

SELECTED ORAL PRESENTATION

Open Access

The genetic structure of south Asian populations as revealed by 650 000 SNPs

Mait Metspalu^{1*}, Gyaneshwer Chaubey¹, Bayazit Yunusbayev^{1,2}, Irene Gallego Romero⁴, Monika Karmin¹, Chandana Basu Mallick¹, Ene Metspalu¹, Sadagopal Shanmugalakshmi⁶, Karuppiiah Balakrishnan⁶, Kumarasamy Thangaraj³, Lalji Singh³, Ramasamy Pitchappan⁵, Toomas Kivisild^{4,1}, Richard Villems¹

From Beyond the Genome: The true gene count, human evolution and disease genomics
Boston, MA, USA. 11-13 October 2010

The analyses of dense marker sets covering the whole genome has revolutionised the field of (human) population genetics. Driven largely by the needs of biomedical research, these new data are helping to unveil our demographic past, exemplified by the study of mtDNA and Y-chromosome variation during the past ~20 years.

We have analysed (Illumina 650K SNPs) over 320 new samples from South and Central Asia and the Caucasus, together with the publicly available databases (HGDP panel and our published data set of ~600 Eurasian samples) and illustrated the power of full genome analyses by addressing two specific questions. (i) What is the nature of genetic continuity and discontinuity between South Asia, Middle East and Central Asia? (ii) What are the genetic origins of the Munda speakers of India? We use principal component and structure-like analyses to reveal the structure in the genome wide SNP data. The most striking feature of the genetic structure of South Asian populations is the clear separation of the Indus valley and southern India populations. The genetic component prevalent in the latter region is marginal in the former and absent outside South Asia. By contrast, the component ubiquitous to Indus valley is also present (~30 - 40%) among Indo-European speakers from Ganges valley and Dravidic speakers in southern India. Furthermore, this component can also be found in Central Asia and the Caucasus as well as in Middle East. We explored possibilities to identify the source region for this genetic component.

Alternative models put the origins of Munda languages speakers either in South Asia (the Munda speakers sport exclusively autochthonous South Asian

mtDNA variants) or in Southeast Asia, where the other Austro Asiatic languages have spread. Y-chromosome variation supports the latter model through sharing of hg O2a in both regions. We show that in addition to the dominant ancestry component being shared between the Indian Dravidic and Munda speakers, up to 30% of Munda speakers retain an ancestry component otherwise prevalent in East Asia. There is no widespread sign of South Asian ancestry component in Southeast Asia. This provides genomic support to the model by which Indian Austro-Asiatic populations derive from dispersal from Southeast/East Asia, followed by an extensive admixture with local Indian populations.

Author details

¹Estonian Biocentre and Department of Evolutionary Biology, University of Tartu, Tartu 51010, Estonia. ²Institute of Biochemistry and Genetics, Ufa Research Center, Russian Academy of Sciences, Ufa 450054, Russia. ³Centre for Cellular and Molecular Biology, Hyderabad, India. ⁴Leverhulme Centre of Human Evolutionary Studies, The Henry Wellcome Building, University of Cambridge, Fitzwilliam Street, Cambridge, CB2 1QH, UK. ⁵Department of Immunology, School of Biological Sciences, Madurai Kamaraj University. ⁶School of Biotechnology, Bharathidasan University, Trichirappalli, India.

Published: 11 October 2010

doi:10.1186/gb-2010-11-S1-O8

Cite this article as: Metspalu *et al.*: The genetic structure of south Asian populations as revealed by 650 000 SNPs. *Genome Biology* 2010 11(Suppl 1):O8.

¹Estonian Biocentre and Department of Evolutionary Biology, University of Tartu, Tartu 51010, Estonia

Full list of author information is available at the end of the article