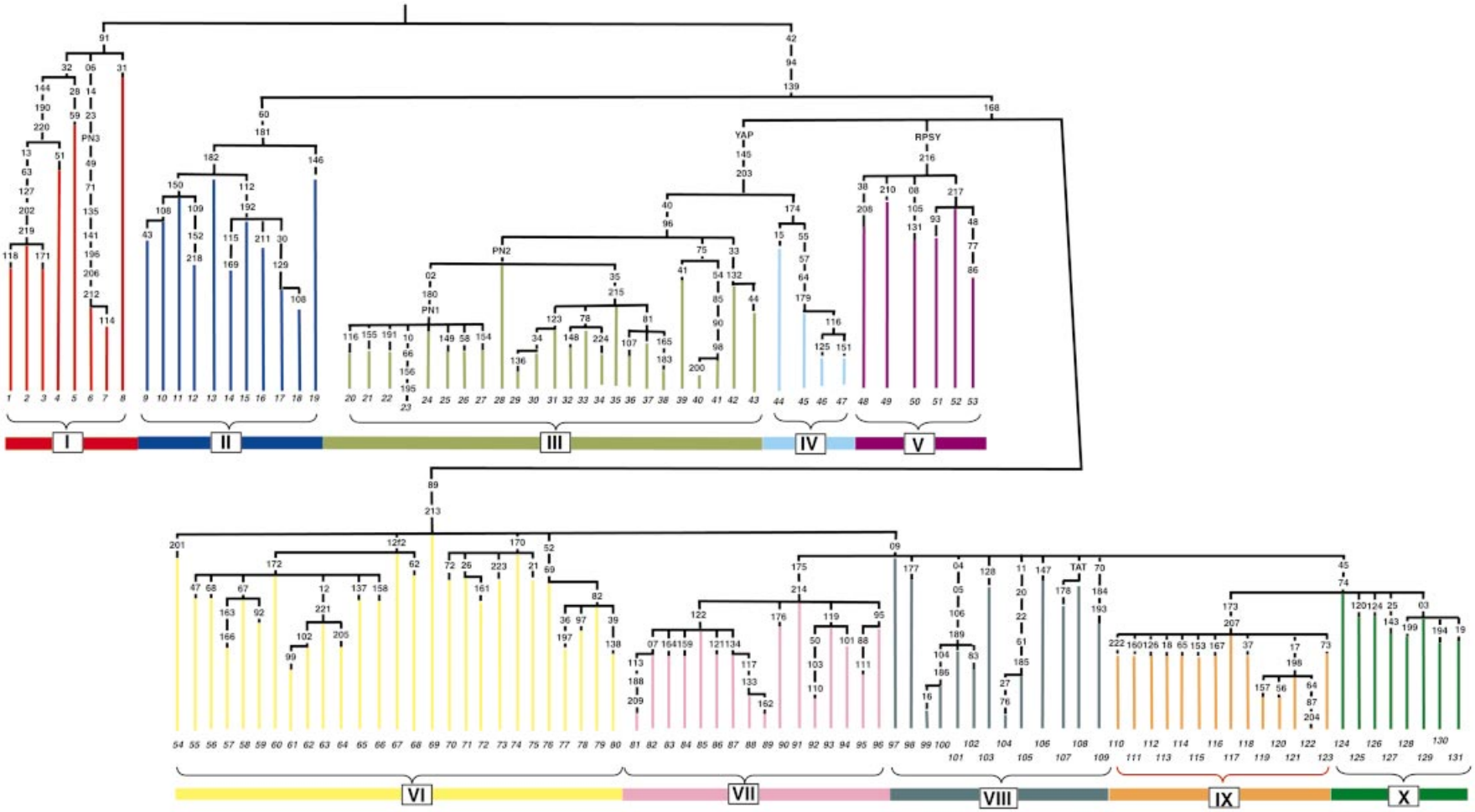

Fig. 1. Maximum parsimony phylogeny of human NRY chromosome biallelic variation. Tree is rooted with respect to non-human primate sequences. The 131 numbered compound haplotypes were constructed from 218 mutations that are indicated on segments. Marker numbers are discontinuous (see text). Haplotypes are assorted into 10 groups (I–X).

great ape sequence data to root the phylogeny. All phylogenetically equivalent mutations whose order cannot be determined are indicated with a slash (i.e. M42/M94/M139). Markers with M numbers > 218 reflect the selective removal of polymorphisms associated with recurrent length variations such as tetra- or pentanucleotide repeats and homopolymer tracts. The determination of the ancestral state for these polymorphisms is uncertain, and (with one exception, M91) they were excluded from the analysis (Underhill *et al.* 2000). The marker panel comprises 125 transitions, 66 transversions, 26 insertions/deletions, plus an Alu element. All polymorphisms except one are biallelic. A double transversion, M116, has three alleles whose derived alleles define quite different haplotypes. Two transitions (M64 and M108) showed evidence of recurrence but cause no ambiguity. No reversions were observed, although one transition, SRY10831 (Whitfield *et al.* 1995), also referred to as SRY1532 (Kwok *et al.* 1996), is known to be a reversion (Hammer *et al.* 1998). It is not included here as we have phylogenetically stable transversion and deletion polymorphisms that unequivocally mimic its patterns.

Haplotypes are partitioned into haplogroups (called Groups I–X) in an attempt to simplify discussion of phylogeography, using the simple criteria of presence or absence of alleles located in the interior of the phylogeny. These discretionary Group designations provide a framework for categorization and discussion of haplotypes. The Y genealogy is composed of 131 haplotypes that delineate the 10 Groups, seven of which are monophyletic. Three groups are polyphyletic, but have related haplotypes defined as follows: the presence of M89/M213 and absence of M9 (Group VI); the presence of M9 and absence of M175/M214 and M45/M74 (Group VIII) or the presence of M45/M74 and absence of M173/M207 (Group X). The contemporary global frequency distribution of the 10 Groups based on > 1000 globally diverse samples genotyped using a hierarchical top down approach is illustrated in Figure 2, which is based upon frequency data given in Underhill *et al.* 2000. Autochthonal variations associated with the NRY, in addition to tracing a common African heritage, resolve numerous population subdivisions, gene flow episodes and colonization events. They show the overall pattern of the progressive succession of Group differentiation and movement across the world reflective of expansions and genetic drift processes.

This composite collection of 218 NRY variants provides improved resolution of extant patrilineages. Additional resolution will occur with the discovery of new delimiting markers. The succession of mutations is unequivocal except in branches defined by two or more markers. While uncertainties related to assessing the effective population size of males make temporal estimates of bifurcation events difficult, age estimates of key nodes have been made assuming a model of population growth (Thomson *et al.* 2000). These indicate a more recent ancestry of the NRY at 59000 years (95% CI = 40000–140000) than previously estimated at 134250 ± 44980 years based on 13 mutational events and constant population size (Karafet *et al.* 1999). Neither demographic model is likely to be realistic, as the palaeoanthropological evidence shows a more complex population history. It should be noted that the lower estimate is considerably younger than the earliest evidence for dispersals of modern humans.

Phylogeography

Intriguing clues about the history of our species can be derived from the study of the geographic distribution of the lineages on the tree in Figure 1, in the approach known as 'phylogeographic' (Avise *et al.* 1987). Such an approach has been previously used for mtDNA networks (Richards *et al.* 1998, 2000; Kivisild *et al.* 1999; Macaulay *et al.* 1999). Figure 3*a–h* depicts the hypothesized chronological geographic distribution of Y Groups from the Isotope Stage 5 interglacial to the Holocene. The underlying assumption of phylogeography is that there is a correspondence between the overall distribution of haplotypes and haplogroups and

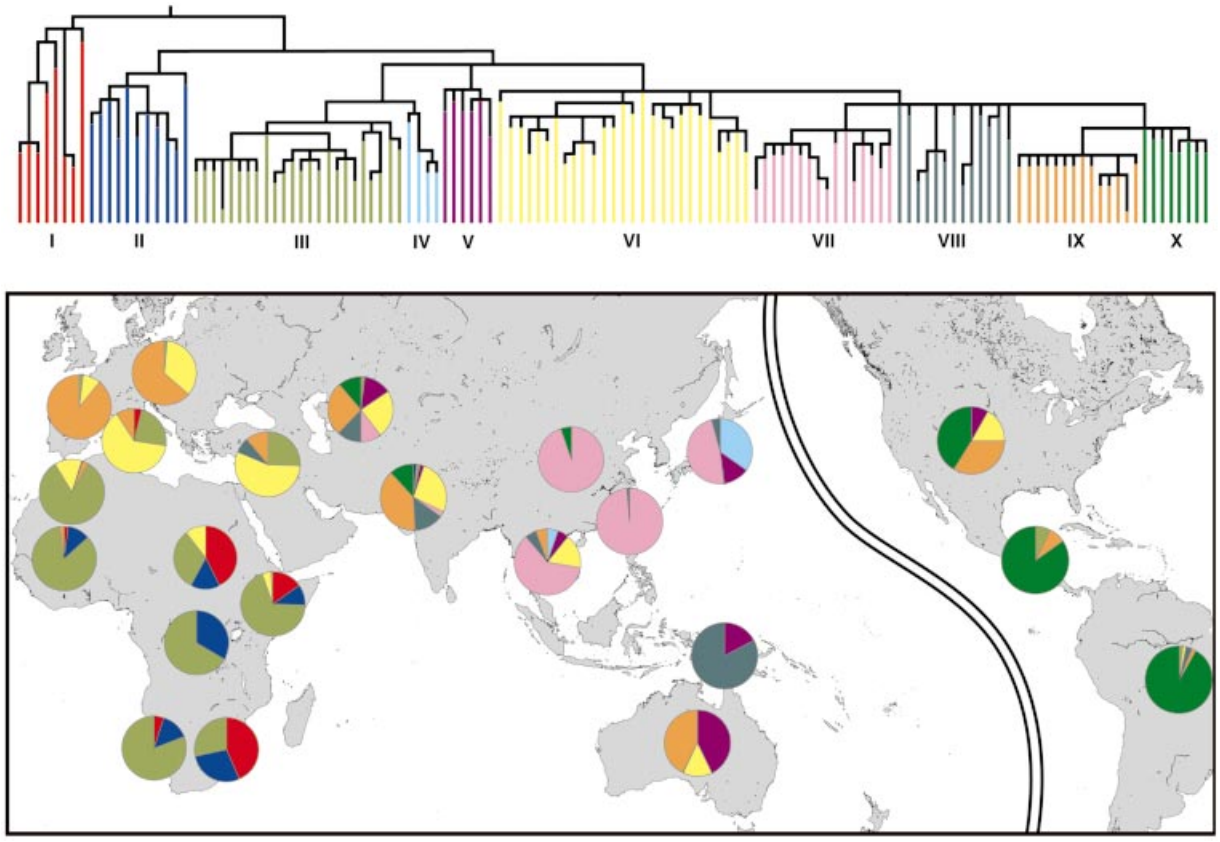

Fig. 2. Contemporary worldwide distribution of Y chromosome groups in 22 regions. Each group is represented by a distinguishing colour. Coloured sectors reflect representative group frequencies. Pacific basin not to scale. With respect to Table 1 of Underhill *et al.* (2000), Hunza and Pakistan + India are combined. In addition the results of Native Americans have been subdivided in North ($N = 14$), Central ($N = 13$) and South ($N = 79$).

