COMMENTARY

Language isolates and their genetic identity: a commentary on mitochondrial DNA history of Sri Lankan ethnic people: their relations within the island and with the Indian subcontinental populations

Gyaneshwer Chaubey

Journal of Human Genetics (2014) 59, 61–63; doi:10.1038/jhg.2013.122; published online 21 November 2013

South Asia is the home to more than a fifth of the world's population, and is thought, on genetic grounds, to have been the first main reservoir in the dispersal of modern humans Out of Africa.^{1,2} Additionally, high level of endogamy within and between various castes, along with the influence of several evolutionary forces and long-term effective population size, facilitate the formation of complex demographic history of the subcontinent.³ Therefore, the ancestry of peopling of the South Asia is a question of fundamental importance in archaeogenetics, linguistics and historical disciplines, and it is not surprising that the number and timing of migrations in and out of South Asia is still vigorously debated.^{2–4} Researches from various disciplines focused on testing the hypothesis that several separate migrations entered to the subcontinent with each migration being associated with different tool technology, linguistic and genetic characteristics.^{2,3,5,6} The mtDNA (mitochondrial DNA) data suggest deep autochthonous diversity with minor sharing with East and West Eurasians,³ whereas, in contrast with this, the recent autosomal data showed substantial similarities of their genome with Caucasus and West Asians.^{4,6} However, at the current resolution, it is unclear that this sharing is extremely ancient or arisen with the arrival of new languages and farming.⁷

Among South Asian countries, India and Pakistan have somewhat better

understanding of their genetic structuring, whereas genetic information from Nepal, Bangladesh, Bhutan, Sri Lanka and the Maldives are either published at the level of forensic data or restricted to few populations. Being at the offshoot of southernmost tip of South Asia and along the proposed southern migration route, the island of Sri Lanka has long been settled by various ethnic groups and may offer a unique insight into initial peopling of the subcontinent. Most importantly, it is one and only precursor of modern humans fossil in South Asia dated back to 37 000 years ago.^{8,9} Therefore, it is important to study prehistoric human settlement of Sri Lanka, and their relation to other adjoining populations.

Though each and every population of South Asia is unique in terms of their genetic structuring, dialect and specific rituals, there are few language isolates who have hypothesized to be the remnants of the ancient settlers and may provide an insight to the initial peopling of the subcontinent.¹⁰ Hence, it is important to study the genome structure of these groups. However, it is also important to note that the language isolates may not be always the genetic isolates. Except from the Andaman islanders, none of the studied South Asian language isolates hitherto were found to be true genetic isolate.^{1,11,12} The linguistic isolate of Sri Lanka is known as Vedda (aka Vadda).^{8,10} Vedda is small hunter-gatherer tribe existing in Northwest province of Sri Lanka. They are called as aboriginal people of Sri Lanka and were suggested to represent the indigenous population of the entire subcontinent.^{8,10}

The first genetic study of Vedda along with other Asian populations suggested their long period of isolation.¹³ However, the analysis of alpha-2-HS-glycoprotein allele frequencies supports the view that the Veddas are biologically most closely related to the Sinhalese.¹⁴ Till date, a high-resolution genetic data was not available from this population and their affinity with other populations of Eurasia remained obscure.

In previous issue, Ranaweera and colleagues⁹ rectified this issue by generating the novel informations from the HVS-I and HVS-II region of mtDNA to date for the Vedda, including other major ethnic groups of Sri Lanka. Through a well-covered sampling strategy, they filled a major geographic 'white spot' inhabited by ~20 million people. Their analysis led to the more precise identification of South Asian-specific indigenous mtDNA haplogroups and a better understanding of the extent of East and West Eurasian admixture among Sri Lankan populations. In addition, for the first time by generating mtDNA data on relic Vedda populations, this study was able to confirm that they exhibit low genetic variability, which is consistent with their small population size and strong effect of genetic drift.

The haplogroup distribution and sharing among different ethnic groups is intriguing.⁹ Majority of mtDNA haplogroups were belonging to South Asian-specific clades (Figure 1). However, the Sri Lankan populations (except Indian Tamils sampled in Sri Lanka) have significantly higher West

Dr G Chaubey is at Evolutionary Biology Department, Estonian Biocentre, Tartu, Estonia
E-mail: gyanc@ebc.ee

Another important element of this study was generating the maternal haplogroup level data from one of the so far genetically virtually unstudied language isolate of South Asia known as Vedda.⁹ This study reports that the Vedda is most distinct among all the ethnic groups, which is likely due to elevated frequency of haplogroups R30, U1 and U7, altogether represent 64% of maternal lineages of Vedda. The haplotype distribution of these haplogroups clearly indicates this scenario as a result of founder effect(s) due to random genetic drift. Owing to this unique haplogroup structuring, Vedda single out as an exceptional tribal population of South Asia, having less than 30% individuals sharing haplogroup M. Majority of their individuals share a branch of haplogroup R30 (that is, R30b2), widespread mainly in the coastal region of South Asia and shows an expansion time between 8 and 24 kya (Figure 1b). Haplogroup U1 haplotypes, which is also reported from Southern India, were found among 12% of Vedda individuals. Haplogroup U7 haplotypes were shared with both North and South Indian populations. The future phylogeographic study of these haplogroups would be able to

