Due to their genetic makeup, Indians are known to be susceptible to common diseases such as diabetes mellitus and coronary heart diseases. It might soon be possible to use biomarkers of genetic susceptibility to identify individuals with either enhanced or reduced risk of disease and to create clinically useful diagnostic as well as therapeutic methods.
In 1858, Charles Darwin proposed the theory of evolution. Based on anatomical similarities, he proposed that the chimpanzee and the gorilla are our closest living relatives and predicted that the earliest ancestors of humans would turn up in Africa. As predicted by Darwin, fossils have indeed been found in Africa. Based on these fossil records and greater genetic diversity within Africa when compared with outside and supported by many genetic markers, Africa is the most likely place of origin and dispersal of the modern humans.

It is largely agreed that modern humans originated about 160,000 years ago in East Africa. East African mega-drought could have caused their exodus out of Africa. Which route did they take?

**Populating India**

The Andaman & Nicobar Islands are in the Bay of Bengal almost equidistant (1500 km) from Kolkata as well as Chennai. The inhabitants of this archipelago are Andaman Islanders. Their origin is unknown.

They are negritos having short stature, dark skin, peppercorn hair and scant body hair similar to African pygmies, in contrast with other Asian pygmy people. Because of this, assumptions were made that they may be related to African pygmies. However, earlier study on blood group showed that they were more related to Oceanic people than Africans.

Our study based on the Y chromosome which is inherited paternally and mitochondrial DNA which is inherited maternally suggested that the native Andamanese belong to a group not previously identified in Asia or Africa suggesting an early split.

These tribes are close to extinction and therefore called vanishing tribes of Andaman Islands. Once upon a time, there were 13 linguistically distinct tribes of which only four survive today with a total population of each tribe not more than two hundred. Only two of these tribes, Jarawa and Sentinelese retain their forest dwelling lifestyle as hunters and gatherers.

Our study on the tribes of Andaman and Nicobar Islands using complete mitochondrial DNA sequences and its comparison with the mitochondrial DNA sequences of the world population, available in the database, lead to...
the theory of southern coastal route of migration to India against the prevailing view of northern route of migration via Middle East, Europe, South East Asia and Australia. This is now generally agreed international opinion. Genetic and archaeological data support a coastally oriented dispersal of modern humans from eastern Africa to southern Asia about 60 to 50 thousand years ago.

Why did they move out of Africa at the time when they did? Africa is a huge continent. It certainly was not overcrowded. East African mega droughts between 135 and 75 thousand years ago, when the water volume of the Lake Malawi was reduced by at least 95%, coincided with the migration of modern humans out of Africa.

Openheimer in his review article stated that “Astronomically modern human left Africa via single southern exit about 70,000 years ago and rapidly spread around the Indian Ocean towards the Antipodes, long before a small branch left a southern Asian colony earlier on the trail to populate Europe and single southern route out of Africa from the Red sea along the Indo-Pacific coast to Australia, including likely extension to China, Japan and New Guinea.” (Figure 1)

Our earlier study revealed that Negrito tribes of Andaman and Nicobar Islands, such as Onge, Jarawa, Great Andamanese and Sentinelese are probably the descendants of the first man to move out of Africa about 65,000 to 70,000 years ago. Analysis of the archaeological assemblages at Jwalapuram in Jurreru river valley of southern India recovered from archaeological industries from the site strongly suggests the presence of modern human in India about 74,000 years ago. This is consistent with our genetic data.

This raised many questions such as: (i) What is the origin of mainland tribal and caste populations? (ii) are there any population(s) in mainland India that are close to Andamanese? (iii) how much...

