

# Can mtDNA Contribute to Disentangle the Afro-Asiatic Querelle?

L. Castri<sup>1</sup>, A. Boattini<sup>1</sup>, S. Sarno<sup>1</sup>, S. De Fanti<sup>1</sup>, E. Flamigni<sup>1</sup>, M. Cioffi<sup>1</sup>, G. Frinchillucci<sup>2</sup>, S. Tucci<sup>1</sup>, P. Garagnani<sup>3</sup>, A. Useli<sup>4</sup>, D. Pettener<sup>1</sup>, D. Luiselli<sup>1</sup>

<sup>1</sup> Area di Antropologia, Dipartimento di Biologia E.S., Università di Bologna. E-mail: alessio.boattini2@unibo.it

<sup>2</sup> Museo Geografico Polare di Fermo

<sup>3</sup> Laboratorio di Immunologia, Dipartimento di Patologia Sperimentale, Università di Bologna

<sup>4</sup> Dipartimento di Zoologia e genetica Evoluzionistica, Università di Sassari

KEY WORDS: Eastern Africa, mtDNA, Afro-Asiatic.

## Introduction

Genetic and linguistic evidences agree that the Eastern side of Africa (EA) is characterized by an incredible degree of "juxtaposed diversity" (Newman, 1995). Afro-asiatic (AA), one of the most widespread African linguistic phyla, reaches its highest levels of internal differentiation in EA. Nevertheless, there is no agreement among Africanists about its place and time of origin. One hypothesis places the homeland in the Levant during the earliest Neolithic, and ties AA spread with agriculture and pastoralism diffusion (Diamond and Bellwood, 2003). Another hypothesis favours its homeland in north-eastern Africa: in that case AA would have spread in pre-agricultural times (Ehret, 1995). From a molecular anthropology point of view, EA is one of the most compelling puzzles in world human diversity, being the resultant of repeated migration events alternated with periods of isolation, accompanied by language shifts and processes of group absorption/split, not to mention its role in the emergence of *Homo sapiens sapiens* (Kisivild *et al.*, 2004; Poloni *et al.*, 2009; Gomes *et al.*, 2010). We aim to contribute to the knowledge of EA genetic history by searching traces of population movements that may be related to AA diffusion.

## Materials and Methods

Materials include 452 HVS-I unpublished sequences from 9 populations settled in EA (35 to 43, Fig.2), among which the first-ever sampled speakers of the AA-Omoti family (Dawro Konta, 35; Hamer, 37) and of the controversial AA-Cushitic Ongota language (36), that, according to Fleming (2006), may be interpreted as a basal branch of the AA tree. Results are compared with those from ~ 4000 EA and AA samples from literature.

Population structure was investigated by means of Spatial Principal Component Analysis (SPCA) and MCLUST. SPCA finds population scores (PCs) maximizing the product of variance and spatial autocorrelation (Moran's I). Scores from PCs (up to ~80% of the total variance) were fed to a model-based clustering algorithm (MCLUST) that chooses the optimal model (according to BIC) and provides the associated classification. Silhouette information measures how much well items (populations) fit to a given clustering. Results (not shown) were used to select - within each of the clusters revealed by MCLUST - the best fitting populations. These samples were grouped (taking into account also geographic and linguistic information) into seven units (Tab.1). Migration parameters among the units were estimated using the software MIGRATE-N by means of a Bayesian approach.

## Results and Discussion

Unit	Geo. Loc	Main linguistic groups	mclust	Pops
I	North-West Africa	AA-Semitic, AA-Berber	A	6, 9, 11, 16
II	Levant	AA-Semitic	A	61, 62, 63, 69
III	Chad Basin	AA-Chadic	B	70, 73, 77, 78
IV	Ethiopia	AA-Cushitic, AA-Semitic	B	30, 32, 33, 50
V	Dawro Konta	AA-Omoti	C1	35
VI	Kenya, Tanzania	Non-AA	C1	40, 41, 43, 47
VII	Kenya, Tanzania	AA-Cushitic	C2	36, 51, 52, 53

A first run of SPCA and MCLUST (Fig.1) on the whole data-set led to the identification of three clusters: A) Northern Africa and the Levant (AA-Berber and AA-Semitic); B) Ethiopia, Chad Basin, Egypt, Yemen (AA-Cushitic, AA-Semitic, AA-Chadic); C) Kenya, Tanzania, Southern Ethiopia (AA-Cushitic and Ongota, AA-Omoti, Nilo-Saharan, Niger-Congo).

A second run (Fig.1) using EA populations only, besides confirming the Ethiopian-Yemenite cluster, was able to ungroup populations from Tanzania,

