J Biosci (2019) 44:70
DOI: 10.1007/s12038-019-9874-6

Peopling of India: Ancient DNA perspectives

K Thangaraj[1]* and Niraj Rai[2]

[1]*CSIR-Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India*

[2]*Birbal Sahni Institute of Palaeosciences, University Road, Lucknow, India*

**Corresponding author (Email, thangs@ccmb.res.in)*

Published online 8 July 2019

To reconstruct and explain patterns of genetic diversity of modern humans, understanding their past and present genetic profile is crucial. While genomes of contemporary people can provide information about present day population structure, analysis of ancient genomes may provide unprecedented insights about the past demographic events that have shaped the contemporary gene pool. Population genetics has recently witnessed an explosion in studies on ancient human population histories, primarily from Europe and America. South Asia has no representation in the ancient genomics literature, despite the wealth of archaeological richness in the form of human skeletal remains that exist in collections all over the country. Representing one-fifth of present day humanity calls for understanding the demographic history of south Asia not merely as a prerequisite but as an urgent need to understand its genetic variations on a global scale. Although the overall picture is taking form, new archaeological and genetic information from the region has started to reveal a more complex scenario of ancient human migrations and admixtures than was ever known before. In this article, we discuss a meaningful insight on the current status of ancient DNA (aDNA) research in India. We have also summarized a few but important aDNA studies, which have been successfully carried out in India. Furthermore, we have highlighted the potential opportunity of aDNA research in the Indian subcontinent.

Keywords. Ancient DNA; archaeology; population genetics; human burials

1. Introduction

The Indian subcontinent is one of the most densely populated parts of the world having enormous linguistic, cultural and genetic diversities, outdone only by the continent of Africa. Phenotypically, some groups in India resemble Africans, some look similar to Europeans, yet others resemble Southeast Asians, and most importantly, a great majority are typical to India. Scholars from different disciplines—history, language, archaeology, anthropology and genetics—have tried to understand how such diversity arose and exists in India. However, most conclusive evidences about the arrival and spread of modern human population in India are based on genetic studies. The diversity in India can be attributed to the migration of human population, starting from the first arrival of mankind on the subcontinent during the Late Pleistocene to prehistoric and historical movements of populations within India, and the movement of ideas and economies, namely, the emergence and spread of agriculture in certain parts of India (Moorjani *et al.* 2013). Population genetics has recently witnessed an explosion in studies on ancient human population histories, primarily from Europe and America (Lazaridis *et al.* 2014; Moreno-Mayar *et al.* 2018; Posth *et al.* 2018). South Asia has no representation in the ancient genomics literature, and not much work has been done on contemporary Indian populations despite the wealth of population diversity and archaeological richness in the form of human skeletal remains that exist in collections all over the country (Chakrabarti 1982). Representing one-fifth of the present day humanity calls for understanding the demographic history of south Asia not merely as a prerequisite but as an urgent need to understand its genetic variations on a global scale. Almost all the genetic studies performed in India are based on the DNA from contemporary populations. It would be of greater interest if data from ancient samples were available, which would provide information about our past and help us in connecting some of the missing links. Although the overall picture is still, at best, emerging, new archaeological and genetic information from the region has started to reveal a more complex scenario of ancient human migrations and admixtures than earlier.

2. Do we have ancient DNA evidence in India?

Unfortunately, we do not have analysable ancient biological remains in India to understand our past. Although archeological evidences suggest that the hominins were present in the Jurreru River valley in Andhra Pradesh in south India, immediately before and after the Youngest Toba Tuff (YTT) eruption, there is, however, complete absence of biological remains. A well-documented prehistoric human settlement in India, the oldest evidence of modern human civilization, is the Mesolithic settlement in the middle Gangetic plains, which dates back to 10,000 years before present (YBP). Archeological excavation at the Mesolithic site of Sarai Nahar Rai in Ganga plains identified the presence of human skeletal remains, hearths, bones of animals, microliths and many more evidences, which suggest the earliest human settlement pattern in India. Following the Mesolithic cultural phase, we have the Harappan civilization throughout the northwestern part of India. Attempts are being made to analyse the biological remains from the Harappan site, and if successful, it would provide valuable information about human settlement, migration, affinities, adaptation and disease pattern in the past.

