Ancient DNA

DNA is the key to unlocking our ancient African past

Alexandra Coutinho, Mário Vicente and Carina Schlebusch
(Uppsala University, Sweden)

Each region of the world, and the human groups living in them, have unique histories of migration, genetic mixing (admixture) and adaptation that have shaped their past. While archaeology has been of extreme value in elucidating this complex, multifaceted past, the genesis of DNA studies has enriched their story, and now ancient DNA (aDNA) has helped answer even more questions surrounding our prehistory. aDNA offers a unique opportunity to access genetic variation of past populations and enables us to contextualize past populations in present-day genetic variation. By linking the past to the present in this way, we now have a deeper understanding of our prehistory, and how our genetic landscape has changed from the past to the present. However, the study of aDNA is complex and there are various factors that need to be considered to yield successful aDNA results.

aDNA gives key insights to our early history

Human evolution and early human prehistory were once areas commonly addressed by the disciplines of archaeology, physical anthropology and linguistics. Although these fields have greatly contributed to what is known of the evolution and spread of Homo sapiens and their cultures, key questions regarding the human past remain unanswered. Integral to the above is a question that archaeologists could never answer, whether changes in material culture, observed in the archaeological record, can be ascribed to ‘Pots’ or ‘People’. In other words, are observed cultural changes a result of the physical migration and possible replacement of one population over another, or alternatively, are they due solely to the spread of ideas?

Genetics has slowly permeated many academic fields as a tool to add perspective to the available scientific research, and archaeology is no exception to this phenomenon. The first published segment of human aDNA was sequenced by Svante Pääbo in 1985 through bacterial cloning and following the advent of the polymerase chain reaction (PCR) method, the field of aDNA research quickly developed.

One of the greatest benefits of aDNA studies is in understanding how population migration and/or population continuity has impacted the genetic landscape of our species through time. Initial aDNA studies on humans focused on mitochondrial DNA (mtDNA) as it is more easily sequenced and there are far more copies of mtDNA than nuclear DNA in the human cell. However, analysis of mtDNA, on its own, still left many questions unanswered.

mtDNA is passed on maternally and, unlike nuclear DNA, does not undergo recombination. It therefore represents the genetic information of a single marker, and ultimately only one single ancestor among our thousands of ancestors. Nuclear DNA recombines, and so includes the genetic information of many markers, and consequently multiple evolutionary histories of individuals going back in time. With the advent of next-generation sequencing (NGS) it is now much easier to also generate DNA sequences of our full nuclear genomes. Unlike older sequencing methods which were ill-equipped to cope with very short DNA fragments, NGS techniques are ideal for sequencing genomes from ancient remains.

The study of aDNA focuses on genetic point mutations, or changes in the DNA sequence of one
nucleotide only. Mutations that have no effect on the individual’s ability to reproduce are neutral and not impacted by natural selection, so their relative frequency in a population is determined by population demographic factors such as population size, migration and admixture between different groups. Different human populations rarely have different sets of point mutations, instead, they carry different frequencies of the same mutation, and this is what creates genetic variation between populations. It is the study of this variation that enlightens human migration and evolutionary history. Yet, along with the advantages of incorporating aDNA into current research comes certain challenges inherent in the characteristics of aDNA that makes it hard to work with.

DNA survival

aDNA is, as its name reflects, ancient and will have degraded over time. Its survival will have been greatly affected by the environment, and in particular, the average temperature in which the remains have been preserved. DNA degrades most rapidly in moist, hot or humid environments or in places with large and frequent fluctuations in temperature. Such environments encourage microorganism and fungal growth and activity, which fragmentizes DNA. The presence of water in the soil, and the influence of soil pH, catalyses chemical deamination, leading to degradation of the DNA, especially towards the ends of DNA fragments. Biochemical studies on DNA suggest that DNA survival is capped at 100,000 years, but in really cold, dry environments such as permafrost or in caves at high altitudes where temperatures are cool and rarely fluctuate, DNA has been successfully extracted from remains as old as 700,000 years.

