

## ***Maternal DNA variation of Pygmy hunter-gatherers and Bantu-speaking agriculturalists from central-west Africa***

Hélène Quach, Francesca Luca, Lolke van der Veen, Jean-Marie Hombert, Lucas Sica, Patrick Mouguiama-Daouda, Alain Froment, Gemma Berniell-Lee, David Comas, Lluís Quintana-Murci

CNRS FRE2849, Institut Pasteur, Paris, France

Two features make central/west Africa an interesting region for human evolutionary genetic studies. First, two groups of populations presenting different lifestyles inhabit this region: the Pygmy hunter-gatherers and the Bantu-speaking agriculturalists. Second, the highest linguistic diversity of the Bantu family is found in Nigeria and western Cameroon, suggesting this area as the « core » region of the Bantu dispersals starting 5000 YBP. Indeed, the great majority of sub-Saharan Africans speak one of > 500 closely related Bantu languages. In the genetic context, central Africa is poorly sampled and maternally-inherited population genetic variation (i.e. mitochondrial DNA, mtDNA) in this region remains largely unknown. We have therefore defined mtDNA variation in more than 1,200 individuals from 20 Bantu-speaking agriculturalists populations and 5 Pygmy hunter-gatherers groups (most of them being also Bantu-speakers) from Gabon and Cameroon. Our main goals were (i) to determine the mtDNA variation in this region and compare it to the global African landscape; (ii) to evaluate the extent to which the transition from hunter-gathering to agriculturalism has influenced human genetic diversity and (iv) to contrast the genetic data obtained in the Bantu-speaking groups with the detailed linguistic classifications obtained, more specifically, in the Ogooué-Ivindo region of Gabon. All individuals were analyzed for the mtDNA hypervariable sequence I (HVS-I) as well as for 25 coding-SNPs chosen on the basis of complete mtDNA African sequences. Overall, Pygmy populations exhibited low levels of diversity, while Bantu agriculturalists showed very high levels of both sequence and haplogroup diversity. Considering the 25 populations as separate groups, the amount of variation due to inter-population diversity turned out to be very low (~1%), indicating that virtually all variation is explained by intra-population diversity. However, when grouping hunter-gatherers and agriculturalists into two separate groups, the between-group diversity reached 16%, indicating strong differentiation of these groups with different lifestyles. Standard neutrality tests suggested that the diversity observed in all Pygmy groups is compatible with constant population sizes, while most agriculturalist populations exhibited signals of populations' expansions. At the phylogeographic level, our results showed that both Pygmies and Bantus share common ancestry, as attested by the shared presence of the L1c lineage. This lineage reach almost fixation in all groups of western Pygmies here studied, mainly in the form of L1c1a and L1c1a1. As to the Bantu-speaking agriculturalists, they also presented high frequencies of the L1c haplogroup (40-60%) but internal variation of this lineage was much higher than in Pygmies, with the three main subclades of L1c (L1c1, L1c2 and L1c3) present in most populations. The fact that L1c exhibits both the highest frequency and internal diversity in this region as compared to the entire African continent point to a central African origin for this lineage. Further, our results suggest that, while gene flow in the Pygmy->Bantu direction may have occurred, gene flow in the Bantu->Pygmy direction has been unlikely on the basis of maternal variation. Lastly, our data have also shed light on the linguistic classification of the different Bantu-speaking groups of Gabon. Detailed linguistic comparisons have evinced two

major linguistic clusters: a “northern” cluster, comprising the B10 (MYENE) and B30 (TSOGO) groups and a “southern” cluster, comprising the B40, B50, B60 and B70 groups. Our genetic data strongly support the linguistic grouping of the MYENE-TSOGO cluster, while the “southern” cluster seems to be less supported. The latter observation could be due to long-term linguistic convergence as a result of prolonged contact with few genetic exchanges among these populations.