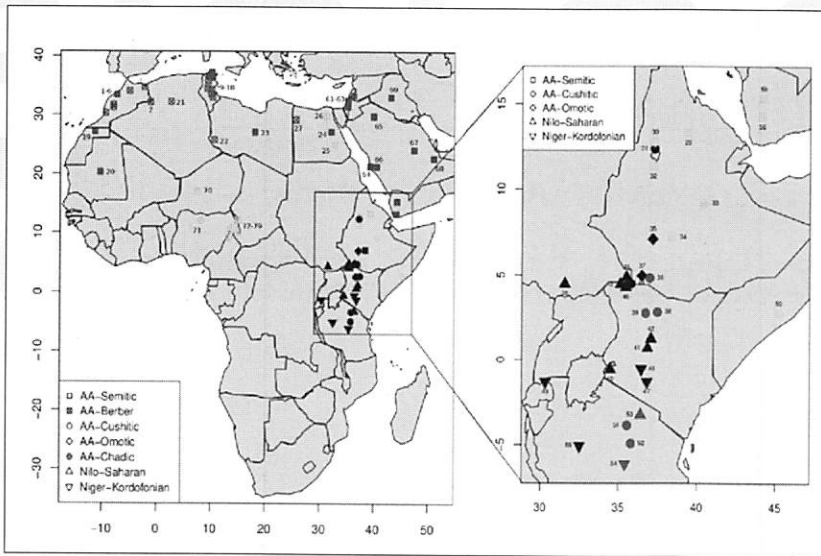


Fig. 1. Model-based clustering of EA and AA populations

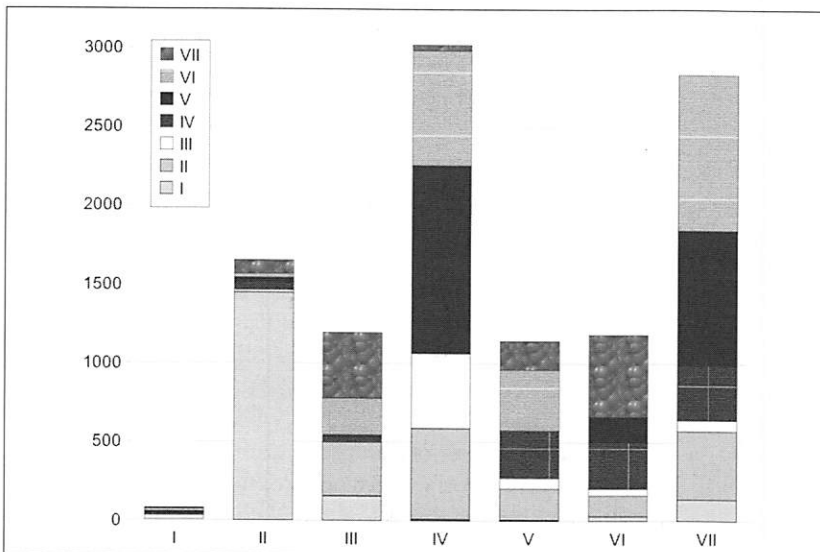


Fig. 2. Migratory contributes ( $n/\mu$ ) to each Unit.

Kenya and Southern Ethiopia into two clusters: C1) Non-AA groups and AA-Omotic (mainly), C2) AA-Cushitic (with Ongota) groups (mainly). Migration parameters estimated with MIGRATE-N for seven population units (Tab.1, Fig. 2) suggest that Ethiopia (IV) looks like the most admixed population, with important contributes from AA-Omotic groups (V) and non-AA Kenya/Tanzania (VI). AA-Cushitic speakers from the same area (VII) show relevant migratory flows from the AA-Omotic groups (V). NW-Africa (I) seems only marginally affected by gene flow, while contributing significantly to the Levant (II) mtDNA pool (II). In conclusion, links between geography, AA distribution and mtDNA genetic structure in EA seem feeble but not negligible. Ethiopia and Kenya/Tanzania are well differentiated by mtDNA, but AA-Omotic groups, despite being affiliated to the cluster C1 (but with a low Silhouette

score), seem to have contributed both to Ethiopian populations (B) and to Kenyan/Tanzanian AA-Cushitic speaking groups (C2). Cluster C1 may be read as the result of introgression of Niger-Congo- and Nilo-Saharan-speaking groups from East. These populations seem to have interrupted a more ancient genetic-linguistic continuity whose remnants is AA-Cushitic-specific cluster C2. The Ongota, despite being a genetic and linguistic outgroup, fit well in the Southern Cushitic cluster (C2). Northern Africa (with some exceptions in Egypt) and the Levant does not show traces of appreciable admixture with EA populations. Finally, AA-Chadic speaking groups are related to EA populations, in particular those from Ethiopia (B). Furthermore, Ethiopia shows links with Egyptian populations.

Summing it up, our data agree well with a origin of AA languages in EA. The Nile Valley corridor, as well as the Southern-Saharan route and the Bab-el-Mandeb strait, are the most likely paths through which AA languages spread in Africa and the Levant. On the contrary it would be difficult to fit these results with a Levantine origin of AA, unless hypothesizing repeated and wide language shifts.

## References

- Diamond J., Bellwood P. 2003. Farmers and Their Languages: The First Expansions. *Science*, 300: 597-603.
- Ehret C. 1995. *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, Tone, Consonants, and Vocabulary*. University of California Press, Berkeley and Los Angeles.
- Fleming H. 2006. *Ongota: A Decisive Language in African Prehistory*. Harrassowitz, Wiesbaden.
- Gomes V., Sánchez-Diz P., Amorim A., Carracedo A., Gusmão L. 2010. Digging deeper into East African human Y chromosome lineages. *Hum. Genet.*, 127(5): 603-613.
- Kivisild T., Reidla M., Metspalu E., Rosa A., Brehm A., Pennarun E., Parik J., Geberhiwot T., Usanga E., Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. *Am. J. Hum. Genet.*, 75(5):752-70. Erratum in: *Am. J. Hum. Genet.*, 2006, 78(6): 1097.
- Newman JL. 1995. *The Peopling of Africa: A Geographic Interpretation*. Yale University Press, New Haven.
- Poloni E.S., Naciri Y., Bucho R., Niba R., Kervaire B., Excoffier L., Langaney A., Sanchez-Mazas A. 2009. Genetic evidence for complexity in ethnic differentiation and history in East Africa. *Ann. Hum. Genet.*, 73(6): 582-600.