Department of Experimental and Health Science
Universitat Pompue Fabra

Genetic Analysis of the prehistoric peopling of Western Europe: ancient DNA and the role of contamination

TESIS DOCTORAL

Mª Lourdes Sampietro Bergua

Evolutionary Biology Unit
Experimental and Health Science Department
Pompeu Fabra University

Barcelona, Noviembre de 2006

Department of Experimental and Health Science
Universitat Pompue Fabra

Genetic Analysis of the prehistoric peopling of Western Europe: ancient DNA and the role of contamination

Mª Lourdes Sampietro Bergua

Memoria presentada por Mª Lourdes Sampietro Bergua para optar al titulo de doctor en Ciencias de la Salud y de la Vida. Esta tesis doctoral ha sido realizada bajo la codirección del Dr Carles Lalueza-Fox (Universidad de Barcelona) y del Dr Jaume Bertranpetit i Busquets (Universitat Pompue Fabra) en la Unidad de Biología Evolutiva, Departamento de Ciencias Experimentales y de la Salud, Universitat Pompeu Fabra. PhD Program in Health and Life Science (2002-2004).

Carles Lalueza-Fox Director	Jaume Bertranpetit i Busquets Director	Mª Lourdes Sampietro Bergua

Barcelona, Noviembre de 2006

A mis padres, a mi hermana
A Oscar

ABREVIATIONS

aDNA: ancient DNA
A: adenina
AMH: anatomically modern humans
AP: apurinic site
Asp: Aspartic
Bp: base pairs
BP:Before present
BSA: bovine serum albumine
C: Citosina
Ct: Cycle threshold parameter
CRS: Cambridge Reference Sequence
DNA: Deoxyribonucleic acid.
EDTA: ethylenediaminetetra-acetic acid
dNTP´s: deoxiribonucleotides triphosphate.
ddNTP´s:dideoxiribonuclleotides triphosphate
G: Guanina
GC: Gas Chromatrography
H strand: heavy strand
H_2O_2: hydrogen peroxide
HX:hypoxantine
HVR: Hyper variable region
KYA: kilo years ago
L strand: Light strand
MS: Mass spectrometry
mtDNA: mitochondrial DNA
MYA: million year ago
MW: molecular weight
N_A: Avogadro´s number
$N_{fe:}$ Effective population size
NUMTs: nuclear mitochondrial sequences
$\cdot O_2$: peroxide radicals
·OH: hydroxy radicals
PC: Principal components
PCA: Principal Component Analysis
PCR: polymerese chain reaction
Ppm: parts per million
PTB: N-phenacyltiazolium bromide
RT-PCR: Real Time PCR
SAM: S-adenosylmethionine
SDS: sodium dodecyl sulphate
SNP: Single Nucleotide Polimorphism
T: Timina
TE: Tris-Edta
Taq polymerase: Thermus aquaticus DNA polymerase
TMCR: Time Most Recent Common Ancestor
UNG: Uracil N-Glycosylase
UV: Ultraviolet light
X-gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

1INTRODUCTION

1.1 HISTORY OF ANCIENT DNA

The history of ancient DNA (aDNA) starts only twenty-two years ago when Higuchi and collaborators (Higuchi et al., 1984) extracted and sequenced, by means of molecular cloning, DNA fragments from a museum specimen of the quagga (an Equid from South Africa that became extinct in the nineteen century). This specimen died 150 years ago. This finding revolutionised the field of molecular biology because it showed that it was possible to retrieve DNA from an organism from the distanct past. A year later, Svante Pääbo (Paabo, 1985), retrieved DNA molecules from 2,500 year old human Egyptian mummies dated using the same methodology,. Nevertheless, retrieving DNA sequences from ancient samples was difficult and methodologically laborious and the number of publications related to the field increased slowly at that time.

The improvement of aDNA research is associated with the discovery in 1987 of the PCR (Polimerase Chain Reaction) technology (Mullis and Faloona, 1987). In contrast to molecular cloning, PCR allows the specific exponential amplification of little amounts of DNA of interest, discarding other DNA sequences from the pool (i.e DNA from micro organisms). Bone and teeth were quickly found to be better sources of aDNA than soft tissues (Hagelberg et al., 1989) and this meant that museums suddenly become recognised as storehouses of preserved genetic information from the past.

Since the discovery of the PCR, the list of publications related to aDNA increased exponentially. High profile journals started publishing studies claiming that aDNA from specimens that were millions years old could be successfully extracted and sequenced. Lindhal denominated this phenomenon “the antidiluvian DNA” (Lindahl, 1993a). Examples of this period are the retrieval of aDNA from organisms preserved in amber (Cano et al., 1993, DeSalle et al., 1992), plants sediments dated from the Miocene (Golenberg et al., 1990) and even the retrieval of the *cytochrome b* mitochondrial gene from a dinosaur bone dated to over 80 million years ago (MYA) (Woodward et al., 1994b). However and, as we will see below (see chapter 1.3) the field is plagued by numerous technical problems; later, some of the most extraordinary claims that had been published so far, like the retrieval of DNA from the dinosaur or from the insects preserved in amber, have been proven to be false (impossible to reproduce or have been shown to derive from an identifiable source of contamination) and severely criticised (Paabo and Wilson, 1991, Lindahl, 1993b, Gibbons, 1994).

Despite this, recent advantages in knowledge about the tempo and mode of DNA template damage, sample contamination and biochemical diagenesis of the DNA molecule have improved aDNA studies to the extent that aDNA is now emerging as a viable scientific discipline (Willerslev and Cooper, 2005). Nevertheless, unimaginable technical improvements apart, studies of DNA sequences should be confined to the past one million years and more probably to the past 100,000 years in order to achieve credible results (Hofreiter et al., 2001b).

Several studies have begun to reveal the potential of aDNA retrieval to record the methods and processes of evolution. These studies provide a unique way to test models and assumptions commonly used to reconstruct patterns of evolution, population genetics and palaeoecological changes (Willerslev and Cooper, 2005). This includes studying the phylogenetic relationships between extinct species with extant species. Australian marsupial wolves (Thomas et al., 1989); New Zealand moa (Cooper et al., 2001), American ground sloth (Hoss et al., 1996a) and *Myotragus balearicus* (Lalueza-Fox et al., 2005a) are examples of about 50 extinct animal species for which this has been done (Paabo et al., 2004). Moreover, studying ancient populations provides the opportunity to track genetic changes in the population over the time. Examples of ancient population genetics are rabbits (Hardy et al., 1995), penguins (Lambert et al., 2002) or mice (Pergams et al., 2003). In addition, the discovery that diverse mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA) sequences may be preserved in permafrost and cave sediments (Willerslev et al., 2003) opens up the exciting possibility of studying ancient organisms even when no macroscopically identifiable remains are present (Paabo et al., 2004). Another source of aDNA investigated so far has been animal or human coprolites (Poinar, 2002). This source of aDNA has yielded insight into the diet and behaviour both humans and animal species in the past (Paabo et al., 2004).

Another important applications of the aDNA retrieval have been the study of the evolution of current pathogens such the bacteria *Mycobacterium tuberculosis* (Zink et al., 2001) or the virus *Yersinia pestis* (Gilbert et al., 2004). This is a potentially very exciting field because the evolution of some pathogens can be expected to be fast enough to allow genetic change to be follow over decades or centuries (Paabo et al., 2004).The study of the origins of domestications of animals such a cattle (Beja-Pereira et al., 2006) and plant species such as maize (Jaenicke-Despres et al., 2003) around 10,000 years ago has been also possible with the development of aDNA techniques.

As we will see below (see chapter 1.6), DNA sequences from mtDNA or cpDNA, that are present in hundreds of copies in each cell, are easier retrievable from ancient specimens than the single genome copies of nuclear DNA sequences. Therefore, this delimits, for example, the ability to resolve phylogenies of species that either diverged recently in time or so rapidly that different parts of the genome have different phylogenies (Paabo et al., 2004). However, a few studies have been reported so far where the successful retrieval of nuclear DNA genes has been possible (Greenwood et al., 1999, Jaenicke-Despres et al., 2003, Poinar et al., 2003, Poinar et al., 2006). The retrieval of nuclear genes of extinct species has opened up the possibility of knowing even the phenotypic characteristics of the species that inhabited our planet in the past (Rompler et al., 2006).

This improvement of the aDNA field in other species contrasts with the obtained on understanding the recent history of the human species (basically due to the big threat associated to modern human contamination when working with ancient human specimens; see chapter 1.3.5); ancient DNA has, however yielded insights into the peopling of the Americas (Stone and Stoneking, 1993, Stone and Stoneking, 1998), peopling of the Caribbean (Lalueza-Fox et al., 2001, Lalueza-Fox et al., 2003), peopling of Central Asia (Lalueza-Fox et al., 2004), peopling of Japan (Oota et al., 1995) or even the peopling of the Canary Islands (Maca-Meyer et al., 2004). Moreover, recently, aDNA has been useful to yield insights into the continuous debate about if the current European populations have a Palaeolithic or a Neolithic origin. The genetic analysis of ancient Neolithic remains from Central Europe supports a Palaeolithic origin of the current European population (Haak et al., 2005). However, one of the most exciting achievements in the human aDNA field was produced when the team supervised by Svante Pääbo recovered 380 base pair (bp) of the hypervariable mtDNA region from a Neanderthal specimen (Krings et al., 1997). This sequence showed that this individual carried a mitochondrial type quite different from those of contemporary humans and that this mtDNA fell outside the genetic variation of modern humans in phylogenetics trees. This result has been subsequently corroborated by more mtDNA sequences from the same individual (Krings et al., 1999) and from sequences of another eleven Neanderthal specimens spread all over Europe (Ovchinnikov et al., 2000, Krings et al., 2000, Schmitz et al., 2002, Serre et al., 2004, Lalueza-Fox et al., 2005b, Lalueza-Fox et al., 2006, Caramelli et al., 2006, Beauval et al., 2005, Orlando et al., 2006).

The breakthrough in ancient DNA sequencing came only one year ago when Krause et al (Krause et al., 2006) sequenced the entire mitochondrial genome (16,770 bp in length) of the

Pleistocene woolly mammoth, *Mammuthus primigenius* by using a powerful variant of the PCR technology known as multiplexing. Multiplex PCRs differs from standard PCRs by simultaneously amplifying multiple genetic targets instead of just one. Consequently, this study indicates that the entire mitochondrial DNA genome from extinct species can potentially be determined with just the same amount of DNA as the used in a standard-single locus PCR.

It seems clear that this fast improvement of the DNA technology will revolutionise the field of aDNA in the near future, and that it is going to be focused mainly on the retrieval of nuclear DNA sequences of extinct species. In fact, last year Margulies et al (Margulies et al., 2005) developed a new sequencing DNA technology that was able to sequence and assemblage 25 million bases in fragments of ~100bp in a four hour run. Taking advantage of this new technique, Poinar et al (Poinar et al., 2006) sequenced 13 million nuclear and mitochondrial base pairs from an extinct well-preserved mammoth dated 28,000 years ago. Recently, Svante Pääbo and his team launched the “Neanderthal genome project”, announcing that using the same DNA technology as Poinar et al did, the entire genome of the first extinct homo species will be available in the next two years.