3. Archeological sites in India

Potential Mesolithic archaeological sites in India are Sarai Nahar Rai, Mahadaha and Damdama in the Vindhyas of middle Gangetic plains. About 40 Neolithic sites were found in the Belan, Adwa, Son, Rihand and the Ganga valleys, which provide information about one of the earliest agricultural practices in India. In the mid-Ganga valley, information about Neolithic culture has also been found at Sohgaura. Neolithic artifacts that include a small number of microliths and handmade potteries have been found in a large number of sites of the mid-Ganga valley, which show continuity and uniformity of cultural material. Of these sites, Imlidih Khurd and Lahurdeva have been excavated properly and proven to be the earliest Neolithic settlement with agricultural practices. Microlithic tools have been identified in Sant Kabir Nagar district, which is near the excavated site close to Ballia district of Uttar Pradesh. Carbon dating of Neolithic levels at Lahurdeva have estimated 6290 ± 160 BCE. Harappan sites of Dholavira, Farmana, Rakhigarhi and Kunal of northwest of India date back to 2800–1900 BCE. Sanauli archaeological site of western Uttar Pradesh shows the first evidence of late Harappan human remains. Human skeletal remains, dating back to 1800 BCE, may provide genetic affinities of populations who moved to Gangetic plains just after the decline of the great Harappan civilization.

4. Emergence of ancient DNA in evolutionary studies

Ancient DNA (aDNA) is an emerging research discipline which allows us to study the DNA extracted from the ancient biological remains. The source of ancient biological remains could be in the form of archaeological materials, museum specimens and other biological substances. The first attempt to extract the aDNA was accomplished in 1984, where researchers have reported traces of DNA from an extinct animal species, quagga (Higuchi *et al.* 1984). The amount of DNA extracted from the quagga specimen was approximately 1% of that expected from fresh muscle. Two clones of quagga DNA contained short mtDNA sequences, and differed by few bases from the mtDNA of a mountain zebra, an existing species of the genus *Equus*. Following this first aDNA report, researchers started claiming that authentic DNA could be isolated from biological materials that were approximately millions of years old. A majority of these claims aimed to retrieve the genetic material from insects that are well preserved in amber. Several plant fossils preserved in sediments dating back to the Miocene were analysed morphologically and attempted for DNA extraction. In 1994, Woodward *et al.* (1994) describe the cytochrome b sequencing results that had been isolated from dinosaur bones dating >80 million years ago. In the late 1990s, two follow-up studies were reported on the DNA sequences of dinosaurs. These ages have been topped by the claimed retrieval of about 250-million-year-old halo bacterial sequences from halite (Vreeland *et al.* 2000; Fish *et al.* 2002).

Consequently, aDNA research has picked up momentum globally as one of the emerging genetic tools to reconstruct the past population history of different species. This revolutionizing area of research unravels several breakthrough events which have taken place in the past and helped us in understanding the migration of people and spread of culture over the period of tens of thousands of years. The cutting-edge aDNA methods are applicable where DNA has degraded to the extent that modern DNA extraction techniques fails due to harsh climatic conditions in India, which affect DNA preservation. While tropical climate is generally not amenable to long-term DNA preservation, recent developments in laboratory techniques have led to the recovery of genomic information from other tropical regions of the world, for example, Africa and Southeast Asia.

With several research publications throughout the world, it was observed that under best preservation conditions, the upper boundary of the successful DNA extraction could be 0.4–1.5 million years with available sequencing technologies (Willerslev *et al.* 2004). With the recent advancements, aDNA has been retrieved from biological materials, as well as sediments from marine areas and lakes. In recent years, it has been proven that DNA extraction is not limited to several-years-old specimens from museums but could be extracted from mummies that date back to thousands of years (Pääbo 1985a, b, 1986).

