

## Kenyan crossroads: migration and gene flow in six ethnic groups from Eastern Africa\*

Loredana Castrì, Paolo Garagnani, Antonella Useli, Davide Pettener & Donata Luiselli

*Dipartimento di Biologia evoluzionistica sperimentale, Laboratorio di Antropologia molecolare, Università di Bologna, via Selmi 3, 40126 Bologna*

e-mail: loredana.castrì@unibo.it

**Keywords** - *mtDNA, African populations, genetic variability, migrations.*

From an anthropological point of view, Kenya is one of the most interesting region of Africa. The territory is close to the place from where most probably modern humans spread all over the world, as suggested by paleoanthropological, archaeological and genetic data (Forster & Matsumura, 2005) and it has most probably played an important role in the maturation phase of mtDNA lineages, as recently proposed (Behar *et al.*, 2008). In the last 5000 years, Kenya has been also involved in different spatial and temporal migratory events (Newman, 1995). Before the arrival of Bantu farmers around the first millennium BC, that deeply influenced the pattern of genetic diversity in sub-Saharan Africa, the region was settled by Sudanic and Cushitic hunter-gatherer groups that probably arrived between the third and second millennium BC living with or replacing the pre-existing Khoi-San and Pygmy groups (Phillipson, 1993; Newman, 1995). Later, in the first millennium AD other groups of eastern and western Nilotic herders got to the region from western and northern areas of Africa (Phillipson, 1993; Newman, 1995). Then, the present pattern of genetic diversity has been shaped by a complex history of migrations and interactions between different populations so that the region was defined by James Newman as “*the most complex ethnolinguistic region in the continent*” (Newman, 1995). At the present, the

territory is inhabited by more than 30 ethnic groups speaking languages belonging to all the 4 different linguistic African families, namely Khoi-San, Niger-Kordofonian, Nilo-Saharan and Afro-Asiatic.

The main goal of our study is to analyse the mtDNA variability in six Kenyan populations in order i) to further contribute to the reference database of African sequences, ii) to assess the pattern of genetic diversity in these six populations and iii) to determine the influence of migratory events that affected the region by estimating gene flow and migration rates among populations.

Oral swabs were collected from 287 Kenyan unrelated individuals along with biodemographic data. The ethnic origin of all individuals and their grandparents was checked by oral interview with the aid of a local African speaker. All individuals gave their informed oral consent to participating to the project.

Collected individuals belong to six ethnic groups (Fig. 1): Turkana, Samburu and Maasai (Eastern Nilotic-speakers); Luo (Western Nilotic-speakers); Rendille and ElMolo (Cushitic-speaking populations). All populations are cattle herders, with the exception of ElMolo that are fishermen (Dyson & Fuchs, 1937).

We re-sequenced the first hypervariable region (HVSI) of mtDNA. Sequences were attributed to

\* L.C. has been awarded with the prize entitled to Michele Belledi, as first author of the best poster presented at the XVII Congress of Italian Anthropologists (Cagliari, 27-29 September 2007, [http://dbs.unica.it/cong\\_ressoai/](http://dbs.unica.it/cong_ressoai/)).



**Fig. 1 – Geographical location of the Kenyan samples.**

the mtDNA haplogroups according to the most recent nomenclature (Torrioni *et al.*, 2006; Behar *et al.*, 2008). Haplotype diversity, nucleotide diversity, the average number of pairwise differences or average mismatch, the goodness-of-fit of mismatch distribution and selective neutrality indices were calculated using Arlequin 2.000 (Excoffier *et al.*, 2005). Gene flow and migration rates between pairs of populations were estimated by means of the software Migrate (Beerli & Felsenstein, 2001) which uses a coalescent theory approach for giving maximum likelihood estimates for effective population sizes and migration rates of  $n$  populations.

Frequency distributions of mtDNA haplogroups are homogenous in the Eastern-Nilotic speaker Turkana, Samburu and Maasai, where the higher number of mtDNA lineages are revealed (Fig. 2). In the Western-Nilotic Luo we observe the presence of some haplogroups that have been suggested to be involved in Bantu farmer migration, such as L0a2, L3e and the central-western haplogroups L1b and L1c (Quintana-Murci *et al.*, 2008). In the Cushitic Rendille and ElMolo a reduction of the number of haplogroups is evident, especially in the ElMolo, but in this groups the subclades of N are well represented. Many of the Kenyan haplotypes belonging to the typical Eastern African haplogroups L0f, L4 and L5, are shared with other Eastern African populations from Ethiopia, Kenya and Tanzania and with

northeastern populations (Kivisild *et al.*, 2004; Tishkoff *et al.*, 2007). Also M1 haplogroup and some subclades of haplogroup N are present in the six populations examined. In particular, M1 is present at a relatively high frequency in the Maasai sample with haplotypes shared with north-eastern African population. This finding is not surprising, since the Nilotic origin of this population. The haplotypes belonging to subclades K, R0 (previously named pre-HV), HV and I are shared mainly with eastern and north-eastern populations (Kivisild *et al.*, 2004; Tishkoff *et al.*, 2007).

