Our Footprints on the Sands of Time

Partha P Majumder and D Balasubramanian

Partha P Majumder is a Professor in the Human Genetics Unit of the Indian Statistical Institute.

D Balasubramanian is Director of Research at the L V Prasad Eye Institute, Banjara Hills, Hyderabad.

Keywords
Ancestry of the Indian genome, mitochondrial DNA and Y chromosome DNA analysis, peopling of India, haplotype analysis.

Introduction

Who are we? Where did we come from? When did we get here? How many communities are we? How are we all related? These questions have been asked by many groups of people, in many places and at different times. Each group has had its own answers – in the form of legends, oracles, divine revelations and collective memories. Some tribes in ancient Middle East believed themselves to be the chosen people of God. The legend of the Great Flood occurs in Judeo-Christian as well as in Hindu traditions. Post-Rigvedic accounts talk of the Primeval Universal Man (*Prajapati*) who created all humanity out of himself, and the four castes out of his different limbs. How historically accurate and verifiable are these oral and written traditions?

Anthropology attempts to look at these issues in a rigorous manner. It deals with the origins, physical and cultural development, racial and ethnic characteristics, as well as social customs and beliefs of mankind. It covers a vast canvas, and uses a large number of diverse disciplines to do so. Archaeology unearths fossils and man-made artifacts, and dates them in a self-consistent manner. Physics, chemistry and geology help in many analytical ways. Modern biology has turned out to be a powerful tool in such understanding of the history of the earth and of organisms present there. We now have the valuable help of genetics in this endeavour – a field that has exploded in the last 50 years. Central to the development of this tool has been the use of the genetic material of organisms, their DNA. It has been possible for scientists to tease out the DNA from fossilized remains of microbes, plants, animals and humans, and use the information contained in the DNA to build family trees. This is anthropology at the molecular level, when the biologist turns into a historian.

Genetically, How Different are We?

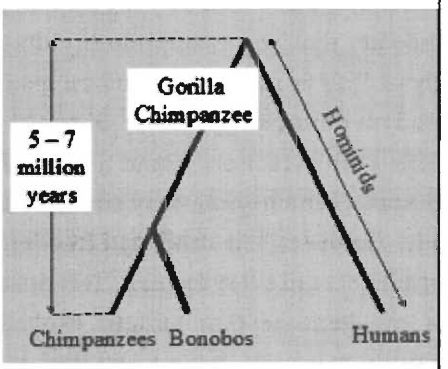
Recent data generated by the human genome sequencing project indicate that any two randomly drawn humans are genetically about 99.9 % identical. Human geneticists who are intensively studying human genomic diversity are engaging themselves with a tiny fraction (about a 10th of a per cent) of the human genome, which some may consider as an insignificant endeavour. However, it is this small fraction of the genome that confers uniqueness to every human. It is primarily on this small fraction of the genome that various evolutionary forces, particularly natural selection, have acted during the period of evolution of modern humans from their most recent common ancestor. Differences in this small genome fraction make some individuals susceptible to a disease, while conferring protection to others from the same disease. Thus, the study of human genomic variation among individuals can help us not only understand the nature and intensity of action of various forces that have modulated our evolutionary course, but also provide valuable data for the understanding of various diseases that afflict us today. Since the human genome comprises about 3 billion nucleotides, this small fraction is actually 3 million nucleotides.

The study of human genomic variation among individuals can help us not only understand the nature and intensity of action of various forces that have modulated our evolutionary course, but also provide valuable data for the understanding of various diseases that afflict us today.

Evolution of Modern Humans and their Entry into India

Our closest relatives are the chimpanzees and gorillas. We have diverged from them about 5-6 million years ago (*Box* 1). Modern humans (*Homo sapiens sapiens*) evolved in Africa (*Box* 2), and came out of Africa and moved to other geographical regions (*Box* 3). Although this is the widely accepted model of human evolution – called the out-of-Africa model, other models have also been proposed (see *Box* 3). India has served as a major corridor for the dispersal of these humans who started out from Africa about 100,000 years ago. The date of entry of modern humans into India remains uncertain. However, modern human remains dating back to the late Pleistocene (55,000-25,000 years before present, ybp) have been found in India. By the

Box 1. Our Closest Ancestors and Phylogenetic Inference



Our story begins about 5-6 million years ago when a population of African apes split into two distinct species, one eventually evolving to *modern human beings* and the other, eventually evolving to *modern chimpanzees*. This is pictorially depicted as a *phylogenetic tree*, which is a graphical representation of the relationships between various taxa, genera, species or population groups.

A phylogenetic tree is usually constructed on the basis of certain observations made on the various taxa (which is the generic term used to denote the 'entities' – genera, species, etc. – among whom we are seeking to reconstruct relationships). The observations can be of various types, measurements of various characteristics of fossils skulls, teeth or bones, or can be measurements of various body characteristics, such as, height, head length, nose height, etc., or can be presence/absence or frequencies of some genetic characteristics, such as, of a specific nucleotide at a specific position in the DNA.

The idea underlying the reconstruction of evolutionary relationships is that as taxa evolve over time, their characteristics diverge. Thus, two taxa that are 'similar' are evolutionarily closer. Conversely, taxa that are dissimilar or 'distant' are also evolutionarily distant. Similar taxa are considered to have diverged from a common ancestor in the recent past, while distant taxa have diverged from a common ancestor in the more distant past. To measure the distance between taxa, various 'distance measures' are used. If we have quantitative measurements (such as, tooth height, head length, etc.) on the taxa under consideration, the most popular measure of distance is the Mahalanobis D^2, which is defined as [1]:

$$D^2 = \sum_{i=1}^{r} \sum_{j=1}^{r} v^{ij} (\bar{x}_i - \bar{y}_i)(\bar{x}_j - \bar{y}_j),$$

where r denotes the number of characters measured, $\bar{x}_i$ and $\bar{y}_i$ denote, respectively, the mean values of the ith character in the two taxa, and v^{ij} denotes the (i, j)th element of the inverse of the $r \times r$ matrix of variances and covariances of the characters (which is assumed to be the same in the two taxa).