5. aDNA studies link Neanderthals and modern humans

It was a breakthrough discovery that Neanderthals, our closest relative in terms of evolution, survived in most parts of Europe and western Asia before vanishing about 30,000 years ago. In the year 2010, Green *et al.* reported a draft sequence of the Neanderthal genome, which possessed more than 4 billion nucleotides. Comparative analysis of the Neanderthal genome with the genome of anatomically modern humans from different parts of the world identifies a large number of genomic regions that might have been influenced by positive selection. It was discovered that Neanderthals shared more genetic components with people of Eurasia than the people of sub-Saharan Africa (Green *et al.* 2010).

With the advent of several aDNA data of Neanderthals across parts of Europe, large number of papers have been published. In 2008 a bone was obtained from Denisova Cave in southern Siberia and the mitochondrial genome was sequenced. It was interesting to note that it symbolizes an unidentified type of hominin mtDNA that shares a common ancestor with anatomically modern humans and Neanderthal mtDNA for about 1 million YBP (years before present). The stratigraphy of the cave suggests that the Denisovan hominin might have lived close in time and space with Neanderthals and with modern humans (Krause *et al.* 2010).

Whole genome data of a Neanderthal skeleton excavated from Caucasus region reveal complex relationships and population history among present day humans, Neanderthals and Denisovans. It also proves that multiple gene flow events might have occurred between Neanderthals, Denisovans and early anatomically modern humans (Prufer *et al.* 2014). Additionally, the Neanderthal genome established a list of substitutions that has been fixed in modern humans (Slon *et al.* 2018).

6. aDNA and south Asian ancestry

As far as ancient genomics is concerned, the prehistoric genomic details of south Asians have not been clear because of the complete absence of aDNA data in this region. However, to fill this gap an attempt has been made to analyse genome-wide data of 362 ancient individuals from eastern Iran, Turan (Uzbekistan, Turkmenistan and Tajikistan), Bronze Age Kazakhstan and south Asia (Narasimhan *et al.* 2018). The data suggest a complex origin of south Asians. It is also documented southward migration from the Eurasian Steppe, associated with the expansion of pastoralist sites from the Steppe to south Asia during 2300–1500 BCE. Using contemporary population samples, we have shown that many modern Indian groups including tribal populations of the Andaman Islands provide genetic evidences for population admixture in India, suggesting that the genetic ancestry of most of the populations of modern India originated from a mixture of two diverse populations, the Ancestral North Indians (ANI) and the Ancestral South Indians (ASI) (Reich *et al.* 2009). Subsequently, our study on populations representing ASI and ANI groups revealed the time of admixture between these two distinct groups (Moorjani *et al.* 2013). Further, to understand the impact of endogamy, we have studied over 250 different ethnic groups from south Asia, and predicted that at least one-third of the populations in south Asia are prone to population-specific recessive disease due to founder effects and opened up new possibilities for directly translatable genetic research in India (Nakatsuka *et al.* 2017).

7. Promise of Harappan and Megalithic aDNA data in Indian prehistory

The relationship of the people of the Indus Valley to those of the Gangetic plain, and indeed to the people of modern India as well as their relationship to the cultures that existed in Asia contemptuously and prior to their existence, is an open question in Indian prehistory. The mature and late Harappan remains from various archaeological sites of India, for example, Rakhigarhi and Sanauli archaeological sites with its associated pottery and burials, could provide an incredibly valuable source of information for addressing both the origins and the decline of the Indus Valley cultures in India. aDNA or the ability to isolate genetic material from the primary burial samples of the Harappan period can allow us to anchor the population relationships and to begin to develop a complete picture of movement of people in prehistoric India.

aDNA researchers have great opportunities to perform the genetic analysis of the Harappan samples and co-analyse these data with Megalithic burials (1500–500 BCE), which are mostly concentrated in the southern part of India. Considering its cultural contexts and progress of modern DNA database, aDNA data can establish a plausible and well-supported outline for south Asian demographic prehistory. Using aDNA data, we can investigate the past and reconstruct the real-time structure of the south Asian populations, which have rightfully been described as pockets of endogamy. The aDNA results will also highlight the genetic relationship between the Harappan cultures and the later Bronze/Iron Age cultures in southern India as well as the genetic relationship between these late Harappan period samples and all modern Indian populations today.