Such kind of projects was unthinkable only a few years ago and it gives us an idea of how fast is moving the aDNA field. However, despite these technical advances, researchers must be aware that reliable results were not obtained unless many precautions and experimental controls are implemented in their studies; especially when working with ancient human remains. Future developments focused mainly on removing or repairing chemical damages in the ancient DNA templates or in discarding all kind of possible modern contaminations in the ancient sample would be very helpful in the development of the aDNA field.

1.2 DNA PRESERVATION

1.2.1 DNA MOLECULE

The DNA molecule is a polymer composed of four nucleic acid bases - two purines (Adenine [A] and Guanine [G]) and two pyrimidines (Citosine [C] and Thymine [T]) - linked to phosporilated sugars by means of *glycosilic bonds*. The phosporilated sugars are 2'deoxirribose units linked to each other by means of *phospodiester bonds* shaping a double helix. Sugar and phosphate backbones are placed outside, whereas the bases are placed inside the DNA molecule. Hydrogen bonds are formed between purines and pyrimidines on opposite chains, always matching A with T and G with C (see
Figure 1).

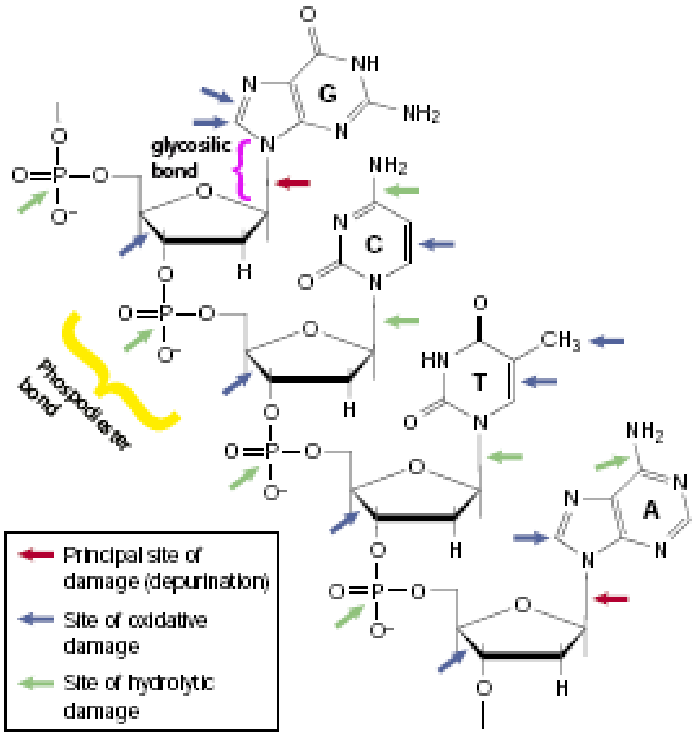


Figure 1: A short segment of one strand of the DNA double helix is shown with the four common bases. Sites susceptible to hydrolytic cleavage are indicated by green arrows and those prone to oxidative damage by blue arrows. Sites susceptible to depurination are pointed with a red arrow. A yellow arrow indicate a phospodiester bond whereas a purple arrow indicates a glycosilic bond G, guanine; C, cytosine; T, thymine; A, adenine. (modified from (Hofreiter et al., 2001b)).

The DNA molecule carries the information for making one organism from one generation to the next one. Because DNA uniquely serves as a permanent copy of the cell genome, changes in its structure have strong impact in the fitness of the cell. DNA damage can block replication or transcription and can result in a high frequency of mutations, consequences that are unacceptable from the standpoint of cell reproduction (Cooper, 1997) .

1.2.2 DNA STABILITY IN VIVO

However, the DNA is one of the least stable molecules within the cells (see Table 1) and, as a consequence, it is labile and prone to many forms of damage. To protect the DNA molecule against that damage, cells have developed structural mechanisms of protection. In vivo, the molecule is protected by water molecules in the major grooves of the double helix. Furthermore, nuclear DNA (but not mtDNA) is linked to histones; these proteins have an important role in the folding of the DNA molecule inside the nucleus of the cells and presumably avoid the exposition of the DNA to the surrounding environment. (Poinar, 2002).

Table 1. Cellular compounds, their susceptibility bonds and groups and their preservation potential. – to ++++ (weakest to strongest), a rough estimate of the preservation potential for an unaltered molecule base upon relative bond strengths (adapted from (Poinar, 2002).

Compounds class	Susceptible bonds	Susceptible groups	Preservation potential
DNA, RNA	Phosphate esters, gycosidic bond, 6 C-C bond pyrimidines	Heterocyclic rings Amino groups, Metil groups	-
Proteins	Peptide bond	Side chain, chiral center	-/+
Carbohydrates	Acetal	Hydroxy, amide	+
Lipids	Ester, ether, amide	Hydroxil, carboxil, ester	++
Cutin	Ester, ether	Hydroxil, carboxil	+++
Lignin	Ether	Metoxil aromatic rings	++++

1.2.2.1 HIDROLYTIC DAMAGE

The DNA molecule is particularly prone to hydrolytic damage – damage related to the presence of water- due to the following reasons:

1.) The **phospodiester bond** is quite labile because of the lack of 2'-OH group in ribose units and subject to quick hydrolytic cleavage generating single-stranded nicks in the double helix. This event take place about once every 2.5 hours in a hydrated system while under dry conditions this rate drop some 20-fold (Lindahl, 1993a, Poinar, 2002).

2.) The **glycosidic bond** is prone to base protonation. Nucleotide bases become a likely leaving group thus causing the cleavage of the bond. The process is termed *depurination* and forms what it is know as an apurinic/apyrimidinic site (AP site). Apurinic sites are much more frequent than the apyrimidinic ones. Once a nucleotide base is released from DNA, the AP site can undergo cleavage of the DNA molecule and thus result in a single-stranded nick (Lindahl, 1993a, Poinar, 2002).

3.) **Deamination**: bases with amino groups such as adenine, cytosine and guanine can undergo deamination resulting in hypoxanthine, uracil, and xanthine, respectively (Poinar, 2002).

1.2.2.2 OXIDATIVE DAMAGE

The oxidative damage is believed to occur through the action of free radicals such as peroxide radicals ($\cdot O_2$), hydrogen peroxide (H_2O_2) and hydroxy radicals ($\cdot OH$) (Lindahl, 1993a). These radicals are endogenously generated by the cell but they may also derive from exogenous sources such as ionising radiation, UV light (UV light can cause also pyrimidine dimmers in the DNA double helix) and cellular processes during bacterial and fungal degradation (Poinar, 2002).

1.2.2.3 NONENZYMATIC DNA METHYLATION

In addition to oxygen, living cells contain several other small reactive molecules that might cause DNA damage and act as an endogenous genotoxic agent. The best characterized is S-adenosylmethonine (SAM), a metil group donor that act mainly over the ring nitrogen's of purine residues. 3-Methyladenine, one of the products derived of this reaction, is a citotoxic that blocks replication. In living cells, this lesion is rapidly repaired (Lindahl, 1993a).

1.2.3 DNA REPAIR

Cells had to develop mechanisms to repair damaged DNA to maintain the integrity of their genomes. We can distinguish two types of mechanisms:

1.2.3.1 DIRECT REVERSAL OF DNA DAMAGE:

Some lesions in DNA can be repaired by direct reversal of the damage. This is the most efficient way of dealing with frequent specific types of DNA damage. Two important examples are the repair of alkilated guanine residues and pyrimidine dimers. The former is repaired by the action of a metyltransferase enzyme, which is present in all eukariotes, including humans. The later is repaired by means of a process called photo- reactivation that uses the energy derived from visible light to break the pyrimidine dimer (Cooper, 1997).

1.2.3.2 EXCISION REPAIR:

Although direct repair is an efficient way of dealing with particular types of DNA damage, excision repair is a more general mean of repairing a wide variety of chemical alterations of the DNA. In excision repair, the damage is recognised and removed. The resulting gap is then filled in by synthesis of a new DNA strand, using the undamaged complementary strand as a template (Cooper, 1997).

1.2.4 DNA DAMAGE AFTER CELL DEATH

Once an organism dies, its decomposition starts due to the unchecked cellular activity of lipases, proteases, amylases and nucleases.
Concerning to DNA, the most important factor in its long-term preservation is the rate at which **nuclease activity** can be stopped. Since the activity of these enzymes is oxygen-dependent, this occurs once the oxygen inside the cell is finished. Despite particular conditions that can allow DNA escape from these enzymes (for example low temperatures (Hoss et al., 1996b)), other processes can damage the DNA molecule. In the cells of a living person, DNA is continually protected from damage by sophisticated repair systems (Cooper, 1997). However, after death, not only these repair mechanisms stop working but also the double helix is subjected to unfavourable environmental conditions that increase the DNA damage rate. As a consequence, the retrieval of DNA from bones and/or other tissues of long-dead organisms can be extremely difficult (Poinar, 2002). The post mortem biochemical modifications that the DNA molecule can suffer are believe to be analogous to those seen in vivo (see former sections) and act via the fragmentation of the molecule's chemical backbone and/or the alteration of individual nucleotide bases (Hoss et al., 1996b).

1.2.4.1 DNA FRAGMENTATION

The most obvious type of post-mortem DNA damage is its fragmentation into small sequences from 100 bp to 500 bp (Paabo, 1989). That type of degradation is due to the action of endonucleases that occur shortly after death (see before), the action of micro organisms and the hydrolytic or oxidative cleavage of phosphodiester bonds (Lindahl, 1993a).

1.2.4.2 NUCLEOTIDE MODIFICATION

Post mortem DNA nucleotide modifications are mainly due to hydrolytic and oxidative damage. Regarding to oxidative damage, Hoss et al (Hoss et al., 1996b) were able to determine the amounts of eight oxidative base modifications in extracted DNA from bones and soft tissues specimens ranging in age from 40 to 50000 years old by means of Gas

chromatography/mass spectrometry (GC/MS). The hydantoin derivatives of pyrimidines compounds were the most frequent (see

Figure 2).

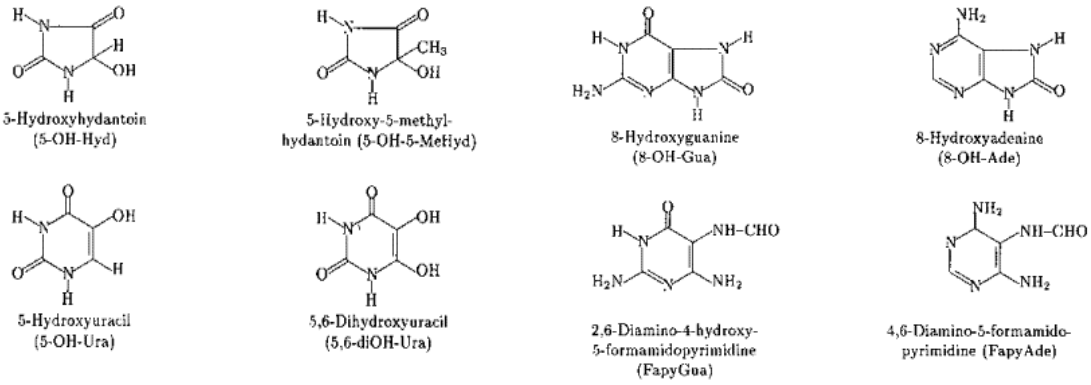


Figure 2: Structure of eight oxidative base modifications detected in ancient DNA extracts (from (Hoss et al., 1996b).