past human movements. The strong geographical signal seen in the Y chromosome data is consistent with this assumption. The interpretative framework should be compared with alternatives, such as continuous gene flow, selection, or the effects of recent events. However, these alternatives have not been formally developed in ways that can be tested against the data, and are less consistent with other lines of evidence.

Groups I and II are restricted to Africa, and are distinct from all other African and non-African chromosomes on the basis of the M168 mutation. In an analogous context, the mtDNA haplogroups L_1 and L_2 are distinguished from other Africans and all non-Africans on the basis of the 3594 mutation (or 3592 *Hpa*I restriction site, Chen *et al.* 2000 and citations therein). Group I is distinguished by the absence of the derived alleles for M42/M94/M139 and presence of M91, while all non-African, as well as the majority of African males, sampled carry the derived alleles. Both Group I and II lineages are diverse and suggest a deeper genealogical heritage than other haplotypes. Representatives of these lineages are distributed across Africa but generally at low frequencies. Populations represented in Groups I and II include some Khoisan and Bantu speakers from South Africa, Pygmies from central Africa, and lineages in Sudan, Ethiopia and Mali. A single Sardinian was in Group I. All members of Group II share the M60 and M181 mutations that are distributed across Africa, with an idiosyncratic occurrence in Pakistan. M182 defines the major sub-clade, although an intermediate haplotype in Mali with the unique M146 mutation still persists. Although not mutually exclusive, some geo-

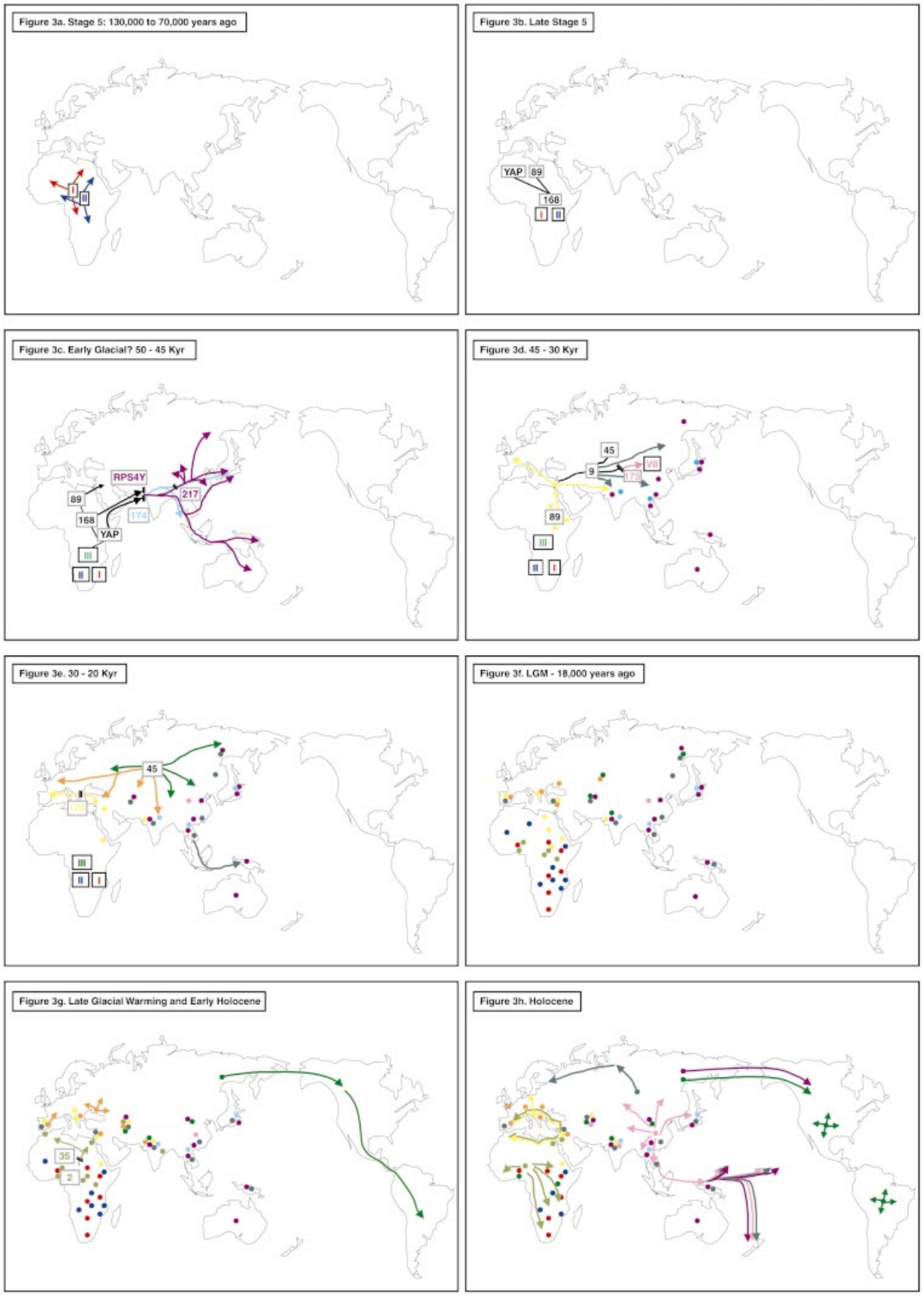

Fig. 3*a–h*. Hypothesized chronology of the global geographic distributions of Y chromosome mutations and groups during geological time periods relevant to the history of anatomically modern humans.

graphic substructure of derived Group II haplotypes is detected. Within the major M182 cluster, M112/M192 lineages reflect mostly central and southern African populations, whereas M150 associated lineages tend to be represented by populations in Sudan, Ethiopia or Mali.

In conclusion, the pattern of Group I and II distributions, their phylogenetic position and accumulated variation is suggestive of an early diversification and dispersal of human populations within Africa, and an early widespread distribution of human populations in that continent. Their patchy distribution, with high frequencies among isolated hunter-gatherer groups and in parts of Ethiopia and Sudan, may be interpreted as the survivorship of some of these ancient lineages through more recent population events. The palaeoanthropological record suggests that during the Stage 5 interglacial (130000–90000 years ago) (Fig. 3*a*), early human populations expanded throughout Africa, north and south of the Sahara, also reaching the Levant (Lahr & Foley, 1998). These population expansions are supported by faunal evidence, which shows the presence of not only modern humans, but east African species in the Middle East at this time (Tchernov, 1994). A last interglacial age for the first pan-African dispersal of humans is much earlier than the presently estimated age of 59000 years ago for the common ancestor of NRY variation. Three considerations should thus be made. First, that as mentioned above, the upper limit of the confidence interval (CI) of the age estimate (140000 years) embraces the period concerned. Second, that the history of human Y-chromosomes is characterized by a reduction of variation, more so than that of female lineages (Shen *et al*. 2000). Thus the initial phase of early human expansions within sub-Saharan Africa between 130000 and 70000 years ago may have witnessed several expansion events, with the extinction of earlier human NRY variation. Support for this is the apparent absence of intermediate haplotypes related to the M42/M94/M139 segment. Finally, the estimation of the age presented here assumes a model of population growth (Thomson *et al*. 2000), while early human population history is more likely characterized by sequential expansions and contractions. The effects of repeated founder effects on the age estimates of mutations may account for the apparent discrepancies. The current age estimates from the Y chromosomes rule out very ancient histories. However, the confidence limits of the molecular estimates are probably less resolved than the palaeoanthropological data, and so are used here to give broad relative frameworks (i.e. order of events) rather than precise bracketing in terms of an absolute chronology.

Out of Africa

The M168 mutation represents the signature of the recent successful modern human migrations across Africa and beyond, and it is at the root of Groups III–X. The geographical distribution of Groups III–X allows us to try to understand some of the major movements that occurred after the human beings left Africa.

The main points considered in that reconstruction are: the early formation of a non-African sub-cluster characterized by mutations RPS4Y/M216, present today among Australians, New Guineans, southeast Asians, Japanese and central Asians (Group V); the shared presence of the derived YAP/M145/M203 alleles in Africans, southeast Asians and Japanese (Groups III and IV); the distribution of a third sub-cluster, characterized by mutations M89/M213, across the entire world with the exception of most of sub-Saharan Africa (Groups VI–X).