On the other hand, if we have data on frequencies of genetic characteristics, such as frequencies of various nucleotides observed at specific positions in the DNA, then a commonly used measure of genetic distance is:

$$D_A = \sum_{k=1}^{r} \left(1 - \sum_{i=1}^{m_i} \sqrt{x_{ik} y_{ik}}\right) / r$$

where m and r are, respectively, the number of different nucleotides observed at the kth position and the number of nucleotide positions on which data have been gathered, and x and y are, respectively, the

Box 1. continued...

frequencies of the ith type of nucleotide at the kth position in the two taxa under consideration. This measure takes a value between 0 (when frequencies of all nucleotides at all positions in the two populations are equal) and 1 (when the two populations share no common alleles). The D distance has its roots in a measure first suggested by Bhattacharyya [2].

If we are seeking to find evolutionary relationships among several taxa, then we first calculate the matrix of distances between all pairs of taxa, using an appropriate distance measure, and then use an algorithm [3] to compute the phylogenetic tree based on the information contained in the distance matrix.

The times of split of the different branches of a phylogenetic tree can also be estimated [4]. These time of split between two taxa is called *divergence time from the most recent common ancestor*, or, also as the *time to the most common recent ancestor* (*TMRCA*). However, estimation of TMRCA requires some assumptions to be made. Framing of the required assumptions is very difficult for characters such as height or weight, which in addition to genetic factors are also largely influenced by the environment. These assumptions are more easily framed and can be tested if one uses genetic data. Further, the evolutionary change of genetic characters can be studied under specific population genetic frameworks. Such statistical frameworks do not exist for the change of morphological characters. Therefore, genetic data are considered to be ideal for evolutionary inferences.

Suggested Reading

[1] P C Mahalanobis, On the generalized distance in statistics, *Proc. Natl. Inst. Sci. India*, Vol. 2, pp. 49-55, 1936.

[2] A Bhattacharyya, On a measure of divergence between two multinomial populations, *Sankhya*, Vol. 7, pp. 401-406, 1946.

[3] There are various algorithms for constructing phylogenetic trees from a distance matrix. Interested readers may consult Chapter 11 of M Nei, *Molecular Evolutionary Genetics*, New York: Columbia Univ. Press, 1987 or Chapter 5 of W -H Li, *Molecular Evolution*, Sunderland, Mass.:Sinauer Associates, Inc., Publishers, 1997.

[4] For the methodologies of estimation and the underlying assumptions, interested readers may consult Chapters 7 and 8 of Li (1987).

middle Paleolithic period (50,000 – 20,000 ybp), humans had spread to many parts of the subcontinent.

Complexity of Evolutionary Reconstruction:

Look at the complexity of the task of reconstruction human evolution in India. India is not a collection of one or two, or even a handful of ethnic groups. The Anthropological Survey of India (ASI) had taken up the heroic and mammoth task of collecting

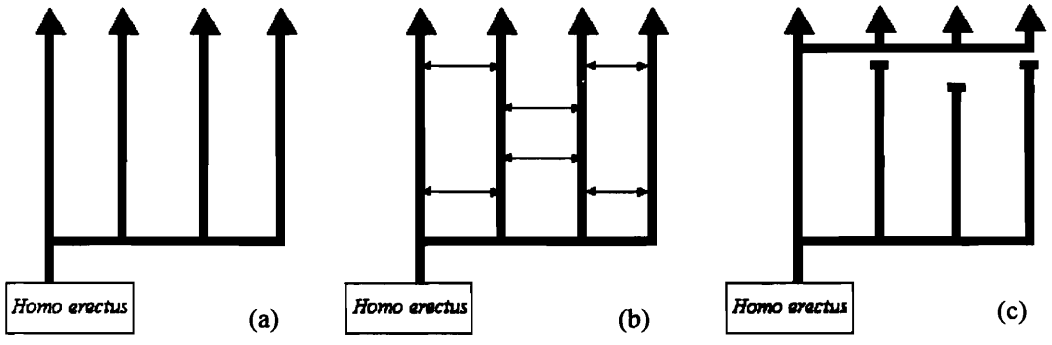
Box 2. Evolution of Anatomically Modern Humans

More than 4 million years ago (mya), one of the species on the evolutionary path to humans began spending most of its time on two feet. That was the genus *Australopithecus*. The most famous fossil specimen of this genus – Lucy – has been classified as the species *afarensis*. This specimen was found in Ethiopia, and its estimated age is about 3.2 million years. From the shape of the pelvis, joints and other anatomical features, we know that the australopithecines walked upright. The upright stance seems to have set in motion a profound evolutionary trend ... hands became free to use for manipulation, including to throw stones at predators; there was gain of visibility over higher underbrush, that enabled spotting of predators before it was too late; and, there was a dramatic increase in brain size, although it is unclear whether this increase was the consequence of bipedality.

The genus *Homo* appears to have arisen about 2 mya. The earliest fossil remains have been found in Africa (dated to about 1.9 mya). Fossil remains of this genus have also been found in Java, dating back to 1.8 mya. Once the first species in the genus *Homo* appeared, it began to spin off new species. So, there were *Homo* species called *habilis*, *ergaster*, *erectus*, *antecessor*, *heidelbergensis*, *neanderthalensis* and, finally, *sapiens*. At least one species of the genus *Homo* did something that the australopiths had never done. Members of *Homo erectus* spread out of Africa into Asia and Europe. *Homo erectus* fossils dating back to over 1 mya have been found in Indonesia.

Anatomically modern humans like ourselves – who are less heavily built, more mobile and with higher cognitive flexibility – arose in Africa about 150,000 years ago. We have been classified into the subspecies *sapiens* of *Homo sapiens*.