And regarding to hydrolytic nucleotide modification, the main modifications described so far are the hydrolytic deamination of thymine or of adenine. The result of these processes is the base change in the DNA backbone to uracil or hypoxantine respectively (Hofreiter et al., 2001b) (see Figure 3).

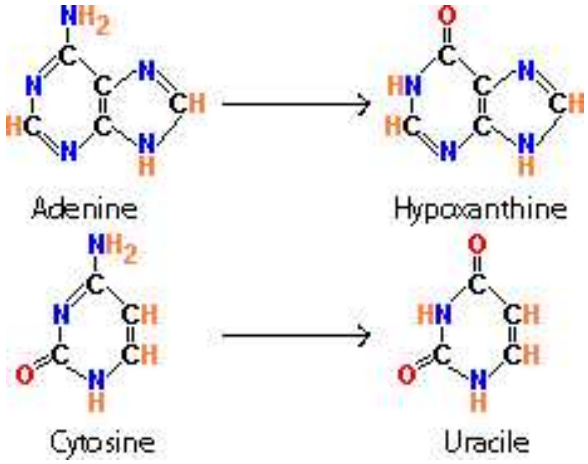


Figure 3: Deamination Cytosine and Adenine to Uracile and Hypoxantine.

1.2.5 DNA SURVIVAL

If the time after death is long enough, the cumulative effects of damage to the DNA could be so extensive that could destroy all the DNA molecules; all these molecules would be transformed to mononucleotides or derivates. Theoretical considerations based on the chemistry and physics of DNA, also taking into account favourable environmental conditions, indicate that it cannot survive any longer than one million years, and probably not longer than 100,000 years in most cases (Lindahl, 1993a). Nevertheless, in exceptional circumstances, this degradation can be significantly reduced. Such conditions include a fast inactivation of the nucleases activity, inhibition of the action of micro organisms, fast desiccation, low temperatures (Hoss et al., 1996a, Smith et al., 2003) and high salt concentration (Hofreiter et al., 2001b). Of such conditions, the temperature is the most important environmental factor in the preservation of the genetic material; due to chemical organic reactions that lead to DNA damage take place with low rate at low temperatures. Hoss et al (Hoss et al., 1996b) showed that a reduction of 20ºC in the average temperature of the fossil remain lead to a reduction of around 10-15 times the frequency of those reactions.

1.3 DNA RETRIEVAL FROM FOSSIL REMAINS

As we have already seen, it is likely that DNA from a fossil remain is undergoing some kind of post mortem DNA damage. Furthermore, fossil remains tend to be rare and precious, and recovering DNA from them implies using destructive techniques. Thus, it seems logical that the first question that an ancient DNA researcher may wonder before starting the study of an ancient sample is: *"How likely is retrieving DNA for this sample?"*. Some empirical rules that can help us on knowing if it is worth trying to extract DNA from a fossil remain has been described. First, DNA preservation from a fossil correlates with the environmental characteristics of the archaeological site where it has been found (temperature, pH...). For instance, the possibility of retrieving DNA from a sample found in permafrost is higher than in a sample found in a desert (Willerslev et al., 2003). Smith et al (Smith et al., 2003) argued that the **thermal history** of hominid fossil is a key parameter for long term survival of bio molecules in the fossil record. The **thermal age** of a hominid fossil was defined as the time taking to produce a given degree of DNA degradation (assuming DNA depurination as the principal mechanism of degradation) when temperature is held constant at 10ºC. This analysis suggests that 17,000 years at 10ºC may be a practical upper limit for DNA survival. Second, the macroscopic appearance of the fossil remain, specially the degree of porosity of the bone (Gilbert et al., 2005). Third, and contrary to what could intuitively be expected, the age of the sample is not a good indicator of the state of DNA preservation (Paabo et al., 1989, Hoss et al., 1996b). In other words, a young fossil remain does not necessarily imply a better DNA preservation than an older fossil remain.

Thus, as we have already seen, many factors are involved in the DNA survival. An indirect practical test has been introduced to compute the degree of damage of a fossil DNA sample (Poinar et al., 1996). It only requires a small amount of the specimen to be analysed (about 50 times less that would be used for a DNA extraction). This test measures changes in the three dimensional structures of aminoacids which correlates with the degree of DNA damage rate (see Figure 4). In all aminoacids used in proteins, except glycine, there are four different chemical groups attached to one carbon atom know as an alpha carbon. These groups are arranged in a tetrahedral shape, with the carbon at the centre. There are two different ways to arrange the groups, which are chemically identical, but mirror images of each other, called **stereoisomers**. Aminoacids in living systems exist only in one estereoisomeric form, called

the L-form. After death, however, transitions to the D-form occur (a process known as **recemization**), and eventually a dynamic equilibrium is reached in which the proportions of L- and D- forms are equal. The rate at which racemization takes place differs for each aminoacid and is dependent on the presence of water, the temperature and the chelation of certain metal ions to proteins. Racemization is thus affected by some of the same factors that affect depurination of DNA, the major hydrolytic reaction responsible of the spontaneous degradation of nucleic acids (Lindahl, 1993a). The racemization of Aspartic Acid (Asp), which has one of the fastest racemization rates, has an activation energy and rate constants over a wide temperature range (at neutral pH) that are similar to those for DNA depurination (Lindahl, 1993a). Poinar et al (Poinar et al., 1996) observed that no DNA sequences could be retrieved from samples in which the D/L Asp ratio was higher than 0.10. Furthermore, given that the amount of fossil aminoacids is extremely small, contamination with modern aminoacids will tend to substantially modify the D/L ratio of all of them. Thus, since the racemization rate of the Asp is faster than any other aminoacid of the same age, a D/L ratio for Asp that is lower than the rate for any other aminoacid should be an indication of contamination by more recent aminoacids.

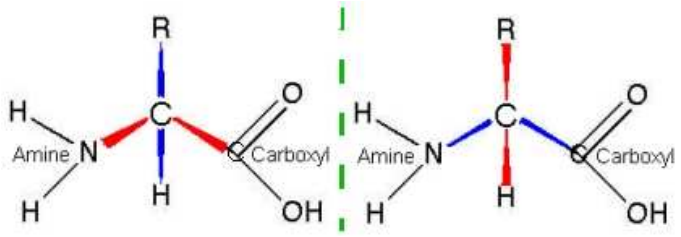


Figure 4: Amino acid racemization. R indicates the amino acid side chain, which varies between aminoacids.

However, Serre et al (Serre et al., 2004) found that the preservation of endogenous DNA in fossils is correlated not only with the degree of aminoacid racemization content but also with the amount and composition of the amino acids in the ancient sample. They found, in addition to the previously defined D/L cut off, that endogenous DNA from a Pleistocene remain can be amplified when the amino acid content is higher than 30,000 parts per million (ppm) and the ratio between glycine to aspartic acid between two and ten (see Figure 5).

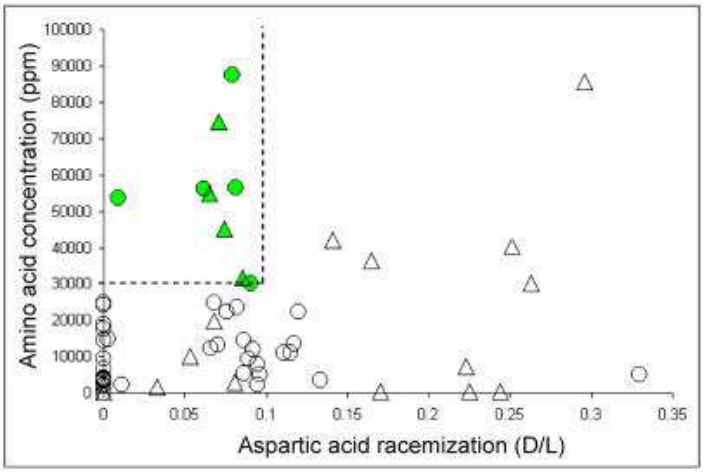


Figure 5: Amino Acid analysis of 64 Hominid remains. For each bone, the extend of Aspartic Acid racemization (D/L) and the aminoacid concentration (ppm) is given. The dash lines delimit the area of amino acid preservation compatible with DNA retrieval. Circles and triangles represents early modern humans and Neandertals respectively (from (Serre et al., 2004).

Nevertheless, DNA retrieval from a well-preserved sample (environmental conditions, positive Asp racemization test and large amino acid concentration) is not going to be an easy task. The aDNA researcher will face to a big amount of technical problems:

1.3.1 DNA COPY NUMBER

The number of DNA molecules is a crucial factor when trying to recover aDNA from an ancient specimen (e.g. (Handt et al., 1994a, Handt et al., 1994b, Krings et al., 1997, Handt et al., 1996). The researcher has to be sure that the number of endogenous molecules in an ancient DNA extract is high enough in order to be confident about the final results. The exponential nature of the PCR process implies that if the starting number of aDNA molecules is extremely low, amplified aDNA molecules could be biased towards particular subsets of copies of putatively damaged template. Thus, conclusions based on these results could be erroneous (see section 1.3.6). Furthermore, knowing the number of molecules in the extract is extremely important due to the problem of modern human contamination (see section 1.3.5); the lower the amount of endogenous DNA molecules in the extract, the more likely the PCR will be contaminated by exogenous DNA (see Figure 6). However, when working with ancient human remains, since the endogenous sequences and the putatively contaminant sequences are very similar, when quantifying them it would not be possible to differentiate between both sources of DNA templates.

The limitation of the number of endogenous DNA molecules is one of the main reasons for using mitochondrial DNA (mtDNA) instead of nuclear DNA (Paabo et al., 2004). Since each cell contains several mitochondria (up to hundreds) and each mitochondria contain several genomes (see section 1.6), it is easier to be successful in amplifying sequences from the mtDNA genome than from the nuclear DNA genome (e.g. (Paabo, 1989, Handt et al., 1994b); the ratio can be 1 to 10,000 (Robin and Wong, 1988).

