

JUST IN

8 **2hrs** Challenges galore as Adityanath completes 100 days as

9 **3hrs** School owner, teacher arrested for stripping girl students

10 **3hrs** No ICC World T20 in 2018, next edition in 2020

11 **3** rur figl

Ground Zero - In-depth reportage from The Hindu

How genetics is settling the Aryan migration debate

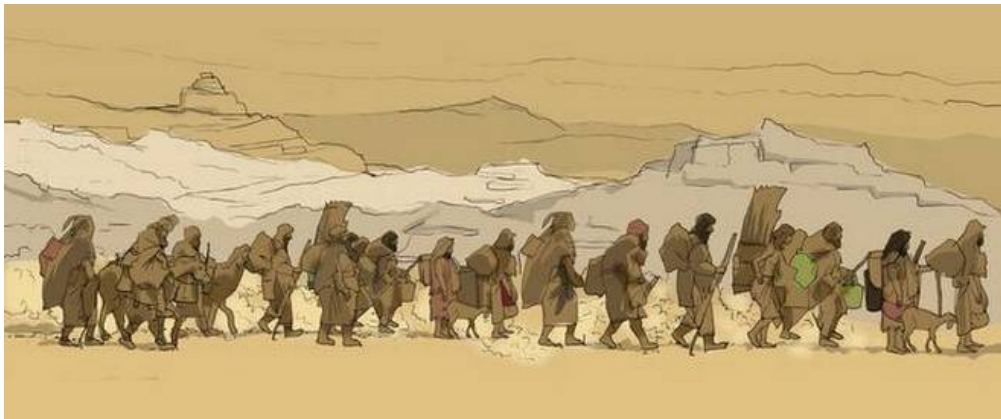


Tony Joseph

JUNE 16, 2017 23:49 IST

UPDATED: JUNE 17, 2017 13:16 IST

SHARE ARTICLE | f 46.18K | t | i | e | o | PRINT | A | A | A



New DNA evidence is solving the most fought-over question in Indian history. And you will be surprised at how sure-footed the answer is, writes Tony Joseph

The thorniest, most fought-over question in Indian history is slowly but surely getting answered: did Indo-European language speakers, who called themselves Aryans, stream into India sometime around 2,000 BC – 1,500 BC when the Indus Valley civilisation came to an end, bringing with them Sanskrit and a distinctive set of cultural practices? Genetic research based on an avalanche of new DNA evidence is making scientists around the world converge on an unambiguous answer: yes, they did.

This may come as a surprise to many – and a shock to some – because the dominant narrative in recent years has been that genetics research had thoroughly disproved the Aryan migration theory. This interpretation was always a bit of a stretch as anyone who read the nuanced scientific papers in the original knew. But now it has broken apart altogether under a flood of new data on Y-chromosomes (or chromosomes that are transmitted through the male parental line, from father to son).

Lines of descent

Until recently, only data on mtDNA (or matrilineal DNA, transmitted only from mother to daughter) were available and that seemed to suggest there was little external infusion into the Indian gene pool over the last 12,500 years or so. New Y-DNA data has turned that conclusion upside down, with strong evidence of external infusion of genes into the Indian male lineage during the period in question.

The reason for the difference in mtDNA and Y-DNA data is obvious in hindsight: there was strong sex bias in Bronze Age migrations. In other words, those who migrated were predominantly male and, therefore, those gene flows do not really show up in the mtDNA data. On the other hand, they do show up in the Y-DNA data: specifically, about 17.5% of Indian male lineage has been found to belong to haplogroup R1a (haplogroups identify a single line of descent), which is today spread across Central Asia, Europe and South Asia. Pontic-Caspian Steppe is seen as the region from where R1a spread both west and east, splitting into different sub-branches along the way.

The paper that put all of the recent discoveries together into a tight and coherent history of migrations into India was published just three months ago in a peer-reviewed journal called 'BMC Evolutionary Biology'. In that paper, titled "A Genetic Chronology for the Indian Subcontinent Points to Heavily Sex-biased Dispersals", 16 scientists led by Prof. Martin P. Richards of the University of Huddersfield, U.K., concluded: "Genetic influx from Central Asia in the Bronze Age was strongly male-driven, consistent with the patriarchal, patrilocal and patrilineal social structure attributed to the inferred pastoralist early Indo-European society. This was part of a much wider process of Indo-European expansion, with an ultimate source in the Pontic-Caspian region, which carried closely related Y-chromosome lineages... across a vast swathe of Eurasia between 5,000 and 3,500 years ago".