The mode of evolution of anatomically modern humans (*Homo sapiens sapiens*) from *Homo erectus* has been a matter of fierce debate. The first interpretation of the fossil record to develop a possible model of the evolution of modern humans was proposed by the palaeoanthropologist Carleton Coon. He proposed the so-called candelabra model (*Fig*ure a), which posited that modern humans evolved almost simulta-



Three models of evolution of modern humans (a) candelabra model, (b) multiregional model [horizontal two-headed arrows indicate genetic admixture between lineages], (c) uniregional (out-of-Africa) model.

Box 2. continued...

neously in different regions of the world – Africa, Asia, Europe and Australia. The main basis of this model was the claim that most ancient fossils show a continuous morphological transition to modern humans. This model was severely criticized both by anthropologists (e.g., Ales Hrdlicka) and geneticists (e.g., Theodosius Dobzhansky), because it was contrary to known modes of species evolution. The candelabra model implied parallel evolution; that is, the biological (genetic) traits that characterize modern humans arose independently four times and also became the predominant (fixed) traits independently in four populations. This was an evolutionary impossibility.

The candelabra model was modified by Milford Wolpoff and his collaborators (1984) who stated that parallelism was the result of substantial intermigration and genetic admixture. This model has come to be known as the multiregional model (*Fig*ure b). Unfortunately, recent quantitative research on human skulls has shown no morphological continuity in the various continents (Salem *et. al.* 1996). Further, the human type with clear similarities to modern humans – the Neanderthal – existed only in Europe and West Asia. Recent studies have shown that there is no evidence of genetic admixture between the Neanderthal and the modern human (Krings *et. al.* 2000).

The now standard model of human evolution is called the uniregional or the out-of-Africa model (*Fig*ure c). This model proposes that modern humans evolved in a single region – Africa – who expanded about 200,000 years ago and arrived in different geographical regions at different points of time. The various forms that evolved from *Homo erectus* in multiple geographical regions all became extinct. It is unclear why these forms became extinct. One possibility is that they were exterminated by modern humans who were better tool makers and users.

Suggested Reading

[1] M Krings, C Capelli, F Tschentscher , H Geisert, S Meyer, A von Haeseler, K Grossschmidt, G Possnert, M Paunovic and S Paabo, A view of Neandertal genetic diversity, *Nature Genetics,* Vol.26, pp. 144-146,2000.

[2] A H Salem, F M Badr, M F Gaballah and S Paabo, The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula, *American Journal of Human Genetics,* Vól.59, pp. 741-743,1996.

[3] M H Wolpoff, X Wu and A GThorne, Modern *Homo sapiens* origins: a general theory of hominid evolution involving the fossil evidence from east Asia, In *The Origins of Modern Humans: A World Survey of the Fossil Evidence,* eds FH Smith and F Spencer Alan R Liss: New York, pp. 411-483, 1984.

and cataloguing information about the linguistic, geographic and sociological features of all ethnic groups across India. The results are an authoritative 43-volume set called *The People of India*, published in 1992. They tell us that we Indians are a mosaic, a patchwork quilt and a rainbow coalition of 450 tribal communities speaking over 750 dialects. These are classified

Box 3. Genetic Support for the Out-of-Africa Model

DNA sequences and genes accumulate changes (mutations) over time. Depending on whether these changes are favoured or disfavoured by natural selection, the changes increase or decrease in their frequencies in a population. With time, selectively advantageous mutations become fixed (that is, their proportion in the population attains the value of unity), while selectively disadvantageous mutations become lost. However, even changes that are selectively neutral evolve in a manner from which it is possible to make many statistical inferences. For example, it is possible to estimate the time when two DNA sequences had a common ancestor. Populations that remain isolated (that is, when there is no gene flow among them) accumulate changes independently, and the populations diverge from each other in their genetic characteristics. Populations that have diverged from a common ancestor for a greater period of time and have remained isolated since divergence show greater genetic dissimilarities than populations that have diverged more recently. Genetic admixture (gene flow) results in promoting genetic similarities among populations.

In addition to the DNA contained in the nucleus, a cell also contains extra-nuclear DNA called mitochondrial (mt) DNA. Unlike the nuclear, autosomal DNA which is derived from both parents of an individual, the mtDNA is derived solely from the mother. Thus, the mtDNA is a record of the female lineage. Using mtDNA sequences from a number of geographically dispersed populations, Cann et al. (1987) reached two important conclusions that supported the out-of-Africa model of human evolution. First, Africans and non-Africans were genetically very distinct, and second, the African gene pool contributed to the gene pools of humans inhabiting other regions. The estimated time back to the most recent common ancestor of modern human mtDNA was 190,000 years (with a large standard error). Further studies (e.g., Vigilant *et. al.* 1991; Ingman *et. al.* 2000) have strongly supported the findings of Cann *et. al.* (1987), although the time estimates have been somewhat variable (albeit in the same ballpark). Similar support to the uniregional model has also been obtained from studies on the Y-chromosomal DNA variation. (The Y-chromosome is transmitted by fathers only to their sons, and hence the Y-chromosomal variation is a record of the male lineage.) Hammer and his colleagues (1998) showed that the oldest types of Y chromosomes were found only in Africa and that the Y-chromosomal variation was the highest in African populations (as is to be expected since in the oldest populations the Y chromosome had the longest time to accumulate mutations). They also estimated that the differentiation of the ancestral type of Y-chromosome into other more recent subtypes began about 147,000 years ago. The out-of-Africa model has also been supported by analysis of autosomal DNA sequence variation (Harding *et. al.* 1997).

Suggested Reading

[1] R L Cann, M Stoneking and A C Wilson, Mitochondrial DNA and human evolution, *Nature*, Vol.325, pp. 31-36, 1987.

[2] M F Hammer, T Karafet, A Rasanayagam, E T Wood, T K Altheide, T Jenkins, R C Griffiths, A R Templeton and S L Zegura, Out of Africa and back again: nested cladistic analysis of human Y chromosome variation, *Molecular Biology and Evolution*, Vol.15 ,pp. 427-441, 1998.

Box 3. continued...