### Table 1: Populations analysed, their linguistic affiliation, social stratification and sampled area (adapted from)

<table>
<thead>
<tr>
<th>Group</th>
<th>Language Family</th>
<th>Traditional Caste or Social Designation</th>
<th>State/Territory</th>
<th>Sampling Location Nearest well known city, town or Island</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kashmiri Pandit</td>
<td>Indo-European</td>
<td>Upper caste</td>
<td>Kashmir</td>
<td>Dras</td>
</tr>
<tr>
<td>Vaish</td>
<td>Indo-European</td>
<td>Upper caste</td>
<td>Uttar Pradesh</td>
<td>Jaunpur</td>
</tr>
<tr>
<td>Srivastava</td>
<td>Indo-European</td>
<td>Upper caste</td>
<td>Uttar Pradesh</td>
<td>Mirzapur</td>
</tr>
<tr>
<td>Sahariya</td>
<td>Indo-European</td>
<td>Lower caste</td>
<td>Uttar Pradesh</td>
<td>Allahabad</td>
</tr>
<tr>
<td>Lodi</td>
<td>Indo-European</td>
<td>Lower caste</td>
<td>Uttar Pradesh</td>
<td>Jhansi</td>
</tr>
<tr>
<td>Satnami</td>
<td>Indo-European</td>
<td>Lower caste</td>
<td>Chhattisgarh</td>
<td>Raipur</td>
</tr>
<tr>
<td>Bhil</td>
<td>Indo-European</td>
<td>Tribal</td>
<td>Gujarat</td>
<td>Ahmedabad</td>
</tr>
<tr>
<td>Tharu</td>
<td>Indo-European</td>
<td>Tribal</td>
<td>Uttarkhand</td>
<td>Nainital</td>
</tr>
<tr>
<td>Meghawal</td>
<td>Indo-European</td>
<td>Lower caste</td>
<td>Rajasthan</td>
<td>Jodhpur</td>
</tr>
<tr>
<td>Vysya</td>
<td>Dravidian</td>
<td>Middle caste</td>
<td>Andhra Pradesh</td>
<td>Anantapur</td>
</tr>
<tr>
<td>Naidu</td>
<td>Dravidian</td>
<td>Upper caste</td>
<td>Andhra Pradesh</td>
<td>Chittoor</td>
</tr>
<tr>
<td>Velama</td>
<td>Dravidian</td>
<td>Upper caste</td>
<td>Andhra Pradesh</td>
<td>Mahaboob Nagar</td>
</tr>
<tr>
<td>Madiga</td>
<td>Dravidian</td>
<td>Lower caste</td>
<td>Andhra Pradesh</td>
<td>Warangal</td>
</tr>
<tr>
<td>Mala</td>
<td>Dravidian</td>
<td>Lower caste</td>
<td>Andhra Pradesh</td>
<td>Hyderabad</td>
</tr>
<tr>
<td>Kamsali</td>
<td>Dravidian</td>
<td>Lower caste</td>
<td>Andhra Pradesh</td>
<td>Kurnool</td>
</tr>
<tr>
<td>Chenchu</td>
<td>Dravidian</td>
<td>Tribal</td>
<td>Andhra Pradesh</td>
<td>Anantapur</td>
</tr>
<tr>
<td>Kurumba</td>
<td>Dravidian</td>
<td>Tribal</td>
<td>Kerala</td>
<td>Palakkad</td>
</tr>
<tr>
<td>Hallaki</td>
<td>Dravidian</td>
<td>Tribal</td>
<td>Karnataka</td>
<td>Uttarakhand</td>
</tr>
<tr>
<td>Santhal</td>
<td>Austro-Asiatic</td>
<td>Tribal</td>
<td>Jharkhand</td>
<td>Santhal Pargana</td>
</tr>
<tr>
<td>Kharia</td>
<td>Austro-Asiatic</td>
<td>Tribal</td>
<td>Madhya Pradesh</td>
<td>Raigarh</td>
</tr>
<tr>
<td>Nyshi</td>
<td>Tibeto-Burman</td>
<td>Tribal</td>
<td>Arunachal Pradesh</td>
<td>Papum Pare</td>
</tr>
<tr>
<td>Ao Naga</td>
<td>Tibeto-Burman</td>
<td>Tribal</td>
<td>Nagaland</td>
<td>Kohima</td>
</tr>
<tr>
<td>Siddi</td>
<td>Dravidian</td>
<td>Tribal</td>
<td>Karnataka</td>
<td>Dharwand</td>
</tr>
<tr>
<td>Onge</td>
<td>Jarawa-Onge</td>
<td>Hunter Gatherer</td>
<td>Andaman &amp; Nicobar</td>
<td>Little Andaman</td>
</tr>
<tr>
<td>Gr. Andamanese</td>
<td>Andamanese</td>
<td>Hunter Gatherer</td>
<td>Andaman &amp; Nicobar</td>
<td>Great Andaman</td>
</tr>
</tbody>
</table>

