A BRIEF OVERVIEW OF THE MITOCHONDRIAL DNA AS MOLECULAR MARKER IN BIOARCHAEOLOGY

IOANA RUSU¹ and BEATRICE KELEMEN¹

SUMMARY. The current paper summarizes the key studies that revolutionized the field of ancient DNA research, with special focus on the mitochondrial genome. Differences between molecular markers used in bioarchaeology are presented with emphasis on their advantages and disadvantages. As previously stated the focus of this review is on the mitochondrial DNA as a molecular marker, commonly used to answer bioarchaeological questions. The essential studies on mitochondrial DNA from ancient specimens that reveal key information about the history of Europeans are illustrated.

Keywords: ancient DNA, bioarchaeology, mitochondrial genome.

Introduction

For centuries, people have been trying to solve the mystery of their origin and to answer questions regarding the evolutionary history of the modern Homo sapiens species. First insights that shed light on the past of the human race were provided by archaeological records. Their major limitation is their small resolution when prehistoric records are studied. These are in many cases degraded over time carrying little or no information. Once the structure, properties and function of the DNA molecule were deciphered a new perspective on the approach to tracing the ancestry of humans deep into the past was soon embraced. Since the genetic material is inherited from previous generations, the history of our species is written in our DNA. Unlike archaeological records, this kind of record does not fade away over time and enables molecular biologists to unravel messages from the past, far beyond the reach of stone paintings or different types of material inventory, for example.

¹ Babeş-Bolyai University, Faculty of Biology and Geology, Interdisciplinary Research Institute on Bio & Nano Sciences, Molecular Biology Center, Cluj-Napoca, România.

Corresponding author: IoanaRusu, Babeş-Bolyai University, Faculty of Biology and Geology, Interdisciplinary Research Institute on Bio & Nano Sciences, Molecular Biology Center, Treboniu Laurian Street 42, 400271, Cluj-Napoca, Cluj, România. Mobile: +40755456580, E-mail: rusu.n.ioana@gmail.com
This idea was emphasized by a scientific paper published in 1987, “Mitochondrial DNA and Human Evolution” (Cann et al., 1987). For many generations of scientists, controversies based on the interpretation of fossil remains regarding the origins of modern humans were intensely debated. Using a bio-molecular approach, this publication (Cann et al., 1987) estimated that the most common mitochondrial ancestor of all modern humans lived about 150,000 years ago in Africa. Supporting the “Out of Africa” theory in favor of the multi-regional scheme, these findings had a great impact in the scientific community (Sykes, 2001).

Brief history

In 1984, Higuchi and his co-workers (1984) managed to retrieve DNA sequences from a museum specimen of quagga (140 years old), an extant member of the genus Equus (Higuchi et al., 1984). By illustrating the potential of long term survival of DNA molecule in ancient remains, this publication had a great impact in various fields of the scientific community. One year later, Alu elements from a 2400 year old Egyptian mummy of a child were molecularly cloned (Pääbo, 1985). It was clear that ancient DNA can be preserved for many hundred of years in archaeological samples, but considering the postmortem degradation of nucleic acids, for how long and in which conditions? The discovery of a human skeleton in the Tyrolean Alps, now widely known as the Ice Man, has provided additional information regarding the preservation of nucleic acids in 5000 years old remains from an extreme, cold environment. Moreover, mitochondrial DNA sequences were successfully retrieved from this specimen and compared to those obtained from present-day people. Interestingly, the Ice Man was genetically linked to contemporary Europeans suggesting that all people carry information about their ancestors in DNA molecules.

A remarkable breakthrough in the field of ancient DNA research is the determination of the almost complete mitochondrial genome sequence of a hominin from northern Spain, dated to over 300,000 years ago (Meyer et al., 2014). The analysis of the oldest DNA ever sequenced revealed surprising facts. The results indicate that it is more likely that the hominin discovered in Spain shares a common ancestor with the Denisovans rather than the Neanderthals, despite the fact that the morphological traits of the skeletal remains suggest the contrary.

Recently, due to technological advances in ancient DNA techniques, a series of archaeological and anthropological controversies can be solved, and at the same time new question arise and need answers. The first technology that revolutionized the field of molecular genetics was the development of Polymerase Chain Reaction (PCR). Since Kary Mullis had this original idea in 1983, this amplification technique was improved and has now become widely used. It enables the production of thousands to millions of copies of a specific DNA segment even from a single template molecule (Bartlett and Stirling, 2003). It is essential for the field of ancient DNA, because usually only a small amount of degraded DNA can be recovered from archaic specimens. More recently, high-throughput DNA sequencing methods were
developed and had a considerable impact in ancient DNA research because significant amounts of sequence data can be generated in a short period of time (Knapp and Hofreiter, 2010). Even though costs for Next Generation Sequencing (NGS) have been significantly reduced in the last few years, it is still not affordable for all laboratories as it also involves powerful bioinformatic means to analyze such large amount of sequence data.