[3] R M Harding, S M Fullerton, R C Griffiths, J Bond, M J Cox, J A Schneider, D S Moulin and J B Clegg, Archaic African and Asian lineages in the genetic ancestry of modern humans, *American Journal of Human Genetics*, Vol.60, pp. 772-789, 1997.

[4] M Ingman , H Kaessmann , S Paabo and U Gyllensten, Mitochondrial genome variation and the origin of modern humans, *Nature*, Vol.408, pp.708-713, 2000.

[5] L Vigilant, A C Wilson and H Harpending, African populations and the evolution of human mitochondrial DNA, *Science*, Vol. 253, pp. 1503-1507, 1991.

into the Austro-Asiatic (AA), Dravidian (DR) and Tibeto-Burman (TB) language families. The tribals make up 8% of India. The non-tribals speak languages that belong to the Indo-European (IE), or Dravidian (DR) families. The IE and DR have contributed in a major way to the development of Indian society and culture. But they are also known to be affected by waves of migration into India since prehistoric times. And then there is the stratification into castes, a phenomenon unique to India.

From the late-1990s, a large number of anthropologists, molecular geneticists and statisticians, working in various universities and institutions in India, undertook studies to reconstruct the complex patchwork quilt using information contained in the DNA molecule. Their early work provided molecular genetic evidence that a major population expansion of modern humans took place within India. Although the period of this demographic expansion remains uncertain, it appears to have taken place 60,000-85,000 ybp. Perhaps this demographic expansion, followed by subsequent migration, resulted in the peopling of parts of South-East Asia and later (50,000-60,000 ybp) of Australia. About 60,000 ybp, there was perhaps another independent expansion of modern humans in southern China, which may have resulted in human migration into India and also into South-East Asia.

Contemporary ethnic India is a land of enormous genetic, cultural and linguistic diversity. It has been shown that, with the exception of Africa, India harbours more genetic diversity than other comparable global regions. The enormous diversity in

Contemporary ethnic India is a land of enormous genetic, cultural and linguistic diversity. It has been shown that, with the exception of Africa, India harbours more genetic diversity than other comparable global regions.

Genes move as people move. As people move, the relative frequencies of the genes – especially those that are selectively neutral – that they originally carried decrease, since new mutations accumulate.

social and cultural beliefs and practices has been well documented and emphasised.

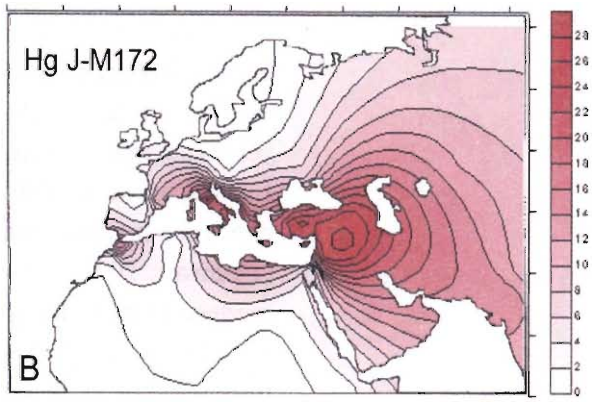
There are considerable differences of opinion among anthropologists regarding the processes of peopling of this subcontinent, and the origins of Indian ethnic groups. During the past several years, scientists have attempted to test some of these hypotheses using genomic data. For generating such data, blood samples of individuals were obtained, with informed consent, from a large number of population groups of India, of diverse geographical, linguistic and ethnic backgrounds. DNA was isolated from each of these blood samples. The DNA samples were then screened for various types of genetic variations. These data were then statistically analysed.

Early Footprints

Demographic expansion of a population that is largely dependent on natural resources results in pressure on the local natural resources which in turn results in the migration of a subset of individuals to newer geographical areas in search of natural resources to ensure their own survival. This process gets repeated as time elapses, and new subpopulations are formed. If the new subpopulations remain maritally isolated, genetic differentiation occurs; that is, the subpopulations genetically diverge from the parental and other subpopulations. When there is demographic contraction of a population, through famines and other such catastrophes, genetic variability in the population rapidly diminishes and it takes a long time to recover the lost variability. These processes have been statistically modelled and their genetic outcomes are statistically predictable (Nei, 1975; Ewens, 1979; Li, 1997).

Genes move as people move. As people move, the relative frequencies of the genes – especially those that are selectively neutral – that they originally carried decrease, since new mutations accumulate. Thus, with the movement of people over geographical space, one expects a decreasing cline of gene

Box 4. Clines of a Genetic 'Marker'



Hg J-M172 is a genetic 'marker'that is located on the Y chromosome. The Y chromosome is passed on from fathers to his sons. This genetic marker arose in the ''fertile crescent'. The fertile crescent is the vast area, sweeping from the eastern end of the Mediterranean Sea and taking in the Tigris and Euphrates rivers, and now comprises the nations of Syria, Iraq and Lebanon. Aside from this marker, agriculture also 'arose'in the fertile crescent. The knowledge of agriculture was carried by farmers from the fertile crescent, first to the adjacent areas and thereafter to other areas of the world. Many of the male farmers in the fertile crescent region possessed the Hg J- M172 on their Y chromosomes, and as they moved they carried this marker with them. When some of them reproduced and left behind children in a new area, this marker was also introduced to the new area. One would, of course, expect that the frequency of this marker in the new area would be lower than it was in the fertile crescent. Thus, over geographical space, one would expect the frequency of this marker to systematically decrease from the epicentre, that is, the fertile crescent. This is what is called as a frequency cline. Thus, the clinal pattern of Hg J-M172 frequency correlates very well with the spread of agriculture from the fertile crescent to other geographical areas.

frequencies from the epicentre of dispersion. The broad natures of these clines follow certain statistical patterns. These statistical predictions and patterns have been successfully exploited by geneticists to trace migration trails back in time (Cavalli-Sforza and Feldman, 2003). (See *Box* 4)