42 SCIENCE REPORTER May 2017
affinity the Indian populations have with Andamanese? (iv) did the Indians contribute to the early human spread?

In order to answer these questions and to explore the ancient history of India, genomic technology was exploited.

**Ancient Roots**

Although the genome sequences of any two unrelated people differ by just 0.1%, they provide information that can help to reconstruct the historical origins of modern populations. They also point to genetic variations that heighten the risk of certain diseases.

After sequencing mitochondrial DNA from 14,000 samples from 240 ethnic populations, two tribes, Kurumba from Kerala (West Coast) and Rajbanshi from West Bengal (East Coast) were found to share some mutations (change of DNA base sequence) with those of Andamanese answering the question that there are populations in mainland India which show some similarity in sharing the sets of mutations.

To shed light on the genetic variability across the Indian subcontinent, 132 Indian samples from 25 groups were analyzed on an Affymetrix 6.0 array of one million Single Nucleotide Polymorphisms (SNPs) spanning the widest range of ancestry in India. Tribal groups were sampled from 13 states and 6 language families (Indo European, Dravidian, Austro-Asiatic, Tibeto-Burman, Great Andamanese and Jarawa-Onge) [Table 1]. Some caste groups mostly from Uttar Pradesh and Andhra Pradesh were sampled to permit comparison of traditionally “Upper” and “Lower” caste groups after providing a comparison of traditionally “Upper” and “Lower” caste groups mostly from Uttar Pradesh and Europeans, while the “Ancestral South Indians” (ASI) are not close to any group outside the subcontinent. It is shown that ANI ancestry ranges from 39-71%, and is higher in traditionally upper caste groups and Indo-European speakers. Groups with only ASI ancestry may no longer exist in mainland India.

Our study has revealed that nearly all Indian groups descend from mixtures of two ancestral populations, ASI and ANI. This applies to traditional “tribes” as well as “castes”. It is impossible to distinguish castes from tribes using the data. The genetics proves that they are not different. This supports the view that castes grew directly out of tribal-like organizations during the formation of Indian society.

The one exception to the finding that all Indian groups are mixed is the indigenous people of the Andaman Islands. The Andamanese appear to be related exclusively to the ASI lineage and therefore lack ANI ancestry. In this sense, they are unique. Understanding their origins provides a window to look into the history of the ASI and the period of thousands of years ago when they diverged from other Eurasians. The Andamanese are the only surviving remnants of the ancient colonizers of South Asia.

**Indo-European Family of Language**

In order to rule India, the British East India Company needed to know its people, their language, their history and culture for which at the instance of Warren Hastings, the then Governor General, the Royal Asiatic Society of Calcutta was established in the year 1784, with Sir William Jones as its President.

William Jones was a linguist and a scholar of Greek, Latin, Sanskrit and a number of other languages. He discovered inexplicable resemblance between Sanskrit, Greek and Latin languages. Considering that Indian civilisation is one of the oldest civilisations in the world, the only convincing explanation for this could have been that Sanskrit had influenced the Greek and Latin.

However, it was beyond their comprehension that a language of conquered people could have even remotely influenced any European language. So, they suggested that there was an ancient parent language of which Sanskrit, Greek and Latin are off-springs and which perhaps no longer exists. Sir William Jones propounded this theory in his Presidential address at a meeting of Royal Asiatic Society held on 2 February 1786. This statement of Jones is pure assumption and not supported by any concrete linguistic archaeological or literary evidence.