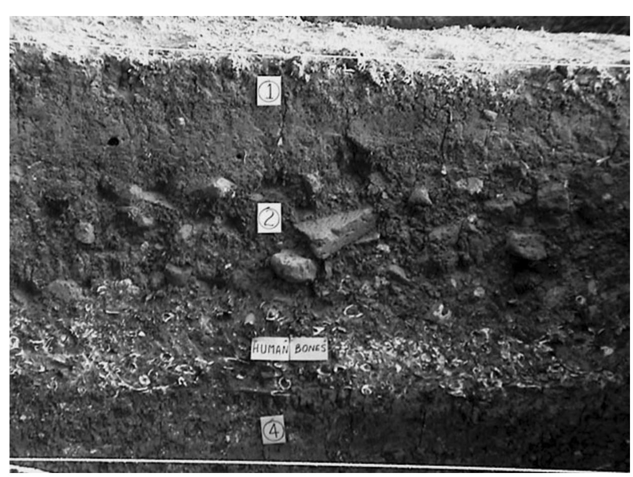

Figure 1. Stacked human bone remains in the well of tower of silence at Sanjan, Gujarat.

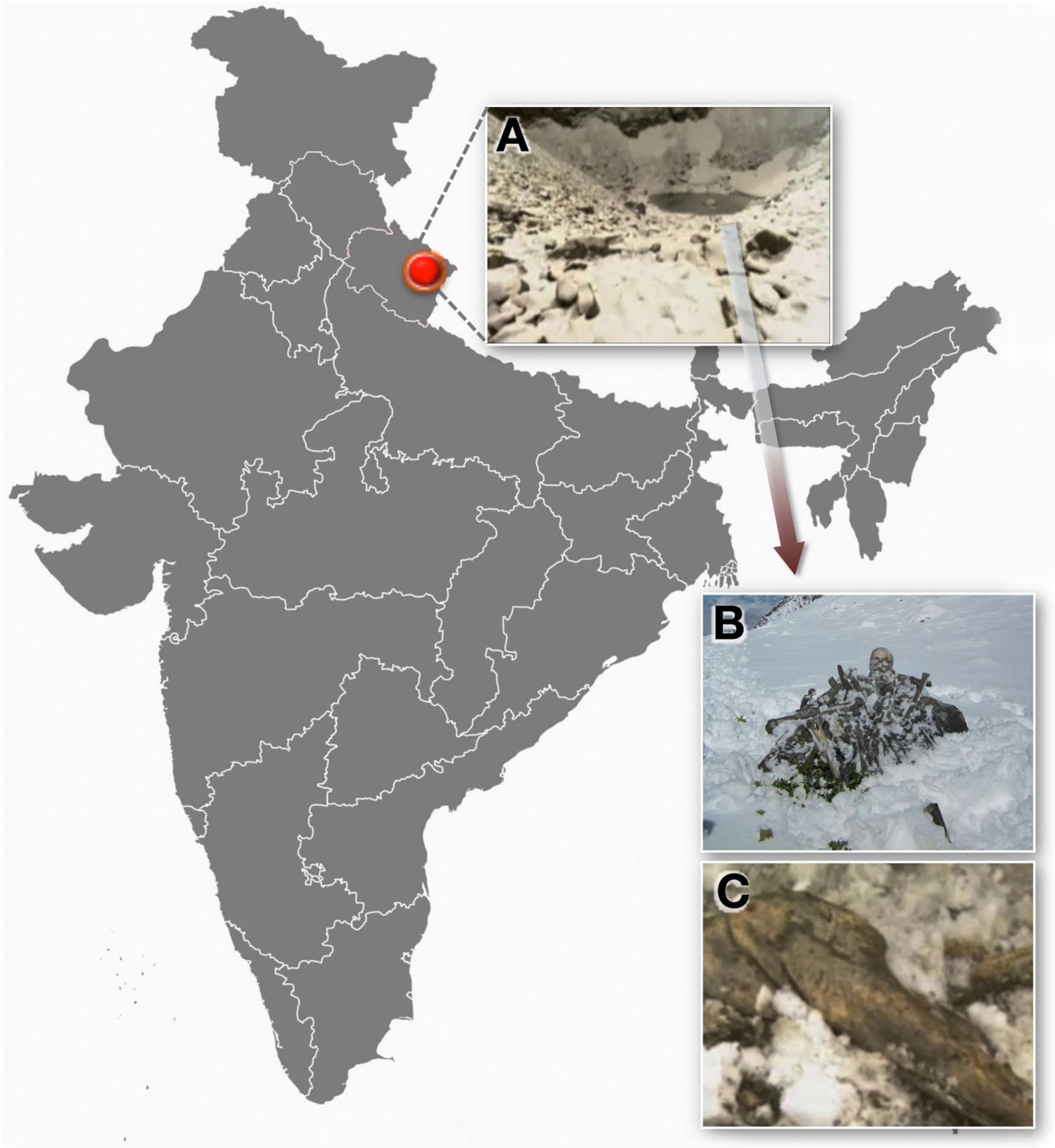

Figure 2. Indian map showing the location of the high-altitude Roopkund lake. (**A**) Close view of Roopkund lake which is frozen almost throughout the year. (**B**) Massive pile of skulls and bones scattered in Roopkund lake and in its vicinity. (**C**) Preserved body of a female in Roopkund lake.

8. aDNA studies in India

A few aDNA studies in India which have been initiated and successfully completed in the recent past are as follows:

8.1 *Mystery of Queen Ketavan of Georgia*

In May 2004, the Archaeological Survey of India, Goa circle, unearthed three bone relics from a church in Goa with an objective to search for the relics of Martyr Queen Ketavan of Georgia. As per the literary records, she was killed by Persian rulers in 1624 and her relics was brought to Goa by two Augustinian friars and buried in St. Augustine church, Goa, in 1627. It was the first successful analysis in the field of aDNA in India by analysing the hypervariable segment I (HVSI), as well as selected diagnostic markers in the mitochondrial DNA (mtDNA) coding region, to establish the authentic identity of the bone samples (Rai *et al.* 2014). The haplogroup affiliation of one of the samples was extremely unique (U1b). The frequency of this haplogroup is absent in modern day Indians and present in Georgia and adjacent regions, which has proved that the sample could be of Queen Ketavan of Georgia.

8.2 *First evidence of ostrich existence in south Asia*

Although ostriches are native to Africa, several geologists and archaeologists have over the years found ostrich egg shell pieces in India, mostly in Rajasthan and Madhya Pradesh. aDNA study of some partly fossilized ostrich egg shells which date back to the late Pleistocene (25,000 years old) was conducted successfully and proved their existence in south Asia until 20,000 YBP (Jain *et al.* 2017). The DNA sequence of ostrich egg shell revealed a 92% identity to the African ostrich species *Struthio camelus*, which was a breakthrough discovery and was not possible on the basis of morphological pattern of the fragile egg shell pieces.

8.3 *aDNA analysis of the first Parsi settlers in India*

The Parsis belong to a small community and their most recent count numbered less than 60,000 individuals. They are Zoroastrian refugees from Iran, who fled Muslim harassment in their homeland and first landed on the coast of western India around 750 AD.