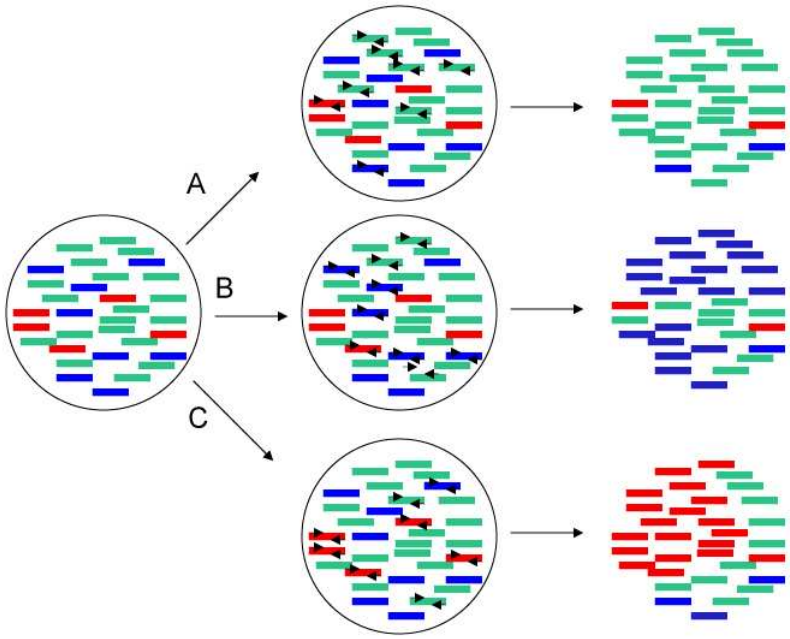


Figure 6: A) PCR result if the first steps of the PCR reaction are based on damaged molecules. B) PCR result if the first step of the PCR reaction are based on contaminants sequences. C) PCR result if the first step of the PCR reaction are based on the endogenous molecules

To find out if the number of DNA molecules is high enough to obtain reproducible results, two types of approach has been done so far. In the first studies (e.g. (Krings et al., 1997), the amount of endogenous DNA molecules were measured by using a molecular biology technique called **quantitative PCR**. The basic principle of the quantitative PCR is the construction of a "competitor construct". This construct is a sequence equal to the endogenous we are trying to quantify. The only difference is that it has a deletion in the sequence. Once we have the construct, several PCRs are set up, each of which carries different dilution of the "competitor construct". The primers are specific for the putative endogenous sequence. As the construct is smaller than the putative endogenous sequence, it is amplified easier in the first steps of the PCR and subsequently the final product will be almost only the construct. When

the concentration of molecules of the putative endogenous DNA is higher than the concentration of the construct, we start having results of the two sizes. Since the concentration of the construct is known in each PCR reaction, the concentration of the putative endogenous sequence can be extrapolated (see Figure 7).

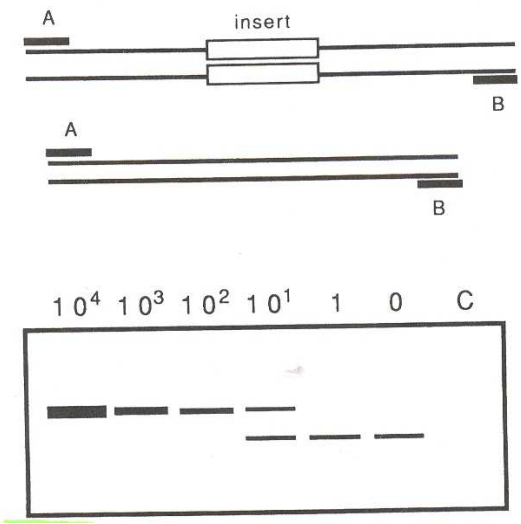


Figure 7: Schematic illustration of quantitative PCR. Above the template from a tissue extract which is amplify by primers A and B and the same template in which an insert has been introduced. To a constant amount of extract, a dilution series of a known amount of the insert template is given. Above the lanes, the number of added molecules is indicated. It can be seen that there are approximately ten copies of the target sequence in the extract added to the PCR. (adapted from (Handt et al., 1994a)

There are two main pitfalls of this technique. On one hand, this technique is tedious and on the other hand the results that we can obtain are only an approximation. For this reason, nowadays another technique is used to quantify the endogenous DNA molecules more accurately. This technique is based on the **RT-PCR (real time PCR) reaction** (Whelan et al., 2003, Alonso et al., 2004, Malmstrom et al., 2005) and it is the most accurate method to quantify a specific segment of DNA so far. In such RT-PCR, a probe lies downstream of the forward primer and has a fluorescent tag at the 5' end and a quencher at the 3' end. As the PCR proceeds, taq polymerase uses its 5'-3' exonuclease activity and destroys the probe thus generating fluorescence, which is proportional to the amount of amplified target DNA present. Hence, if fluorescence is rapidly detected, then large amounts of DNA are present and vice versa when release is small (see Figure 8). Despite the exquisite sensitivity of this reaction, it is necessary to use some form of calibration in order to know the number of target molecules (see (Alonso et al., 2004) in order to know how to construct a mtDNA standard).

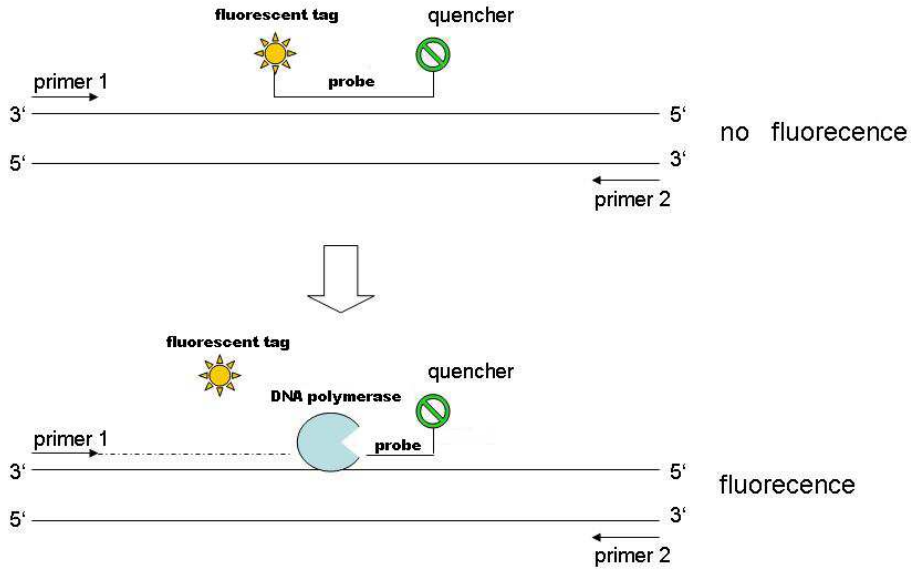


Figure 8: Real Time PCR experiment. A probe lies downstream of the forward primer and has a fluorescent tag at the 5' end and a quencher at the 3' end. As the fluorescent tag and the quencher are closed, the fluorescent emited by the tag is annulled by the quencher.Since the taq polymerase has exonuclease 3'-5' activity, when it is elongating and find the probe it is able to digest it. When the probe is degested, the fluorescent tag and the quencher are far away so, fluorescent is emitted. The fluorescent emmited is proportional to the number of molecules in the sample.

It is known that if a PCR starts from 1,000 or more molecules, then an experiment does not need to be repeated to verify that nucleotide misincorporation does not influence the final result. Nevertheless, several repetitions become necessary when fewer molecules initiate a reaction. Hofreiter et al, (Hofreiter et al., 2001a) claimed that at least two independent amplification products are needed when the extracts of ancient specimens contain few template molecules in the PCR. In the case of obtaining discrepancy results then, at least, one more amplification should be performed to determine which of the two sequences is reproducible (see Figure 9). They argued that when such precautionary measurements are taken, errors induced by damage to the DNA template are unlikely to be more frequent than 0.12% even under the unlikely scenario where each amplification starts from a single template molecule.

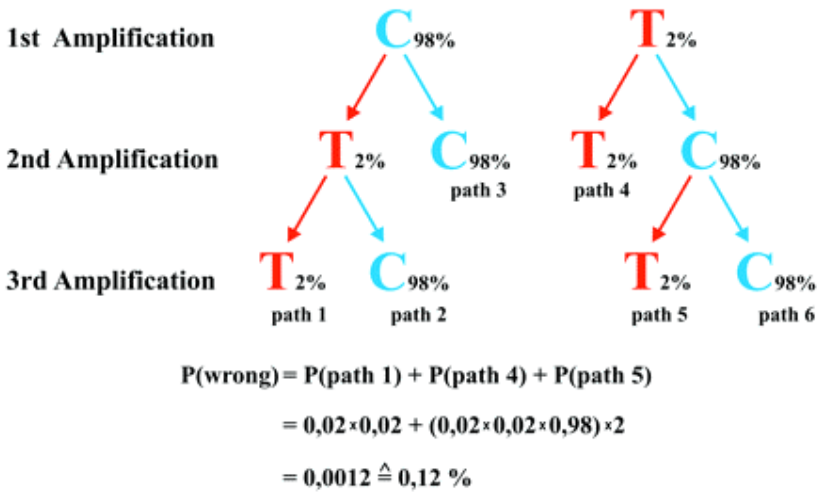


Figure 9: Schematic illustration of the strategy where two or three amplifications are used to determine an ancient DNA sequence. The correct base (C) is blue and the incorrect base (T) red. An error rate of 2% is assumed and the probability of determining an incorrect base is the sum of the three paths that arrive at the incorrect base after two or three amplifications (from (Hofreiter et al., 2001a).

1.3.2 DNA FRAGMENTATION

Several studies have been published so far (Paabo, 1989, Handt et al., 1994a, Handt et al., 1994b, Poinar et al., 2006) showing that it is almost impossible to obtain long amplification products from an ancient template. Furthermore, it has been observed an inverse relationship between amplification efficiency and length of the amplification products (e.g. (Handt et al., 1994a, Malmstrom et al., 2005). In practice, this means that long sequences can only be retrieved by means of short sequences using **overlapping primers** (see Figure 10). The maximum length of the overlapping amplification will depend on the state of preservation of the sample. For instance, Lalueza et al (Lalueza-Fox et al., 2006) were able to amplify 300 bp of the mitochondria control region of a Neanderthal specimen by using overlapping fragments of only 70-80 bp. In contrast, Cooper et al in 2001 (Cooper et al., 2001) were able to amplify the complete mitochondrial genome sequence of two extinct moas as a series of 400-600bp amplification product.

