High resolution phylogeographic study of mtDNA macrohaplogroup M in South Asia.

(progress report)

[1]Mait Metspalu, [1]Gyaneshwer Chaubey, [1]Erik Prank, [1]Anu Solnik, [1]Monika Karmin, [1]Ene Metspalu, [1]Georgi Hudjasov, [1]Eva-Liis Loogväli, [2]Doron Behar, [3]Juan J Sanchez, [4]Phillip Endicott, [5]Sadagopal Shanmugalakshmi, [5]Karuppiah Balakrishnan, [6]Ramasamy Pitchappan, [7]Toomas Kivisild, [1]Richard Villems

1Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of Tartu and Estonian Biocentre, Riia 23, Tartu, Estonia, 2Molecular Medicine Laboratory, Rambam Health Care Campus, Haifa, Israel, 3National Institute of Toxicology and Forensic Science, Canary Islands Delegation, Campus de Ciencias de la Salud, 38320 La Laguna, Tenerife, Spain, 4Henry Wellcome Ancient Biomolecules Centre, Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom, 5School of Biotechnology, Bharathidasan University, Trichirappalli, Tiruchirappalli 620 024, India, 6Department of Immunology, School of Biological Sciences, Madurai Kamaraj University, Madurai, 625 021, India, 7Leverhulme Centre of Human Evolutionary Studies, The Henry Wellcome Building, University of Cambridge, Fitzwilliam Street, Cambridge, CB2 1QH, United Kingdom

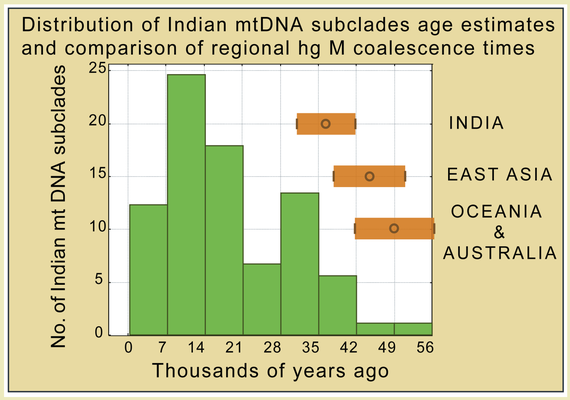


Figure 1a. The overall younger coalescence estimate for the South-Asian hg M lineages relative to East Asia and Oceania is confirmed also in this study. The distribution of the coalescence estimates (based on the synonymous clock) for all individual South Asian hg M & N subclades reveals a peak around 31KYA and an event of few coalescence events around the LGM some 25KYA .

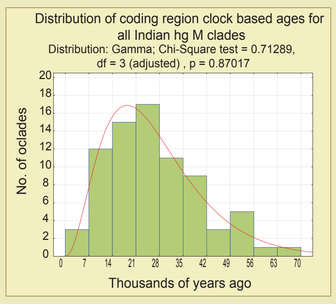
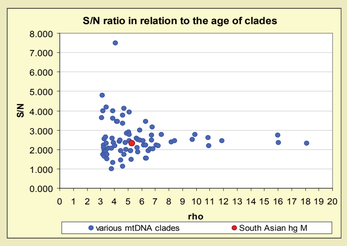


Figure 1b. However, coding region "all sites" clock based age estimates for all South Asian hg M clades show a distribution not different from gamma distribution.



We have sequenced 75 human mitochondrial genomes that reveal new subclades within the South Asian domain of haplogroup M.

We analysed the distribution of coalescence events within the South Asian mtDNA phylogeny.

We developed a Snapshot genotyping panel to investigate the phylogeography of the main branches of the SA hg M phylogeny. We were able to simultaneously genotype for 26 polymorphisms in altogether 2300 samples from different geographic regions of India. The Snapshot genotyped positions are shown in red in Figure 3. The resulting South Asian mtDNA haplogroups spatial frequency distribution data is shown on Figure 2 and 4.

