

RESEARCH HIGHLIGHT

# The jigsaw puzzle of our African ancestry: unsolved, or unsolvable?

Chiara Batini\* and Mark A Jobling

## Abstract

A revised root for the Y chromosome phylogeny further fragments the picture of modern human origins that can be reconstructed from genetic, linguistic and archaeological data.

'Where do we come from?' 'When did the first humans live?' 'Apparently we're all supposed to be Africans, how is that possible?' These are some of the questions that people ask, and to which a variety of researchers in anthropology, archaeology and genetics try to give reasonable answers. The publication of a revised root for the Y chromosome phylogeny by Scozzari and colleagues [1] now contributes new genetic evidence on the dating and geographical origins of early modern humans.

Yes, we do all come from Africa, or, more precisely, the population ancestral to all modern humans lived south of the Sahara Desert, probably around 200 thousand years ago (kya). This conclusion is supported by several approaches, among which the study of modern human genetic variation has played a major role. A key finding is that, among global samples, the highest DNA diversity is found among African populations, and this diversity declines with migration distance from Africa. Furthermore, when 'gene trees' are constructed from sequence variants of particular loci, in most cases the earliest branches are found within African populations.

Within the great continent of Africa, however, the details are much less clear. Was there only one ancestral population that expanded from a specific area? If so, where was it? What was its size? Those who study African prehistory know that answers to these questions are much more difficult to find: this is a complex jigsaw puzzle for which the final picture and the number of pieces are still unknown.

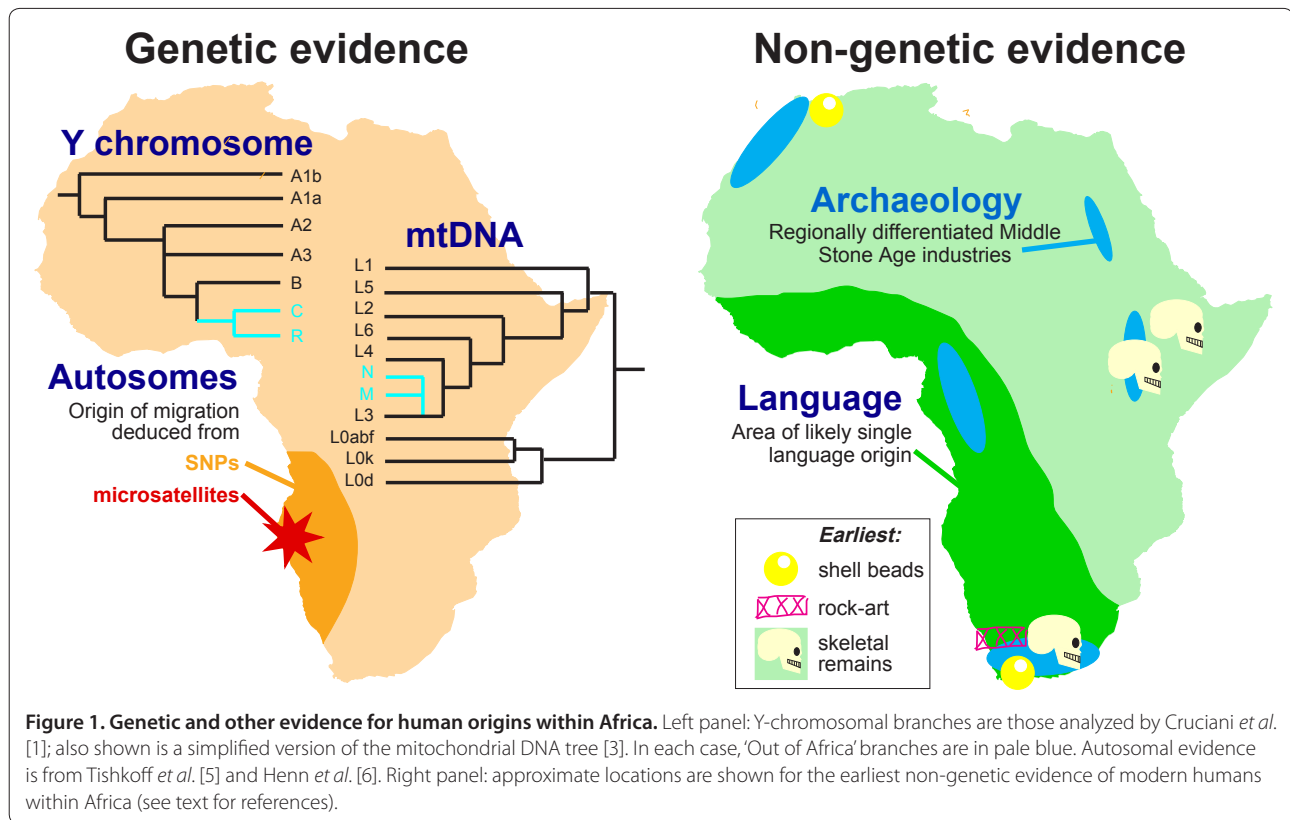
## A time and place for our Y-chromosomal roots