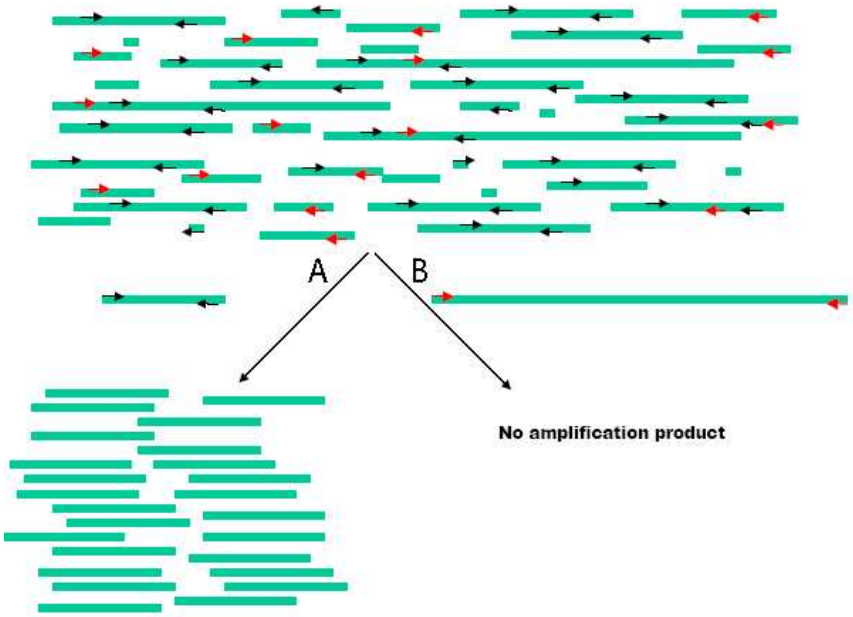


Figure 10: A representation of fragmented strands. A) A situation in which we try to amplify an enough short fragment so both primers can hybridise and therefore the PCR amplifiction is possible. B) A situation in which we don not get any amplification product due to one of the primers can not hybridise.

1.3.3 JUMPING PCR

It is a phenomenon that can occur during the PCR reaction due to the presence of fragmentation in the molecule templates; it occurs when one template "recombines in vitro" with another that shares a similar sequence during the PCR reaction. It was described for the first time in 1990 by Svante Pääbo and collaborators (Paabo et al., 1990). They designed specific pairs of templates partially sharing the same sequence and tried to amplify them by a pair of primers, one primer specific to one template and the other primer specific to the other template (so neither of the templates contained both primer sites). They showed that the only way to obtain PCR products was by fragmenting the templates with restriction enzymes, so jumping PCR happens (Paabo et al., 1990). Furthermore, they found that lesions such as breaks or AP sites can cause the extending primer to jump to another similar template during the PCR. Consequently, amplification products from damaged templates such as archaeological DNA could be made up of a high proportion of chimerical molecules between endogenous and contaminant sequences or endogenous with post mortem damage and without post mortem damage (see Figure 11). Special care has to be taken into account when

analysing such results, since the chimerical products of this reaction could be erroneously taken as a novel ancient DNA sequence.

Molecular cloning and sequencing of multiple clones is highly recommended in order to detect such kind of phenomenon in our sample; these analyses will help us to sort out the different types of molecules present in the amplification products and distinguish chimeric products from those that are not. Nevertheless, it has been argued that the presence of a jumping PCR phenomenon in the PCR products could be used as an indication that the DNA may be of ancient origin (Handt et al., 1994a).

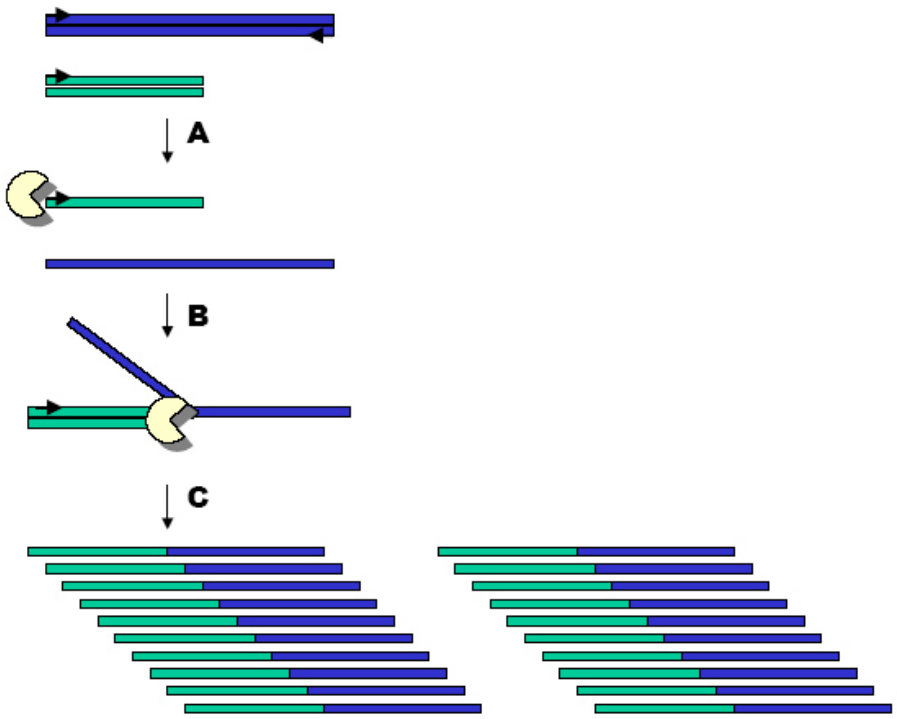


Figure 11: A schematic representation of a jumping PCR event. Two similar sources of DNA molecules are represented, althought one of them is fragmented (the green one). A) DNA polymerase start elongating the fragmented molecule. B) It can be possible that when the DNA polymerase arrives at the end of the fragmented fragment, can jump to the other source of DNA molecues (green bars) and continue elongating it. C) Therefore, at the end a lot of recombining molecues are generated.

1.3.4 INHIBITORS

It is well-known for the aDNA community that when nucleic acids are extracted from ancient specimens, the extracts often contain components that can inhibit the activity of the DNA polymerises (Hanni et al., 1995); furthermore, it is known that that inhibition affects longer amplification fragments more than shorter fragments (Pusch and Bachmann, 2004). However, after the extraction protocol and in order to avoid the presence of these inhibitors, the resulting aqueous phase is carefully removed (as carryover of organic solvents to subsequent extraction and amplification stages may inhibit PCR), and desalted and concentrated using usually filtration (Cooper et al., 2001). While this procedure usually removes possible PCR inhibitors of less than 30,000 molecular weight (MW), in some samples (such as these buried in the soil and paleofaeces) it is common to find some co-extraction of PCR inhibitors along with the DNA. One simple solution to deal with this problem is the incorporation of bovine serum albumin (BSA) (or other molecules that bind to inhibitory chemicals) to PCR reactions (Paabo et al., 1988). BSA binds to inhibitors and thus prevents them from inhibiting the polymerase enzymes activity. Another way of stopping the activity of these inhibitors wide-used by the aDNA community is by making serial dilutions of the extracts until we find one dilution for which the PCR is successful (e.g. (Paabo, 1989, Woodward et al., 1994a, Binladen et al., 2006).

Nevertheless, little is known about the nature of those inhibitors; it has been proposed that they are soil components such us humic or fluvic acids (Paabo et al., 1989) or sub products derived from organic reactions, such the volatile compound products of the Maillard reaction (Poinar, 2002). The Maillard reaction is defined as the reaction of carbonyl groups on reducing sugars (such as those that belong to the backbone skeletal of the DNA double helix) with the primary amines of aminoacid and forms a structure known as **cross links**. It has been demonstrated that treatments with the reagent PTB (N-phenacyltiazolium bromide) breaks the cross links DNA-protein and allows DNA sequences to be amplified from some ancient remains that otherwise could not be amplified (Poinar, 2002).

1.3.5 CONTAMINATION

We define a contaminant sequence as an exogenous (and probably modern) DNA molecule in the pool of aDNA sequences. Contamination is one of the most common pitfalls in aDNA studies and represents one of the biggest threads when working with samples from ancient humans. The problems regarding ancient human DNA and modern human contamination, often detected of more than one haplotype of one single individual has been known several years ago(Handt et al., 1994a, Handt et al., 1996, Kolman and Tuross, 2000, Gilbert, 2005b).
As the retrieval of tiny amounts of DNA from ancient remains is a multi-step process, contaminants can enter in multiple stages:

1.) **Sample Handling**: This kind of contamination is extremely important in the study of ancient human remains (the main objective of the current thesis) because usually cannot be monitored or controlled (only it can be monitored in contemporaneous excavations). Unprotected handling of the remains may impregnate the samples with the handlers sweat or skin cells, and thus exogenous DNA could penetrate into the remains(Gilbert, 2005a). Bone and teeth remains are extremely porous and thought very susceptible to contamination by handling. The pulp cavity of the tooth is directly connected to the exterior by numerous dental tubules. Therefore, once a tooth is washed or the root is directly handled, the impervious nature of enamel is of no help in limiting contamination by handling (Gilbert, 2005b). Mercury porosimetry demonstrates that the minimum total interconnected porosity in human bones is higher than 8% of the bone volume, the majority of witch is derived from the Harvesian canals(Gilbert, 2005b). We can distinguish different potential sources of handling contaminants depending on the step of DNA retrieval:

   - Archaeologists and anthropologists: they are responsible of recovering, washing and macroscopically analysing the fossil remains and interacts directly with the fossil material.
   - Geneticists: Once the sample arrives to the aDNA laboratory, the genetic team must decontaminate the sample using bleach or UV light before starting the DNA extraction. In this step it is compulsory the use of gloves,

mask face and coverall inside the aDNA laboratory in order to avoid contamination by handling.

2.) **Extraction procedure**: Contaminants can be introduced in the aDNA extract during the preparation of all the reagents that are going to be used in the extraction process as well as in the place were the extraction is going to be performed. In order to avoid this kind of contamination, it is very important that the place where the aDNA extraction is carried out is physically separated from the main laboratory (molecular biology laboratory) with positive air pressure, UV light at night and continually bleach cleaning of the bench surfaces (Cooper and Poinar, 2000). In addition, the manipulation of all reagents must be done in a flow cabinet and the use of coverall, gloves, facemask and sterile tips is compulsory.

3.) **PCRs setting up**: Contaminants can be introduced in this stage by two main ways: First, due to the low specificity of the primers. PCR reaction are based upon the premise that primers will bind to specific loci not present in unrelated organism, and thus will selectively amplify only the DNA of interest under specific PCR conditions (e.g. annealing temperature) However, there could be DNA from some soil micro organisms present in the aDNA extract. The low amount of aDNA forces using unspecific conditions (e.g. high number of cycles and low temperature) in the PCR. For this reason, it could be that the primers can also bind to those similar sequences that are present in the extract and therefore amplify them. Second, contaminants can be introduced while researchers are setting up the PCR reaction. As in the extraction step, contaminants can be in the used reagents and in the environment where researchers are working. All the precautions that are described in the previous step are compulsory here as well. The setting up of the PCR reaction must be done in the ancient laboratory and then carried to the main laboratory. Material interchange between the aDNA laboratory and the main laboratory should be strictly forbidden because the risk of contamination is extremely high due to the large amount of amplicons that are generated in the molecular laboratory (from 2^{20} to 2^{30} copies in each PCR).