All Y-chromosomes that are not exclusively African contain the M168 mutation, which may have originated within an East African population as a sub-group of Group II. M168 lineages evolved into three distinct sub-clusters: one which acquired an Alu insertion (YAP = DYS287) and the equivalent M145/M203 nucleotide substitutions, and two other lineages, defined by the distinct mutations RPS4Y/M216 and M89/M213. The destiny of these three sub-clusters represents a deep structuring of

Y-chromosome diversity outside Africa. However, before considering their history across the world, its contextual origin within Africa is important. We mentioned above the existence of palaeoanthropological evidence for an early pan-African/Levantine expansion of modern humans during the last interglacial period, which we associate with the expansion of Groups I and II lineages. The subsequent African history from the archaeological and fossil records is poorly known, but the palaeoclimatic record shows that the onset of glacial climates 70000 years ago was accompanied by fragmentation of African environments and isolation of both northwest and northeasternmost Africa from each other and the south. Lahr & Foley (1994, 1998) suggested that it had been during this period of fragmentation and isolation that African populations acquired variation that was then exported out of Africa independently through multiple dispersals of different African groups. In other words, that part of the diversity found outside Africa today was the magnification of a process of diversification within Africa between 90000–50000 years ago (Fig 3*b*). This reconstruction received support from the ''Weak Garden of Eden'' model, based on the pairwise mismatch distribution of mtDNA lineages showing the contraction of ancestral African diversity into separate groups, who would have then expanded independently within and beyond Africa at slightly different times (Sherry *et al.* 1994). The NRY data gives resolution to these models. The evolution of the M168 mutation into four separate clusters (YAP/M145/M203/M40/M96; ancestors of YAP/M145/M203/M174; M89/M213 and ancestors of RPS4Y/M216 lineages) is consistent with a process of population sub-division in Africa of a M168 population, prior to the main dispersal events into Eurasia 50000–40000 years ago. The age of the M168 mutation, representing the last common ancestor of all non-African human Y-chromosomes, has been estimated to be 40000 years (95% CI 31000–79000) (Thomson *et al.* 2000). As with the age of the common ancestor of NRY variation discussed above, this estimate is young in relation to the events recorded in the archaeological and fossil records. However, it embraces within its confidence interval the period in which the models above postulate an African population fragmented and differentiated into distinct sub-clusters that later dispersed out of the continent. The recent age of the M168 mutation is evidence that most modern extant Y-chromosomes trace their ancestry to African forefathers who left Africa relatively recently and eventually replaced completely archaic Y-chromosome lineages in Eurasia. Suggested departure routes for these dispersals include passages via the Horn of Africa to India and the Middle East Levantine corridor (Cavalli-Sforza *et al.* 1994; Lahr & Foley, 1994).

The evolution of NRY diversity within Africa

The YAP/M145/M203 Group III

Sub-Saharan African populations today are characterized by the presence of four NRY Groups (I, II, III, and VI), of which Group III is the most frequent. The lineage that acquired the YAP/M145/M203 polymorphisms in Africa divided into two sub-clusters. One is found today in Africa and the Mediterranean, defined by the M40/M96 mutations (Group III), where M40 = SRY4064 (Whitfield *et al.* 1995). The other sub-cluster is found in Asia and is defined by the M174 mutation (Group IV) (discussed in the next section).

Group III lineages are found at high frequencies in Africa, and relatively high frequencies in the Middle East and southern Europe (characterized by the M35/M215 mutations) with occasional occurrences in Central Asia, Pakistan and America (Hammer *et al.* 1997, Hammer & Horai, 1995, Qamar *et al.* 1999; Underhill *et al.* 2000). Considerable haplotype differentiation within Group III is observed. Most notably, the PN2 transition (Hammer *et al.* 1997) unites two high frequency sub-clades, defined by M2/PN1/M180 mutations in sub-Saharan Africa, and M35/M215 in north and east Africa, the Mediterranean basin and Europe. The widespread distribution of these two sub-clades,

which together account for 80% of Group III lineages, is considered to be the result of recent events. The M2 transition (Seielstad *et al.* 1994) and its analogues, PN1 (Hammer *et al.* 1997) and M180, are linked to the RFLP 49f Ht4, found in high frequency throughout Africa. The wide distribution of this sub-clade in sub-Saharan Africa probably reflects the Bantu agricultural expansion in the last three thousand years (Passarino *et al.* 1998; Scozzari *et al.* 1999). The expansion of Bantu farmers would have been largely accompanied by the replacement of other Y-chromosomes. The extent of founder effects associated with the recent expansion of Group III lineages is illustrated by the M191 mutation, which occurs in $\sim 40\%$ of the M2/M180/PN1 clade members. Furthermore, the low frequency of lineages within Groups I and II and of the 20% minority of the haplotypes within Group III that lack the PN2 mutation, distinguished by either the M33 or M75 mutations and confined to Africa, is evidence of the impact of the Bantu expansion which overwhelmed the pre-existing African NRY chromosome diversity. This is not revealed by the pattern of mtDNA diversity, which indicates a persistence of mtDNA haplotypes, suggesting a larger effective population size of African women versus African men. However, it finds reflection in the sub-Saharan African fossil record, which shows greater early Holocene/late Pleistocene morphological diversity than at present (Lahr, unpublished results).

The M35/M215 sub-clade cluster of haplotypes fragments a lineage (Ht 4) described previously (Hammer *et al.* 1997). We suggest that a population with this sub-clade of the African YAP/M145/M203/PN2 cluster expanded into the southern and eastern Mediterranean at the end of the Pleistocene (Fig. 3*h*). These lineages would have been introduced then from the Middle East into southern Europe (and to a lesser extent northern India and Pakistan) by farmers during the Neolithic expansion.

In East Africa some of the least differentiated YAP/M145/M203/M40/M96 lineages are observed (Ht 28, 35, 39). The M89/M213 mutations are at the root of Groups VI–X. However, Group VI as presently known, remains polyphyletic, differentiated from the others (Groups VII–X) by the absence of the M9 mutation. We suggest that the M89/M213 mutations shared by East Africans and most Eurasian and Amerindian world populations occurred in a northeast African population carrying the M168 lineage. Similarly, East Africa has been indicated as the source of mtDNA haplogroup M (Quintana-Murci *et al.* 1999) and superhaplogroup U (the oldest lineage with the 16223 transition) (Kivisild *et al.* 1999). Thus, part of this early population carrying the M89/M213 mutations is still represented within northeast African NRY diversity, while most of its descendants are found outside of Africa, having dispersed via the Levantine corridor to Eurasia 45–30 K years ago (Fig. 3*d*). This Eurasian M89/M213 ancestral population would have later diversified into further sub-clusters of Group VI, as well as Groups VII, VIII and a lineage carrying the derived M45/M74 alleles, which originated Groups IX and X (discussed below).

The colonization of Australo-Melanesia and formation of early Asian outliers

The RPS4Y/M216 Group V

RPS4Y/M216 lineages are found in Asia, Australo-Melanesia and North America (Bergen *et al.* 1999; Karafet *et al.* 1999). We suggest that a M168 African population dispersed from the Horn of Africa via a coastal or interior route (50–45 K years ago; Walter *et al.* 2000) towards southern Asia, where the RPS4Y/M216 mutations probably originated (Fig. 3*c*). Descendants of this dispersal reached southeast Asia and were also the first to colonize the Sahul landmass – New Guinea and Australia. The latter are characterized by the lineages M38/M208 (ht48) and M210 (ht49). One RSP4Y/M216 lineage acquired the M217 mutation, which spread through central and eastern Asia, also reaching Japan (where individuals with RPS4Y/M216/M8/M105/M131 are much less frequent than RPS4Y/M216/M217 lineages), and later North America. Today, with the exception of

Australia, and to a lesser extent New Guinea, the RPS4Y/M216 lineages have a relic distribution. Significant similarities between Group V lineages in Australia and New Guinea and chromosome 21 MX1 haplotype 2 have been observed (Jin *et al*. 1999).

The YAP/M145/M203/M174 Group IV

Group IV lineages (YAP/M145/M203/M174) are exclusively Asian. Interestingly all members carry the ancestral alleles for M40 and M96. On the basis of YAP and M40 and nested cladistic analysis, this pattern has been interpreted as evidence for an Asian origin of the YAP insertion mutations and a subsequent back migration to Africa, followed by a major expansion (Altheide & Hammer, 1997; Hammer *et al*. 1998). This interpretation requires reassessment. The presence of M174 in the phylogeny underscores the difficulty of basing directionality on the absence of a character state. The M174 data taken alone would support an African origin of the YAP/M145/M203 polymorphisms, as the M174 ancestral allele is found exclusively in Africa. The relatively shallow time-depth of the Y phylogeny (Thomson *et al*. 2000) suggests that the extinction rate of Y haplotypes is high, which might account for the apparent absence of any YAP/M145/M203 chromosomes in Africa that carry the ancestral allele SRY4064 (= M40). Thus the current apparent lack of less derived precursors haplotypes within both Africa and Asia precludes the localization of the geographical origin of YAP. Our results from nested cladistic analysis (unpublished) indicate a range expansion, but not the migration direction required to localize the origin of YAP. Additional clarification concerning the geographical origin of the YAP insertion may come from the possible eventual discovery of intermediate haplotypes. There is no other evidence, either from the other NRY groups, mtDNA (Quintana-Murci *et al*. 1999) or palaeoanthropology (Lahr & Foley, 1998), of a Paleolithic migration back to Africa. We suggest that a population carrying only the YAP/M145/M203 polymorphism dispersed from Africa through the Horn towards southern Asia (Fig. 3*c*), similar to the Asian dispersal of RPS4Y/M216 lineages discussed above (also probably at a similar time). YAP/M145/M203/M174 lineages are today mostly confined to Japan and Tibet, where they occur at high frequencies, with fewer found scattered throughout southeast Asia (Su *et al*. 1999). The chromosome 21 MX1 haplotype 6, which correlates with Group III lineages, also occurs in East Asia, including Japan (Jin *et al*. 1999). This parallels the shared common ancestry of Y chromosome related lineages between Africa (Group III) and East Asia (Group IV). Similar to Group V lineages, Group IV haplotypes have a relic distribution in Asia.