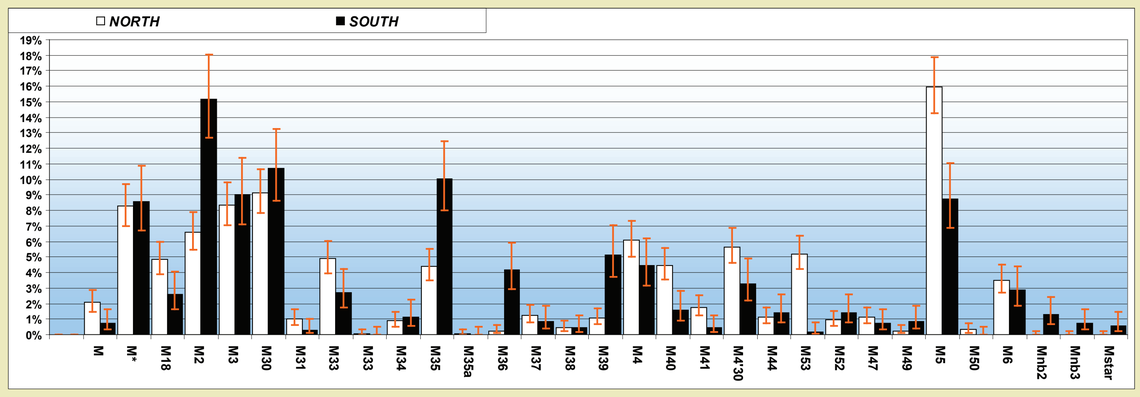


Figure 4. While majority of South Asian-specific hg m subclades do not show significant frequency difference along the north to south axis of the subcontinent, there are a few marked exceptions. Hgs M2 and M5 are the most frequent single haplogroups in southern and northern India, respectively. The spreads of neither haplogroup are however restricted to either region. This is characteristic to most haplogroups that show frequency differences between north and south of the subcontinent. However, we have found two basal branches of hg M that albeit at low overall frequency do show up only in the south of the subcontinent.

Figure 5. One of the possible reasons for the younger age estimates for the South Asian hg M lineages as compared to these of the other continent-specific founders might be more constant and larger population size in South Asia throughout prehistory. Under these circumstances one could expect that natural selection would have more power to eliminate the slightly deleterious non-synonymous substitutions. However, the S/N ratio in the Indian hg M phylogeny (Fig. 3) lies right in the middle of the S/N variation of worldwide mtDNA clades.

