

## COMMENTARY

# Male lineages in the Himalayan foothills: A commentary on Y-chromosome haplogroup diversity in the sub-Himalayan Terai and Duars populations of East India

Qasim Ayub

*Journal of Human Genetics* (2011) 56, 813–814; doi:10.1038/jhg.2011.114; published online 6 October 2011

Of the estimated 7 billion humans, more than 1.2 billion reside in India. This mass of humanity comprises many diverse ethnicities, linguistic and religious groups. There is evidence of modern human inhabitation in the Indian sub-continent soon after their departure from Africa ~50–60 000 years before present, and the land has witnessed several subsequent human migrations and invasions that have shaped its unique cultural, social and genetic structure.

The Y chromosome has a well-established phylogeny, and the variants located on its single copy region have been extensively used to give a male-specific evolutionary perspective. Continuing with this trend in the previous issue of this journal Debnath *et al.*<sup>1</sup> examines the Y chromosomal variation in 375 male individuals belonging to 10 sub-Himalayan Indian populations and compare them with existing data sets.

The populations were sampled from the Terai-Duar savannah and grasslands in the present-day state of West Bengal (recently renamed 'Paschim Banga'). This ecoregion lies between Nepal and Bhutan, and comprises the districts of Darjeeling, Jalpaiguri and Cooch Behar. Due to the climatic conditions in this part of the world, ancient human fossils are scanty, but a number of Mesolithic settlements have been found west of this region in the Ganges plains that indicate the presence of hunter gatherers since 10 000 years before present.<sup>2</sup>

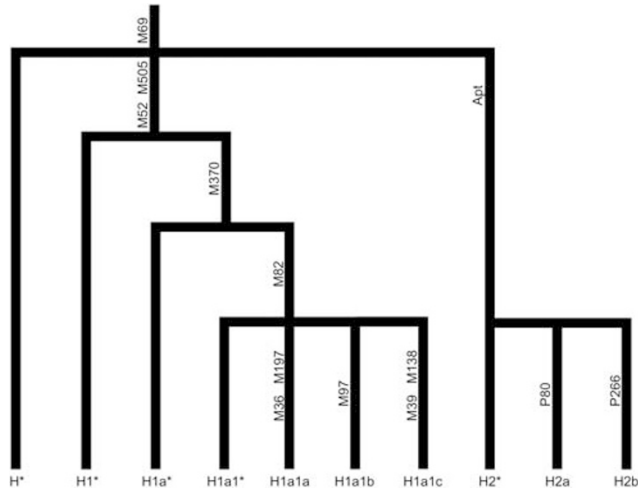
The populations that were analysed included representatives from all four linguistic groups of Indo-European, Dravidian, Tibeto-Burman and Austro-Asiatic speakers that are found in the region. They included three Mundari-speaking Austro-Asiatic, four Tibeto-Burmese and a Dravidian tribe, and two Indo-European castes. On the basis of 76 bi-allelic Y markers, the authors demonstrate extensive genetic admixture among the linguistic groups and male gene flow from neighbouring populations, and from North and Southeast Asia. They conclude that geographical proximity instead of linguistic affinity is a better predictor of genetic relationships in this region, a finding that is in agreement with previously published work on populations from South Asia.<sup>3</sup>

The current study found an extremely diverse sub-Himalayan male gene pool. Three major haplogroups (O, H1a\* and R1a1\*) were shared across all four language families and accounted for 79% of the sampled population. The Indo-European castes were the most diverse, but shared many haplogroups with the Tibeto-Burmese. The latter retained a high frequency of O3a3c1 (M117-derived) Y chromosomes that distinguishes them from their Austro-Asiatic neighbours that lack this haplogroup and are predominantly O2a (derived for P31 and M95), which has been reported previously.<sup>4</sup>