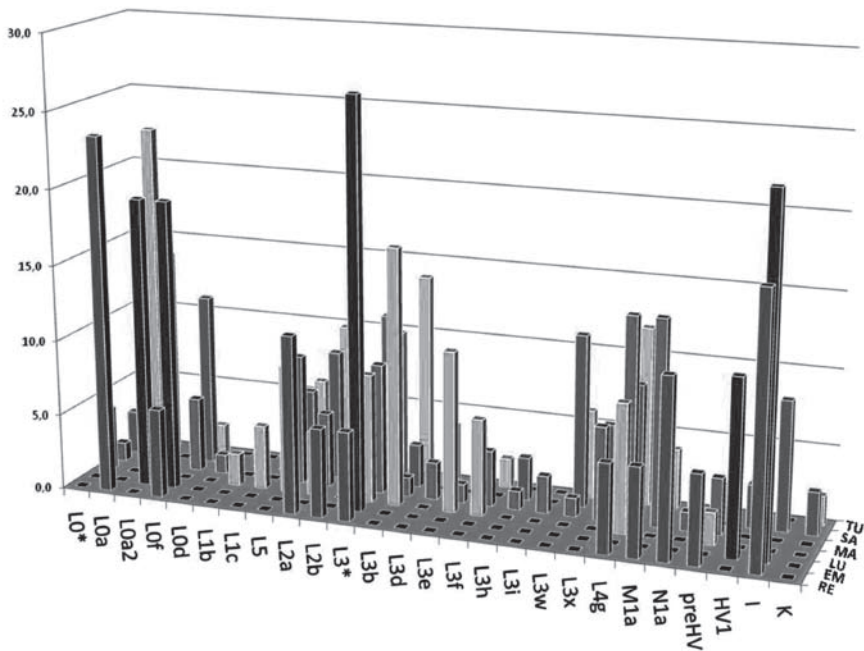
As for the diversity and demographic patterns, all Kenyan populations display levels of haplotype diversity (0.98-0.99), nucleotide diversity (0.029-0.030) and mean numbers of pairwise differences (10.33-10.87) in the range of other Eastern African populations with signatures of population expansion identified by significance of neutrality tests and unimodal mismatch distributions. Only the ElMolo group are characterized by a low haplotype diversity (0.88), close to that observed in Khoisan hunter-gatherers, and by the nonsignificance of neutrality tests and the multimodal mismatch distribution indicating small population size and strong genetic drift.

Analysis of molecular variance shows that the populations are genetically differentiated.  $\Phi_{st}$  is significant in all analyses (0.085-0.099,  $p < 0.0001$ ), but the average level of divergence increases when we compare tribes speaking languages of different families ( $\Phi_{ct}$  values 0.026 vs 0.009, significant only in the linguistic grouping criteria analysis). At a more micro-geographic level (9 Kenyan populations) the values of genetic divergence between population decrease ( $\Phi_{st}$  between 0.0019 and 0.0036), and the AMOVA index  $\Phi_{ct}$  is never significantly different from 0, regardless of the grouping criteria considered.

The migration rate estimates (i.e. number of migrants per generation) between pairs of populations, namely the six Kenyan populations and northern, north-eastern and central-western African populations, are reported in Table 1. It is evident the presence of strong gene flows from northern and north-eastern populations towards

**Tab. 1 - Migrate analysis on African and Kenyan populations. Short chains=10 (used trees 1000/100000); long chains=3 (used trees 10000/1000000); average values on 3 independent multiple runs. The highest values of gene flow between Kenyan populations are underlined; the highest values of gene flow between Kenyan and the other African regions are in italic. The arrows indicate the direction of gene flow.**