In an email exchange, Prof. Richards said the prevalence of R1a in India was "very powerful evidence for a substantial Bronze Age migration from central Asia that most likely brought Indo-European speakers to India." The robust conclusions of Professor Richards and his team rest on their own substantive research as well as a vast trove of new data and findings that have become available in recent years, through the work of genetic scientists around the world.

What's happened very rapidly, dramatically, and powerfully in the last few years has been the explosion of genome-wide studies of human history based on modern and ancient DNA, and that's been enabled by the technology of genomics

Peter Underhill, scientist at the Department of Genetics at the Stanford University School of Medicine, is one of those at the centre of the action. Three years ago, a team of 32 scientists he led published a massive study mapping the distribution and linkages of R1a. It used a panel of 16,244 male subjects from 126 populations across Eurasia. Dr. Underhill's research found that R1a had two sub-haplogroups, one found primarily in

and the technology of ancient DNA...” David Reich, Geneticist and professor, Harvard Medical School

Europe and the other confined to Central and South Asia. Ninety-six per cent of the R1a samples in Europe belonged to sub-haplogroup Z282, while 98.4% of the Central and South Asian R1a lineages belonged to sub-haplogroup Z93. The two groups diverged from each other only

about 5,800 years ago. Dr. Underhill’s research showed that within the Z93 that is predominant in India, there is a further splintering into multiple branches. The paper found this “star-like branching” indicative of rapid growth and dispersal. So if you want to know the approximate period when Indo-European language speakers came and rapidly spread across India, you need to discover the date when Z93 splintered into its own various subgroups or lineages. We will come back to this later.

So in a nutshell: R1a is distributed all over Europe, Central Asia and South Asia; its sub-group Z282 is distributed only in Europe while another subgroup Z93 is distributed only in parts of Central Asia and South Asia; and three major subgroups of Z93 are distributed only in India, Pakistan, Afghanistan and the Himalayas. This clear picture of the distribution of R1a has finally put paid to an earlier hypothesis that this haplogroup perhaps originated in India and then spread outwards. This hypothesis was based on the erroneous assumption that R1a lineages in India had huge diversity compared to other regions, which could be indicative of its origin here. As Prof. Richards puts it, “the idea that R1a is very diverse in India, which was largely based on fuzzy microsatellite data, has been laid to rest” thanks to the arrival of large numbers of genomic Y-chromosome data.

Gene-dating the migration

Now that we know that there WAS indeed a significant inflow of genes from Central Asia into India in the Bronze Age, can we get a better fix on the timing, especially the splintering of Z93 into its own sub-lineages? Yes, we can; the research paper that answers this question was published just last year, in April 2016, titled: “Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences.” This paper, which looked at major expansions of Y-DNA haplogroups within five continental populations, was lead-authored by David Poznik of the Stanford University, with Dr. Underhill as one of the 42 co-authors. The study found “the most striking expansions within Z93 occurring approximately 4,000 to 4,500 years ago”. This is remarkable, because roughly 4,000 years ago is when the Indus Valley civilization began falling apart. (There is no evidence so far, archaeologically or otherwise, to suggest that one caused the other; it is quite possible that the two events happened to coincide.)

The avalanche of new data has been so overwhelming that many scientists who were either sceptical or neutral about significant Bronze Age migrations into India have changed their opinions. Dr. Underhill himself is one of them. In a 2010 paper, for example, he had written that there was evidence “against substantial patrilineal gene flow from East Europe to Asia, including to India” in the last five or six millennia. Today, Dr. Underhill says there is no comparison between the kind of data available in 2010 and now. “Then, it was like looking

into a darkened room from the outside through a keyhole with a little torch in hand; you could see some corners but not all, and not the whole picture. With whole genome sequencing, we can now see nearly the entire room, in clearer light.”

Dr. Underhill is not the only one whose older work has been used to argue against Bronze Age migrations by Indo-European language speakers into India. David Reich, geneticist and professor in the Department of Genetics at the Harvard Medical School, is another one, even though he was very cautious in his older papers. The best example is a study lead-authored by Reich in 2009, titled “Reconstructing Indian Population History” and published in *Nature*. This study used the theoretical construct of “Ancestral North Indians” (ANI) and “Ancestral South Indians” (ASI) to discover the genetic substructure of the Indian population. The study proved that ANI are “genetically close to Middle Easterners, Central Asians, and Europeans”, while the ASI were unique to India. The study also proved that most groups in India today can be approximated as a mixture of these two populations, with the ANI ancestry higher in traditionally upper caste and Indo-European speakers. By itself, the study didn’t disprove the arrival of Indo-European language speakers; if anything, it suggested the opposite, by pointing to the genetic linkage of ANI to Central Asians.