In conclusion, we suggest that Group IV and V lineages in Asia represent the descendants of two early dispersal events of African populations from the Horn of Africa to southern Asia, (Fig. 3*c*), a dispersal event that would have also taken the mtDNA M haplogroup to India. Alternatively, the distribution of Group IV and V lineages in Asia could reflect a single dispersal event of a population carrying both Group IV and V lineages, the former in low frequencies, facilitating its subsequent extinction in most descendant populations except Japan and Tibet, where drift would have increased its frequency. This southern route of dispersal from East Africa to India and beyond does not seem to have been used very frequently, either by modern humans or archaic hominids (Lahr & Foley, in press), but rather in particular windows of time when climate and geography allowed, possibly related to the combination of low sea-levels and light monsoonal regimes. The early human groups that used this route around 50000 years ago (taking the earliest occupation of Australia as the endpoint of this dispersal) were not restricted to coastal areas, and must have successfully colonized the Asian mainland, as shown by the distribution of surviving Group IV and V lineages. However, they would have been largely replaced by more recent population events associated with the subsequent expansion of Group VIII throughout all of Asia, Group VII lineages in the east Asia, and Group IX lineages in west Eurasia.

The colonization of Eurasia from the Levantine corridor

The M89/M213 Group VI and M9 related Groups VII–X

The third large sub-cluster of M168 lineages is characterized by the M89/M213 mutations at the root of Groups VI–X. As discussed above, this sub-cluster is suggested to have evolved in East Africa, from where it dispersed to Eurasia through the Levantine corridor around 45000 years ago. This dispersal would have also involved several mtDNA haplogroups characterized by the 10873 C→T mutation (Quintana-Murci *et al.* 1999). The palaeoanthropological record clearly shows the expansion of modern humans from the Levant at this time, an expansion that is strongly characterized morphologically and archaeologically by the first Upper Palaeolithic cultures, as well as evolutionarily by the extinction of the Middle Eastern and European Neanderthals. We would suggest that at the root of this major Eurasian expansion, from which all main non-African populations (with the exception of Australia) derive, lies in a population of northeast African origin carrying M89/M213 lineages.

It should be noted that, although having a very strong technological signature, the history of the Upper Palaeolithic is a complex one. Most archaeologists associate the Upper Palaeolithic with the well-known sequence of cultures in Europe beginning with the Aurignacian. However, earlier Upper Palaeolithic assemblages characterize the first Levantine sites (Klein 1999), as well as possibly the first modern human sites in Europe (Kuhn & Stiner, 1998), while in Siberia and south and eastern Asia, early Upper Palaeolithic cultures also lack the European traits identified with the Aurignacian and Gravettian industries (Klein, 1992; Foley & Lahr, 1998). This early widespread distribution of variable Upper Palaeolithic cultures throughout Eurasia suggests a very fast expansion from which later populations derive, as well as highlighting the potential extinction of some populations within this early structure, as later expansions occurred over the range of earlier ones. This complexity helps us to understand the differentiation of M89/M213 lineages within Eurasia.

We suggest that a population carrying the M89/M213 mutations dispersed from Africa to the Middle East, from where it originally expanded west, north and east around 40000 years ago (Fig. 3*d*). Only ~3% of Group VI related lineages do not have any known mutations (ht 69) which may represent the earliest spread of this lineage from the Middle East. The ht 69 lineage is observed in the Middle East, Pakistan and C. Asia. Current data indicate that its presence in N.E. Africa is uncertain. A probable western expansion of M89/M213 Levantine populations would have taken Group VI ht 69 lineages to Europe as the earliest Upper Palaeolithic occupation of the area. However its appearance in Europe is very low (0.2%), indicating that few of these lineages have survived to the present, possibly having been replaced in Europe by related M170 lineages. The very early (40–50000 years) mtDNA U5 lineages survive in about 7% of European women (Richards *et al.* 1998, 2000). The eastwards expansion of Levantine M89/M213 populations would have taken Group VI lineages characterized by the M52/M69 mutations to India, an expansion that may correspond to the spread of mtDNA haplogroup U2 and U7 (Kivisild *et al.* 1999). An expansion towards Central Asia or the Caucasus would have given rise to a M89/M213 population that acquired the M9 mutation. This mutation lies at the root of Groups VII–X, all Eurasian and Amerindian in distribution. The population carrying the M9 mutation must have expanded widely (Fig. 3*d*), differentiating into several lineages carrying independent mutations (the polyphyletic Group VIII), and two discrete populations one in north Asia characterized by the M45/M74 mutations (at the root of Groups IX and X) and one in northern China characterized by the M175/M214 mutations (Group VII). In summary, the early diversification of a M89/M213 population in Eurasia between 40 and 30000 years ago would

have given rise to at least six Y-chromosome populations:

(i) an early European Upper Palaeolithic group whose Y-chromosomes later became nearly extinct (Group VI), but whose existence is evidenced archaeologically and in mtDNA diversity;
(ii) a Levantine population carrying undifferentiated M89/M213 lineages (Group VI), which later acquires the M170 mutation;
(iii) a northern Indian/Pakistan population carrying the M52/M69 mutations on a M89/M213 lineage (Group VI);
(iv) a widespread Asian population, characterized by several polyphyletic lineages carrying the derived M9 allele of the M89/M213 lineages (Group VIII);
(v) a discrete north Asian population, characterized by the derived M9 and M45/M74 alleles of the M89/M213 lineages (ancestor of Groups IX and X);
(vi) a discrete northeast Asian population, characterized by the derived M9, M175/M214 alleles of the M89/M213 lineages (ancestor of Group VII).

Tying these populations more specifically into the known archaeological record remains problematic for a number of reasons, the primary one being that the very early upper Palaeolithic is poorly differentiated, and in particular is not nearly as well known in Asia as it is in Europe.

From this ancestral Eurasian modern human genetic structure, later populations expanded. The geographical expansion of M89/M213/M9 lineages lacking the derived M45/M74 and M175/M214 alleles (Group VIII) should not be described as a single event, but rather as the independent formation and fragmentation of populations carrying this group of lineages throughout most of Asia, displacing the earlier Group IV and V lineages towards outlying positions (typically highlands and islands). Among these lineages is the TAT transition (Zerjal *et al.* 1997), which marks the Uralic migrations into northern Europe. Another lineage, characterized by M4/M5/M106/M189 (hts99–102), eventually reached New Guinea (but apparentely not Australia, see Table 1 in Underhill *et al*, 2000, and Forster *et al.* 1998), where it came to predominate over the earlier Group V Y-chromosomes.

Five major demographic events characterize the subsequent Y chromosome genetic history of Eurasian populations. Chronologically, the first of these relates to the expansion and differentiation of the northern Asian M89/M213/M9/M45/M74 lineages. Part of this population, characterized by the M173/M207 mutations (Group IX), expands westwards around 30000 years ago (Fig. 3*e*), reaching Europe, the Caucasus, the Middle East, Central Asia and northern India-Pakistan. This population expansion around 30000 years ago gives rise to the Upper Palaeolithic Aurignacian, or Gravettian, or both (Semino *et al.* 2000). Another group within this M89/M213/M9/M45/M74 population differentiates, acquiring a number of mutations in separate lineages (M120, M124, M25/M143, M3) that define Group X. These lineages also spread widely across the Asian northern steppes, as well as in Central Asia and northern India-Pakistan (Fig. 3*e*). Group IX, and to a lesser extent Group X, lineages become the most frequent ones in western Asia and India-Pakistan. The Group X lineage characterized by the M3 mutation later spreads from Siberia to the Americas. A geographic pattern similar to M45/M74 has been found in mtDNA haplogroup X, although this has a very low frequency in both Europe and the Americas (Brown *et al.* 1998).

The NRY data also disclose an expansion from the Levant into the Mediterranean basin around 25–20000 years ago (Fig. 3*e*). This expansion would have been of a population carrying Group VI lineages from which M170 arose *in situ* in Europe, a descendant of the early M89/M213 population in the area. This event, which apparently also brought to Europe the mtDNA haplogroup H (Torroni *et al.* 1998), does not have an archaeological signature. The distribution of the chromosome 21 MX1 haplotype 8 in Europe

and W. Asia, but not E. Asia (Jin *et al.* 1999), appears to coincide with Group VI lineages, especially if we assume its distribution in Oceania and America is a consequence of recent gene flow. Levantine populations at the time manufactured Upper Palaeolithic tools similar to those in Europe (probably introduced by the expansion of earlier Group VI M52/M69 (now at low percentages) and Group IX populations in the area), and neither their tools, nor their particular Group IX lineages, would be distinguishable from those in Europe at the time.

These expansion events were followed by a period of significant population contraction associated with the LGM 18–16000 years ago (Fig. 3*f*). These contractions are well represented in the archaeological record (Sofer & Gamble, 1990), and in Europe led to the formation of discrete refugia. Equally, the European archaeological record shows the extent of the subsequent demographic expansions from these discrete refugia as conditions ameliorated (Housely *et al.* 1999). We suggest that these Mesolithic expansion events are reflected in the relative frequencies of the two main sub-groups of Group IX lineages in Europe today (Semino *et al.* 2000).

The expansion of Neolithic farmers from the Middle East into Europe is also represented in the NRY data, although suggesting a relatively localized area of impact. As mentioned before in relation to African NRY history, a Mesolithic population carrying Group III lineages with the M35/M215 mutation expanded northwards from sub-Saharan to north Africa and the Levant (Fig. 3*g*). The Levantine population of farmers that dispersed into Europe during and after the Neolithic carried these African Group III M35/M215 lineages, together with a cluster of Group VI lineages characterized by M172 and M201 mutations (Fig. 3*h*). By integrating p49f RFLP polymorphism data, it was shown that the p12f 8 kb, M172 and M201 lineages share a common ancestor (Semino *et al.* 2000). The last important demographic event of Eurasian genetic history disclosed by NRY data relates to eastern Asian populations. One of the main derivatives of the M9 Asian cluster is Group VII, defined by the M175 and M214 mutations. This group has achieved very high frequency in East Asia, with occurrences in Central Asia and Polynesia. The M122 mutation within this cluster defines a dominant set of lineages (hts 81–89) with widespread distribution in East Asia (Su *et al.* 1999). Haplotypes falling into this group have not been found in either Europe or those areas colonized by Asians in the Pleistocene – Australia, New Guinea and the Americas. Group VII lineages, which seem to have originated in Northern China (Su *et al.* 2000a), may reflect the impact of millet and rice agriculture on East Asian Mongoloid demographic history (Cavalli-Sforza *et al.* 1994), displacing to a great extent all other NRY variants with a clinal frequency from the expected Chinese area of origin (Fig. 3*h*). Mimicking earlier events, the relic distribution of Group VIII lineages in eastern and southeast Asia, found today together with Groups IV and V lineages in population outliers, would be the result of the recent expansion of Group VII lineages. This major demographic impact of Asian agriculturists is also reflected in the fossil record, which shows the expansion of distinctive Mongoloid cranial features associated with the spread of farmers from 7000 years ago (Brown, 1999). The history of Group VII lineages in eastern Asia seems analogous to that of the M2/M180/PN1 sub-clade (hts 20–27) of African Group III, which has been interpreted as marking the Bantu agricultural expansion (Passarino *et al.* 1998). It is interesting to note the greater number of high frequency subclades in East Asia, such as those haplotypes associated with M95, M119, M134, SRY465 (= M176) compared to Africa. This is suggestive of the maintenance of greater genetic diversity, possibly due to greater population size, of the groups involved in the East Asian agricultural expansion with respect to the case of the Bantus.