After anatomically modern humans evolved in Africa about 200,000 years ago, a small population of about 1,000 individuals (that is, a tribe), most probably from East Africa, expanded and spread throughout much of Africa about a 100,000 years ago. Genetic data indicate further migration into Asia, which perhaps took place along two routes, between 80,000 and 40,000 years ago. They reached the Arabian peninsula, around the Persian Gulf, along the shorelines of Iran and Pakistan, south along the Indian coast, north again to the mouth of the Ganges, to South-East Asia. From South-East Asia the migration trail bifurcated to the north and the south; the southern trail took

After anatomically modern humans evolved in Africa about 200,000 years ago, a small population of about 1,000 individuals (that is, a tribe), most probably from East Africa, expanded and spread throughout much of Africa about a 100,000 years ago.

modern humans to Australia and Oceania, while the northern trail took them to China, Japan and eventually the Americas. Archaeological data indicate that after the initial appearance of modern human fossils in the Middle East about 130,000–90,000 years ago, modern humans did not reappear in that region, or in Europe, until 50,000-60,000 years ago, although there is evidence for modern humans in Australia about 60,000 years ago (Tatersall, 2002). This is consistent with the possibility of modern humans having taken the southern exit route. Genetic evidence in support of the southern exit route comes from mtDNA. The evidence is not decisive yet (see *Box* 5).

Box 5. Southward Ho!

It is now established beyond reasonable doubt that anatomically modern humans evolved in Africa between 200,000 and 150,000 years ago. A small population of about 1,000 individuals, most probably from East Africa, expanded and spread throughout much of Africa about 100,000 years ago. Genetic data indicate further migration into Asia, and later to other parts of the world. Which route did the first humans take when they moved out of Africa? The easiest would have been a walking trail along the river Nile, across the Sinai Peninsula ('northern exit route'). Then, modern humans could have gone towards Europe or into India. Modern humans from East Africa could have taken an alternative route – across the Red Sea and along the shoreline of Saudi Arabia ('southern exit route'); to the best of our knowledge, this possibility was first proposed by Masatoshi Nei and Arun Roychoudhury in 1993. Or, they may have used both routes.

It is now clear that the northern exit route was used and that this exit took place around 45,000 years ago (Lahr and Foley, 1994, 1998; Mountain, 1998). On the other hand, fossil evidence indicates that southern Australia was definitely inhabited by anatomically modern humans 45,000 years ago; northern Australia and South-East Asia were perhaps inhabited even earlier (Tatersall, 2002). There is no conceivable way that modern humans could have reached Australia at about the same time that they exited through the northern route.

Was the southern exit route used prior to the northern exit route? Could humans have moved along the shoreline of Saudi Arabia into India, and reached Australia along the shorelines of the Arabian Sea and the Bay of Bengal through North-East India, South-East Asia and Borneo? This would have been a more difficult route, as it would require crossing of straits and seas. Of course, tree trunks could have been used as rafts and as primitive boats. There have been no appropriate fossil finds to answer whether a southern exit route was used and existing genetic evidence has only been suggestive (Forster *et. al.* 2001; Endicott *et. al.* 2003; Kivisild et al. 2003). Indeed, some have interpreted the genetic evidence as indicating the absence of a southern exit route (Cordaux and Stoneking, 2003).

Box 5. continued...

Since genes move as people move, the commonly used method to trace trails of human migration is to identify specific genetic signatures in the source population and look for these signatures in extant populations along the suspected migration route. This has not yet been possible in respect of the putative southern exit route. However, two recent studies (Thangaraj et al., 2005; Macaulay et al., 2005) have found some ancient genetic signatures in population isolates residing in the Andaman and Nicobar Islands and in South-East Asia. The individuals belonging to these populations possess Negrito morphological features (frizzly hair, thick everted lips, dark complexion, for example). These populations are therefore candidate relics of the ancient humans who moved out of Africa and populated other global regions. If anatomically modern humans had indeed used the southern exit route as they moved to populate Australia, then the Andaman Islands and South-East Asia could have lain on their trail. Thangaraj et al. (2005) have found two mitochondrial DNA (mtDNA) 'signatures' – sub-haplogroups that they have termed M31 and M32 – among the Onge and Great Andamanese of the Andaman and Nicobar Islands. These sub-haplogroups arose from haplogroup M – found in high frequencies among extant populations of Ethiopia and mainland India – about 65,000 years ago. They did not find the M31 and M32 sub-haplogroups in mainland India. At the same time, Macaulay et al. (2005) report two mtDNA sub-haplogroups that they have termed as M21 and M22, in the Orang Asli population of Southeast Asia. These sub-haplogroups also branched off from the M haplogroup about 60,000 years ago.

Mitochondrial DNA evidence indicates that the macrohaplogroup L3 arose in East Africa about 85,000 years ago, from which arose the haplogroups M, N and R that are widely co-distributed throughout Asia. Various sub-haplogroups, including M21, M22, M31 and M32, branched off from these ubiquitous haplogroups. The time estimates of these sub-haplogroups are consistent with the southern exit route hypothesis, and indicate that the southern exit route may have been used before the northern exit route.

Do the data presented in the two studies provide convincing evidence in support of the southern exit route hypothesis? We think that the data are consistent, but not necessarily clinching. Ideal evidence would comprise identification of a specific genetic signature (or, a set of signatures) in the source population and finding it in extant populations along the postulated southern exit route. Can we reasonably expect to find such evidence? For two reasons, we think not. First, the source-specific genetic signatures may have been lost during the time course of evolution by the accumulation of mutations and because of genetic drift. We may, therefore, never find such signatures in extant populations. It may, of course, be possible to identify some signatures among extant populations that were early derivatives of the 'original' signatures. If it is found that the frequencies of these derived signatures decrease from around the exit point as one moves towards Australia along the postulated trail, then such a pattern (a frequency cline) would comprise strong evidence in favour of the southern exit route hypothesis. Second, albeit unlikely, it could be that the modern humans who took the southern exit route did not leave any genes along their trail. This is unlikely because even if they were moving in a band, some members would surely have been left behind at various points on their march towards Australia from out of Africa. The search for genetic evidence to test the southern exit route hypothesis will continue and will continue with greater vigour stimulated by the interesting findings of Thangaraj *et. al.* (2005) and Macaulay *et. al.* (2005).