Regardless of the environment, encapsulation within bone ‘protects’ DNA from extracellular factors. Its survival varies according to the type of bone from which it is extracted. Overall, DNA has been found to be better preserved in areas of the skeleton where the bone is compact, such as the core of long bones. For example, in 2012, Meyer et al. were able to obtain sufficient DNA from a phalange which belonged to a Denisovan individual, an archaic human that lived 40,000 years ago in central Asia, to represent the entire genome 30 times over (30× coverage). Studies have further shown that generally, DNA extracted from the extremely compact petrous part of the temporal bone in the skull as well as DNA from tooth cementum is in a much better state of preservation than in other parts of the skeleton (Figure 1).

The challenges of working with aDNA

The successful extraction, sequencing and analysis of aDNA is not without its challenges. Even with recent advancements in DNA technologies, issues such as the preservation conditions of a sample, possible contamination of aDNA with modern DNA, and other damage to the aDNA sequence makes reliance on the authenticity of an aDNA sequence difficult. The compromised and damaged nature of aDNA, contamination from DNA in the surrounding environment (from the soil and microorganisms) and more importantly, from the people who handle and work with the remains, make it challenging to retrieve and sequence endogenous aDNA. While bioinformatics approaches are now able to discriminate between endogenous and contaminant DNA, the more contamination present in the sample, the smaller the fraction of endogenous DNA that will be returned for a certain amount of sequencing.

In order to minimize any further risk of contamination, precautions are taken in the laboratory where samples are processed to keep the surrounding environment as free from DNA as possible (Figure 2). When extracting DNA from bone and tooth material, all samples are irradiated under a UV light to destroy any surface contamination. The top layer of sample bone is drilled away to further ‘decontaminate’ the bone, and bleach is used to clean the topmost layer of bone before drilling even begins. Bone and tooth powder are only extracted from the innermost layers of bone and teeth, and negative controls are used throughout the extraction and library building process to ensure that no contamination has been introduced by the laboratory personnel. Protective gear is also worn to prevent further contamination (Figure 2).

Given their nature, many aDNA samples come from museum material, hence sampling must occasionally take place on-site. In such cases, portable aDNA sampling tents are used to create a make-shift sterile environment to sample ancient material (Figure 2C). It is important to note that despite compliance with the precautionary steps described before, resulting raw sequenced DNA data almost always includes a mixture of environmental DNA, possible contaminant human DNA from, e.g., excavators and/or curators, and endogenous DNA.

Raw sequence aDNA data undergoes initial processing to prepare it for population genetic analysis. This processing involves filtering the data, aligning the data against the reference human genome to discard
Ancient DNA

Any environmental DNA contaminants and conducting different analyses to obtain a measure of DNA authenticity. This is where deamination damage on the ends of DNA fragments may serve as an indication of authenticity. Analysis of the mtDNA and Y-chromosomes of individuals also verify the authenticity of the sample and whether it has been contaminated, as having more than one mitochondrial or Y-chromosome type present in a sample would indicate contamination either from modern DNA or another aDNA sample.

After pre-processing, aDNA data may then be analysed with data from modern-day individuals to facilitate population history inferences. Principal component analysis (PCA) is a data summary technique that is frequently used for visualizing aDNA variation in the context of modern-day genetic variation. In the case of aDNA, ancient individuals are projected onto the background of the modern-day sample set to determine not only how ancient samples cluster together into possible population groups, but to which modern populations these ‘clusters’ are most related. PCA analysis reduces multiple dimensions, or in this case, the information from many point mutations across many individuals, to its principal components, that is the data pattern that best explains the variation seen in a dataset. The output is in the form of a PCA ‘plot’ where the axes can be represented by the first two ‘principal components’ of the data (but other PCs can also be represented). The example in Figure 3 is a PCA of the genetic variation in modern-day African populations (coloured circles) and data from the remains of various ancient African individuals from whom DNA could be extracted and analysed (skull symbols). The PCA in Figure 3B represents a 2D map reflecting the genetic relatedness of ancient and modern-day individuals to each other and can be compared with the geographic map (Figure 3A) of where the individuals/populations originate geographically.