The colonization of the Americas

Amerindian populations today show a variety of Y-chromosome groups, but many of these, such as Group III, Group VI, Group VIII and

Group IX lineages, are likely to be the result of recent gene flow since colonial times. This recent gene flow is more apparent in North American aboriginal groups. Most Amerindians have GroupX/M3 lineages, the precursor of which has been identified in Siberia (Santos *et al.* 1999). North American groups also show small frequencies of Group V lineages associated with the NaDene speaking populations (Karafet *et al.* 1999, Ruhlen, 1998). It remains possible that some Group VIII lineages may have been part of a polymorphic northeast Asian/Siberian population that dispersed to the Americas, or may reflect the survivorship of a few relic lineages of independent dispersal events in the late Pleistocene. Archaeological data in the Americas has proven inconclusive in relation to this question. Most Paleo-Indian American sites reflect the expansion of the Clovis culture into the Americas 11000 years ago, although a few sites in South America, dated to between 13000–11000 years ago, show different archaeological traditions (Fiedel, 2000). The fossil data is equally complex, as Paleo-Indian remains show greater morphological diversity than is observed among recent groups (Steele & Powell, 1993, Neves & Pucciarelli, 1991).

The colonization of Polynesia

The Polynesian islands were the last part of the world colonized by humans in the last 3000 years. The archaeological record suggests that the ancestors of Polynesians originated in Taiwan and southeast Asia, dispersed south reaching New Guinea, and then further across the Pacific (Bellwood, 1989). Excluding lineages with signatures of recent European gene flow, Polynesian islanders have NRY lineages from Groups V–VIII (Karafet *et al.* 1999; Su *et al.* 2000*b*). Their Group V and VIII lineages derive from New Guinean populations, supporting the view that the East Asian ancestral group (carrying Group V lineages) assimilated genetic diversity in New Guinea (Kayser *et al.* 2000; Underhill *et al.* 2001). This is consistent with mtDNA Polynesian data, which shows the existence in low frequency of Melanesian lineages (Lum *et al.* 1994).

Final remarks

In conclusion, we have shown that the distribution of Y chromosome diversity is able to describe some of the most important human prehistoric movements, and delineate the composition of modern populations. In most cases, the portrait of NRY phylogeography is bolstered by archaeological and other genetic data, especially from the mtDNA locus. However, the narrative given by the two uniparentally inherited genetic systems are sometimes different, due to the different demographic forces modulated by gender. The comparison of both systems will provide better insights into the prehistory of our species.

Supported in part by NIH grants GM28428 and GM55273 (L. L. C-S). We thank all the men who donated DNA. We thank R. G. Klein, M. Ruhlen and G. L. Cavalleri for helpful discussions. We thank C. Edmonds, K. Prince and D. Soergel for assistance with graphic illustrations. We thank E. Bosch and J. Bertanpetit for a more representative summary of frequency data for Moroccan samples.

References

Altheide, T. K. & Hammer, M. F. (1997). Evidence for a possible Asian origin of YAP+Y chromomsomes. *Am. J. Hum. Genet.* **61**, 462–466.

Avise, J., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. (1987). Intraspecific phylogeography: the molecular bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* **18**, 489–522.

Bellwood, P. (1989). The colonization of the Pacific: some current hypotheses. In: *The colonization of the Pacific: a genetic trail* (eds. A. V. S. Hill & S. W. Serjeantson), pp. 1–59. Oxford: Oxford University Press.

Bergen, A. W., Wang, C.-Y., Tsai, J., Jefferson, K., Dey, C., Smith, K. D., Park, S.-C., Tsai, S.-J. & Goldman, D. (1999). An Asian-native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet.* **63**, 63–80.

Bianchi, N. O., Bailliet, G., Bravi, C. M., Carnese, R. F., Rothhammer, F., Martínez-Marignac, V. L. & Pena, S. D. J. (1997). Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. *Am. J. Phys. Anthropol.* **102**, 79–89.

Brown, P. (1999). The first East Asians? In: *Interdisciplinary perspectives on the origin of the Japanese* (ed. K. Omoto), pp. 105–124. Kyoto: International Centre for Japanese Studies.

Brown, M. D., Hosseini, S. H., Torroni, A., Bandelt, H.-J., Allen, J. C., Schurr, T. G., Scozzari, R., Cruciani, F. & Wallace, D. C. (1998). mtDNA haplogroup X: an ancient link between Europe/western Asia and North America? *Am. J. Hum. Genet.* **63**, 1852–1861.

Cavalli-Sforza, L. L., Menozzi P. & Piazza, A. (1994). Chapter 2. In: *The history and geography of human genes*. Princeton, New Jersey: Princeton University Press.

Chen, Y.-S., Olckers, A., Schurr, T. G., Kogelnik, A. M., Huoponen, K. & Wallace, D. C. (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.

Fiedel, S. J. (2000). The peopling of the New World: present evidence, new theories, and future directions. *J. Archaeol. Res.* **8**, 39–103.

Foley, R. (1998). The context of human genetic evolution. *Genome Res.* **8**, 339–347.

Foley, R. & Lahr, M. M. (1998). Mode 3 technologies and the evolution of modern humans. *Camb. Archaeol. J.* **7**, 3–36.

Forster, P., Kayser, M., Meyer, E., Roewer, L., Pfeiffer, H., Benkmann, H., Brinkmann, B. (1998). Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. *Mol. Biol. Evol.* **15**, 1108–1114.

Hammer, M. F. & Horai, S. (1995). Y chromosomal DNA variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**, 951–962.

Hammer, M. F., Karafet, T., Rasanayagam, A., Wood, E. T., Altheide, T. K., Jenkins, T., Griffiths, R. C., Templeton, A. R. & Zegura, S. L. (1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**, 427–441.

Hammer, M. F., Spurdle, A. B., Karafet, T., Bonner, M. R., Wood, E. T., Novelletto, A., Malaspina, P., Mitchel, R. J., Horai, S., Jenkins, T. & Zegura, S. L. (1997). The geographic distribution of human Y chromosomes. *Genetics* **145**, 787–805.

Housley, R. A., Gamble, C. S., Street, M. & Pettitt, P. (1997). Radiocarbon evidence for the lateglacial human recolonisation of Northern Europe. *Proceedings of the Prehistoric Society* **63**, 25–54.

Jin, L., Underhill, P. A., Doctor, V., Davis, R. W., Shen, P., Cavalli-Sforza, L. L. & Oefner, P. J. (1999). Distribution of haplotypes from a chromosome 21 region distinguishes multiple prehistoric human migrations. *Proc. Natl. Acad. Sci. USA* **96**, 3796–3800.

Jorde, L. B., Bamshad & M. Rogers, A. R. (1998). Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *BioEssays* **20**, 126–136.

Kayser, M., Brauer, S., Weiss, G., Underhill, P. A., Roewer, L., Schiefenhovel, W. & Stoneking, M. (2000). Melanesian origin of Polynesian Y chromosomes. *Current Biology* **10**, 1237–1246.

Karafet, T. M., Zegura, S. L. Posukh, O., Osipova, L., Bergen, A., Long J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., Wiebe, V., Griffiths, R. C., Templeton, A. R. & Hammer, M. F. (1999). Ancestral Asians source(s) of new world Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64**, 817–831.

Kivisild, T., Bamshad, M. J., Kaldma, K., Metspalu, M., Metspalu, E., Reidla, M., Laos, S., Parik, J., Watkins, W. S., Dixon, M. E., Papiha, S. S., Mastana, S. S., Mir, M. R., Ferak, V. & Villems, R. (1999). Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Current Biol.* **9**, 1331–1334.

Klein, G. K. (1992). The archeology of modern human origins. *Evolutionary Anthropology* **1**, 5–14.

Klein, G. K. (1999). *The human career: human biological and cultural origins*. Chicago, IL: University of Chicago Press.

Kuhn, S. L. & Stiner, M. C. (1998). The earliest Aurignacian of Riparo Mochi (Liguria, Italy). *Current Anthropology* **39**, S175–S189.

Kwok, C., Tyler-Smith, C., Mendonca, B. B., Hughes, I., Gerkovitz, G. D., Goodfellow, P. N. & Hawkins, J. R. (1996). Mutation analysis of the 2kb 5í to SRY in XY females and XY intersex subjects. *J. Med. Genet.* **33**, 465–468.

Lahr, M. M. & Foley, R. (1994). Multiple dispersals and modern human origins. *Evol. Anthropol.* **3**, 48–60.

Lahr, M. M. & Foley, R. (1998). Towards a theory of modern human origins: geography, demography and diversity in recent human evolution. *Yrbk of Phys. Anthropol.* **41**, 137–176.

Lum, K., Rickard, O., Ching, C. & Cann, R. L. (1994). Polynesian mitochondrial DNAs reveal three deep maternal lineage clusters. *Hum. Biol.* **66**, 567–590.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonnè-Tamir, B., Sykes, B. & Torroni, A. (1999). The emerging tree of west eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* **64**, 232–249.

Neves, W. A. & Pucciarelli, H. M. (1991). Morphological affinities of the first Americans: an exploratory analysis based on early South American human remains. *J. Human Evolution* **21**, 261–273.

Owens, K. & King, M. C. (1999). Genomics views of human history. *Science* **286**, 451–453.

Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M. & Santachiara-Benerecetti, A. S. (1998). Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. *Am. J. Hum. Genet.* **62**, 420–434.

Qamar R., Ayub Q., Haliq S., Mansoor A., Karafet T., Mehdi S. Q. & Hammer M. F. (1999). African and Levantine origins of Pakistani YAP+ Y chromosomes. *Hum. Biol.* **71**, 745–755.

Quintana-Murci, L., Semino, O., Bandelt, H.-J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. (1999). Genetic evidence of an early exit of Homo Sapiens sapiens from Africa through eastern Africa. *Nat. Genet.* **23**, 437–441.

Richards, M. B., Macaulay, V. A., Bandelt, H.-J. & Sykes, B. C. (1998). Phylogeography of mitochondrial DNA in western Europe. *Ann. Hum. Genet.* **62**, 241–260.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A. & Bandelt,

H.-J. (2000). Tracing European Founder Lineages in the Near Eastern mtDNA Pool. *Am. J. Hum. Genet.* **67**, 1251–1276.

Ruhlen, M. The origin of the Na-Dene. (1998). *Proc. Natl. Acad. Sci. USA* **95**, 13994–13996.

Santos, F. R., Pandya, A., Tyler-Smith, C., Pena, S. D., Schanfield, M., Leonard, W. R., Osipova, L., Crawford, M. H. & Mitchell, R. J. (1999). The central Siberian origin for native American Y chromosomes. *Am. J. Hum. Genet.* **64**, 619–628.

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, D. C. & 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.

Seielstad, M. T., Hebert, J. M., Lin, A. A., Underhill, P. A., Ibrahim, M. D., Vollrath, D., & Cavalli-Sforza L. L. (1994). Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.* **3**, 2159–2161.

Semino, O., Passarino, G., Oefner, P., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiæ, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A. S., Cavalli-Sforza L. L. & Underhill, P. A. (2000). The genetic legacy of Palaeolithic Homo sapiens sapiens in extant Europeans: a Y-chromosome perspective. *Science* **290**, 1155–1159.

Shen, P., Wang, F., Underhill, P. A., Franco, C., Yang, W.-H., Roxas, A., Sun, R., Lin, A. A., Hyman, R. W., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner P. J. (2000). Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* **97**, 7354–7359.

Sherry, S. T., Rogers, A. L., Harpending, H., Soodyall, H., Jenkins, T. & Stoneking, M. (1994). Mismatch distributions of mtDNA reveal recent human population expansions. *Hum. Biol.* **66**, 761–775.

Shinka, T., Tomita, K, Toda, T., Kotliarova, S. V., Lee, J., Kuroki, Y., Jin, D. K., Tokunaga, K., Nakamura, H. & Nakahori, Y. (1999). Genetic variations on the Y chromosome in the Japanese population and implications for modern human Y lineage. *J. Hum. Genet.* **44**, 240–245.

Soffer, O. & Gamble, C. (eds.) (1990). *The world at 18000 BP*, Vol. 1: *High latitudes*. London: Unwin Hyman.

Steele, D. G. & Powell, J. F. (1993). Paleobiology of the first Americans. *Evol. Anthropol.* **2**, 138–146.

Su, B., Xiao, J., Underhill, P., Deka, R., Zhang, W., Akey, J., Huang, W., Shen, D., Lu, D., Luo, J., Chu, J., Tan, J., Shen, P., Davis, R., Cavalli-Sforza, L., Chakraborty, R., Xiong, M., Du, R., Oefner, P., Chen, Z. & Jin, L. (1999). Y-chromosome evidence for a northward migration of modern humans into eastern Asia during the last ice age. *Am. J. Hum. Genet.* **65**, 1718–1724.

Su, B., Jin, L., Underhill, P. A., Martinson, J., Saha, N., McGarvey, S. T., Shriver, M. D., Chu, J., Oefner, P., Chakraborty, R. & Deka, R. (2000*a*). Polynesian origins: insights from the Y chromosome. *Proc. Natl. Acad. Sci. USA* **97**, 8225–8228.

Su, B., Xiao, C., Deka, R., Seielstad, M. T., Kangwanpong, D., Xiao, J., Lu, D., Underhill, P., Cavalli-Sforza, L. L., Chakraborty, R. & Jin, L. (2000*b*). Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum. Genet.* **107**, 582–590.

Sun, C., Skaletsky, H., Rozen, S., Gromoll, J., Nieschlag, E., Oates, R. & Page, D. C. (2000). Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum. Mol. Biol.* **9**, 2291–2296.

Tchernov, E. (1994). New comments on the biostratigraphy of the Middle and Upper Pleistocene of the Levant. In: *Late Quaternary chronology and paleoclimates of the Eastern Mediterranean* (eds O. Bar-Yosef & R. S. Kra), pp. 333–350. Tuscon: Radiocarbon.

Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. & Feldman M. W. (2000). Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **97**, 7360–7365.

Torroni, A., Bandelt, H. J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M. L., Bonnè-Tamir, B. & Scozzari, R. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am. J. Hum. Genet.* **62**, 1137–1152.

Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., Kauffman, E., Bonnè-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J. R., Mehdi, S. Q., Seielstad, M. T., Wells, R. S., Piazza, A., Davis, Feldman, M. W., Cavalli-Sforza, L. L. & Oefner, P. J. (2000). Y Chromosome Sequence Variation and the History of Human Populations. *Nat. Genet.* **26**, 358–361.

Underhill, P. A., Passarino, G., Lin, A. A., Marzuki, S., Cavalli-Sforza, L. L. & Chambers, G. (2001). Maori origins, Y chromosome haplotypes and implications for history in the Pacific. *Human Mutation* (in press).

von Haeseler, A., Sajantila, A. & Pääbo, S. (1996). The genetical archaeology of the human genome. *Nat. Genet.* **14**, 135.

Walter, R. C., Buffler, R. T., Bruggemann, J. H., Guillaume, M. M., Berhe, S. M., Negassi, B., Libsekal, Y., Cheng, H., von Cosel, R., Neraudeau, D. & Gagnon, M. (2000). Early human occupation of the Red Sea coast of Eritrea during the last interglacial. *Nature* **405**, 65–69.

Whitfield, L. S., Sulston, J. E. & Goodfellow, P. N. (1995). Sequence variation on the human Y chromosome. *Nature* **378**, 379–380.

Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Rower, L. & Santos, F. R. (1997). Genetic relationships of Asians and Northern Europeans revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60**, 1174–1183.

APPENDIX I. Description of 218 NRY markers.

Marker No.	Nucleotide Change	Position[1] (bp)	Primers (5′-3′) Forward	Reverse	Reference[4]
M1	→+Alu	3	caggggaagataaagaaata	actgctaaaaggggatggat	1
M2	A→G	168	aggcactggtcagaatgaag	aatggaaaatacagctcccc	2
M3	C→T	181	taatcagtctcctcccagca	aaaattgtgaatctgaaatttaagg	3
M4	A→G	88	tcctaggttatgattacagagcg	acgtcttgtaaacatgacaagg	3
M5	C→T	73	gggtttatactgacctgccaatgtt	ttattgggaactttcagggg	3
M6	T→C	37	cactaccacatttctggttgg	cgctgagtccattctttgag	3
M7	C→G	236	actgtgagcgagctgaaaat	gcagccttgtgaaccaatta	3
M8	G→T	137	cccacccacttcagtatgaa	aggctgacagacaagtccac	3
M9	C→G	68	gcagcatataaaactttcagg	aaaacctaactttgctcaagc	3
M10	T→C	156	gcattgctataagttacctgc	taataaaaattgggtcaccc	3
M11	A→G	44	tctctctgtctgtctctccctcc	gagcataaacaagaacttactgagc	3
M12	G→T	286	actaaaacaccattagaaacaaagg	ctgagcaacatagtgacccc	3
M13	G→C	157	tcctaacctggtggtctttc	gttggataaaatcatggctca	3
M14	T→C	180	agacggttagatcagttctctg	tagataaaagcacattgacacc	3
M15	GG→TT, +insGTACAGAGA	108	acaaatcctgaacaatcgc	tgcatgtggttaaaatttcc	3
M16	C→A	38	tgttatgtcatttgaacccag	ccgtgtgttgctgggctgt	3
M17	delG	68	ctggtcataacactggaaatc	tgaacctacaaatgtgaaactc	3
M18	insAA	63	ctggtcataacactggaaatc	tgaacctacaaatgtgaaactc	3
M19	T→A	131	ctggtcataacactggaaatc	tgaacctacaaatgtgaaactc	3
M20	A→G	118	gattgggtgtcttcagtgct	cacacaacaaggcaccat	3
M21	A→T	357	cttttatttctgactacaggg	aacagcagatttgagcagg	3
M22	A→G	129	agaagggtctgaaagcaggt	gcctactacctggaggctt	3
M23	A→G	159	tctctaacttctgtgagccac	ggaaaaactaaactctaaatctct	13
M25	G→C	121	aaagcgagagattcaatccag	ttttagcaagttaagtcaccagc	13
M26	G→A	68	ccagtggtaaagttttattacaattt	ttcacagtaagcaggcaatcc	4
M27	C→G	398	cggaagtcaaagttatagttactgg	cactgctgttctgtctaccaca	13
M28	T→G	277	gcttacttgggacacaggct	agagaagttgtcatacgataatgg	13
PN3	G→A	47	gacaatgtgagtaggagaagcatc	gttgctgggctgtctaggaat	5
M30	G→A	132	gaaccagacaatacgaaatagaag	tttagcggcttatctcattacc	13
M31	G→C	71	gaaccagacaatacgaaatagaag	tttagcggcttatctcattacc	13
M32	T→C	166	ttgaaaaaatacagtggaac	caagtgtttaaggatacaga	13
M33	A→C	180	ttgaaaaaatacagtggaac	caagtgtttaaggatacaga	13
M34	G→T	131	cacttcacatttgtttttagg	agtcattatttagtcattccag	13
M35	G→C	168	taagcctaaagagcagtcagag	agagggagcaatgaggaca	13
M36	T→G	74	agatcatcccaaaacaatcataa	aaggctgaaatcaatccaatctg	13
M37	C→T	203	cagattggattgatttcagcctt	agcatacaaaaaaaaaaaactgc	13
M38	T→G	146	cagtttttagagaataatgtcct	ttaaagaaaagaaaagcagatg	13
M39	delC	236	cagtttttagagaataatgtcct	ttaaagaaaagaaaagcagatg	13
M40	G→A	62	gtataataggctgggtgctg	catgagttcaaatgattctt	6
M41	G→T	117	gtataataggctgggtgctg	catgagttcaaatgattctt	13
M42	A→T	297	aaagcgagagattcaatccag	ttttagcaagttaagtcaccagc	4
M43	A→G	77	actaaaacaccattagaaacaaagg	ctgagcaacatagtgacccc	13
M44	G→C	263	ctggcaccttctgatattttgag	tgtgatttctatgtgtttgaggac	13
M45	G→A	109	gctggcaagacacttctgag	aatatgttcctgacaccttcc	4
TAT	T→C	28	gactctgagtgtagacttgtga	gaaggtgccgtaaaagtgtgaa	7
M47	G→A	395	agatcatcccaaaacaatcataa	aaggctgaaatcaatccaatctg	13
M48	A→G	160	aaacaatatgtatgctaattttgct	tcaatgtaaatgttagtataaggatg	13
M49	T→C	229	cggcaacagtgaggacagt	tgcttcaggagatagaggctc	4
M50	T→C	175	cggcaacagtgaggacagt	tgcttcaggagatagaggctc	4
M51	G→A	33	gagcctctatctcctgaagc	tgactgctctgttgcgaca	13
M52	A→C	477	actgtagcatggtcatctagggtg	gacgaagcaaacatttcaagagag	13
M54	G→A	164	cctcctctggtctgggttt	tgtgtcaggactggttccat	4
M55	T→C	228	cgtaggcgtttgacagcag	cctttcttcgtaatcctccc	4
M56	A→T	39	ccagaaactgaagtacaaatgc	tctcattgctgcctctcttt	4
M57	insA	132	attgggaggaagtggtttctg	gttctgtatctttccattatttgc	13
M58	G→A	224	tctctaacttctgtgagccac	ggaaaaactaaactctaaatctct	13
M59	A→C	279	cggcaacagtgaggacagt	tgcttcaggagatagaggctc	13
M60	insT	243	gcactggcgttcatcatct	atgttcattatggttcaggagg	4