However, this theoretical structure was stretched beyond reason and was used to argue that these two groups came to India tens of thousands of years ago, long before the migration of Indo-European language speakers that is supposed to have happened only about 4,000 to 3,500 years ago. In fact, the study had included a strong caveat that suggested the opposite: “We caution that ‘models’ in population genetics should be treated with caution. While they provide an important framework for testing historical hypothesis, they are oversimplifications. For example, the true ancestral populations were probably not homogenous as we assume in our model but instead were likely to have been formed by clusters of related groups that mixed at different times.” In other words, ANI is likely to have resulted from multiple migrations, possibly including the migration of Indo-European language speakers.

The spin and the facts

But how was this research covered in the media? “Aryan-Dravidian divide a myth: Study,” screamed a newspaper headline on September 25, 2009. The article quoted Lalji Singh, a co-author of the study and a former director of the Centre for Cellular and Molecular Biology (CCMB), Hyderabad, as saying: “This paper rewrites history... there is no north-south divide”. The report also carried statements such as: “The initial settlement took place 65,000 years ago in the Andamans and in ancient south India around the same time, which led to population growth in this part. At a later stage, 40,000 years ago, the ancient north Indians emerged which in turn led to rise in numbers there. But at some point in time, the ancient north and the ancient south mixed, giving birth to a different set of population. And that is the population which exists now and there is a genetic relationship between the population within India.” The study, however, makes no such statements whatsoever – in fact, even the figures 65,000 and 40,000 do not figure in it!

This stark contrast between what the study says and what the media reports

said did not go unnoticed. In his column for *Discover* magazine, geneticist Razib Khan said this about the media coverage of the study: “But in the quotes in the media the other authors (other than Reich that is - ed) seem to be leading you to totally different conclusions from this. Instead of leaning toward ANI being proto-Indo-European, they deny that it is.”

Let's leave that there, and ask what Reich says now, when so much new data have become available? In an interview with *Edge* in February last year, while talking about the thesis that Indo-European languages originated in the Steppes and then spread to both Europe and South Asia, he said: “The genetics is tending to support the Steppe hypothesis because in the last year, we have identified a very strong pattern that this ancient North Eurasian ancestry that you see in Europe today, we now know when it arrived in Europe. It arrived 4500 years ago from the East from the Steppe...” About India, he said: “In India, you can see, for example, that there is this profound population mixture event that happens between 2000 to 4000 years ago. It corresponds to the time of the composition of the Rigveda, the oldest Hindu religious text, one of the oldest pieces of literature in the world, which describes a mixed society...” In essence according to Reich, in broadly the same time frame, we see Indo-European language speakers spreading out both to Europe and to South Asia, causing major population upheavals.

The dating of the “profound population mixture event” that Reich refers to was arrived at in a paper that was published in the *American Journal of Human Genetics* in 2013, and was lead authored by Priya Moorjani of the Harvard Medical School, and co-authored, among others, by Reich and Lalji Singh. This paper too has been pushed into serving the case against migrations of Indo-European language speakers into India, but the paper itself says no such thing, once again!

Here's what it says in one place: “The dates we report have significant implications for Indian history in the sense that they document a period of demographic and cultural change in which mixture between highly differentiated populations became pervasive before it eventually became uncommon. The period of around 1,900–4,200 years before present was a time of profound change in India, characterized by the de-urbanization of the Indus civilization, increasing population density in the central and downstream portions of the Gangetic system, shifts in burial practices, and the likely first appearance of Indo-European languages and Vedic religion in the subcontinent.”

The study didn't “prove” the migration of Indo-European language speakers since its focus was different: finding the dates for the population mixture. But it is clear that the authors think its findings fit in well with the traditional reading of the dates for this migration. In fact, the paper goes on to correlate the ending of population mixing with the shifting attitudes towards mixing of the races in ancient texts. It says: “The shift from widespread mixture to strict endogamy that we document is mirrored in ancient Indian texts.”

So irrespective of the use to which Priya Moorjani et al's 2013 study is put, what is clear is that the authors themselves admit their study is fully compatible with, and perhaps even strongly suggests, Bronze Age migration of Indo-

European language speakers. In an email to this writer, Moorjani said as much. In answer to a question about the conclusions of the recent paper of Prof. Richards et al that there were strong, male-driven genetic inflows from Central Asia about 4,000 years ago, she said she found their results “to be broadly consistent with our model”. She also said the authors of the new study had access to ancient West Eurasian samples “that were not available when we published in 2013”, and that these samples had provided them additional information about the sources of ANI ancestry in South Asia.

