Ancient Indian roots?

A study published in the February issue of *The American Journal of Human Genetics* (Sengupta *et al* 2006) provides the most thorough investigation to date of the genetic origins of the Indian people. It suggests that the origins are largely indigenous, dating back to the Palaeolithic period, and that the genetic influence of the Indo-European-speakers who established the caste system was small. How have they come to these conclusions?

We humans seem to have an insatiable curiosity about our origins. Surprisingly, perhaps, we know quite a lot about the distant past: populations all over the world originated in Africa, with those outside Africa being largely if not entirely derived from the exodus of a small number of Africans around 50,000-60,000 years ago (Jobling *et al* 2004). We also know a lot about the last few centuries from historical records. But we know much less about the period in between. And a lot happened: the climate fluctuated enormously, including a glacial maximum ("ice age") around 20,000 years ago, while after ~10,000 years ago the way of life of most people changed from obtaining their food by hunting and gathering to producing it by growing crops or herding animals. Historical records from the last few thousand years document migrations, wars and plagues.

In the Indian subcontinent, the Palaeolithic archaeological record is sparse, but the Neolithic period left abundant traces in the Indus Valley from ~6,500 BC onwards, culminating in the Harappan Civilization ~2,600–1,500 BC (Kenoyer 1998). Towards the end of this time, Indo-European speakers appeared and established the caste system that long dominated much of Indian society. The cultural importance of these transitions is undisputed, but did they involve large-scale replacement of the existing people by outsiders, bringing first Neolithic practices and perhaps Dravidian languages from the Fertile Crescent, and later Indo-European languages, iron and the horse from Central Asia? Or did the people already in India change their habits and languages? Studies of genes can now help to answer these questions, and the Y chromosome, which is passed from father to son, provides a particularly clear, although male-specific, perspective.

The first such genetic investigations were interpreted as suggesting extensive movements into India (Quintana-Murci *et al* 2001), and, more specifically, a Central Asian origin for caste, but not tribal, groups (Bamshad *et al* 2001; Basu *et al* 2003; Cordaux *et al* 2004). Other studies, however, suggested that both tribal and caste populations could be derived largely from the earliest Palaeolithic settlers (Kivisild *et al* 2003). So do castes and tribes share a common origin? Where did the Dravidian language arise? Did the migration of Central Asians to India that had such a strong linguistic and social impact leave a genetic footprint as well?

Sengupta *et al* (2006) have now reinvestigated these questions using a large new dataset. They analysed 728 Indian samples (17 tribes and 19 castes from different geographical regions and social/linguistic categories) with 71 slowly-evolving Y-chromosomal binary markers and 10 faster-evolvingY-microsatellites. For comparison, they also typed 176 Pakistani and 175 East Asian males with the same markers. They then used a phylogeographic approach that took into account both the geographical distribution of the main lineages and the variation within them to infer the direction and timing of their spread.

They conclude that, while some lineages did move in from the outside, many of the major ones are likely to have arisen within India. Four lineages (J2a-M410, J2b-M12, O2a-M95 and O3e-M134) probably came from elsewhere, but five major ones (C5-M356, F*-M89, H-M69, L1-M76 and R2-M124) were likely to have originated within the subcontinent, and the origins of one common lineage (R1a1-M17) could not be determined. Furthermore, they calculated from the diversity within the lineages that most of them had been around for 9,000–20,000 years or so.

The O lineages are largely restricted to a few tribal populations, and Sengupta *et al*. agree with previous suggestions that they probably came from East Asia; the J2 lineages may have come from West Asia. The indigenous lineages, however, are widespread and common in both tribal and caste populations, implying that tribes and castes have a shared origin, and that the genetic roots of most Indians must trace back to India some 9,000 years ago or more. Consequently, the Indo-European-speakers entering ~3,600 years ago can

have contributed few Y-chromosomal lineages. In addition, Sengupta *et al* (2006) suggest an origin for Dravidian speakers in peninsular India rather than the Indus Valley or outside the subcontinent.