APPENDIX I. (*cont.*)

Marker No.	Nucleotide Change	Position[1] (bp)	Primers (5′-3′) Forward	Reverse	Reference[4]
M61	C→T	98	attggattgatttcagccttc	attttattttctgtgttccttgc	13
M62	T→C	60	actaaaacaccattagaaacaaagg	ctgagcaacatagtgacccc	13
M63	G→A	43	ctcttcccttggttcctattc	ggtggatgcttctgcctaa	4
M64	A→G	279	tatagaccctgactactcaagagaa	ggtagccctttccagtctgt	4
M65	A→T	152	ttctgatgccagcttgttcg	gctacgggattaaagtaaccttg	13
M66	A→C	135	ctgtgtaacaccatcaagtgc	acatcttctggagacatacttcc	4
M67	A→T	377	ccatattctttatactttctacctgc	gtcttttcacttgttcgtggac	4
M68	A→G	268	ccatattctttatactttctacctgc	gtcttttcacttgttcgtggac	13
M69	T→C	222	ggttatcatagcccactatactttg	atctttattccctttgtcttgct	4
M70	A→C	45	ggttatcatagcccactatactttg	atctttattccctttgtcttgct	4
M71	C→T	197	ttgaattatagtcccttgcctc	gcttcttttctgtgatggctg	4
M72	A→G	157	ttgaattatagtcccttgcctc	gcttcttttctgtgatggctg	4
M73	delGT	260	cagaataataggagaatttttggt	attttccttattttctaagcagc	4
M74	G→A	195	atgctataataactaggtgttgaag	aattcagcttttaccacttctgaa	4
M75	G→A	296	gctaacaggagaaataaattacagac	tattgaacagaggcatttgtga	4
M76	T→G	339	tagaagtagcagattgggagagg	cctgataaaatgaaaaaaatggtc	13
M77	C→T	129	cttttcttcccttagctgttcc	gcaacaatgacaggtcaacc	13
M78	C→T	197	cttcaggcattatttttttttggt	atagtgttccttcacctttcctt	13
M81	C→T	147	acttaatttatagtttcaatccctca	ttcatggagatgtctgtatctgg	4
M82	delAT	179	ctgtactcctgggtagcctgt	aagaacgattgaacacactaactc	4
M83	C→T	120	gggaaaggagttatccagaaa	aatgacccatgcttacttagc	4
M85	C→A	437	aacagaattatcaggaaaaggttt	gcaatatacgtttctgcttcca	4
M86	T→G	85	tcccattatttgctatatttgct	tttctcatttaacttttctgaccc	4
M87	T→C	277	tcccattatttgctatatttgct	tttctcatttaacttttctgaccc	4
M88	A→G	166	attctagggtcaggcaactagg	tgtttgttctattctatggtcttcc	4
M89	C→T	347	agaagcagattgatgtcccact	tccagttaggagatcccctca	4
M90	C→G	170	tgatgtttcttcagtctttgagg	aagatgaatcctgctaccacc	4
M91	delT	360	gagcttggactttaggacgg	aaactttaaggcacttctggc	4
M92	T→C	340	ttgaatttcccagaattttgc	ttcagaaactggttttgtgtcc	4
M93	C→T	459	aacaaaacaaaacaaaaatactgaa	ggttcacttgaagatagtttaggtta	13
M94	C→A	227	cacatggagaacagagaaatgc	cttgtgaaatgttgtgaaagtgg	4
M95	C→T	172	gagtggaaatcaagatgccaag	gggctttctgtaacctggaga	4
M96	G→C	70	gttgccctctcacagagcac	aaggtcactggaaggattgc	13
M97	T→G	355	gttgccctctcacagagcac	aaggtcactggaaggattgc	13
M98	G→C	158	gaatggggtgttacatggaga	cctattatgagagacctgttttcc	13
M99	delA	96	gaatggggtgttacatggaga	cctattatgagagacctgttttcc	13
M101	C→T	154	tcacagcagcttcagcaaa	ataaaaattagactctgtgttactagc	13
M102	G→C	301	aaactgggacacttgtaatgaat	gttttttgagtccttgtttattctt	4
M103	C→T	259	cagtaagtgaactcacacataattcc	ccagttttatttcagtttcacagc	4
M104	A/G→A[2]	162	gaacttgtcgggaggcaat	tgatacacttcctcctttagtgg	13
M105	C→T	478	gggaggcaacctaagaaag	aggtgaacgcattctgtcat	4
M106	A→G	411	gggaggcaacctaagaaag	aggtgaacgcattctgtcat	4
M107	A→G	298	caaaagcactcgggttcct	ctttactcccacttatgcaacg	4
M108	T→C	40	agatggagccagcagaaag	acacagatgaattgaatgatggt	4
M109	C→T	264	gggtatcaaatgtcttcaacct	gggaatttcctgctacttgc	13
M110	T→C	241	cagggaaggaccgtaaaagg	atgtttatcatgtgcagtaaaggtt	4
M111	delTT	188	aatcttctgcaaagggttcc	cagctacaaaacaaaatactggac	13
M112	G→A	286	acttttttccaacagttatttttga	tatatttcttgatgatgagaccaat	13
M113	A→G	112	acttttttccaacagttatttttga	tatatttcttgatgatgagaccaat	13
M114	T→C	387	ttaccacacagttgagtagttctaaa	caaataaaattggagcggtta	13
M115	C→T	201	agtttacagtcacatcaatttgga	tccttttttctaccatatctctgttt	13
M116	A→C, A→T	176	aagtatgacttatgaagtacgaagaaa	attcagttagattttacaatgagca	13
M117	delATCT	142	aagtatgacttatgaagtacgaagaaa	attcagttagattttacaatgagca	13
M118	A→T	109	attctaagtttcacttcctgatcc	tagttccctaaattacgctacctc	13
M119	A→C	224	gaatgcttatgaatttcccaga	ttcacacaatatacaagatgtattctt	13
M120	T→C	224	gagcttggactttaggacgg	aaactttaaggcacttctggc	13

APPENDIX I. (*cont.*)