The demographic history of a population cannot be established by studying just a single locus. Therefore, to find out the signatures of ancient as well as recent admixture to reconstruct the demographic history for the Sri Lankan populations (including linguistic isolate Vedda), along with the complete mtDNA informations, high-resolution Y-chromosomal and high coverage complete genome resequencing data are essential.

This study was supported by the EU through the European Regional Development Fund through the Centre of Excellence in Genomics to Estonian Biocentre and University of Tartu.

- 1 Thangaraj, K., Chaubey, G., Kivisild, T., Reddy, A. G., Singh, V. K., Rasalkar, A. A. *et al.* Reconstructing the origin of Andaman Islanders. *Science* **308**, 996 (2005).
- 2 Mellars, P., Gori, K. C., Carr, M., Soares, P. A. & Richards, M. B. Genetic and archaeological perspectives on the initial modern human colonization of southern Asia. *Proc. Natl Acad. Sci. USA* **110**, 10699–10704 (2013).
- 3 Chaubey, G., Metspalu, M., Kivisild, T. & Villems, R. Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays* **29**, 91–100 (2007).
- 4 Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
- 5 Chaubey, G., Metspalu, M., Choi, Y., Mägi, R., Romero, I. G., Soares, P. *et al.* Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Mol. Biol. Evol.* **28**, 1013–1024 (2011).
- 6 Metspalu, M., Romero, I. G., Yunusbayev, B., Chaubey, G., Mallick, C. B., Hudjashov, G. *et al.* Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. *Am. J. Hum. Genet.* **89**, 731–744 (2011).
- 7 Diamond, J. & Bellwood, P. Farmers and their languages: the first expansions. *Science* **300**, 597–603 (2003).
- 8 Kennedy, K. A. in *To What Extent Were Prehistoric Sri Lankans Isolated from the Indian Mainland?* *Biotic*

- and *Archaeological Considerations* (eds Abraham, S., Gullapalli, P., Raczek, T. & Rizvi, UZ) 35 (Left Coast Press, CA, 2013).
- 9 Ranaweera, L., Kaewsutthi, S., Tun, A. W., Boonyarit, H., Poolsuwan, S. & Lertrit, P. Mitochondrial DNA history of Sri Lankan ethnic people: their relations within the island and with the Indian subcontinental populations. *J. Hum. Genet.* **59**, 28–36 (2014).
 - 10 Driem, G. V. *Languages of the Himalayas* (Brill, Leiden, 2001).
 - 11 Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K. E., Rasmussen, S., Albrechtsen, A. *et al.* An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science* **334**, 94–98 (2011).
 - 12 Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
 - 13 Harihara, S., Saitou, N., Hirai, M., Gojobori, T., Park, K. S., Misawa, S. *et al.* Mitochondrial DNA polymorphism among five Asian populations. *Am. J. Hum. Genet.* **43**, 134–143 (1988).
 - 14 Umetsu, K., Yuasa, I., Yamashita, T., Saito, S., Yamaguchi, T., Ellepola, S. B. *et al.* Genetic polymorphisms of orosomucoid and alpha-2-HS-glycoprotein in Thai, Sri Lankan and Paraguayan populations. *Jinrui. Idengaku. Zasshi.* **34**, 195–202 (1989).
 - 15 Dissanayake, V. H., Samarakoon, P. S., Scaria, V., Patowary, A., Sivasubbu, S. & Gokhale, R. S. The Sri Lankan Personal Genome Project. *The Sri Lankan Personal Genome Project* **2**, 4–8 (2011).
 - 16 Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Rohl, A. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
 - 17 Behar, D. M., Metspalu, E., Kivisild, T., Rosset, S., Tzur, S., Hadid, Y. *et al.* Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS One* **3**, e2062 (2008).
 - 18 Chaubey, G., Karmin, M., Metspalu, E., Metspalu, M., Selvi-Rani, D., Singh, V. K. *et al.* Phylogeography of mtDNA haplogroup R7 in the Indian peninsula. *BMC Evol. Biol.* **8**, 227 (2008).
 - 19 Rani, D. S., Dhandapany, P. S., Nallari, P., Govindaraj, P., Singh, L. & Thangaraj, K. Mitochondrial DNA haplogroup 'R' is associated with Noonan syndrome of south India. *Mitochondrion* **10**, 166–173 (2010).