Box 5. continued...

Suggested Reading

[1] R Cordaux and M Stoneking, South Asia, the Andamanese, and the genetic evidence for an "early" human dispersal out of Africa, *Am. J. Hum. Genet.*, Vol.72, pp.1586-90, 2003.

[2] P Endicott, M T Gilbert, S Stringer, C Lalueza-Fox, E Willerslev, A J Hansen and A Cooper, The genetic origins of the Andaman Islanders, *Am. J. Hum. Genet.*, Vol.72, pp.178-84, 2003.

[3] P Forster, A Torroni, C Renfrew and A Röhl, Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution, *Mol. Biol. Evol.*, Vol.18, pp.1864-1881, 2001.

[4] T Kivisild, S Rootsi, M Metspalu, S Mastana, K Kaldma, J Parik, E Metspalu, M Adojaan, H V Tolk, V Stepanov, M Golge, E Usanga, S S Papiha, C Cinnioglu, R King, I Cavalli-Sforza, P A Underhill, and R Villems, The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations, *Am. J. Hum. Genet.*, Vol.72, pp.313-32, 2003.

[5] M M Lahr and R A Foley, Multiple dispersals and modern human origins, *Evolutionary Anthropology*, Vol.3, pp.48-60, 1994.

[6] M M Lahr and R A Foley, Towards a theory of modern human origins: geography, demography and diversity on recent human evolution, *Yearbook of Physical Anthropology*, Vol.41, pp.137-176, 1998.

[7] V Macaulay, C Hill, A Achilli, C Rengo, C Clark, W Meehan, J Blackburn, O Semino, R Scozzari F Cruciani, A Taha, N K Shaari, J M Raja, P Ismail, Z Zainuddin, W Goodwin, D Bulbeck,H-J Bande, S Oppenheimer, A Torroni and M Richards, Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes, *Science*, Vol.208, pp.1034-1036, 2005.

[8] J L Mountain, Molecular evolution and modern human origins, *Evolutionary Anthropology*, Vol.7, pp. 21-37, 1998.

[9] M Nei and A K Roychoudhury, Evolutionary relationships of human populations on a global scale, *Mol. Biol. Evol.*, Vol.10, pp.927-43, 1993.

[10] I Tatersall, The case for saltational events in human evolution, In *The Speciation of Modern Homo sapiens*, ed TJ Crow, Oxford University Press, Oxford, pp.192-218, 2002.

[11] K Thangaraj, G Chaubey, T Kivisild, A G Reddy, V K Singh, A A Rasalkar and L Singh, Reconstructing the origin of Andaman Islanders, *Science*, Vol.208, p.996, 2005.

The second route was through the Middle East, Arabia or Persia to Central Asia, from where migration occurred in all directions reaching Europe, East and North-East Asia about 40,000 years ago. These migration trails are summarized in *Figure* 2. There is substantial genetic and archaeological evidence in support of migration through this 'northern exit route' (Lahr and Foley, 1994, 1998; Mountain, 1998).

Fundamental Genomic Unity of India

Cells contain bodies (organelles) called mitochondria, each of

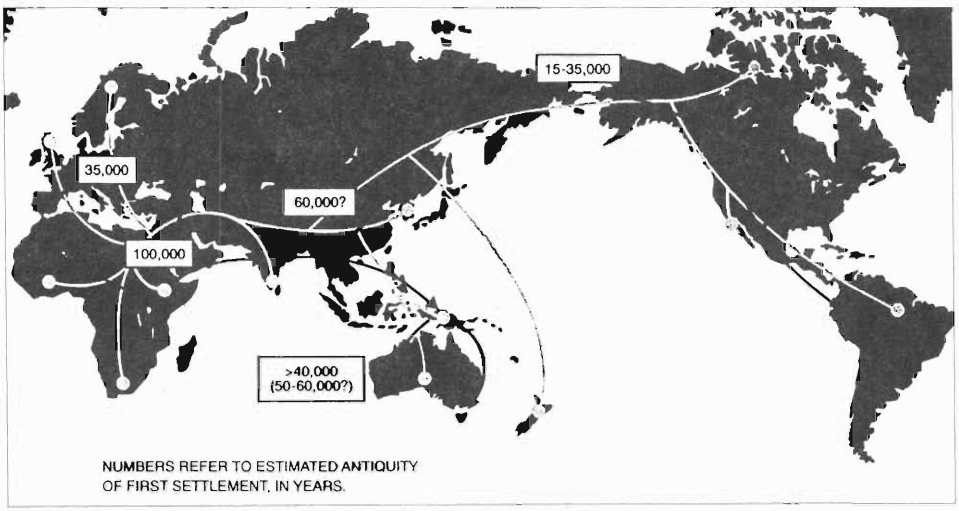


Figure 2. Simplifed diagram of human population movements out of Africa.

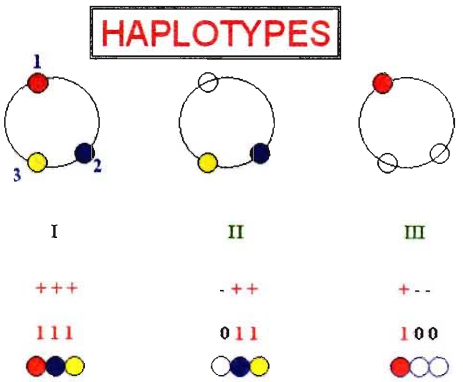
which carries a tiny circular piece of DNA, separate from the DNA in the cell's nucleus. While nuclear DNA is inherited equally from both parents, mitochondrial (mt) DNA is inherited only from the mother. As mtDNA is thereby passed on in an unbroken chain from mother to daughter, data on mtDNA have been extremely useful in the study of human evolution, including prehistoric migrations and demographic events such as sudden population size expansion or extreme reduction.

Mitochondrial DNA samples collected from individuals belonging to over 50 ethnic populations from different geographical regions of India have been studied. These populations were chosen to include both tribal, caste and religious groups at different levels of social hierarchy and belonging to different linguistic families.