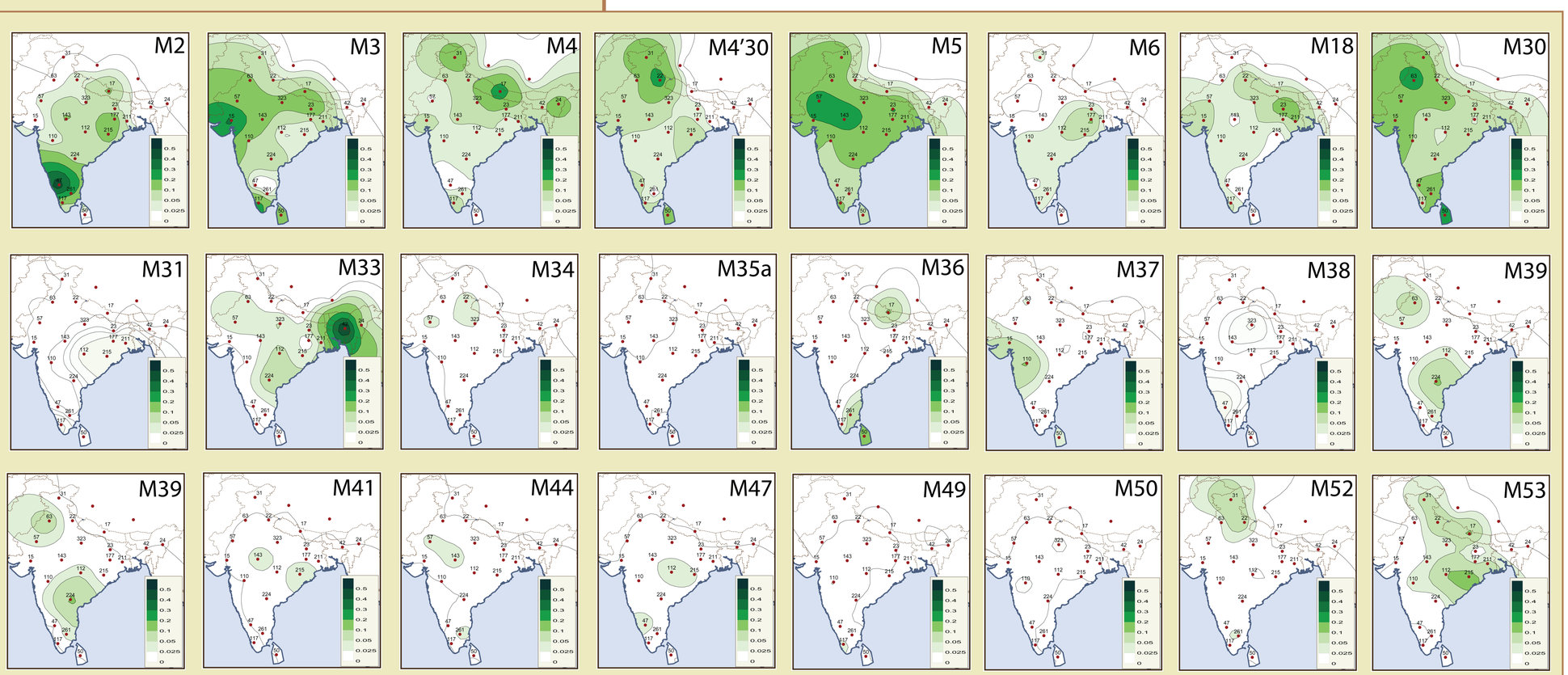


Figure 2. Spatial frequency distributions of the different South Asian- specific hg M subclades.

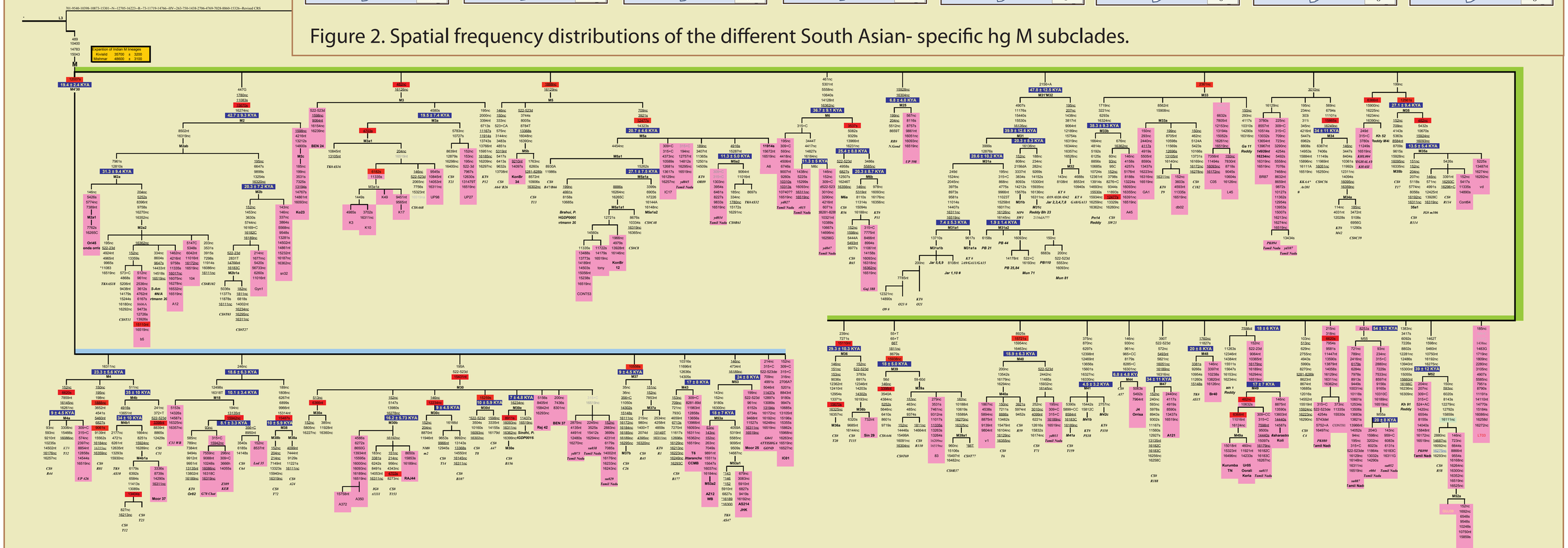


Figure 3. Phylogenetic reconstruction of 189 complete mitochondrial genomes belonging to South Asian specific subclades of haplogroup M. The 75 newly sequenced genomes are shown in pink. In comparison with the phylogenies of East Asian and West Eurasian- specific mtDNA lineages the most striking feature of the South Asian mtDNA tree is the relatively less nested structure manifested in the high number of lineages rooted directly in the pan-Eurasian founder haplotype M. As argued by other studies less severe bottlenecks during the demographic past of South Asia might explain this phenomenon. We note also that the complete mitochondrial sequences from until now somewhat underrepresented Tamil-Nadu state tend to fall outside the established set of South Asian hg M lineages. Instead they form new basally rooted haplogroups like for example M55 and M49.