1.3.6 MISCODING LESSION

Unfortunately, although the majority of the damage that the DNA molecule can suffer after cell death inhibits the activity of PCR enzymes (for example, strand fragmentation or intermolecular cross-links; see sections 1.3.2 and 1.3.4), a small proportion of damage events do not inhibit the polymerase activity but generate *miscoding lesions*. Miscoding lesions are defined as base modifications in the amplified sequence that change the appearance of aDNA template (Fattorini et al., 1999) and potentially generate misleading haplotype analysis (Gilbert et al., 2003a, Gilbert et al., 2003b). It has been reported several times (for example, see (Paabo, 1989, Hofreiter et al., 2001a, Binladen et al., 2006) that the majority of miscoding lesions arise from the deamination of C to U or from A to hipoxantina (HX); this process is particularly fast in the case of cytosine (Lindahl, 1993a). When a DNA molecule containing such lesions is used as a PCR template, C→T and G→A transitions will be introduced erroneously in the final sequences. However, because either of the complementary DNA strands can be sequenced after amplification when using Sanger sequencing technology, each of these transitions can produce two observable phenotypes (see Figure 12). Hansen et al (Hansen et al., 2001) termed each set of miscoding lesions as *type I* (A→G and T→C) and *type II* (G→A and C→T).

Interestingly, the recent development of the sequencing-by-synthesis technology (see (Margulies et al., 2005) offers the possibility of going deeper into the nature of those miscoding lesions (Gilbert et al., 2006, Stiller et al., 2006). The advantage of this new technique is that the nature of the date generation process is such that DNA sequence data can be assigned to individual, originally single-stranded molecules (Gilbert et al., 2006). Basically, DNA extracted from an ancient remain is ligated to biotynilated linkers and single DNA strands are attached to Sepharose beads, amplified by PCR and subjected to pyrosequencing (Margulies et al., 2005). Therefore, each read sequence derives from one single-stranded DNA molecule, and the read sequence allows the actual template strand to be inferred. (Gilbert et al., 2006) using this new technology demonstrated thought comparative analyses on 390 965 bp modern chloroplast and 131 437 bp ancient woolly mammoth sequence date that type 2 (G→A and C→T) miscoding lesions represent the majority of damage derived miscoding lesions. Furthermore, they suggested, in contrast to previous studies (Paabo et al., 1989, Hofreiter et al., 2001a, Gilbert et al., 2003a, Binladen et al., 2006) that the predominant cause of Type II transition is not the cytosine to uracil deamination, but the degradation of guanine to a derivative that is misread by the polymerase as an adenine.

Nevertheless, Stiller et al (Stiller et al., 2006) using the same DNA technology as (Gilbert et al., 2006) found that the predominant cause of Type II transitions are the cytosine residues when analysing an ancient mammoth remain. They suggest that the lesion affecting cytosine residues is very likely to be by deamination.

Given the importance of the control region of human mtDNA studies (see chapter 1.6), Gilbert et al (Gilbert et al., 2003a, Gilbert et al., 2003b) studied the post mortem damage distribution of this region. Contrary to what a priori is expected, they found that postmortem DNA damage is not randomly distributed across the control region, but there are "hotspots" of postmortem damage. In other words, there are some positions that are more prone to suffer postmortem damages than others. Furthermore, DNA damage occurs preferentially at those positions that have been detected as fast evolving in human population studies. Gilbert et al (Gilbert et al., 2003b) hypothesise that these mutational hotspots may be hyper mutable because they are more exposed to environmental damage than the others or they are not binding sites of proteins that can protect them. They concluded that postmortem damage in the control region could difficult population genetic analysis of ancient humans due to the impossibility to distinguish whether a mutation is due to a postmortem damage or it was already in the sequence.

The use of the enzyme UNG (Uracil N-Glycosylase) has been highly recommended in order to avoid the presence of miscoding lesions in the final sequence. This enzyme excises the uracil bases in the DNA molecule (produced due to the deamination of cytosine) and therefore it is able to reduce the apparent G→C and A→T mutations and the subsequent errors in the sequence results (Hofreiter et al., 2001a).

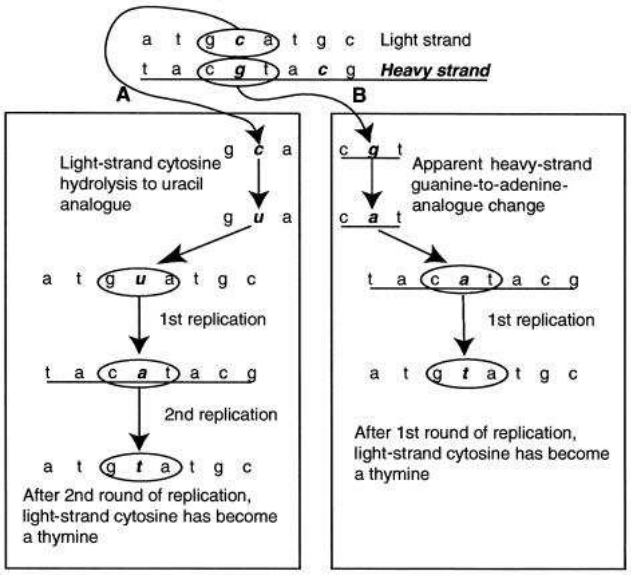


Figure 12: Determination of a strand of origin for postmortem-DNA-damage events by using type 2 (C→T/G→A) transitions as an example. *A*, L-strand C→T transitions after two cycles of amplifications, resulting in a permanent L-strand change. *B*, A theoretical H-strand G→A change, producing the L-strand phenotype of C→T change following one cycle of amplification. However, since a direct G→A postmortem modification is chemically impossible, the example depicted in this panel is not possible. Thus, all C→T changes observed on the L strand must have occurred as L-strand C→T postmortem damage, and all G→A changes on the L strand must have occurred as H-strand C→T postmortem damage (From (Gilbert et al., 2003a)

1.3.7 MOLECULAR CLONING

Molecular cloning is a well-known molecular technique in aDNA field. It allows us to detect sequence heterogeneity in a single PCR reaction, which can be associated to jumping PCR, contamination events and/or miscoding lesions by endogenous DNA damage or Taq polymerase errors. Cloning PCR products consist in inserting a single PCR product inside a bacterial plasmid, which is subsequently inserted into a bacterial cell (*Escherichia.coli*). After growing up the bacteria, a posterior screening will allow us to select those bacteria that carry the insert. Colonies derived from successfully transformed cells can be identified through blue/white screening due to the disruption of the plasmid's β-galactosidase gene (which metabolises X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) into a blue product). Therefore, white colonies carry the insert (these colonies are not able to metabolise X-gal and therefore β-galactosidase gene has been disrupted because of the insert) whereas blue colonies do not.

Since each bacterium incorporates a single amplicon in its genome, postmortem DNA damage, jumping PCR and contamination can be assessed (see Figure 13) by screening a reasonable number of clones (see(Bower et al., 2005).

Nowadays, when working with ancient templates (mostly when working with human ancient templates) cloning each PCR product is crucial to detect sequence heterogeneity in the extract.

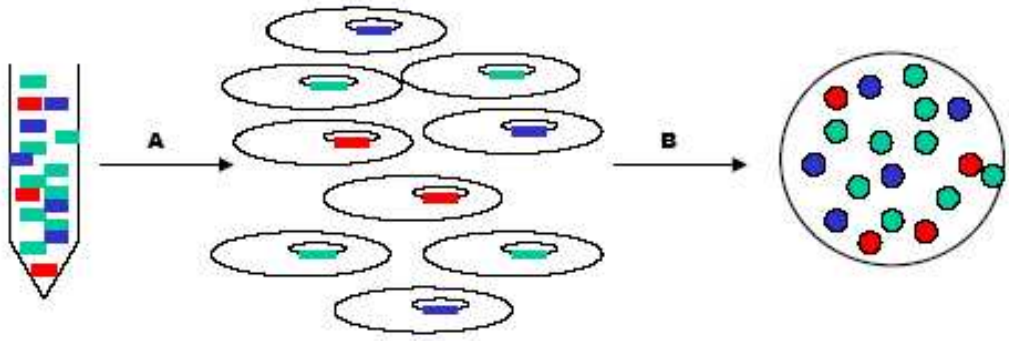


Figure 13: An schematic representation of cloning. Products are amplified from three different templates (e.g: endogenous, contaminants and damaged molecules).A) PCR products are ligated into bacterial plasmids. B.) Bacterias are plated in LB agar plates with X-gal and IPTG only those that have incorporated the insert will be able to metaboliced x-gal.

1.4 AUTHENTICITY CRITERIA

As we have seen (see section 1.3), aDNA research presents extreme technical problems (specially when working with human remains), which can lead to erroneous conclusions.
In the nineties, several aDNA works published in high profile journal were proved to be erroneous (see section 1.1). It was in 1994 when Hand et al (Handt et al., 1994a) published the first list of authenticity criteria that should be followed by all researchers in order to make aDNA a respectable and credible scientific field. Cooper and Poinar updated that list in 2000 (Cooper and Poinar, 2000) and then Paabo et al in 2004 (Paabo et al., 2004):

1.) Physically isolated work areas: the best case scenario is having the aDNA laboratory (where the extractions and the setting up of the PCRs reactions take place) and the main laboratory (where all the PCR amplifications are run) in different buildings in order to avoid contaminations from former amplified products.
2.) Negative control extractions and amplifications to detect sporadic or low-copy number contaminations during each stage. Positive controls should generally be avoided as they provide a contamination risk.
3.) Appropriate molecular behaviour: Due to DNA degradation, PCR amplification strength should be inversely related to product size. In general, if shorter fragments are not easier amplified than longer ones, it is an indication that the source of DNA is likely to be a modern contamination (Paabo et al., 1989, Handt et al., 1994a). Therefore, very long PCR products (> 500 bp) are suspected to be contaminants.
4.) Reproducibility: Multiple PCRs and extractions from the same sample should yield consistent results. First, they are useful to detect contamination of a particular extraction or amplification and second, nucleotide misincorporation leading to consisted changes can be detected only when multiple amplifications are performed. Overlapping primers are also highly recommended in order to detect for instance NUMTs (see section 1.6.1)
5.) Cloning of amplification products and sequencing of multiple clones: In order to detect post mortem damage, contamination, jumping PCR events and to unravel heterogeneities in the PCR products (see section 1.3.7).

6.) Independent replication: separate samples of a specimen are extracted and sequenced in independent laboratories in order to detect intra laboratory contamination.
7.) Biochemical preservation: Biochemical assays of macromolecular preservation serves two purposes: First, they support the claim that a specimen is well enough preserved to allow the preservation of DNA. Second, they may be used as rapid screening techniques to identify specimens that, according to their general state of preservation, may contain DNA. Several techniques have been suggested (see section 1.3), although the most widely used is the analysis of the rate of racemization of Aspartic aminoacid (Poinar et al., 1996).
8.) Quantization of DNA templates by means of quantitative PCR or RT-PCR (see section 1.3.1). If a large number of molecules is present (>1000), there is no need to perform several amplifications since consistent changes are extremely unlikely to occur. However, when the number of starting templates is low, several amplifications are needed in order to exclude the possibility of sporadic contamination.
9.) Associated remains: In studies of human remains where contamination is especially problematic, the presence of similar DNA targets survive in associated faunal material is one additional supporting evidence of DNA preservation because it shows that the environment is favourable to it.