Recently, a new piece of the puzzle has turned up. In a paper in the *American Journal of Human Genetics*, Rosaria Scozzari and colleagues present a revised structure for the deepest part of the tree of the male-specific region of the Y chromosome [1]. The novelty of their study is that they took a large-scale sequencing approach, allowing the unbiased ascertainment of many previously undescribed variants. They sequenced 206 kb of Y-chromosomal DNA in seven males, including members of the previously deepest-rooting major branches A and B, which are almost exclusively African and reach their highest frequencies in scarce hunter-gatherer populations. Previously, the root lay between these two branches, but it has now moved to a new position that is within A, between A1b and A1a, from which descend all other Y branches (Figure 1).

In addition to proposing a new root for the human Y chromosome phylogeny, Scozzari and colleagues also use their data to argue for a revised view of the dating and geographical origins of early modern humans. The 138 SNPs discovered provide sufficient stable molecular information to be used for dating estimates, yielding a time to the most recent common ancestor (TMRCA) of  $141.5 \pm 15.6$  kya; this is considerably older than previous estimates, which were more recent than 100 kya. In addition, a population survey of more than 2,000 males shows the earliest sub-branches A1b and A1a to be rare (8/2,204 individuals analyzed) and restricted to North and Central Africa. This geographical distribution is probably the most controversial issue, in the context of previous evidence, but is consistent with a recent lower-resolution study showing that the early sub-branches of both A and B show a mainly Central African distribution [2].

Due to the paucity of fossil remains and the poor conditions for preservation, no study of ancient DNA has yet been carried out on African samples, so all we have is the variation found among extant populations. But, where are the rest of the pieces, do they fit with the new Y-chromosomal evidence, and what picture is beginning to emerge?

\*Correspondence: cb334@le.ac.uk  
Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK



### Different genetic markers and methods can point to different origins

Maternally inherited mitochondrial DNA (mtDNA) was one of the first genetic markers to be analyzed. The pioneering study that introduced the concept of 'mitochondrial Eve' paved the way for a long series of phylogenetic, population and simulation studies in African populations. The most recent continent-wide phylogenetic study is a high-resolution dissection of African-specific mtDNA variation [3]. An analysis of more than 600 complete 16.5 kb mitochondrial genomes places the root of the mtDNA tree between L0 and L1'6 and yields a TMRCA of  $200 \pm 13$  kya, based on variation of the coding region only. Within L0, two sub-branches (L0d and L0k) are characteristic of Khoisan-speaking populations of southern Africa, while the rest are found at low frequencies among southern, eastern and western populations, with the exception of one sub-branch of L0a found specifically in Eastern Pygmies [3,4]. Distinguishing between a southern, an eastern or a south-eastern origin of the ancestral female populations seems to be a very delicate issue; however, because of the deep ancestry of most of the early sub-branches, a highly structured female population is inferred for these early *Homo sapiens* communities.

Two recent explorations of autosomal variation have contributed some other interesting missing pieces. An analysis of 1,327 microsatellite and indel markers in 121 African populations has inferred 14 ancestral population clusters, four of which correspond to the four hunter-gatherer groups: Western and Eastern Pygmies from Central Africa, Khoisan speakers from southern Africa, and Hadza from Tanzania [5]. A more recent study of 55,000 SNPs in 12 populations has shown a very similar general pattern [6]. All hunter-gatherer groups (with the exception of the two groups of Pygmies) separate into different clusters, suggesting that the structure observed today is the result of an ancient separation among their ancestral populations. Both studies also attempt to infer the geographical origin of the expansion of modern humans in Africa. In summary, they both assume that moving away from the origin will affect, in the first case, the level of genetic variation (calculated as the variance of the specific microsatellites, which should decrease with distance [5]), and, in the second case, the levels of linkage disequilibrium in the population (calculated through the statistic  $r^2$ , and expected to increase with distance [6]). Both studies find a clear signal of origin in south-west Africa, in the current territory of Khoisan-speaking populations in Namibia.

Thus, despite the hope that genetics might provide definitive answers on ancestry, the analysis of different genomic regions points to different places within Africa. Perhaps we should not be surprised about this: the Y chromosome and mtDNA are independent loci with their own histories and phylogenies, which need not necessarily lead back to the same places, and the same times. Also, for each independent Y or mtDNA, there are four of any autosome, and this low effective population size means that they are particularly susceptible to stochastic effects. This genetic drift can be exacerbated by sex-specific behaviors of men and women. By contrast, the autosomal evidence is based on many independent loci, and uses statistical inferences that do not involve phylogenetic trees. The comparison of independently evolving regions of the genome analyzed at different resolutions with different inference methods might not yield a coherent picture of the origin of modern humans in Africa.