Despite the fact that significant technological progress was achieved in the field of ancient DNA research, the investigation of genomes from archaeological samples is still challenging due to contamination with exogenous DNA. For this reason, strict guidelines have to be followed (Poinar, 2003).

**Molecular markers**

The study of genetic variations represents a powerful tool that helps us form a relatively complete image of human evolution, as major demographic events leave traces in the human genome mainly by inducing changes in allele frequencies. These imprints are recorded by the modern human genomes because they have been inherited from previous generations. Initial molecular markers used to assess the genetic variation among populations were the blood group systems (Jorde *et al*., 1998). Although, in this manner a perspective of the evolutionary history of our species was provided, it was soon remarked that a single system could only open a small window into the past. It was also observed that using only one genetic system, such as the ABO blood groups, can sometimes lead to same gene frequency in two populations, just by chance, rather than common ancestry (Sykes, 2001). To eliminate such misleading results several genetic systems are now analyzed, often in a multidisciplinary approach.

There are several distinct categories of genetic markers used for unraveling the past history of humans. One group includes nuclear markers such as Single Nucleotide Polymorphisms (SNPs) and Short Tandem Repeats (STR) which have different mutation rates. The evolution of STR, also known as microsatellites, is considerably higher than that of SNPs which make them more suitable for assessing recent history and less informative for ancient evolutionary events (Jorde *et al*., 1998). STRs can be found on the Y chromosome and are frequently used for the investigation of paternal ancestry. Mitochondrial DNA is used in a similar way to trace back maternal lineages (Bouwman *et al*., 2008; Cafer, 2010).

**Properties of mitochondrial DNA**

Besides being maternally inherited in most multicellular organisms, mitochondrial DNA possesses some unique features which makes it appropriate for a large variety of studies, including human evolution, migration or establishing relationships between past populations.
Human mitochondrial DNA (mtDNA) is a double-stranded, circular molecule of 16569 base pairs (Anderson et al., 1981). This small genome is located in the cell’s energy organelles, mitochondria, and it is presumed to have originated from bacteria that were taken inside advanced cells as endosymbionts, hundreds of millions of years ago (Sykes, 2001). The mitochondrial genome contains genes necessary for the production of energy and a non-coding fragment named the control region. Other mitochondrial genes were slowly transferred to the nucleus in the course of evolution (Pakendorf and Stoneking, 2005).

Unlike any nuclear genes which are present in the human cells in only two copies, there is a multitude of mitochondria in most cells. Since one of the major difficulties for ancient DNA studies is the post-mortem degradation of nucleic acids (Höss et al., 1996; Pääbo et al., 2004; Willerslev and Cooper, 2005), a high copy number of mtDNA makes it more accessible mean for tracking matrilineal ancestry.

Another difference between the two types of genomes is their mutation rate, which is higher in mtDNA, being especially high in the control region (Brown et al., 1979). This elevated substitution rate offers an insight on the developments that lead to the actual human gene pool over time (Cann et al., 1987).

Nuclear DNA, which is inherited from both parents is subjected to recombination processes. In contrast, human mtDNA is inherited only from one parent, the mother, while paternal mitochondria are lost during fertilization. There were controversies regarding the lack of genetic exchange that had significant impact for mtDNA based phylogenetic studies (Hagelberg et al., 1999). In a particular case where the recombination of mtDNA was proposed to explain the presence of a rare mutation at high frequency in separate mtDNA lineages from a small island, the explanation turned out to be just a sequence alignment error (Hagelberg et al., 2000). Currently, it is suggested that the main cause for homoplasyis is the presence of mutational hotspots which occur preferentially at hypervariable sites (Stoneking, 2000; Galtier et al., 2006).

All these characteristics make mtDNA a suitable molecular marker used for the screening of the genetic history of humans. Assuming that neutral mutations accumulate in time with a constant rate, it has been used as a molecular clock, timing evolutionary events. In this context, Sykes reconstructed the evolutionary history of Europeans and suggested that all modern members originated from seven distinct maternal lineages (Sykes, 2001). The mutation rate estimated for the human mitochondrial DNA is quite different from that reported in pedigree and phylogenetic studies. This is caused by several factors, such as: the presence of mutational hot spots, genetic drift, selection, and the lack of detection of high levels of homoplasy in phylogenetic studies (Howell et al., 2003). The assumption that a uniform mutation rate exists is considered an oversimplification which doesn’t perfectly reflect the true evolutionary and demographic history of our species (Pakendorf and Stoneking, 2005; Pulquério and Nichols, 2007; Galtier et al., 2009).
What is analyzed?