Mainly papers related to ancient humans or Neandertals remains (e.g. (Caramelli et al., 2003, Vernesi et al., 2004) fulfil completely the list since its publication, whereas papers related to extinct animals do not (e.g. (Shapiro et al., 2004, Bunce et al., 2005). In studies related to ancient animals, the phylogenetic criterion is more deterministic because it is easier, in this particular case, to discard modern contamination. Furthermore, this stringent list has been severely criticised by some authors (see (Gilbert et al., 2005, Hebsgaard et al., 2005, Bandelt, 2005)that argue that getting a reliable result does not only mean that the researchers have only fulfil the nine criteria but that the researcher should be more cognitive and self-critical in their results. In other words, instead of checking if the list of criteria has been completely fulfilled, researchers should pay more critical attention to the way the data was obtained and why the results should be considered authentic in the particular context and conditions of the samples analysed.

Some current results (see (Malmstrom et al., 2005) claimed that the list of the well-known nine authenticity (Cooper and Poinar, 2000) criteria should be reviewed and subsequently updated specially when working with ancient human’s remains.

1.5 ORIGIN AND MAINTENEANCE OF THE CURRENT HUMAN GENETIC DIVERSITY

It is clear that humans are all different from each other. The diversity that we observe between individuals is due to both genetic and environmental factors. Current estimates of the human genetic diversity say that if we take two not related individuals at random, approximately 1 of each 1000 nucleotides will be different (Reich et al., 2002). These genetic differences exist as well between human populations. Contrary to the what would be expected based on the phenotypical variability observed between human populations, different studies (Excoffier et al., 1992, Barbujani et al., 1997, Romualdi et al., 2002, Jorde et al., 2000) point out that if the human individuals are hierarquicaly clustered in populations and continents, approximately 80% of the variance of the model is explained because of differences between individuals of the same population. Only 5% to 10% is explained because of differences between populations of the same continent and a 10% to 15% of the variance is explained by genetic differences between continents. This result has been traditionally used as a prove of the lack of sense of clustering the individuals according to races (Kittles and Weiss, 2003)

The current genetic variation we observe in the human populations is the result of the complex interaction between four evolutive forces: mutation, natural selection, genetic drift and migration. Disentangling the effect of each one in shaping the genetic variability of our genome can help us on understanding both past demographic and selective events that occurred in the human species.

1.5.1 MUTATION

Mutation is defined as a structural change in the DNA molecule. It is the ultimate source of genetic variation and thus, allows the evolution to be possible (Crow, 1997). The term mutation covers a broad range of structural events: from substitution of a single base to insertions and deletions of a few bases or even chromosomal rearrangements. The molecular mechanisms that generate that big spectrum go from chemical mechanism (for instance, cytosine deamination), physical mechanisms (for instance, breaking of the double helix for generating an insertion) or enzymatic (for instance, slippage of the DNA polymerase).

These mutations could be generated in whatever cell type that form an organism but only those that appear in the germ line could have the opportunity of passing on to the next generation. Mutations that occur in the rest cell types (they are called somatic mutations), although could be associated to particular phenotypes (for instance, cancer) will not have direct evolutive consequences since they will not pass to the next generation.

1.5.2 NATURAL SELECTION

Mutation generates genetic diversity whereas natural selection shapes directionally that variation. It is defined as a directional change in the frequency of a mutation in the descendents by increasing or decreasing the fitness (ability of a genotype to survive and reproduce) of the individuals that carry such mutation when interacting with the environment. We can distinguish three different types of selection: 1.) **positive selection,** a mutation is positively selected when increases the fitness of the carriers in comparison to the other individuals due to that mutation give them some advantage in the adaptation to the environment; 2.) **negative selection,** a mutation is negative selected when reduces the fitness of the carriers, that means that the carriers has less probability of having descendents that the non carriers; 3.) **balancing selection,** a mutation is balanced selected when the heterozygous have higher probability of having descendents than the homozygotes.
Classic population genetics states that in the case of positive selection a mutation will increase its frequency in the population towards its fixation if the environmental conditions are always favourable and in the case of negative selection the mutation will tend towards its elimination. In the case of balancing selection the frequency of the two variants will reach equilibrium. Therefore, positive and negative selections reduce the genetic variation whereas balancing selection increases it.

1.5.3 GENETIC DRIFT

Genetic drift is defined as the random fluctuation of allele frequencies in a finite population due to stochastic variations in the contribution of each individual to the next generation. The magnitude of the genetic drift is inversely proportional to the number of chromosomes that pass to the next generation. We define this number as the **effective population size**. In

contrast to the two former evolutive forces, genetic drift affects the whole variability in our genome and it is not locus specific. We can distinguish two demographic processes where genetic drift has an important role in shaping the genetic diversity: founder event and bottleneck. **Founder event** is produced when a small set of individuals from a bigger population is separated towards the colonization of new regions. This has two main consequences: first, this small group of individuals could carry allelic frequencies that are not representative of the original population and second, there are big chances that the allelic frequencies could fluctuate until the population reaches a large size. As a consequence, common alleles in the original population could have a small frequency in the new population or even could disappear. On the contrary, rare alleles in the original population could increase their frequency in the new population, even if they have deleterious effects in the individuals that carry them. **Bottleneck** is produced when there is a drastic reduction in the number of individuals in a population. Its consequences in the population are similar to that found in the founder event.

1.5.4 MIGRATION

Migration refers to the movement of the individuals (migrants) from one area to another. Colonization is a special case of migration and is the process of movement into previously unoccupied land.

It is a demographic event able to modulate the genetic diversity of populations due to the movement of individuals from one population to another. The continuous gene flow (outcome when a migrant contributes to the next generation in their new location) between populations tends to homogenise the amount of genetic diversity present in both populations before starting the migration movement, resulting in a global population or meta-population. The fingerprint of the migration will be observed only when the allelic frequencies differ between both populations and the homogenisation process has not yet finished.

The current human populations are the result of complex colonization and migration events as well as the result of demographic expansions and episodes of genetic drift during the colonization of the continents (Barbujani and Goldstein, 2004).

1.6 THE MITOCHONDRIA

The mitochondria is a small cellular organelle placed on the cytoplasm of almost all eukaryotic cells. They are the sites where the oxidative metabolism is placed and thus they are responsible for generating most of the ATP molecules derived from the breaking of organic molecules in the cells. The number of mitochondria varies depending on the cell type: those requiring a lot of energy, such as nerve and muscle cells, contain thousands whereas other cell types may content only a few hundred as the sperm (Jobling, 2004).

In contrast to other cell organelles, mitochondria contain their own DNA, which encodes some of their bio molecules. Lynn Margullis in 1990 (Margulis et al., 1990) proposed the endosymbiotic theory of the mitochondria evolution, currently widely accepted by the scientific community. According to this theory, the mitochondria derive from a bacterium that about 1.5 billon years ago was introduced into a proto-eukariotic cell in a symbiotic relationship. The bacteria gave energy to the cell, whereas the cell provided a safe environment to the bacteria. This symbiotic relationship lead to the lost of autonomy of both organisms; moreover, the safe environment where the bacteria lived made easy loosing duplicated functions that shared with the eucaryotic host cell. Thus, the bacteria lost the genes associated to these functions and nowadays the majority of the proteins that the mitochondria needs are imported from the nucleus of the cell (Wallace et al., 1997). Furthermore, a continuous transfer and insertion of mtDNA sequences from the mitochondria to the nuclear genome has been described (Bensasson et al., 2003). These insertions are known as NUMTS (nuclear mitochondrial sequences) and analysis of the human genome has revealed the existence between 250 and 600 of such insertions of variable length (Pakendorf and Stoneking, 2005). NUMTS are considered as molecular fossils due to the molecular evolution in the nucleus is slower than in the mitochondria (see below) and have been used for a phylogenetic approach to date the origin of modern humans (Mishmar et al., 2004).

The prokaryotic origin of the mitochondria left its fingerprints in some features of the mtDNA: it is a circular genome, it is not protected by histones, it has discrete origins of replication, genes have no introns and it has a different genetic code than the nuclear genome(Jobling, 2004).

1.6.1 HUMAN MITOCHONDRIAL DNA

Anderson and collaborators (Anderson et al., 1981) were the first to obtain the complete sequence and the gene organization of the human mtDNA. This sequence was called *The Cambridge Reference Sequence (CRS)* and it is currently used as the standard reference. The human mtDNA is a circular double-stranded molecule of 16569 bp in lenght that codes for 13 subunits of the oxidative phosporilation system, 2 ribosomal RNAs (12S and 16S rRNAs) and 22 transfer RNAs (tRNA) uses for the protein synthesis in the mitochondria (see Figure 14). Two strands compose the human mtDNA, the H strand (heavy strand) that is rich in guanine and the L strand (light strand) that is rich in cytosine. The majority of the human mtDNA is coding DNA, with the exception of a fragment of 1100 bp that has mainly regulatory functions (Pakendorf and Stoneking, 2005). This region is known by different names: the control region (because its regulatory functions), the D-loop (because the structure that forms when the mtDNA is replicating) and the hyper variable region (because of its high mutation rate). The hyper variable region (HVR) has a mutation rate that is higher than in the rest of the mtDNA and therefore it has been widely used to study population relationship at the intra-specific level (e.g.(Comas et al., 1998, Salas et al., 2002, Plaza et al., 2003). The numeration of the CRS were first established by Anderson and Collaborators and then slightly modified by Andrews et al (Andrews et al., 1999). It arbitrarily starts in the replication origin of the H strand, in the middle of the control region. According to this numeration, the control region is expanded from the position 16024 to 16569 (HVR I) and continues from the position 1 to 579 (HVR II).