He made important identifications of words in the Romanic of Gypsy languages with Sanskrit. This is not surprising as our recent findings based on Y Chromosome Short Tandem Repeat (STR) network (Haplotype H1a1a-M82) lineages clearly show closest connection of Romani with the traditional scheduled caste and scheduled tribe populations of northwestern India from where they migrated to Europe about 1400 years ago. Their origin can be traced back to their ancestors in Tamil Nadu about 24000 years ago from where they migrated.

Max Muller, who was one of the first to apply the Aryan name to the Indo-European concept, indentified the racial-linguistic entity as racially white and was instrumental in the formulation of the racial theory of Indian civilization.

‘Arya’, is a Sanskrit word meaning cultured and noble. Never in the entire history of Sanskrit literature Arya has been used to indicate a biological race. The discredit for using it, as such, goes to Max Muller. It caused maximum damage to Europe where Hitler lapped
it up to strengthen German nationalism and brought brutality and massacre to its people through the Second World War.

Max Muller conjectured that people living somewhere in Central Asia, near Caspian Sea, and speaking Proto Indo-European language constituted the Aryan race. Aryans migrated westward to Europe, while the other group known as Indo-Iranians migrated southwards to Iran.

A group of these Indo-Iranians termed as Indo-Aryans moved to India. In India, they invaded and defeated the indigenous people (Dravidian) around 15th Century BC, pushed them to south India, occupied ‘Sapt-Sindhu’ (the land of seven rivers, their territory in India, occupied ‘Sapt-Sindhu’ (the land of seven rivers, their territory in India and later composed Rig-Veda in 12th Century BC there.

However, the Aryan invasion theory was then modified into the Aryan Migration Theory. According to this theory the Aryans migrated to India from Iran via Afghanistan peacefully, in groups of various sizes over a period, and in due course gave their language and culture to the indigenous people living there.

This theory is actually based on conjecture. It is illogical to even think that small groups of a few hundred or even a thousand people coming from foreign land could replace the language and culture of large populations of indigenous people settled there for thousands of years.

Trautmann has extensively written against the thinking that saw race in terms of colour: “The three primary races are Caucasian, the Mongolian and the Negroid. Both Aryans and Dravidians are related branches of the Caucasian race. Biologically both the North and South Indians are of the same Caucasian race, only when close to the equator the skin becomes darker. Caucasian race is not the white race. It can be of any colour from pure white to almost pure black with every shade of brown in between.”

The kings of South India like the Chola and the Pandya dynasties relate their lineages back to Manu. The Matsya Purana more over makes Manu, the progenitor of all the Aryans, originally a south Indian king, Satyavrata. This makes the Aryans of North India descendants of Dravidian kings. The two cultures are so intimately related that it is difficult to say which came first. This is consistent with our research findings that mixture between different populations was rampant until endogamy (marrying within the local community) and cast system developed. This supports the view of one school of thought that the Aryans and Dravidians are part of the same culture. However, it contradicts the second school of historians such as Max Muller who for the first time applied the Aryan name to the Indo-European concept.

Our recent study revealed that well after the establishment of agriculture in the subcontinent, an original mixture event of ANI and ASI that contributed to all present-day Indians, the period of around 1,900-4,200 years BP (before present), witnessed the shift from widespread mixture to strict endogamy. This is the period during which the first appearance of Indo-European languages and Vedic religion in the sub-continent was witnessed and characterised by the de-urbanization of the Indus civilization. The Rig Veda, the oldest text in India has sections that are believed to have been composed at different times. The oldest part does not mention the caste system at all. The four-class (verna) system, comprised of Brahmans, Ksatiryas, Vaisyas and Sudras is mentioned only in the part of the Rig Veda that was likely to have been composed centuries later for example the law code of Manu that forbade inter marriage between castes. This is consistent with our genetic findings.