aDNA in Africa

African aDNA studies are few to date, as early aDNA research focused on Europe, and early DNA sequencing technology was not yet developed enough to be able to handle samples from warm climates. However, in recent years molecular genetics methods and tools have continued to improve, so that not only have researchers had more success with remains preserved in poor conditions, but aDNA has been successfully extracted from far more ancient remains compared with even a decade ago. At the time of writing, eight studies in total have published ancient genomic DNA results from Africa, four studies with aDNA from North Africa and four studies with aDNA from sub-Saharan Africa.

Human population history in North Africa has a unique standing in Africa. Modern-day groups are largely related to Eurasian and Middle Eastern populations with minimal levels of genetic contributions from sub-Saharan Africa (blue dots, Figure 3). This was suggested to be the result of back to Africa migrations during the Neolithic period, and migrations related to the introduction of farming practices to North Africa. aDNA studies of early Neolithic (~7000 before present [BP]) Moroccan remains (blue skulls, Figure 3) indeed found the individuals to be most genetically similar to Anatolian farmers and Natufians from the Middle East, suggesting a potential early westward migration of these groups. However, a more recent study on 15,000 BP remains from Morocco demonstrated that northern Africa received significant amounts of gene flow from Eurasia predating the start of the Holocene. Gene flow from the south, across the Sahara into North Africa, was limited, and seemed to have occurred only recently. This was apparent as low levels of genetic ancestry from sub-Saharan groups in ancient North African individuals dated to the Neolithic. Mummies from the easternmost part of North Africa also demonstrated that admixture between sub-Saharan Africans and northern Africans...
was recent, in that people from ancient Egypt (~3400 BP) displayed less sub-Saharan admixture compared with present-day Egyptians.

Modern and aDNA studies on sub-Saharan Africa indicated that the history of this part of the continent may be divided into two very different phases. Before the invention and spread of farming practices in Africa, hunter-gatherer groups were related in an ‘isolation-by-distance’ fashion influenced by geography. This means that groups were most closely related to their neighbours, and less related to groups that were geographically further away. This stands in stark contrast to the large population movements that followed the invention of farming practices in Africa. As farming populations rapidly spread across the whole continent, their genetic signature ‘erased’ the ‘isolation-by-distance’ pattern observed in hunter-gatherer populations. Currently, it is believed that three regions in Africa developed agriculture independently of one another: the Sahara/Sahel (around 7000 BP), the Ethiopian highlands (~7000–4000 BP) and western Africa (~5000–3000 BP). The Nile River Valley is thought to have adopted agriculture (~7000–8000 BP) from the Neolithic transition in the Middle East (~10,000–11,000 BP). From these centres of origin, farming practices spread to the rest of Africa, with domesticated animals reaching the southern tip of Africa around 2000 BP and crop farming around 1800 BP.

**Clarifying African prehistory**

The first ancient nuclear genome to be sequenced from sub-Saharan Africa was that of a 4500-year-old Ethiopian individual (named Mota, Figure 3). This genome revealed that East African populations younger than 4500 years (brown dots and brown skulls, Figure 3) were influenced by a single or multiple migrations from outside Africa back into the continent of populations who were genetically similar to early Neolithic farmers from western Eurasia. The PCA demonstrates this (Figure 3B), where current-day East African populations (brown dots) and ancient East African individuals from farming contexts (brown skulls) lie between the Mota individual (a representative non-admixed East African hunter-gatherer) and non-Africans (blue dots). This study clearly indicated migration from Eurasia back into East Africa after 4500 years ago, and showed that current-day East African groups’ ancestries link them to their East African ancestors as well as Eurasian groups that migrated back into Africa. This finding was confirmed multiple times in subsequent aDNA studies that included ancient East African individuals.