MCMC estimates										
Population	$\Theta$	2Nm [x=receiving population]								
[x]		Rendille→x	Samburu→x	Turkana→x	Maasai→x	ElMolo→x	Luo→x	N Afr→x	NE Afr→x	CW Afr→x
Rendille	0.08452	-----	2,93	1,57	3,74	<u>7,61</u>	3,36	0,84	2,40	1,21
Samburu	0.06729	5,56	-----	9,63	<u>13,07</u>	6,86	5,85	<i>7,80</i>	<i>15,49</i>	4,33
Turkana	0.03787	1,19	2,13	-----	<u>12,53</u>	6,89	12,09	<i>13,99</i>	<i>12,04</i>	<i>10,18</i>
Maasai	0.05115	11,40	9,43	<u>13,84</u>	-----	1,54	0,75	1,93	6,78	4,47
ElMolo	0.00454	0,17	0,09	0,48	0,30	-----	1,00	0,39	0,61	0,98
Luo	0.06665	4,47	12,34	8,66	13,08	7,48	-----	<i>7,39</i>	<i>5,13</i>	<i>13,61</i>
N Afr	0.89478	67,46	80,98	127,50	40,92	50,43	32,03	-----	217,16	239,49
NE Afr	0.02444	12,41	11,02	14,64	13,71	10,60	15,20	11,61	-----	14,54
CW Afr	0.01230	1,63	1,18	1,44	0,91	1,31	1,63	1,33	1,13	-----



**Fig. 2 - mtDNA haplogroup frequency distribution in the examined Kenyan populations. The colour version is available online at the JASs website.**

Nilo-Saharan-speaking populations of Kenya (Turkana, Samburu and Maasai) probably due to their common origin. It is also interesting to note the high contribution of central-western Bantu-speaking populations to Turkana and in particular, towards Luo, which indeed are characterized by a high percentage of haplogroups typical of Bantu-speaking groups. Finally, we note the lowest values of exchanged migrants in Elmolo which confirms their genetic isolation.

In conclusion, a genetic homogeneity of the Kenyan populations can be pointed out from phylogeographic and AMOVA analyses, especially at micro-geographical level, although the presence of some differences in the haplogroup distribution. Continuous gene flow between populations resulted in a genetic homogeneity larger than that expected on the basis of the ethnic conflicts affecting the region. On the other hand, the ethnic and language differences between tribes and geography acted as genetic barrier to gene flow at macro-geographical level, as showed by the AMOVA results on all Eastern African populations. Additional data from Y-chromosome and autosomal markers are needed to clarify the genetic role of this region in the spread of modern humans in Eurasia.

## References

- Beerli P. & Felsenstein J. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. *Proc. Natl. Acad. Sci. USA*, 98:4563-4568.
- Behar D.M., Vilems 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 R.S., Rosset S. & The Genographic Consortium. 2008. The Dawn of Human Matrilineal Diversity. *Am. J. Hum. Genet.*, 82: 1130-1140.
- Dyson W.S., and Fuchs V.E. 1937. The Elmolo. *Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 67: 327-338.
- Excoffier L., Laval G., and Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform. Online*, 1:47-50.
- Forster P., Matsumura S. 2005. Evolution. Did early humans go north or south? *Science*, 308:965-6.
- Kivisild T., Reidla M., Metspalu E., Rosa A., Brehm A., Pennarun E., Parik J., Geberhiwot T., Usanga E. & Vilems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am. J. Hum. Genet.*, 75:752-70.
- Newman J.L. 1995. *The Peopling of Africa: A Geographic Interpretation*, Yale University Press, New Haven.
- Phillipson D.W. 1993. *African archaeology*. Cambridge University Press, Cambridge.
- Quintana-Murci L., Quacha H., Harmanta C., Luca F., Massonnet B., Patin E., Sica L., Mougouama-Daoudad P., Comas D., Tzur S., Balanovsky O., Kidd K.K., Kidd J.R., van der Veen L., Hombert J.M., Gessain A., Verdu P., Froment A., Bahuchet S., Heyer E., Dausset J., Salas A. & Behar D.M. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. *Proc. Natl. Acad. Sci. USA*, 105:1596-1601.
- Tishkoff S.A., Gonder M.K., Henn B.M., Mortensen H., Knight A., Gignoux C., Fernandez-Pulle N., Lema G., Nyambo T.B., Ramakrishnan U., Reed F.A. & Mountain J.L. 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. *Mol. Biol. Evol.*, 24:2180-2195.
- Torrioni A., Achilli A., Macaulay V., Richards M. & Bandelt H.J. 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet.*, 22:339-45.

Editor, Giovanni Destro-Bisol