Various parts of the human mitochondrial DNA are analyzed, depending mostly on the focus of the study. Until a few years ago, the majority of research included the investigation of the mitochondrial control region. This non-coding region of the mitochondrial genome comprises two highly polymorphic segments, HyperVariable Regions I and II (HVR-I, HVR-II), and thus provides a good image on the diversity of the mitochondrial genome. In order to obtain data on the variability of the human mitochondrial genome, two types of markers are frequently used: restriction fragment length polymorphisms (RFLPs) and sequence analysis of PCR products. It was observed that DNA molecules recovered from ancient human remains are highly degraded and fragmented due to biological and chemical processes that occur postmortem (Gilbert et al., 2003). Therefore, the amplification of ancient DNA sequences is restricted to short fragments of approximately 100-200 base pairs (Pääbo et al., 2004). The strategy to overcome this limitation is to amplify short overlapping segments that span the hypervariable regions (Gabriel et al., 2001). For this reason, it is considered that for the classification of mtDNA variation into haplogroups, in case of large population studies from archaeological remains, RFLP analysis is more suitable (Izaguirre and De La Rua, 2002).

Data from only the HVRI sequence allows a haplogroup prediction, which becomes more accurate when HVRII is also considered. For a more sensitive haplogroup discrimination, which is especially important in forensic cases, single point mutations from the coding region are also screened (Köhinemann et al., 2009). These informative SNPs can be determined by RFLP analysis (Torroni et al., 1996) or by SNaPshot assay, a multiplex method (Brandstätter et al., 2003). Also, to distinguish more precisely between sequences assigned to haplogroup H and non-H haplogroup, a short segment comprising the defining nucleotide for haplogroup H (position 7028) can be additionally sequenced besides HVRI (Dissing et al., 2007).

The study of the mitochondrial genome has a great importance for the medical field as it has been revealed that multiple SNPs can be associated with a variety of degenerative diseases (MITOMAP: http://www.mitomap.org, 2013). For example, a homoplasmic mutation that occurs at position 4336 in the nucleotide sequence might be a contributing factor in Alzheimer and Parkinson diseases (Egensperger et al., 1997). Homoplasmy appears when all copies of the mitochondrial genome are identical, while be heteroplasmic mutations are present only in some of the copies of the mitochondrial genome (Wallace, 1994). Variations in some mitochondrial genes are associated with Leber Hereditary Optic Neuropathy and contribute, together with nuclear genetic factors, to the expression of the disease (Taylor and Turnbull, 2005). In this light, the analysis of variations in the sequence of mitochondrial genes is essential, as it might be used for the diagnosis of some genetic disorders.
HVRI sequences from ancient human remains are frequently studied to determine the maternal relationships between individuals (Dissing et al., 2007). In order to avoid inconclusive results due to ambiguous haplotype assignment, additional diagnostic point mutations from the coding region are sometimes screened (Rudbeck et al., 2005; Deguilloux et al., 2014). Being maternally inherited, data from mitochondrial genomes can’t provide a complete picture of the biological relationships among individuals. As a consequence, autosomal markers, usually STRs, and the non-recombining region of the Y chromosome are analyzed to complete the information about closely related individuals recovered from archaeological sites (Keyser-Tracqui et al., 2003).

Recent advances in molecular genetic technologies have enabled direct access to information from the whole mitochondrial genome. Complete mitochondrial genome sequences were reconstructed from ancient specimens, including a 38000 year old hominid, identified as Neanderthal (Green et al., 2008). Whole mitochondrial genome sequences were analyzed to gain a more comprehensive idea about human population history. They were used to construct phylogenetic trees which seemed to be based on a more accurate estimate of the mutation rate than the previously reported ones, which considered only the differences in the control region (Ingman et al., 2000). However, it is still uncertain whether whole mitochondrial genome sequencing is always the best approach considering the costs involved. An alternative to gain insights on the past of human populations is to supplement control region sequences with those mutations from the coding region that contain informative nucleotide polymorphisms (Pakendorf and Stoneking, 2005).

How is the mtDNA sequence analyzed?