The study also provides additional support for the revised Y haplogroup H phylogenetic tree<sup>5</sup> and reports an additional single-nucleotide polymorphism on this branch (Figure 1). Thirteen mutations are now associated with 10 H haplogroups. The H1 lineage is thought

to have an indigenous origin in peninsular India and decreases in frequency outside the subcontinent.<sup>6</sup> The haplogroup frequency is ~22% in India, 12% in Nepal, 3.9% in Pakistan and <1% in Iran, Turkey, Middle East, Central and Southeast Asia. High frequency of this haplogroup in European Roma Gypsies (~17%) is taken as an evidence of their Indian ancestry. Within India, the haplogroup frequency differs widely among linguistic groups being highest in Dravidian castes and tribes and lowest among Tibeto-Burman speakers, and it is tempting to speculate that it may be a signature associated with native Dravidians. However, one must remember that these linguistic differences evolved much more recently in comparison with the observed Y haplogroup distributions, and it is highly unlikely that they would be linked to any particular Y haplogroup. A major shortcoming of the study is that no Y short tandem repeats were genotyped, and so, we can only observe patterns established by distribution of slowly evolving bi-allelic polymorphisms, which date back to the Palaeolithic period and may have reached high frequency in a given population due to genetic drift. Issues related to more recent and interesting events such as the displacement of Dravidians within the sub-continent cannot be addressed with such markers. Y short tandem repeats variance within certain lineages such as H1 (M52-derived) Y chromosomes could have been used to estimate coalescence times and provide a phylogeographic approach to origins of this and other lineages. Similarly, O3a3c1 haplotypes could have been used to address

Q Ayub is at the The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK.  
E-mail: qa1@sanger.ac.uk



**Figure 1** Revised Y chromosome haplogroup H parsimony tree based on 13 biallelic polymorphisms. The name of each polymorphism is shown along the branches, and haplogroup names are given at the tips of each branch.

lingering questions about the arrival of Austro-Asiatic and Tibeto-Burman speakers in this sub-Himalayan region, although dating these events is still fraught with uncertainty.

Insights into male diversity from resequencing complete Y chromosomes is now technological and economically feasible,<sup>7</sup> and it is hoped that such indigenous Indian samples will be made available to the wider genomic community to complement the data sets being generated by the 1000 Genomes Project,<sup>8</sup> which lack such collections but includes

five populations with South Asian ancestry (Bengalis from Bangladesh, Punjabis from Pakistan, and expatriate Sri Lankan Tamil, Indian Telugu and HapMap3 Gujarati in Houston). Genotyping the thousands of Y single-nucleotide polymorphisms generated by such studies will not only improve our resolution of the spatial and temporal distribution of the Y phylogeny, but also eventually the availability of whole genomic sequences from indigenous populations in this region will enable us to better model demographic history of South Asia and help

unravel the distinction between linguistics, castes and tribes in this part of the world.

- 1 Debnath, M., Palanichamy, M. G., Mitra, B., Jin, J.-Q., Chaudhuri, T. K. & Zhang, Y.-P. Y-chromosome haplogroup diversity in the sub-Himalayan Terai and Duars populations of East India. *J. Hum. Genet.* **56**, 765–771 (2011).
- 2 Petraglia, M. D. & Allchin, B. Human evolution and culture change in the Indian subcontinent. In: *The Evolution and History of Human Populations in South Asia: Inter-disciplinary Studies in Archaeology, Biological Anthropology, Linguistics and Genetics* (eds Petraglia M. D. and Allchin B.) 1–22 (Springer: Dordrecht, 2007).
- 3 Ayub, Q. & Tyler-Smith, C. Genetic variation in South Asia: assessing the influences of geography, language and ethnicity for understanding history and disease risk. *Brief. Funct. Genomic. Proteomic.* **8**, 395–404 (2009).
- 4 Basu, A., Mukherjee, N., Roy, S., Sengupta, S., Banerjee, S., Chakraborty, M. *et al.* Ethnic India: A genomic view, with special reference to peopling and structure. *Genome Res.* **13**, 2277–2290 (2003).
- 5 Fornarino, S., Pala, M., Battaglia, V., Maranta, R., Achilli, A., Modiano, G. *et al.* Mitochondrial and Y-chromosome diversity of the Tharus (Nepal): a reservoir of genetic variation. *BMC Evol. Biol.* **9**, 154 (2009).
- 6 Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C.-E. T. *et al.* Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am. J. Hum. Genet.* **78**, 202–220 (2006).
- 7 Xue, Y., Wang, Q., Long, Q., Ng, B. L., Swerdlow, H., Burton, J. *et al.* Human Y-chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* **19**, 1453–1457 (2009).
- 8 Altshuler, D., Durbin, R. M., Abecasis, G. R., Bentley, D., R., Chakravarti, A., Clark, A. G. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).