



Category: Miscellaneous

A genomic insight into the origin and dispersal of Austroasiatic speakers in South and Southeast Asia

Debashree Tagore¹, Farhang Aghakhanian², Rakesh Naidu², Partha P. Majumder^{1,3}, Maude E. Phipps² and Analabha Basu¹

¹National Institute of Biomedical Genomics, Kalyani 741251, INDIA

²Jeffrey Cheah School of Medicine and Health Sciences, Monash University (Malaysia), Selangor, MALAYSIA

³Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata 700035, INDIA

Presenting author: dt1@nibmg.ac.in

Abstract

India and Southeast Asia are home to diverse linguistic groups; the Austroasiatic language group being one of them. The Austroasiatic speakers live in scattered settlements in these regions. What led to such dispersed distribution over this vast geographical space is yet to be resolved. Our work is aimed at reconstructing the migration route of early Austroasiatic settlers and examines their relationship with other linguistic groups. We genotyped 511 unrelated individuals from India and Malaysia out of which 189 were Austroasiatic. The rest belonged to Indo-European, Dravidian, Tibeto-Burman and Austronesian language families. Jarawa and Onge populations from Andaman and Nicobar Islands were also included. Our genotype data was combined with that of 940 individuals from HGDP dataset. We analyzed nearly 0.3 million autosomal SNPs and found that allele frequency correlation between Malaysian Austroasiatics and Indian Tibeto-Burmans was slightly higher ($R^2 = 0.77$) than with Indian Austroasiatics ($R^2 = 0.72$). Principal Components Analysis revealed that Malaysian Austroasiatic clustered closer to Tibeto-Burman than to Indian Austroasiatic. Similar clustering pattern was obtained by fineSTRUCTURE cluster dendrogram. The ADMIXTURE analysis inferred genetic component that is modal to the Malaysian Austroasiatic, is also significantly higher amongst Tibeto-Burman than Indian Austroasiatic ($P < 2.117e-10$), indicating that genetic distance correlates better with geography than language. Studying segments which were Identity by descent between individuals belonging to two different linguistic groups; i.e. Austroasiatic and Tibeto-Burman, we found Tibeto-Burman sharing larger number of segments with Malaysian Austroasiatic, but overall smaller in size. On the other hand the segments shared between the two Austroasiatic populations (India and Malaysia) are comparatively larger in size ($P = 0.034$) but smaller in number. Our analyses indicate that Malaysian Austroasiatic and Tibeto-Burman initially split from a common ancestor. Then a small group of individuals separated from Malaysian Austroasiatic giving rise to the present day Indian Austroasiatic. Treemix and D-statistics analysis provided evidence for gene flow between Malaysian Austroasiatic and Tibeto-Burman post split. Meanwhile, the southward migration of East Asians resulted in an extensive genetic exchange between East Asians and Tibeto-Burman as was evident in our ADMIXTURE analysis. This subsequent genetic exchange might have shaped the present day language structure.

Citation: Tagore, D., Aghakhanian, F., Naidu, R., Majumder, P.P., Phipps, M.E. and Basu, A. A genomic insight into the origin and dispersal of Austroasiatic speakers in South and Southeast Asia [Abstract]. In: Abstracts of the NGBT conference; Oct 02-04, 2017; Bhubaneswar, Odisha, India: Can J biotech, Volume 1, Special Issue, Page 138. <https://doi.org/10.24870/cjb.2017-a124>