How can you determine when and where a lineage originated? And how does the origin and spread of a lineage relate to what we think of as the origin of a population? These are rather contentious issues. According to the simplest way of thinking, current high frequency and high diversity may mark the place of origin of a lineage; but high frequency can also arise by genetic drift, and high diversity by admixture. The time can be calculated in several ways, and a wide range of mutation rates can be used, so molecular dates are much less certain than archaeological ones. In thinking about the second question, we can paraphrase the Italian geneticist Guido Barbujani: imagine that at some time in the future Indian astronauts colonise Mars, and geneticists then type their Y chromosomes. We may well find that their lineages date back to 9,000–20,000 years ago. But we would not be wise to infer that they have been living on Mars for 9,000 years.

These difficulties apply to all studies of this kind, and Sengupta *et al* (2006) take more trouble than previous authors in inferring the starting points of their lineages, so their conclusions deserve serious consideration. It would, however, have been interesting to have seen a wider range of methods applied to the estimation of the dates, and to have a better feel for their uncertainties. And identifying the geographical travels of these lineages over that last 9,000 years remains a formidable problem.

The genetic data are only one element in our complete understanding of the complex peopling of India. Geneticists have now suggested a new model, and it needs to be tested using archaeological, linguistic and historical data. Debates about the Palaeolithic or Neolithic origins of the European gene pool, or the timing of the entry to the Americas, have raged for decades, and this one is probably not over yet.

References

Bamshad M, Kivisild T, Watkins W S, Dixon M E, Ricker C E, Rao B B, Naidu J M, Prasad B V, Reddy P G, Rasanayagam A, Papiha S S, Villems R, Redd A J, Hammer M F, Nguyen S V, Carroll M L, Batzer M A and Jorde L B 2001 Genetic evidence on the origins of Indian caste populations; *Genome Res.* **11** 994–1004

Basu A, Mukherjee N, Roy S, Sengupta S, Banerjee S, Chakraborty M, Dey B, Roy M, Roy B, Bhattacharyya N P, Roychoudhury S and Majumder P P 2003 Ethnic India: a genomic view, with special reference to peopling and structure; *Genome Res.* **13** 2277–2290

Cordaux R, Aunger R, Bentley G, Nasidze I, Sirajuddin S M and Stoneking M 2004 Independent origins of Indian caste and tribal paternal lineages; *Curr. Biol.* **14** 231–235

Jobling M A, Hurles M E and Tyler-Smith C 2004 *Human evolutionary genetics* (New York, Abingdon: Garland Science)

Kenoyer J M 1998 *Ancient cities of the Indus valley civilization* (Karachi: Oxford University Press)

Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk H V, Stepanov V, Golge M, Usanga E, Papiha S S, Cinnioglu C, King R, Cavalli-Sforza L, Underhill P A and Villems R 2003 The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations; *Am. J. Hum. Genet.* **72** 313–332

Quintana-Murci L, Krausz C, Zerjal T, Sayar S H, Hammer M F, Mehdi S Q, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling M A, Tyler-Smith C and McElreavey K 2001 Y-chromosome lineages trace diffusion of people and languages in southwestern Asia; *Am. J. Hum. Genet.* **68** 537–542

Sengupta S, Zhivotovsky L A, King R, Mehdi S Q, Edmonds C A, Chow C E, Lin A A, Mitra M, Sil S K, Ramesh A, Usha Rani M V, Thakur C M, Cavalli-Sforza L L, Majumder P P and Underhill P A 2006 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** 202–221

Denise R Carvalho-Silva
Tatiana Zerjal
Chris Tyler-Smith*
The Wellcome Trust Sanger Institute,
Wellcome Trust Genome Campus,
Hinxton, Cambs. CB10 1SA, UK
**(Email, cts@sanger.ac.uk)*

*e*Publication: 20 February 2006