

https://dspace.org/

Genetic variation in Khoisan-speaking populations from southern Africa

Schlebusch, Carina Maria

2011-02-01
http://hdl.handle.net/10539/8992

Downloaded from a DSpace repository. https://dspace.org/

ABSTRACT

The San and Khoe people currently represent remnant groups of a much larger and widely distributed population of hunter gatherers and cattle herders, respectively, who had exclusive occupation of southern Africa before the arrival of Bantu-speaking groups in the past 1,200 years and sea-borne immigrants within the last 350 years. This project made use of mitochondrial DNA (mtDNA), Y-chromosome DNA and autosomal DNA markers to examine the population structure of various San and Khoe groups and to reconstruct their prehistory. The groups included in the study consists of six different Khoe-San groups (‡Khomani, Nama, Khwe, !Xun, /Gui + //Gana + Kgalagari and Ju\'hoansi), four different Coloured groups and five other population groups that were included in the comparative analysis.

For the mtDNA study a minisequencing technique was successfully developed which allowed the assignment of mtDNA lineages into the 10 global mtDNA macro-haplogroups. Haplogroups were further resolved using control region sequence data obtained from both hypervariable regions (HVR I and HVR II). Using this approach 538 individuals (both males and females) were screened and their mtDNA types were resolved into 18 haplogroups encompassing 245 unique haplotypes. In addition, 353 males were examined for Y-chromosome DNA variation using 46 bi-allelic Y-chromosome markers and 12 Y-STR markers. The Y-chromosomes in the sample were assigned into 29 haplogroups (using bi-allelic variation) following the nomenclature initially recommended by the Y-chromosome Consortium and resolved into 268 unique haplotypes (Y-STR variation). To assess the level of autosomal variation, 220 genome wide autosomal SNPs were typed in 352 individuals. These SNPs were combined in different datasets and analysed using two different approaches allowing for genotype and haplotype analyses. Data from these three marker systems were analysed using different analytical methods (distance based phylogenetic analysis, network analysis, dating of lineages, principal components analysis, phylogeographic analysis, AMOVA analysis, population structure analysis, and population genetic summary statistics) to asses the ancestral associations and the genetic affinities of the various San, Khoe and Coloured populations.

The most striking observation from this study was the high frequencies of the oldest mtDNA haplogroups (L0d and L0k) and Y-chromosome haplogroups (haplogroups A and B) found in Khoe-San and Coloured groups. The sub-haplogroups were, however, differentially distributed in the different Khoe-San and Coloured groups which suggested different demographic histories.

The current distribution of Khoe-San groups comprises a wide geographic region extending from southern Angola in the north to the Cape Province (South Africa) in the south. Linguistically Khoe-San groups are also divided into northern Khoisan-speaking groups (Ju division) and southern Khoisan-speaking groups (Tuu division) with an additional linguistic group (Khoe) associated with some Khoe-speaking San groups in Botswana and the Khoe herders of South Africa and Namibia (such as the Nama). For all three genetic marker systems, northern groups (Ju speaking - !Xun, Ju\'hoansi and Khoe-speaking San - /Gui + //Gana) grouped into one cluster and southern groups (historically Tuu speaking - ‡Khomani and Coloured groups) grouped into a second cluster with the Khoe group (Nama) clustering with the southern Khoe-San and Coloured groups.

The Khwe genetic profile was very different from the other Khoe-San groups. Although high proportions of Bantu-speaking admixture were identified in the Khwe group, they also contained a unique distribution of other mtDNA and Y-chromosome lineages. A previously published theory suggested that, based on the presence of a specific E-M35 Y-chromosome haplotype, the Khwe might be descendants of an east African pastoralist group that introduced the pastoralist culture to a region located in the present day northern Botswana. This pattern also mirrors what archaeologists have found with respect to the introduction of pastoralism to southern Africa. The theory was further supported and elaborated on in the present thesis. Considering the frequency and distribution of E-M35, the highest frequency (46%) was found in the Khwe with a present-day distribution in northern Botswana and southern Angola while a decrease in frequency is observed towards the south with low frequencies (<10%) in the Karoo Coloured groups. Conversely, none of the mtDNA (female) L0k and L0d lineages observed in the Khwe group were observed in the southern Khoe-San and Coloured groups. From these observations a theory was proposed that after introduction into the region of northern Botswana, the

southwards spread of pastoralism was not a clear-cut demic or cultural diffusion. Rather some male individuals integrated with the southern tribes and took with them the pastoralist practice and likely also their Khoe-language.

Altogether this thesis presented new insights into the multifaceted demographic history that shaped the existing genetic landscape of the Khoe-San and Coloured populations of southern Africa.