Our study has provided evidence for mixture in India between about 1,900 and 4,200 years BP. This does not support the theory of Aryan invasion or migration from West Eurasia into India during this period. In fact on the contrary, a recent study searched for evidence in West Eurasian groups most closely related to the ANI ancestors of Indians but failed to find any evidence for shared ancestry between the ANI and groups in West Eurasia within the past 12500 years.

The ANI-ASI mixture was so widely spread that it affected not only traditionally upper-caste groups, but also traditionally lower-caste and isolated tribal groups. This can be satisfactorily explained by assuming that mixture between ANI and ASI occurred much before the development of caste system in India. It is possible that the ANI and ASI were living in or near South Asia for a long period prior to their mixture. These evidences strongly reject the theory of Aryan invasion/migration.

In early Holocene, South Asia was a continent of hunter gatherers. During and after Neolithic period, agriculture developed especially in the fertile river valley. The oldest evidence of agriculture comes from the fertile crescent, the centre for demic diffusion (about 11 thousand years ago). However, new archaeological work shows that plants were domesticated independently in many parts of the globe including India.
The global distribution of Y chromosome sub-haplo groups J2-M172 has been associated with Neolithic demic diffusion in Eurasia. Two branches of J2-M127, J2a-M410 and J2h-M102 make a considerable part of Y-chromosome gene pool of the Indian subcontinent. Contrary to demic spread and single wave of agricultural expansion in Eurasia, our study has revealed an ancient arrival of these haplogroups into Indian subcontinent.

Remarkable presence of J2a-M410 among tribal groups inhabited in the remote geographical regions strongly dismisses the earlier belief of it being caste-specific and associated with demic diffusion of Neolithic farmers in North Africa and Eurasia from Mesopotamia (Iraq and Syria) and as a result of gene influx from Western Asia.

**Genomics and Public Health**

Our study has important medical implications. It provides evidence that nearly all Indian groups including traditional “tribes” and “castes” are admixture of ASI and ANI populations. There was rampant mixing of North and South populations.

However, about 1,900-4,200 years ago strict endogamy (marrying within the tribe and caste) developed. These have serious impacts on occurrence of genetic diseases. The myosin binding protein – C3 (MYBPC 3 gene) – is involved in about 45% of the total cardiomyopathy. This is associated with inheritable cardiomyopathies in the Indian patients in which 25 base pairs of intron between exon 32 and 33 are deleted causing skipping of wild type exon 33 during the translation of the protein. Presence of both copies of this mutated gene (homozygous) is lethal – the affected child dies before birth or at a very early age.

In heterozygous condition, where only one of the two genes has the deletion mutation (heterozygous), the individual lives up to the age of 50 to 55 years and then dies because of sudden cardiac arrest without showing any prior symptoms. This mutation is found throughout the Indian sub-continent and is absent in all other countries particularly the western countries. Our detailed study revealed that this mutation originated about 33,000 years ago. Its distribution shows the relatedness of these populations2. We have developed PCR-based screening for carrier detections.

We discovered that many groups in modern India descended from a small number of founding individuals and have since been genetically isolated from other groups. This means that India is genetically not a single large population, but instead is best described as many smaller isolated populations.

There are a number of genetic diseases caused by the same gene in Caucasians and Indians but the mutations found in that gene are different between the two populations. Therefore, it is a must that mutations in those genes are worked out in Indian populations for prenatal diagnosis, carrier detection and genetic counselling. We cannot use the data from western populations to extrapolate for our populations.

Our study predicts that the incidence of recessive genetic diseases will be higher in many groups in India. Further study of these groups should lead to rapid discovery of genes that cause devastating diseases and would help in the clinical care of individuals and their families who are at risk.