One by one, therefore, every single one of the genetic arguments that were earlier put forward to make the case against Bronze Age migrations of Indo-European language speakers have been disproved. To recap:

1. The first argument was that there were no major gene flows from outside to India in the last 12,500 years or so because mtDNA data showed no signs of it. This argument was found faulty when it was shown that Y-DNA did indeed show major gene flows from outside into India within the last 4000 to 4,500 years or so, especially R1a which now forms 17.5% of the Indian male lineage. The reason why mtDNA data behaved differently was that Bronze Age migrations were severely sex-biased.

2. The second argument put forward was that R1a lineages exhibited much greater diversity in India than elsewhere and, therefore, it must have originated in India and spread outward. This has been proved false because a mammoth, global study of R1a haplogroup published last year showed that R1a lineages in India mostly belong to just three subclades of the R1a-Z93 and they are only about 4,000 to 4,500 years old.

3. The third argument was that there were two ancient groups in India, ANI and ASI, both of which settled here tens of thousands of years earlier, much before the supposed migration of Indo-European languages speakers to India. This argument was false to begin with because ANI – as the original paper that put forward this theoretical construct itself had warned – is a mixture of multiple migrations, including probably the migration of Indo-European language speakers.

Connecting the dots

Two additional things should be kept in mind while looking at all this evidence. The first is how multiple studies in different disciplines have arrived at one specific period as an important marker in the history of India: around 2000 B.C. According to the Priya Moorjani et al study, this is when population mixing began on a large scale, leaving few population groups anywhere in the subcontinent untouched. The Onge in the Andaman and Nicobar Islands are the only ones we know to have been completely unaffected by what must have been a tumultuous period. And according to the David Poznik et al study of 2016 on the Y-chromosome, 2000 B.C. is around the time when the dominant R1a subclade in India, Z93, began splintering in a “most striking” manner, suggesting “rapid growth and expansion”. Lastly, from long-established archaeological studies, we also know that 2000 BC was around the time when the Indus Valley civilization began to decline. For anyone looking at all of these data objectively, it is difficult to avoid the feeling that the missing pieces of

India's historical puzzle are finally falling into place.

The second is that many studies mentioned in this piece are global in scale, both in terms of the questions they address and in terms of the sampling and research methodology. For example, the Poznik study that arrived at 4,000-4,500 years ago as the dating for the splintering of the R1a Z93 lineage, looked at major Y-DNA expansions not just in India, but in four other continental populations. In the Americas, the study proved the expansion of haplogroup Q1a-M3 around 15,000 years ago, which fits in with the generally accepted time for the initial colonisation of the continent. So the pieces that are falling in place are not merely in India, but all across the globe. The more the global migration picture gets filled in, the more difficult it will be to overturn the consensus that is forming on how the world got populated.

Nobody explains what is happening now better than Reich: "What's happened very rapidly, dramatically, and powerfully in the last few years has been the explosion of genome-wide studies of human history based on modern and ancient DNA, and that's been enabled by the technology of genomics and the technology of ancient DNA. Basically, it's a gold rush right now; it's a new technology and that technology is being applied to everything we can apply it to, and there are many low-hanging fruits, many gold nuggets strewn on the ground that are being picked up very rapidly."

So far, we have only looked at the migrations of Indo-European language speakers because that has been the most debated and argued about historical event. But one must not lose the bigger picture: R1a lineages form only about 17.5 % of Indian male lineage, and an even smaller percentage of the female lineage. The vast majority of Indians owe their ancestry mostly to people from other migrations, starting with the original Out of Africa migrations of around 55,000 to 65,000 years ago, or the farming-related migrations from West Asia that probably occurred in multiple waves after 10,000 B.C., or the migrations of Austro-Asiatic speakers such as the Munda from East Asia the dating of which is yet to be determined, and the migrations of Tibeto-Burman speakers such as the Garo again from east Asia, the dating of which is also yet to be determined.

What is abundantly clear is that we are a multi-source civilization, not a single-source one, drawing its cultural impulses, its tradition and practices from a variety of lineages and migration histories. The Out of Africa immigrants, the pioneering, fearless explorers who discovered this land originally and settled in it and whose lineages still form the bedrock of our population; those who arrived later with a package of farming techniques and built the Indus Valley civilization whose cultural ideas and practices perhaps enrich much of our traditions today; those who arrived from East Asia, probably bringing with them the practice of rice cultivation and all that goes with it; those who came later with a language called Sanskrit and its associated beliefs and practices and reshaped our society in fundamental ways; and those who came even later for trade or for conquest and chose to stay, all have mingled and contributed to this civilization we call Indian. We are all migrants.

Tony Joseph is a writer and former editor of BusinessWorld. Twitter: @tjoseph0010