Based on data at several variable positions (polymorphic loci) on the mtDNA, these DNA molecules can be classified into haplotypes (*Box* 6). Suprisingly, of the several thousand possible haplotypes, just one haplotype accounted for about 50 % of all mtDNA molecules. It can, therefore, be inferred that this is the most ancient haplotype in Indian populations. Further, in the vast majority of the populations studied, this haplotype was the most frequent.

Box 6. Haplotypes and Haplogroups

The human mitochondrial DNA (mtDNA) is a circular string of about 16,000 nucleotides A, T, G or C (like a necklace of beads) that is present in thousands of copies in each cell, but is outside of the cell nucleus. There are specific nucleotide positions at which not all individuals have the same nucleotide. These positions are called polymorphic. Given below is a schematic representation of the mtDNA of three individuals I, II and III.



Three specific positions (1, 2 and 3) are marked on each mtDNA. At the first position, the individuals I and II possess a specific nucleotide ('the red bead'), while the individual III possesses a different nucleotide (a 'white bead'). This position is, therefore, polymorphic. (Strictly, a position is said to be polymorphic in a population if at least 5% of the individuals carry the rarer - that is, the less frequent – nucleotide.) Similarly, positions 2 and 3 are also polymorphic. The mitochondrial DNA is transmitted intact (the entire necklace) by the mother to all her children (the mtDNA of the father is not transmitted to any of his children). Thus, considering all the three positions, one can designate the mtDNA of individual I as +++ or 111 corresponding to the red, blue and yellow beads at these three positions. Technically, +++ or 111 (any of these notations may be used) is the haplotype of individual I. Similarly, individuals II and III belong to haplotypes –++ (111) and +– – (100), respectively.

Groups of 'similar' haplotypes are called haplogroups. Haplotypes I and II above are more similar; hence, they belong to the same haplogroup. (Technically, similarity is based on the evolutionary relationships between two haplotypes. But we shall not discuss that here.)

Haplotypes serve as signatures. The extensive sharing of one or two haplotypes, out of the several thousand possible ones, by population groups across India, irrespective of their geographical location, habitat, linguistic affinity or social proximity, reveals a fundamental unity of mtDNA lineages in India, in spite of the extensive cultural and linguistic diversity. We believe that there was a relatively small number of founding female lineages in India. Ethnic differentiation took place subsequently through a series of demographic expansions, geographic dispersal and social groupings.

Footprints from Central and West Asia

Based on some evolutionary considerations, the mtDNA

haplotypes can be further grouped into broader classes, called haplogroups (HGs). The frequency patterns of these HGs are highly variable across geographical regions and ethnic groups. These patterns have proven useful in the study of human population movements and evolution. Two of these HGs, HG-M (which is specific to the Asian, particularly India, region) and HG-U (which occurs in high frequencies among Caucasian groups, including those of Central and West Asia) are the most relevant to us. HG-M was found to be the most frequent – 71.4% of the individuals in the pooled sample belonged to this HG. The frequency (51.11%) of this HG was found to be significantly lower among Tibeto-Burman tribals compared to the Austro-Asiatic (76.27%) and the Dravidian (76.66%).

Indian tribal populations, particularly the Austro-Asiatics and Dravidians, harbour very high frequencies of the Asian, but not of the Caucasian, HG.

HG-U was also found to occur in most populations in India. The frequency of this HG in the pooled tribal sample was about 10%. Considerable differences in the frequencies of this HG were observed among Austro-Asiatic (13.56%), Dravidian (9.17%) and Tibeto-Burman (6.7%) tribals; these differences were, however, not statistically significant at the 5% level.

Thus, Indian tribal populations, particularly the Austro-Asiatics and Dravidians, harbour very high frequencies of the Asian, but not of the Caucasian, HG. Further, although the Caucasian HG-U occurs in about 10% of the Indian tribals, when this HG is split into subclusters (based on co-occurrence of other variants in the mtDNA), it was found that most Indian tribals (77.3%) belong to subcluster U2i. This, subcluster occurs infrequently in other Caucasian populations. The implication of this finding is that some of the so-called "Caucasian HGs" may be heterogeneous, and their occurrence in a region may not reflect their having been brought into the region through Caucasian admixture. Thus, not all of the HG-U lineages may have been brought into India through Caucasian admixture; those that belong to subcluster U2i may actually be indigenous to India.

As the antiquities of the tribal populations are far greater than the time of presumed entry (3,000-4,000 ybp) of Indo-Aryan

Prehistoric, historic and linguistic evidences have suggested that West Asian and central Asian gene pools have contributed to the Indian gene pool, primarily the north-Indian gene pool.

speakers ('Caucasians') in India, current data support an earlier conclusion that an ancestral type of mtDNA that led to the evolution of the HG-U was introduced in India preceding the arrival of Indo-Aryan speakers in India. Thus, although there are indications of admixture of indigenous Indian populations with Indo-Aryan speaking immigrants, the extent of admixture may have been smaller than is usually contended.

Lessons from Y-Chromosomal DNA Variation

Prehistoric, historic and linguistic evidences have suggested that West Asian and Central Asian gene pools have contributed to the Indian gene pool, primarily the North Indian gene pool. Since many contend that immigrants from Central and West Asia were predominantly males, we also sought to find appropriate genetic signatures on the DNA of the Y-chromosome, the male chromosome which sons inherit from their fathers. Based on their genetic variation, the Y-chromosomes can be grouped into various Y-HGs (HGs). HG-3 is the most frequent (35 to 58%) haplogroup in North Indian populations, while HG-9 is the most frequent (33 to 57%) in West Asian populations. Globally, the peak HG-9 frequency is found in the Caucasus-Anatolia region.