A study that sequenced 16 ancient Africans across East and southern Africa showed that ancient herders carrying this mixed East African-Eurasian ancestry migrated all the way down into southern Africa. These...
groups introduced herding practices to the south of the continent (brown skulls and dots in eastern and southern Africa, Figure 3A and B). The ancient genomes also revealed how hunter-gatherer genetic ancestry looked before the spread of farming through the continent. The ancestry of hunter-gatherer populations of East and southern Africa (yellow to red skull cline; a gradation in biological features over geographic space, on PCA – Figure 3B) fit on a cline, where neighbouring groups were more related to one another than groups further away. Subsequent African demographic history was ever more complex, with repeated gene flow between different groups, and varying levels of population replacement by especially western African Bantu-speaking farmers.

The Bantu expansion is one of the largest expansion events of farmers globally and began around ~5000–3000 BP in western Africa (in the region of current eastern Nigeria and western Cameroon). It is visible in the archaeological record via increased sedentism, the spread of agricultural practices and the use of iron. Today, the majority of sub-Saharan Africans speak one of ~500 closely related Bantu languages despite their distribution over an area of ~500,000 km². Earlier genetic studies indicated that the current distribution of Bantu-speaking populations is largely a consequence of the movement of people (demic diffusion) rather than a diffusion of only language. This has been confirmed by aDNA studies; i.e., where ancient remains found in Iron Age archaeological contexts in East and southern Africa (grey skulls on Figure 3A) group genetically with current-day West African populations on the PCA (Figure 3B).

The effects of the Bantu expansion may also be seen in aDNA studies on southern Africa. A study that sequenced the nuclear genomes of seven ancient southern African individuals, three dating to the Late Stone Age (2000 years old – red skulls in southern Africa – Figure 3A), and four dating to the Iron Age (300 to 400 years old – grey skulls in southern Africa – Figure 3A) found that the Later Stone Age individuals were related to current-day Khoe-San hunter-gatherer individuals (red dots) and the Iron Age individuals to current-day West Africans (grey dots). The study confirmed extensive population replacement in southern Africa, where Later Stone Age ancestors of the Khoe-San hunter-gatherers were replaced by incoming Bantu-speaking farmer groups with West African genetic ancestry. Interestingly, one of the 2000-year-old individuals related to Khoe-San groups yielded enough ancient DNA to reach 13x coverage, a very high yield with regards to aDNA. When using this high coverage ancient individual as a reference, it was revealed that all modern Khoe-San groups have 9–22% admixture from the admixed group of Eurasian-East African ancestry that introduced herding into southern Africa. This can be noted on the PCA (Figure 3B) as a shift in modern-day Khoe-San groups (red dots) towards the East African side of the PCA. The high coverage ancient individual was further compared with other Africans to re-estimate the time when the first modern human groups genetically diverged from one another. Whereas past research estimated a divergence time of 100,000 years, (200,000, using the updated human mutation rate) using the non-admixed Stone Age individual pushed back the divergence time to between 260,000–350,000 years, towards the genesis of the Middle Stone Age when humans became morphologically and behaviourally modern.

African genomes and future aDNA studies in Africa will continue to clarify the picture of our deep genetic history and bring us closer to answering the questions of human origins in Africa. Concurrently, it will continue to clarify the large-scale but complex movements associated with the spread of farming practices in Africa.

**Future of aDNA**

aDNA methods are continually evolving, evidenced by research advances in the past decade alone. Progress has been made on wet lab techniques to allow sequencing of DNA samples from areas prone to bad DNA preservation, such as Africa. In areas where DNA preservation is favourable, such as Europe, there are now data from hundreds of ancient individuals to bolster bioinformatic analyses and increase the amount of reliable information on the analysed ancient populations. As aDNA is becoming more common as a means of interpreting human history, researchers continue to improve and advance dry lab and population genetic analyses techniques to be able to process and handle aDNA sequence data in an optimized way.

