Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages

Liane Fendt[1], Gabriela Huber[1], Alexander W. Röck[1], Bettina Zimmermann[1], Martin Bodner[1], Rhena Delport[2], Konrad Schmidt[3], Walther Parson[1*]

[1]Institute of Legal Medicine, Innsbruck Medical University, 6020 Innsbruck, Austria

[2]Department of Family Medicine, School of Medicine, University of Pretoria, Pretoria 0028, South Africa

[3]Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, 6020 Innsbruck, Austria

[*]Correspondence: walther.parson@i-med.ac.at; Tel. +43 512 9003 70640, Fax +43 512 9003 73640

Keywords:

Mitochondrial DNA, Africa, Khoe-San, Haplogroup, EMPOP

Abstract

Here we provide 129 complete mitochondrial control region sequences of indigenous Khoe-San individuals from Angola to contribute to the still underrepresented pool of data from Africa. The dataset consists of exclusively African lineages with a majority of Sub-Saharan haplogroups. The probability of a random match was calculated as 0.09. The data set comprises 21 haplotypes occurring more than once and 17 unique haplotypes. Upon publication, haplotypes were incorporated in the EMPOP database (www.empop.org; EMP00069) [1].

Population

Southern Africa's indigenous herding Khoe and hunter-gathering San are together classified under the collective term Khoe-San. Both groups display the so-called click consonants as common linguistic feature [2 - 6]. During South Africa's military involvement in the Namibian War of Independence from 1966 to 1988 the South African forces also recruited indigenous people who originated from Angola. Those who fought on the side of the South African Defense Force were moved to Schmidtsdrift in the Northern Cape Province in 1990 after the fights [7]. The Schmidtsdrift community in South Africa now comprises mainly Ju-speaking !Xu (Kung) and Khoe-speaking Khwe that are reported to have intermarried to some degree [8]. Before their movement from Angola to South Africa, the !Xu subsisted as stock farmers in central Angola whereas the Khwe lived as cultivators in the southeast region of the country [9, 10]. The !Xu and the Khwe trace their ancestral lineage back to the San people that once inhabited South Africa.

Only few Angolan mtDNA sequences were available in the literature, which were typed for the first and second hypervariable segment of the control region [11-13]. The blood samples for this study came from 129 healthy randomly drawn volunteer donors of to our knowledge unrelated individuals collected in the Schmidtsdrift community. Individuals identified themselves as !Xu (113) and Khwe (13). Ethnicities of three individuals remained unknown. Informed written/oral consent was obtained from all human subjects. Ethic approval was obtained by the Hans Snykers Institute for the collection of samples for both biochemical and genetic studies on the Bushmen from the Faculty of Medicine, University of Pretoria Ethics Committee.

DNA extraction, amplification and sequencing

Genomic DNA was extracted from peripheral blood as described in [14]. Full mitochondrial control region (CR) was amplified, sequenced and interpreted as reported in [15].

Data analysis

According to our in-house data quality management process, the resulting consensus sequences were inspected independently by two different analysts using the sequence analysis software Sequencher (Version 4.8) and reviewed by a third scientist. Consensus sequences covered a common reading frame from position 16024 to 576 and were reported as differences to the rCRS [16] following updated nomenclature guidelines for mtDNA [17]. Haplogroups were assigned according to Phylotree, build 12 [18]. Within our Khoe-San sample set the random match probability was calculated as the sum of squared CR frequencies, disregarding length variants at positions 16193, 309 and 573. The haplotypes from this study will be available on the EMPOP database (www.empop.org) upon publication (EMPOP accession number EMP00069).

Results

Observed CR haplotypes and diversity indices

The entire control region analysis revealed exclusively typical Africa-specific L lineages with a majority of Sub-Saharan origin (**Table S1)**. Within our dataset, we found 90.7% L0 haplotypes, hence it is remarkable that 97.4% of all L0 haplotypes (or 88.4% of the observed Khoe-San dataset) belonged to the Khoe-San specific L0d/L0k cluster which is reported to account for ~ 60% in observed Khoe-San populations [19]. Among these, the most frequent haplogroups within the Angolan Khoe-San samples were L0d1c1 (33.3%) as well as L0k1 (27.1%). However, we identified a minor contribution of 9.3% of haplogroups L2 and L3 within the Khoe-San. The absence of lineages L1, L5, L6, and L4 in our dataset corresponds to their documented main dispersals within North/East –Africa (L5, L6, L4) and Central to Southeast Africa (L1) [20-22].

The most frequent haplotypes within our Khoe-San dataset were 73G-146C-152C-195C-198T-247A-315.1C-456T-498DEL-523DEL-524DEL-16167T-16187T-16189C-16223T-

16230G-16234T-16242T-16243C-16311C-16519C (haplogroup L0d1c1) with 29 occurrences, 73G-146C-152C-189G-195C-198T-207A-247A-315.1C-523DEL-524DEL-16166C-16172C-16187T-16189C-16209C-16214T-16223T-16230G-16278T-16291G-16311C-16519C (haplogroup L0k1) with 18 occurrences and 73G-146C-152C-195C-198T-247A-309.1C-315.1C-456T-498DEL-523DEL-524DEL-16167T-16187T-16189C-16223T-16230G-16234T-16242T-16243C-16311C-16497G-16519C (halpogroup L0d1c1) with 10 occurrences.

