

Reconstructing Indian Population History

India has been underrepresented in genomewide studies of human variation. Here we report the genotyping of 25 ethno-linguistically diverse Indian populations at hundreds of thousands of genetic polymorphisms. We provide strong evidence for two ancient populations, genetically divergent, which are ancestral to many Indian groups today. One, the "Ancestral North Indians" (ANI), is genetically close to Middle Easterners, Central Asians, and Europeans, while the other, the "Ancestral South Indians" (ASI), is not close to any modern population outside India and the subcontinent. By introducing methods that can estimate ancestry without accurate ancestral populations, we show that ANI ancestry in India ranges from 39-71%, and is significantly higher in upper caste and Indo-European speakers. Populations with only ASI ancestry may no longer exist in mainland India. However, we show that the indigenous Andamanese are unique in being an ASI-related population without ANI-related ancestry, indicating that the peopling of the islands by modern humans must have occurred before ANI-ASI gene flow on the mainland. Allele frequency differences between caste populations in India are much larger than population differences in Europe, reflecting founder effects with little subsequent gene flow. Based on this observation, we predict that there will be an excess of recessive diseases in India, different in each population, which should be possible to screen genetically.