The ancient site of Sanjan (N 20°11′59.6, E 72°48′00.2) was first discovered in 2001. An extensive excavation revealed hundreds of skeletal remains as well as the evidence of a large city that existed between the 8th and 13th centuries. Many well-established structures were found during the excavation, which has proven the existence of Zoroastrian community, for example, the tower of silence—a mortuary structure consisting of a large circular wall enclosing a raised platform, at the center of which there is a dry well. This is strictly and uniquely Zoroastrian in nature. The dead body is exposed to the natural elements (devoid of garments and ornaments) on this platform and after a time period, the osseous remains are swept into the central well (figure 1). Thus, this structure strongly suggests the settlement of a Parsi community that would be the clinching evidence of Parsi occupancy at Sanjan during the 8th to the 13th centuries. If so, this site was the oldest Parsi settlement ever to be excavated. aDNA analysis of 21 individual skeletons revealed an Indian-specific maternal gene pool, while a majority of the samples had Middle East and European-specific haplogroups. Further aDNA analysis of skeletal remains and whole genome data of modern Parsi samples across India and Pakistan proved their origin as Parsi and migration to India during the 7th and 8th centuries (Chaubey *et al.* 2017) (figure 2).

8.4 *Genetic makeup of Harappans*

The Harappan civilization was spread over large parts of south Asia between 3000 and 1800 BCE and was among the first large-scale societies of the ancient world. It is assumed that there was a gap in the history of south Asia and its discovery provided unbroken history of this part of the world. The available archaeological data from a few excavated sites, for example, Rakhigarhi in Haryana state and Dholavira in Gujarat state in the northwest India, clearly demonstrate that the Harappan civilization developed gradually from the beginning of settled life dated around 7000 BCE. Globally, people have always been curious to know more about this great and unique civilization. In the absence of aDNA, however, it has remained a mystery how the people of the Harappan civilization formed from earlier groups and contributed to later ones. Excavations at Rakhigarhi and Dholavira explored human burials (figure 3). Obtaining sufficient DNA from these skeletal remains, and rigorous scientific sampling and analysis of the same, would lead to a breakthrough to understand the prehistory of India.

Figure 3. Primary burial of an individual from mature Harappan site showing sleeping position with pots to the north of the head (courtesy Professor Vasant Shinde).

9. Conclusion

Despite the rich genetic resources available in India, successful exploration of ancient specimens in the subcontinent is severely lacking. To address this issue and to provide a much more diverse picture of the ancient genetic diversity globally as well as to address critical questions in ancient Indian history, ancient DNA data, along with carbon dating information about their chronology from within the Indian subcontinent, is needed. Human bones recovered from archaeological contexts tend to have a low endogenous DNA content. However, recent studies have demonstrated that the petrous portion of the temporal bone, the densest bone in the skeleton, can provide far higher endogenous aDNA. Furthermore, it is possible to obtain endogenous ancient DNA from petrous bones from hot environments, suggesting that skeletal material from south Asia, previously deemed unsuitable for aDNA analysis, could now yield invaluable information.

Through the genetic analysis of ancient south Asian human remains, there is potential opportunity to examine broad patterns of population diversity and migration on the continent. Major hunter-gatherer, agriculturalist and pastoralist expansions proposed on the basis of existing linguistic and archaeological data are of particular interest.

aDNA from the Indian subcontinent will greatly enrich our understanding of ancient history in the region overall. There are many important questions about our prehistory; for example, who were the people before the establishment of the Indus Valley civilization, who were the custodians and builders of the great Harappan civilization and what happened after the collapse of the Harappan civilization, how did they contribute to their ancestry of the modern day Indian population, etc.? Using the state-of-the-art aDNA approaches and collaborative studies with Indian archaeologists and anthropologists, we can go back to the past and reconstruct the population history of India.

With all these questions, this is the opportune time to analyse the human skeletal remains from different archaeological sites of the Indian subcontinent. Interdisciplinary research would also bring together geneticists and archaeologists and serve as a pilot for future large-scale studies on the Indian population history.