Examples of some Founder events leading to population-specific diseases in India include (i) pseudo cholinesterase deficiency in the Vysya community of Andhra Pradesh; (ii) familial Madras motor neuron disease (FMMND) in Tirupathi–Chittoor, southern India; (iii) sickle cell anemia in some north & north-east Indian populations, including the tribes of Chhattisgarh, Madhya Pradesh, Orissa, etc.; (iv) Handigodu syndrome in Sagara taluk of the Shimoga district of Karnataka; (v) predominance of breast cancer in the Parsi community, and (vi) sudden cardiac arrest in the South Asians.

Since about 70% of the Indian population lives in villages and most of the health-statistics of the country does not include these areas, we have established the ‘Genome Foundation (GF)’ which aims to utilize the knowledge available in the public domain for the welfare of the underprivileged.

More than 6,000 human diseases are currently classified as resulting from the action of a single mutant gene. It is indeed unfortunate that only after there is an affected child in the family that parents or relatives get concerned with the genetic nature of the disease. Prospective approaches are required for combating disorders of high prevalence and with known genetic basis such as thalassemias, sickle cell anemia, haemoglobin e disease, muscular dystrophy, neurodegenerative diseases, mental retardation, haemophilia, etc.; and this demands a high level of awareness among the general public as well as the clinicians.

Thalassemia was completely eradicated in Sardinia by adopting mandatory screening and genetic counselling for the whole population. In the absence of specific treatment and gene therapy being a long cherished goal, it is very important to follow the concept of molecular diagnosis, carrier detection, genetic counselling, pre-pregnancy monitoring, pre-implantation genetic diagnosis and prenatal diagnosis.

The identification of genetic defects(s) in the patients can be utilized as a handle to track the inheritance of defective gene in the foetus by performing foetal sampling procedure during early pregnancy. In this backdrop, genetic diagnosis has immense potential with particular stress on prospective screening and counselling for common disorders such as thalassemias, muscular dystrophies, etc.

Several diseases prevalent in India are yet to be identified and reported; and their genetic basis is unknown. For example, in our preliminary survey of the rural part of U.P in northern India, we observed a family whose female members, over four generations, do not have fingers and toes; families having several male and female individuals affected by muscular dystrophy and many families which have only disabled children born.

Due to their genetic makeup,
Indians are known to be susceptible to common diseases such as diabetes mellitus and coronary heart diseases. It might soon be possible to use biomarkers of genetic susceptibility to identify individuals with either enhanced or reduced risk of disease and to create clinically useful diagnostic as well as therapeutic methods. Such a programme would be a quantum leap forward in preventive medicine, practiced on genotype-specific basis.

**Precision Medicine**

It is perhaps very pertinent to point out that the concept of personalised medicine based on each person’s prakriti (Vata, Pita and Kapha) or constitution to maintain balance between mind and spirit, for the first time in the world was laid out in Ayurveda about 4000-200 B.C. Recently it has been suggested that “the phenotypic classification of India’s traditional medicine has a genetic basis and its Prakrity-based practice in vogue for many centuries resonates with personalized medicine”.

All diseases are genetic. Finding a disease-gene enables one to predict those who are at risk even before they develop the disease, or those whose children might be at risk even before conception. Much more common diseases such as heart disease, cerebral vascular disease, emphysema, bronchitis, pneumonia, influenza, cancer, diabetes, suicide, liver diseases all of these are genetic. The diagnosis of disease-predisposing genes will alter the basic practice of medicine in future.

Why do some drugs work better in some patients than in others? Some drugs may even be highly toxic to certain individuals. With the advances in Pharmacogenetics (Pharmacogenetics), it can be used to optimize drug prescription (Pharmacogenetics). For example, a cancer drug called ‘genetinib’ causes dramatic shrinkage of tumour in only about 10% of lung cancer patients who carry mutations in the EGF (epidermal growth factor) receptor gene which is tyrosine kinase that regulates proteins by attaching phosphate groups to them. The non-responders do not carry this mutation.

The key question however is whether the genomic medicine/personalized medicine will improve the health of all people or whether it will just widen the technology gap between the rich and the poor. The Genome Foundation is, therefore, making efforts to take this technology to the rural populations.

---

**References**