Interestingly the frequency of this haplogroup is highest (23.5%) among caste Brahmins and is lower (17.1%) among caste Rajputs, though strictly speaking, these cannot be considered as well-defined genetic groups. The HG-9 haplogroup may have been brought into India by the Indo-European speakers from West Asia. Historically, Brahmins were considered to be the torch-bearers and promoters of Aryan rituals. Therefore, it is likely that this group had the highest genetic contact with the Aryan-speaking peoples. This is consistent with the high frequency of HG-9 observed among them. This haplogroup may have percolated to the Rajputs, either through admixture with Brahmins or directly with the Aryan-speaking immigrants. It is noteworthy that HG-9 is absent among some caste groups considered to be indigenous. The observation that HG-9 occurs in a lower

frequency (10.5 %) among Muslims, compared to the Brahmins and Rajputs, is consistent with their known social history.

Haplogroup-3, the most frequent haplogroup in India, is widely found elsewhere in Asia too, except East Asia, and is virtually absent in Africa and the Americas. HG-3 is also found in high frequencies in Central Asia (Russia and the Altai region) and East Europe (Poland and Hungary). It appears that this haplogroup arose in Central Asia about 7,500 ybp, and the distribution of this haplogroup reflects a recent and major population expansion within Eurasia. We have found that the frequency of HG-3 in Uttar Pradesh is quite high (35 to 58%). These provide clear indications of population movements from Iran and Central Asia into India.

Another evidence, albeit somewhat indirect, of possible admixture of northern Indian populations with Central/West Asians, has also been obtained. This evidence was obtained from a recent study of disease-related genomic variations. The HIV-1 virus, the causative agent of AIDS, uses certain receptors on the surface of target cells to gain entry into them. A mutant form of the receptor gene *CCR5*, however, does not allow the HIV-1 entry into the cell. The frequency of this mutant form of the gene is known to be high among Caucasian populations, including population of Central and West Asia.

DNA samples of about 1500 individuals, drawn from 40 diverse ethnic populations of India, were screened. It was found that the mutant form is completely absent or only sporadically present in most populations. However, among the Muslims of Uttar Pradesh, the frequency of the mutant form was 5.4%, which may be due to admixture with immigrants from Central and West Asia.

Epilogue

We, therefore, see that information contained in our DNA molecules can be profitably used to estimate genetic diversity, trace population movements and reconstruct human evolution. We would also like to point out that such data can potentially be

"We all should know that diversity makes for a rich tapestry, and we must understand that all the threads of the tapestry are equal in value no matter their color; equal in importance no matter their texture."

Address for Correspondence
Partha P Majumder
Indian Statistical Institute
Kolkata 700 108, India
Email: ppm@isical.ac.in

D Balasubramanian
L V Prasad Eye Institute
Hyderabad 500 034, India
Email:dbala@lvpei.org

used to stigmatize groups of people. (Recall: Claims of Aryan superiority and the Nazi holocaust.) We would, therefore, like to underscore that even though such population genetic data have social implications, interpretations and use of these data have to be done with extreme caution. Maya Angelou has said in her collection of essays and stories entitled *Wouldn't Take Nothing for My Journey Now* "It is time for the preachers, the rabbis, the priests and pundits, and the professors to believe in the awesome wonder of diversity so that they can teach those who follow them. It is time for parents to teach young people early on that in diversity there is beauty and there is strength. We all should know that diversity makes for a rich tapestry, and we must understand that all the threads of the tapestry are equal in value no matter their color; equal in importance no matter their texture."

Suggested Reading

[1] A Basu, N Mukherjee, S Roy, S Sengupta, S Banerjee, M Chakraborty, B Dey, M Roy, B Roy, N P Bhattacharyya, S Roychoudhury and P P Majum-der, Ethnic India: a genomic view, with special reference to peopling and structure, *Genome Research*, Vol. 13, pp. 2277-2290, 2003.

[2] N P Bhattacharyya, P Basu, M Das, S Pramanik, R Banerjee, B Roy, S Roychoudhury and P P Majumder, Negligible gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms, *Genome Research*, Vol.9, pp. 711-719, 1999.

[3] L L Cavalli-Sforza and M W Feldman, The application of molecular genetic approaches to the study of human evolution, *Nature Genetics* Supplement, Vol. 33, pp.266-275, 2003.

[4] F-C Chen and W-H Li, Genomic divergences between human and other hominoids and the effective population size of the common ancestor of human and chimpanzee, *American Journal of Human Genetics*, Vol. 68, No.2, pp. 444-456, 2001.

[5] W J Ewens, *Mathematical Population Genetics*, Springer-Verlag, Berlin, 1979.

[6] M M Lahr and R A Foley, Towards a theory of modern human origins: geography, demography and diversity on recent human evolution, *Yearbook of Physical Anthropology*, Vol.41, pp.137-176, 1998.

[7] W-H Li, *Molecular Evolution*, Sinauer, Sunderland,1997.

[8] J L Mountain, Molecular evolution and modern human origins, *Evolutionary Anthropology*, Vol.7, pp. 21-37, 1998.

[9] M Nei, *Molecular Population Genetics and Evolution*, North-Holland, Amsterdam, 1975.

[10] O'hUigin C, Y Satt, N Takahata and J Klein, Contribution of homoplasy and ancestral polymorphism to the evolution of genes in anthropoid primates, *Molecular Biology and Evolution*, Vol.19, pp. 1501-1513, 2002.

[11] S Roychoudhury, S Roy, B Dey, M Chakraborty, M Roy, B Roy, A Ramesh, N Prabhakaran, M V Usha Rani, H Vishwanathan, M Mitra, S K Sil and P P Majumder, Fundamental Genomic Unity of Ethnic India is Revealed by Analysis of Mitochondrial DNA, *Current Science*, Vol.79, pp. 1182-1192, 2000.

[12] Y Satta, J Klein and N Takahata, DNA archives and our nearest relative: The trichotomy problem revisited *Molecular Phylogenetics and Evolution*, Vol. 14, pp. 2590-275, 2000.

[13] I Tatersall, The case for saltational events in human evolution, In *The Speciation of Modern Homo sapiens*, ed TJ Crow, Oxford University Press, Oxford, pp.192-218, 2002.