The intrapopulation probability for a random match (RMP) was calculated 0.09 for the entire control region when disregarding length variants at positions 16193, 309 and 573. Among the 129 samples we found 38 haplotypes whereof 21 occurred more than once and 17 were unique.

Point heteroplasmy

We found seven point heteroplasmic transitions in a total of six samples (499R, 16284R twice, 204Y three times) concurring with heteroplasmic sites observed in more than 5,000 global population samples [23] whereas one instance of heteroplasmy (16327Y) has not been reported there.

Acknowledgements

The authors would like to thank Harald Niederstätter and Anita Kloss-Brandstätter for technical assistance and useful discussions.

Reference List

1. W. Parson, A. Dür, EMPOP - a forensic mtDNA database, Forensic Sci. Int. Genet. 1 (2007) 88-92.

2. J. Greenberg, The languages of Africa, Int. J. Am. Linguistics 29 (1963) 1: part II.

3. T. Güldemann, M. Stoneking, A historical appraisal of clicks: a linguistic and genetic population perspective, Ann. Rev. Anthropol. 37 (2008) 93-109.

4. C. König, Khoisan languages, Language and Linguistics Compass 2 (2008) 996-1012.

5. C. Ehret, African languages: an introduction, in: B. Heine, D. Nurse (Eds.), Language and history, Cambridge University Press, 2000, pp. 272-297.

6. A. Mountain, The first people of the cape, David Philip Claremont, South Africa, 2003

7. J. Sharp, S. Douglas, Prisoners of their reputation? The veterans of the "Bushmen" battalions in South Africa, in: P. Skotnes, (Ed.), Miscast: Negotiating the presence of the Bushmen, Cape Town: University of Cape Town Press, 1998, pp. 323-330.

8. P. Mitchell, Genetics and southern African prehistory: an archaeological view, J. Anthropol. Sci. 88 (2010) 73-92.

9. J. Rocha, Bantu-Khoisan interactions at the edge of the Bantu expansions: insights from southern Angola, J. Anthropol. Sci. 88 (2010) 5-8.

10. A. Barnard, Kinship, language and production: a conjectural history of Khoisan social structure, Africa: J. Int. African Inst. 58 (1988) 29-50.

11. A Beleza, L Gusmão, A Amorim, Á Carracedo, A Salas, The genetic legacy of western Bantu migrations, Hum Genet 117 (2005) 366-375.

12. S. Plaza, A. Salas, F. Calafell, F. Corte-Real, J. Bertranpetit, Á. Carracedo, D. Comas, Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola, Hum Genet 115 (2004) 439-447.

13. M. Coelho, F. Sequeira, D. Luiselli, S. Beleza, J. Rocha, On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola, BMC Evo Biol 9 (2009) 80.

14. S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, Nucleic Acids Res. 16 (1988) 1215.

15. W. Parson, H.-J. Bandelt, Extended guidelines of mtDNA typing of population data in forensic science, Forensic Sci. Int. Genet. 1 (2007) 13–19.

16. R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.

17. H.-J. Bandelt, W. Parson, Consistent treatment of length variants in the human mtDNA control region: a reappraisal, Int. J. Legal Med. 122 (2007) 11–21.

18. M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 1039 (2008) e386-e394.

19. D.M. Behar, R. Villems, H. Soodyall, J. Blue-Smith, L. Pereira, E. Metspalu, R. Scozzari, H. Makkan, S. Tzur, D. Comas, J. Bertranpetit, L. Quintana-Murci, C. Tyler-Smith, R.S. Wells, S. Rosset and The Genographic Consortium. The dawn of human matrilineal diversity, Am. J. Hum. Genet. 82 (2008) 1130-1140.

20. A. Salas, M. Richards, T. De la Fe, M.V. Lareu, B. Sobrino, P. Sánchez-Diz, V. Macaulay, Á. Carracedo, The making of the African mtDNA landscape, Am. J. Hum. Genet. 71 (2002) 1082–1111.

21. L. Quintana-Murci, C. Harmant, H. Quach, O. Balanovsky, V. Zaporozhchenko, C. Bormans, P.D. van Helden, E.G. Hoal, D.M. Behar, Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture, Am. J. Hum. Genet. 86 (2010) 611-620.

22. L. Vigilant, R. Pennington, H. Harpending, T.D. Kocher, A.C. Wilson, Mitochondrial DNA sequences in single hairs from a southern African population, Proc. Natl. Acad. Sci. USA 86 (1989) 9350-9354.

23. J.A. Irwin, J.L. Saunier, H. Niederstätter, K.M. Strouss, K.A. Sturk, T.M. Diegoli, A. Brandstätter, W. Parson, T.J. Parsons, Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples, J. Mol. Evol. 68 (2009) 516–527.