As DNA sequencing becomes ever more affordable, and the ability to sequence more ancient genomes becomes possible, scientists are coming to a deeper understanding of the human species and every facet of what makes us who we are. aDNA research continues to reveal more nuanced, complex stories of human prehistory and human evolution. These stories help us to understand the present, and better predict how we may react in the future in an ever-changing environment.

**DEFINITIONS**

**Admixture:** when two ancestral populations genetically mix to form a hybrid descendant population.

**Demography:** the study of life-history and different characteristics of a population, such as kinship, size and migration.

**Demic diffusion**: a demographic term referring to a migratory model of population diffusion into and across an area that had been previously uninhabited by that
group, possibly, but not necessarily, displacing, replacing or intermixing with a pre-existing population. A term opposed to cultural diffusion.

**BP (before present):** date abbreviation where the number of years is counted before the present, the present being 1950.

**Holocene:** the period in history covering the last 11,000 years, where climate was unusually stable and warm.

**Neolithic:** an archaeological time period that defines the use of new stone tools associated with the development of farming, pottery manufacture and the time when humans started to form settlements.

**Genetic divergence:** the process where two or more populations become genetically isolated from one another.

**Cline:** gradations in biological features over geographic space.

---

**Further reading**

- Nielsen, R., Akey, J.M., Jakobsson, M., Pritchard, J.K., Tishkoff, S. and Willerslev, E. (2017) Tracing the peopling of the world through genomics. *Nature* **541**, 302–310
- Skoglund, P. and Mathieson, I. (2018) Ancient genomics of modern humans: the first decade. *Annu. Rev. Genomics Hum. Genet.* **19**, 381–404
- Schlebusch, C.M. and Jakobsson, M. (2018) Tales of human migration, admixture, and selection in Africa. *Annu. Rev. Genomics Hum. Genet.* **19**, 405–428
- Schlebusch, C.M., Malmström, H., Günther, T. et al. (2017) Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655
- Skoglund, P., Thompson, J.C., Prendergast, M.E. et al. (2017) Reconstructing prehistoric African population structure. *Cell* **171**, 59–71

---

**Alexandra Coutinho** has always wanted to be Indiana Jones but settled instead for studies in aDNA at the Adelaide Centre for Ancient DNA (ACAD) in Australia, followed by graduate studies in the Human Evolution Program at Uppsala University. While this field has allowed her the opportunity to explore our human past, Alex feels that such discoveries are worthless if they are not shared with the rest of the world. She plans to begin a career as a science communicator to do just that. If you too would like your science to be shared with others, send her an email at danielaxana@yahoo.com.au

**Mário Vicente** has a long-lasting interest in Africa; the passion for population genetics history came later. After his master’s in Biological Anthropological Science at the University of Cambridge, he moved to Sweden where he is finishing his PhD in Human Evolution and Genetics at Uppsala University. Mário’s research mainly focuses on the demographic history of sub-Saharan Africa, with special interest in the genetic history of hunter-gatherers and the spread of pastoralism across the continent. Email: mario.vicente@ebc.uu.se

**Carina Schlebusch** is Associate Professor in Human Evolution at Uppsala University. She heads a research group investigating human evolutionary history with a focus on Africa. Carina was born in South Africa and completed her PhD at the University of the Witwatersrand, Johannesburg. Her thesis dealt with the genetic history of southern African population groups. In 2009, she moved to Uppsala University where she completed a Wenner Gren fellowship postdoc, after which she started her own research group. The Schlebusch group uses genetic research as a tool to investigate African population history and migrations of people within the African continent. Email: carina.schlebusch@ebc.uu.se