Marker No.	Nucleotide Change	Position[1] (bp)	Primers (5′-3′) Forward	Reverse	Reference[4]
M121	delAGAAA	183	gagcttggactttaggacgg	aaactttaaggcacttctggc	13
M122	T→C	73	tggtaaactctacttagttgccttt	cagcgaattagattttcttgc	13
M123	G→A	161	tggtaaactctacttagttgccttt	cagcgaattagattttcttgc	13
M124	C→T	246	tggtaaactctacttagttgccttt	cagcgaattagattttcttgc	13
M125	T→C	301	gccaccctcttatgcctct	tcaggagttatgtgaggaccc	4
M126	delAATA	277	gccaccctcttatgcctct	tcaggagttatgtgaggaccc	4
M127	C→A	372	tgaaaggaaatcagtgtaagagc	tgtaaaggtgactgtagttcagca	13
M128	delCA	314	acttttccaacagttatttttga	tatatttcttgatgatgagaccaat	13
M129	G→A	221	aatggcttactacaaagaacatttc	tacacggtctctaccaaagaaga	13
RPS4Y	C→T	41	tatctcctcttctattgcag	ccacaaggggaaaaaacac	8
M131	del9bp	93	cacacccagaatacaataatttt	ctttttattctacaagtgtacattttc	13
M132	G→T	482	aacagaattatcaggaaaaggttt	ttttacttgttcgtgtactttcaa	4
M133	delT	116	tgaaatggaaatcaataaactcagt	ccttttcttttttctttaacccttc	13
M134	delG	54	agaatcatcaaacccagaagg	tctttggcttctctttgaacag	13
M135	insC	150	tgaaatggaaatcaataaactcagt	ccttttcttttttctttaacccttc	13
M136	C→T	196	atgtgaagacaacactgtgtgg	ttgtggtagtcttagttctcatgg	4
M137	T→C	289	tggtaaactctacttagttgccttt	cagcgaattagattttcttgc	13
M138	C→T	291	aacttccaaaaactgtgaaaagatt	gattaaagcagccccttcag	13
M139	delG	397	ttactgataatgccatattgttttg	ttctcagacaccaatggtcct	13
M141	T→A	51	catcttaaaatacatttcatagcttt	gcttactattaggtctagcatcct	13
M143	G→T	246	atgctataataactaggtgttgaag	aattcagcttttaccacttctgaa	13
M144	T→C	342	agcacaagggtcacattgag	aggacaaggctttttgttgtt	4
M145	G→A	166	ttcagcaagagtaagcaagagg	ccttttggatcatggttctt	13
M146	A→C	141	gaatggggtgttacatggaga	cctattatgagagacctgttttcc	13
M147	insT	113	gtattctggggcaattttagg	ttgatacaagaggttattttaagca	13
M148	A→G	314	aacagaattatcaggaaaaggttt	ttttacttgttcgtgtactttcaa	13
M149	G→A	469	aacagaattatcaggaaaaggttt	ttttacttgttcgtgtactttcaa	13
M150	C→T	146	gcagtggagatgaagtgagac	cctactttcccccttcttctg	13
M151	G→A	209	acttaatttatagtttcaatccctca	ttcatggagatgtctgtatctgg	13
M152	C→T	101	aagctattttggtttcctttca	gccttgtgtgggtatgattg	13
M153	T→A	427	ttactgataatgccatattgttttg	ttctcagacaccaatggtcct	13
M154	T→C	252	acttaatttatagtttcaatccctca	ttcatggagatgtctgtatctgg	13
M155	G→A	251	tctctaacttctgtgagccac	ggaaaaactaaactctaaatctct	13
M156	A→G	147	ttcagcaagagtaagcaagagg	ccttttggatcatggttctt	13
M157	A→C	176	gctggcaagacacttctga	aatatgttcctgacaccttcc	13
M158	G→A	77	tgaaatggaaatcaataaactcagt	ccttttcttttttctttaacccttc	13
M159	A→C	89	attggattgatttcagccttc	attttattttctgtgttccttgc	13
M160	A→C	251	cagaataataggagaatttttggt	attttccttattttctaagcagc	13
M161	C→A	111	tcacagcagcttcagcaaa	ccttttggatcatggttctt	13
M162	C→C/T[2]	202	gaacttgtcgggaggcaat	tgatacacttcctcctttagtgg	13
M163	A→C	168	gcagcatataaaactttcagg	aaaacctaactttgctcaagc	13
M164	T→C	329	tagaagtagcagattgggagagg	cctgataaaatgaaaaaaatggtc	13
M165	A→G	132	aaagcgagagattcaatccag	ttttagcaagttaagtcaccagc	13
M166	G→A	53	tggtaaactctacttagttgccttt	cagcgaattagattttcttgc	13
M167	C→T	130	cgcggctttgaatttcaagctctg	ccagggccccgagggactctt	9
M168	C→T	371	agtttgaggtagaatactgtttgct	aatctcataggtctctgactgttc	4
M169	T→C	97	agtttgaggtagaatactgtttgct	ccagggccccgagggactctt	4
M170	A→C	327	tgcttcacacaaatgcgttt	ccaattactttcaacatttaagacc	4
M171	G→C	440	agtttgaggtagaatactgtttgct	ccagggccccgagggactctt	4
M172	T→G	197	ttgaagttacttttataatctaatgctt	ataatttattactttacagtcacagtgg	4
M173	A→C	191	aagaaatgttgaactgaaagttgat	aggtgtatctggcatccgtta	4
M174	T→C	219	acatctcagatcgttgtttggt	aaaaagccatgcaattacctg	4
M175	delTTCTC	79	ttgagcaagaaaaatagtaccca	ctccattcttaactatctcaggga	4
M176	C→T	52	gccgaagaattgcagtttgc	ccaatgttacccgattgtcc	10
M177	C→T	363	tttaacattgacaggaccag	gtgttggttctcctgtaaag	6
M178	C→T	220	taagcctaaagagcagtcagag	agagggagcaatgaggaca	13
M179	C→T	316	attatgcagaattaagatgaccag	attatgcagaattaagatgaccag	4
M180	T→C	402	acactactgtgctgtaatttgtgaa	tggtaagatttctctacatgaacag	4

APPENDIX I. (*cont.*)

Marker No.	Nucleotide Change	Position[1] (bp)	Primers (5′-3′) Forward	Reverse	Reference[4]
M181	T→C	130	gcttttatttattctactttgttttt	aacaatgaccaattcttttttcat	4
M182	C→T	38	tattcaaagacttaaagcagtggtta	ggaatcatcttgtaactaagttatcct	4
M183	A→C	324	actgggtaaatatgactatgattgag	ttccttttaacctattattactttcc	4
M184	G→A	62	cactttattttagtctgtgtcttttttc	aaacttagtaacatctatttctcctct	4
M185	C→T	89	ggagtacctatcactgaatgtgc	gtcattcatttctgcttggaac	4
M186	delG	63	ttgcatttactgttctagagagttct	ccacagagcaagactccatc	4
M188	C→T	185	gtattccctttgaagaaacatattg	aagtccatcacaagttaatttttttc	4
M189	G→T	191	actctcagcttatgtttgtcattg	gcttttggtacactctgctttt	4
M190	A→G	73	ctctgtcacaagtaaggaaatgat	ctgtccattggtagccttttt	4
M191	T→G	342	ttgcatttgtcatggttggt	gccaggataatttttttgtattttc	4
M192	C→T	202	catgggctgctgacatttt	aaatccttttggtttgtttgttt	4
M193	insCAAA	57	gcctggatgaggaagtgag	gccttctccatttttgacct	4
M194	T→C	101	gcctggatgaggaagtgag	gccttctccatttttgacct	4
M195	A→G	430	ccactcagctttcctcagg	cgttcgtttagtcataagatcg	4
M196	C→G	330	ttagacaacttactactttgatgtcct	taaacattacatgagaaattgctgt	4
M197	T→C	105	tcagacagtttagttggttacttcc	aagacacatcctcagctaatcttt	4
M198	C→T	45	tgaggtggaatgtatcagtatacc	tgatttcaaggatttgttagtctt	4
M199	insT	405	tgaggtggaatgtatcagtatacc	tgatttcaaggatttgttagtctt	4
M200	G→A	318	ggcttacacttgcagactttg	ggagaaatgtacaagagtctaaacc	4
M201	G→T	136	tatgcatttgttgagtatatgtc	gttctgaatgaaagttcaaacg	4
M202	T→G	259	ggaattgcagggtttaagc	gccacccacctgtaatcc	4
M203	G→C	108	gagtgccaagctgaggatga	tccttgcagccgctgaggag	4
M204	T→G	234	gtcaggctaacagactcctc	tgaagaggagtctgttagcctg	4
M205	T→A	78	gtataatactgtggttggaaagca	ccaaactatgtgataataaatggg	4
M206	T→G	31	gtataatactgtggttggaaagca	ccaaactatgtgataataaatggg	4
M207	A→G	79	aggaaaaatcagaagtatccctg	caaaattcaccaagaatccttg	4
M208	C→T	352	ataaatacaaaatcacctgatggat	ttaaacagcgaaattactaacaaaa	4
M209	A→G	471	cactgtcttccacaatggttg	aggtgattttgtatttatcttccc	4
M210	A→T	461	cactgtcttccacaatggttg	aggtgattttgtatttatcttccc	4
M211	C→T	381	caattcactatttgaggaatcca	gaagtctctgatttatttggcag	4
M212	C→A	234	tataatcaagttaccaattactggc	ttttgtaacattgaatggcaaa	4
M213	T→C	290	tataatcaagttaccaattactggc	ttttgtaacattgaatggcaaa	4
M214	A→G	404	tattacaaaatatggaaacaaggc	gaaatgccacttcactccag	4
M215	A→G	163	gtaaaactcagatatatacatcccatg	aaaaaaaaagaatcactatcttaacg	4
M216	C→T	54	ctcaaccagtttttatgaagctag	gagagtctgaactaatgtgtcttgt	4
M217	A→C	219	gcttatttttagtctctcttccat	acctgttgaatgttacatttcttt	4
M218	C→T	380	ttgtgagtttttttccatcaatc	tttattgacgatggtattagaagag	4
M219	T→C	232	ttgtgagtttttttccatcaatc	tttattgacgatggtattagaagag	4
M220	A→G	367	ttgtgagtttttttccatcaatc	tttattgacgatggtattagaagag	4
M221	G→A	200	gggaaatgtgaaaggaaaata	ttaaacttttataacactgacagaaac	4
M222	G→A	175	ggtgatggatgaggagtaaaaa	cattcaagatcccagaactgtc	11
M223	C→T	67	ttcagcaagagtaagcaagagg	ccttttggatcatggttctt	this study
M224	T→C	193	cttcaggcattatttttttggt	atagtgttccttcacctttcctt	this study
SY1187	12f2	10kb	caaattgtaagacaaggacatgc	gcacgttgtgcacatgtactc	12
SY1193	12f2	8kb	caaattgtaagacaaggacatgc	tgtccatgttggatgaggaa	12
PN1	C→T	78	tcacataatttcattttccc	tagtcctctccttattaacg	5
PN2	C→T	94	ggtaacacccataaaggttg	ttcactaccagcctaagtac	5

[1] Nucleotide position from 5′ end of forward primer.
[2] Duplicated locus with most men carrying both the A and the G allele at position 162.
[3] Duplicated locus with most men carrying the C allele only at position 202.
[4] Reference codes: 1: Hammer & Horai, 1995; 2: Seielstad *et al* 1994, 3: Underhill *et al* 1997; 4: Shen *et al* 2000; 5: Hammer *et al* 1997; 6: Whitfield *et al* 1995; 7: Zerjal *et al* 1997; 8: Bergen *et al* 1999; 9: Bianchi *et al* 1997; 10: Shinka *et al* 1999; 11: Sun *et al* 1999; 12: Sun *et al* 2000; 13: Underhill *et al* 2000.