References

Chakrabarti DK 1982 The development of archaeology in the Indian subcontinent. *World Archaeol.* **13** 19

Chaubey G, Ayub Q, Rai N, Prakash S, Mushrif-Tripathy V, Mezzavilla M, Pathak AK, Tamang R, *et al.* 2017 'Like sugar in milk': reconstructing the genetic history of the Parsi population. *Genome Biol.* **18** 110

Fish SA, Shepherd TJ, McGenity TJ and Grant WD 2002 Recovery of 16S ribosomal RNA gene fragments from ancient halite. *Nature* **417** 432–436

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, *et al.* 2010 A draft sequence of the Neandertal genome. *Science* **328** 710–722

Higuchi R, Bowman B, Freiberger M, Ryder OA and Wilson AC 1984 DNA sequences from the quagga, an extinct member of the horse family. *Nature* **312** 282–284

Jain S, Rai N, Kumar G, Pruthi PA, Thangaraj K, Bajpai S and Pruthi V 2017 Ancient DNA reveals late Pleistocene existence of ostriches in Indian sub-continent. *PLoS One* **12** e0164823

Krause J, Fu Q, Good JM, Viola B, Shunkov MV, Derevianko AP and Paabo S 2010 The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464** 894–897

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, *et al.* 2014 Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513** 409–413

Moorjani P, Thangaraj K, Patterson N, Lipson M, Loh PR, Govindaraj P, Berger B, Reich D and Singh L 2013 Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93** 422–438

Moreno-Mayar JV, Vinner L, de Barros Damgaard P, de la Fuente C, Chan J, Spence JP, Allentoft ME, Vimala T, *et al.* 2018 Early human dispersals within the Americas. *Science* **362** 1–11

Nakatsuka N, Moorjani P, Rai N, Sarkar B, Tandon A, Patterson N, Bhavani GS, Girisha KM, Mustak MS, Srinivasan S, Kaushik A, Vahab SA, Jagadeesh SM, Satyamoorthy K, Singh L, Reich D and Thangaraj K 2017 The promise of discovering population-specific disease-associated genes in south Asia. *Nat. Genet.* **49** 1403–1407

Narasimhan VM, Patterson NJ, Moorjani P, Lazaridis I, Mark L, Mallick S, Rohland N, Bernardos R, *et al.* 2018 The genomic formation of south and central Asia. *bioRxiv*

Pääbo S 1985a Molecular cloning of ancient Egyptian mummy DNA. *Nature* **314** 644–645

Pääbo S 1985b Preservation of DNA in ancient Egyptian mummies. *J. Archaeol. Sci.* **12** 7

Pääbo S 1986 Molecular genetic investigations of ancient human remains. *Cold Spring Harb. Symp. Quant. Biol.* **51** 441–446

Posth C, Nakatsuka N, Lazaridis I, Skoglund P, Mallick S, Lamnidis TC, Rohland N, Nagele K, *et al.* 2018 Reconstructing the deep population history of central and south America. *Cell* **175** 1185–1197

Prufer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, *et al.* 2014 The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505** 43–49

Rai N, Taher N, Singh M, Chaubey G, Jha AN, Singh L and Thangaraj K 2014 Relic excavated in western India is probably of Georgian Queen Ketevan. *Mitochondrion* **14** 1–6

Reich D, Thangaraj K, Patterson N, Price AL and Singh L 2009 Reconstructing Indian population history. *Nature* **461** 489–494

Slon V, Mafessoni F, Vernot B, de Filippo C, Grote S, Viola B, Hajdinjak M, Peyregne S, Nagel S, Brown S, Douka K, Higham T, Kozlikin MB, Shunkov MV, Derevianko AP, Kelso J, Meyer M, Prufer K and Paabo S 2018 The genome of the offspring of a

Neanderthal mother and a Denisovan father. *Nature* **561** 113–116
Vreeland RH, Rosenzweig WD and Powers DW 2000 Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* **407** 897–900
Willerslev E, Hansen AJ, Ronn R, Brand TB, Barnes I, Wiuf C, Gilichinsky D, Mitchell D and Cooper A 2004 Long-term persistence of bacterial DNA. *Curr. Biol.* **14** R9–R10
Woodward SR, Weyand NJ and Bunnell M 1994 DNA sequence from Cretaceous period bone fragments. *Science* **266** 1229–1232