### Anthropological evidence for early humans in different African regions

Of course, genes are not the only source of evidence. Early sub-Saharan human remains have been found in eastern Africa, with dates between 150 and 200 kya, and more recently in southern Africa (Figure 1; see [7] for the most ancient). However, preservation of ancient bones is limited in most of Central Africa because of soil acidity. Here, archaeology may help. There is general agreement that early *Homo sapiens* communities can be associated with Middle Stone Age industries. These have been found all over Africa, but dating has been difficult in many cases, and mainly based on radio-carbon, which has a time limit of about 40 kya. However, a few finds have been dated beyond this limit, and the general observed pattern is a clear regional differentiation at around 100 kya (for a review, see [8]). Early examples of geometric rock art have been found in South Africa (70 to 100 kya), while the earliest shell beads have been reported from South (75 kya) and North Africa (82 kya) [9] (Figure 1). Therefore, material culture seems to suggest that by 70 to 100 kya the ancestral population to modern humans was already fragmented into well-differentiated cultural units. So: many puzzle pieces, but a very complex picture, which does not seem to give a straightforward answer.

Even language shows, globally, the influence of our African origin: the diversity of word sounds (phonemes), like that of genes, decreases with migration distance from there. Within the African continent the most likely location of single language origin inferred from individual languages encompass an area from western to southern Africa [10] (Figure 1).

### An unsolved, and perhaps unsolvable, puzzle

Massively parallel sequencing approaches are likely to contribute to this area. The recent Y chromosome analysis [1] represents the death throes of 'conventional' large-scale DNA sequencing: a total of approximately 1,400,000 bp were sequenced using traditional methods, but now newer technologies should allow access to many megabases of Y-chromosomal DNA (paralogous sequences permitting). So the Y chromosome tree will become vastly more complicated, but also more resolved, and more able to yield reliable dates. New whole-genome sequences from Africa will also clarify our views of origins there. Unfortunately, despite new advances, results from ancient DNA seem unlikely to help us in Africa.

So, how is the jigsaw puzzle of our African origins coming along? Many pieces have been found up to now by different researchers, and the picture built with them seems to be that of a fragmented population, living throughout the whole continent and differentiated into clear cultural units. Signals based on high-resolution autosomal variation in a reduced number of populations point to a south-western origin, while analyses of uniparental markers shows prevalence of early tree branches either in north Central Africa (Y chromosome) or in the south east (mtDNA). However, some pieces may be lost, and some may simply not exist: so a final coherent picture may never be completed, and different overlapping layers may be needed to represent the complexity of our evolutionary past.

#### Abbreviations

bp, base pair; kya, thousand years ago; mtDNA, mitochondrial DNA; SNP, single nucleotide polymorphism; TMRCA, time to the most recent common ancestor.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

The authors' work is supported by the Wellcome Trust (grant number 087576).

Published: 28 June 2011

#### References

1. Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, Scozzari R: A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. *Am J Hum Genet* 2011, **88**:814-818.
2. Batini C, Ferri G, Destro-Bisol G, Brisighelli F, Luiselli D, Sánchez-Diz P, Rocha J, Simonson T, Brehm A, Montano V, Elwali NE, Spedini G, Eugenia D'Amato M, Myres N, Ebbesen P, Comas D, Capelli C: Signatures of the pre-agricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. *Mol Biol Evol* 2011, in press.
3. Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S, Genographic Consortium: The dawn of human matrilineal diversity. *Am J Hum Genet* 2008, **82**:1130-1140.
4. Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, Comas D: Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Mol Biol Evol* 2011, **28**:1099-1110.
5. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo

- JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, Williams SM: **The genetic structure and history of Africans and African Americans.** *Science* 2009, **324**:1035-1044.
6. Henn BM, Gignoux CR, Jobin M, Granka JM, Macpherson JM, Kidd JM, Rodriguez-Botigou L, Ramachandran S, Hon L, Brisbin A, Lin AA, Underhill PA, Comas D, Kidd KK, Norman PJ, Parham P, Bustamante CD, Mountain JL, Feldman MW: **Hunter-gatherer genomic diversity suggests a southern African origin for modern humans.** *Proc Natl Acad Sci U S A* 2011, **108**:5154-5162.
7. McDougall I, Brown FH, Fleagle JG: **Stratigraphic placement and age of modern humans from Kibish, Ethiopia.** *Nature* 2005, **433**:733-736.
8. McBrearty S, Brooks AS: **The revolution that wasn't: a new interpretation of the origin of modern human behavior.** *J Hum Evol* 2000, **39**:453-563.
9. Mellars P: **Why did modern human populations disperse from Africa ca. 60,000 years ago? A new model.** *Proc Natl Acad Sci U S A* 2006, **103**:9381-9386.
10. Atkinson QD: **Phonemic diversity supports a serial founder effect model of language expansion from Africa.** *Science* 2011, **332**:346-349.

doi:10.1186/gb-2011-12-6-118

**Cite this article as:** Batini C, Jobling MA: **The jigsaw puzzle of our African ancestry: unsolved, or unsolvable?** *Genome Biology* 2011, **12**:118.