The mitochondrial genetic variants of modern people and of past populations can be classified into distinct haplogroups. Similarly, mitochondrial sequences that share the same mutations are grouped into a certain haplogroup suggesting that they all have a common ancestor. To screen for the divergence points among various sequences, these are compared to the first complete sequence of a human mitochondrial genome, the revised Cambridge reference sequence (Andrews et al., 1999). Considering mitochondrial sequence variation from three distinct European populations, Torroni and his collaborators revealed that almost 100% can be grouped into 10 different haplogroups (H, I, J, K, M, T, U, V, W, X) (Torroni et al., 1996). To define these European haplogroups, mutations in the mitochondrial coding region where determined using restriction fragment length polymorphisms.

The evolutionary relationships between human populations are represented as phylogenetic networks in which the haplogroups are major branching points. Such phylogenetic trees were initially built by using information from the non-coding
region because of its high mutational rate (Richards et al., 1996; Macaulay et al., 1999). A refined, comprehensive, phylogenetic tree that provides a high resolution on global mitochondrial DNA diversity, from the most recent common matrilineal ancestor of all humans is currently available online (van Oven et al., 2009; http://www.phylotree.org).

The mitogenomic data from Homo neanderthalensis enabled the reconstruction of the genetic sequence of the “Mitochondrial Eve”. It was proposed that the Revised Cambridge Reference Sequence which belongs to the European haplogroup H2a2a should be replaced with the Reconstructed Sapiens Reference Sequence (RSRS) (Behar et al., 2012). At this point, this approach is still controversial because the switch to a new reference sequence may cause misinterpretations (Bandelt et al., 2014). Still, in a recent study aimed to assess the key factors that lead to the current diversity and distribution of haplogroup H, mitochondrial variation observed in ancient DNA recovered from archaeological samples was genotyped against RSRS (Brotherton et al., 2013).

**Ancient DNA and European mitochondrial genome diversity**

The evolutionary history of modern humans is governed by several factors which include changes in the environment, culture and the geographic distribution of populations. Major demographic events contributed to the current European gene pool by altering the allele frequencies. The analysis of genetic variations in multiple modern populations sheds light on the geographic dispersal of our ancestors, that shaped the actual distribution of populations and their genetic structure. Recent demographic events and population genetic processes obscure some aspects of the human evolutionary history (Lacan et al., 2013). Direct access to ancient DNA offers invaluable, additional, information to the one revealed by present-day markers, leading to a better understanding of our complex genetic history (Der Sarkissian et al., 2013).

In Europe, several major population movements contributed to the actual patterns of genetic diversity (Soares et al., 2010; Pinhasi et al., 2012). These include the first colonization of the continent, movements due to the last glacial maximum and the transition from a hunter-gatherer way of life to farming, during the Neolithic expansion. The reason for the abundance of studies that focus on ancient mitochondrial DNA from archaeological remains is that they explain how these major events influenced the complex evolutionary history of Europe (Hervella et al., 2012; Brandt et al., 2013; Brotherton et al., 2013). Ancient DNA studies are also performed to identify more recent migration patterns, establish population affinities, origins and evaluate possible maternal relationships (Bogácsi-Szabó et al., 2005; Núñez et al., 2011; Ottoni et al., 2011; Deguilloux et al., 2014).
Genetic investigation of archaeological specimens dated to medieval times have the potential to improve our knowledge about past societies, as substantial changes in population movement occurred during the Middle Ages. Since the southeastern part of Europe, including Romania, was one of the key routes in population dispersal during medieval times, the analysis of mitochondrial genetic variation of past populations from this area, almost unrepresented in current literature, could fill a very large knowledge gap and allow us to identify some of the biological, socioeconomic and cultural factors that, in the end, influenced how the European population looks today from a genetic point of view.

Conclusions

All things considered, mitochondrial DNA is a remarkable molecular marker, widely used in forensics, medical studies and bioarchaeology due to its unique characteristics. A common tendency in the bioarchaeology field is to use multidisciplinary approaches and to integrate in the archaeological context, the information revealed by genetic studies. Even though, some of the questions regarding the past of the human species and its populations cannot find answers, the mitochondrial DNA remains the most straightforward tool used to unravel some of the puzzle pieces of the evolutionary history of humans.

Acknowledgments

The authors would like to acknowledge the project "Genetic Evolution: New Evidences for the Study of Interconnected Structures (GENESIS). A Biomolecular Journey around the Carpathians from Ancient to Medieval Times." (CNCSIS-UEFISCDI_PNII_PCCA_1153/2011, contract 229/2013).

REFERENCES


I. RUSU, B. KELEMEM