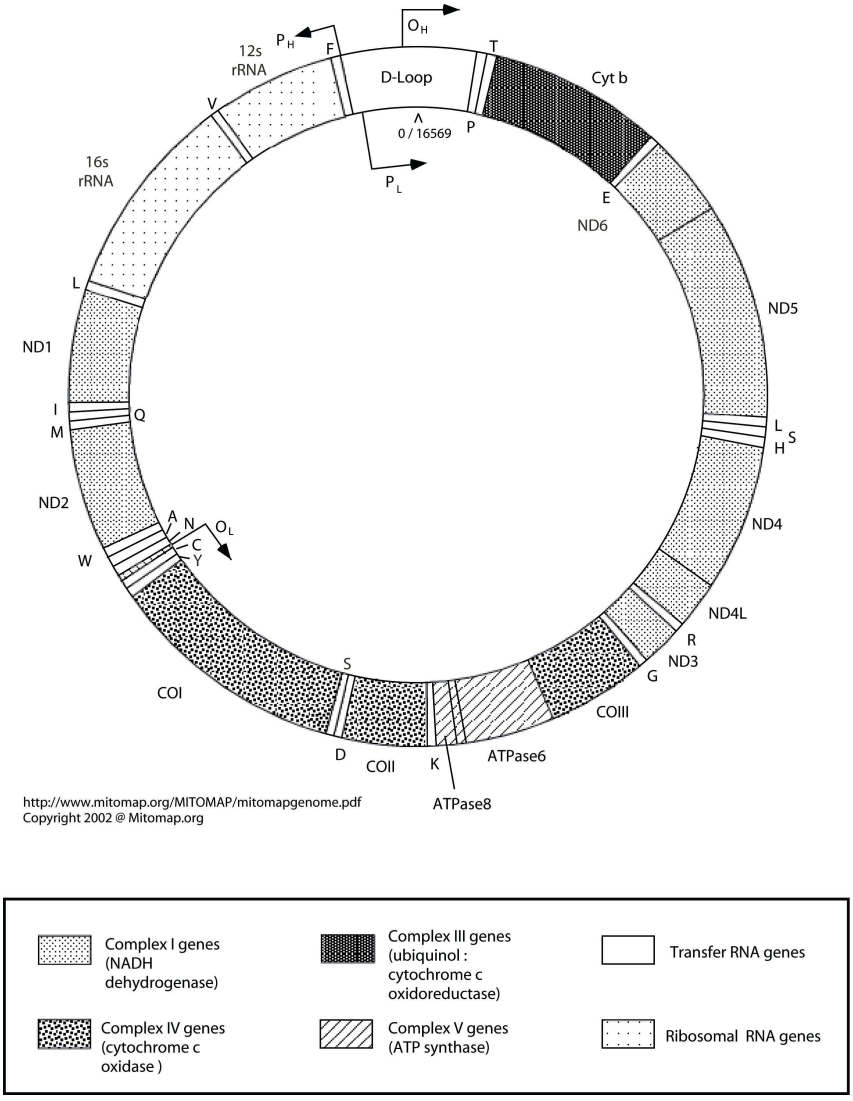


Figure 14: Human Mitochondrial DNA (Adapted (MITOMAP: A Human Mitochondrial Genome Database. http://www.mitomap.org)

Since the first study in human mtDNA variation was published (Brown, 1980), a large number of papers have used this genetic marker to study the genetic variation in humans and making inferences about the human evolution (e.g.(Cann et al., 1987, Vigilant et al., 1991, Ingman et al., 2000) . This widespread use is due to the unique features that the mtDNA shows, making it of particular interest for these kinds of studies:

INTRODUCTION

1.) MATERNAL INHERITANCE

The mtDNA is transmitted from one generation to another via maternal inheritance (e.g. (Wallace et al., 1999). Oocytes contain around 100,000 mitochondria while sperm contain only about 50-75 localized in the midpiece, between the sperm head and tail. Clearly, even if fertilization involve a complete mixing of paternal and maternal mtDNA molecules, the contribution of the father to the zigote's pool of mtDNA would be relatively small due to its small number of mitochondria compared to the oocytes. In addition, it is known that paternal mtDNA is marked for destruction in the oocyte by ubiquitination (Sutovsky et al., 1999, Sutovsky et al., 2000). However, some rare cases have been recently reported where the paternally inherited of mtDNA has been proved (Schwartz and Vissing, 2002); nevertheles, all these cases have been always related to pathology in the maternal mtDNA and thus, it is widely accepted the maternal inheritance of mtDNA.

This maternal mode of inheritance is one of the great advantages of mtDNA in human evolution studies due to:

1.) As it is an haploid genome, does not recombine and therefore enables researcher to trace back through time its genealogy via maternal lineages without the confusing effects of bi-parental inheritance and recombination inherent to nuclear DNA (Pakendorf and Stoneking, 2005).
2.) All the individuals from the same maternal lineage have the same sequence. Thus, mutations that differentiate two mitochondrial lineages are due to mutations that have been accumulated over time since these two lineages splited.

2.) HIGH COPY NUMBER

It is estimated that inside each cell there are between 1,000 and 10,000 mitochondria genomes (Robin and Wong, 1988). This large number of copies contrasts with the two copies of any given nuclear DNA locus. This propriety, together with the extra nuclear location of the mtDNA, makes easier to extract mtDNA than nuclear DNA for genetic analysis. Therefore, mtDNA is the main choice when analysing aDNA and particular forensic samples (e.g. hairs without root) (Pakendorf and Stoneking, 2005).

Nevertheless, despite that all the mtDNA copies came from the mtDNA copies present in the oocyte, there is no reason for what the multiple copies of mtDNA within an individual should

be all identical. In other words, some mtDNA molecules can suffer a mutation event when they are in a particular cell. **Heteroplasmy** is defined as the existence of different mtDNA sequences within an individual (e.g. (Comas et al., 1995, Bendall and Sykes, 1995). Current estimates indicate that about 14% of the population has a second mtDNA sequence at a frequency of at least 1% (Tully et al., 2000). In general, the overall genetic homogeneity of mtDNA within individuals is due to a substantial bottleneck that occurs during the oogenesis (Poulton and Marchington, 2002).

3.) HIGH MUTATION RATE

The mutation rate of mtDNA is 6 to 17 times higher than that of nuclear genes (Wallace et al., 1987). Current estimates show a rate of 0.017×10^{-6} substitutions per site per year for the whole mitochondria genome excluding the control region (Ingman et al., 2000). However, the mutation rate in the control region is even higher because it is not a codifying region (Tamura and Nei, 1993). Recent estimates on pedigrees suggest the rate at approximately 0.0043 per generation (Sigurgardottir et al., 2000).

The high mutation rate of the mtDNA compared to the nuclear DNA has been explained by the presence of three main factors: 1.) mtDNA has not proteins like histones that protect the nuclear DNA. 2.) As the oxidative metabolism is placed on the mitochondria, mtDNA is surrounded by free radicals and therefore it is more exposed to oxidative damage. 3.) The mitochondria has a DNA repair system less efficient than the nucleus so that the possible changes that are produce in the mtDNA sequence are less prone to be repaired than those changes in the nucleus (Brown et al., 1979).

It is important to take into account that not all sites mutate at the same rate. This happens both in the coding region and in the control region. A *recurrent mutation* is defined as a mutation that due to the high mutation rate of a particular site arose independently more than one time in different mtDNA lineages.

4.) LACK OF RECOMBINATION

Although until year 1999 it was considered like a fact that the mtDNA did not recombine, in that year were published several papers (e.g. (Awadalla et al., 1999, Hagelberg et al., 1999) claiming evidence for recombination in human mtDNA. The main point of these papers was that the human mtDNA has too much homoplasy (homoplasy means the generation of the same state by independent means). However, posterior studies (e.g. (Arctander, 1999, Jorde and Bamshad, 2000, Kumar et al., 2000) showed that the statistical tests, phylogenetic analyses and experimental design used in the former studies were wrong and then the possible recombination in human mtDNA was called into question.

Since the presence of paternal mtDNA in the mitochondria is a very rare phenomenon (see before), recombination in the mitochondria should not be a major issue. In the absence of heteroplasmic DNA molecules in the mitochondria of the oocyte, any recombination between two mtDNA copies would result in mtDNAs that are exactly the same.

1.7 HUMAN POPULATION HISTORY

1.7.1 HUMAN AS A PRIMATE SPECIES

Humans are primates. We are closely related to chimpanzees from which we have diverged from 5 to 7 million years ago (MYA) (see Figure 15); this close relationship is reflected in the amount of shared genome between both species. Humans share 98.76% of its genome with the chimpanzee species(Paabo, 2003).

The 5 to 7 million years of hominid evolution since humans and chimpanzees shared a common ancestor lead to develop specific human features in the human linage. The brain size and the bipedal locomotion are the most important developments (Carroll, 2003). These features are listed in Table 2

Table 2: Selected traits that distinguish humans from other apes (adapted from Carroll et al 2003)

Selected traits that distinguish humans from other apes
Body shape and torax
Cranial properities (brain case and face)
Relative brain size
Relative limb lenght
Long ontogeny and lifespan
Small canine teeth
Skull balanced upright on vertebral column
Reduced hair cover
Elongated thumb and shortened fingers
Dimensions of the pelvis
Presence of a chin
S-shaped spine
Language
Advance tool making
Brain topology

1.7.2 THE FOSSIL RECORD OF HOMINIDS

Nowadays, there are several described ancient species belonging to hominids (see Figure 15). Regarding to fossil record, our ancestors appear to have existed at low population densities and were seldom fossilized. Teeth, cranium and mandible are the most frequent human fossil

remains. Other bones are rarely preserved and soft body tissues hardy ever (except in mummies) (Jobling, 2004).

The finding of new hominid species is always a matter of controversy; opinions about the number of genera and species and their relationships in the homo linage differ considerably between experts in the field, and changes over time.

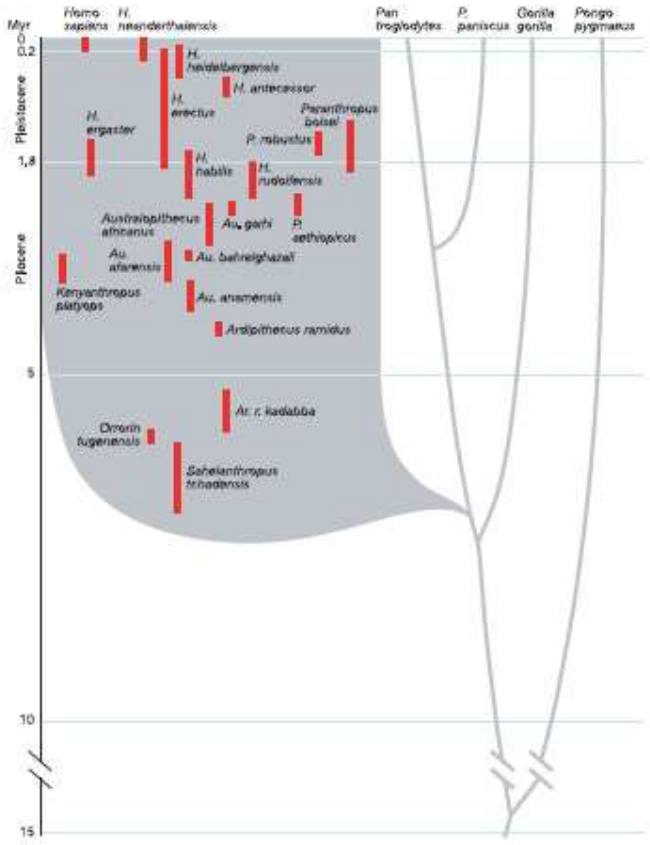


Figure 15: Timescale and phylogeny of hominids. Ape relationships are shown in grey for the chimpanzee (pan troglodytes), bonobo (P.paniscus), gorilla and orangutan (Pongo pygmaeus). The approximate times of divergence are derived from molecular data. The phylogenetic relationships between hominids (shaded) are uncertain. The solid red bars denote the time span of the fossil species and/or the uncertainly of fossil ages. (Adapted from (Carroll, 2003)

The oldest hominid fossil remains have been found in Africa, thus suggesting an African origin of the hominid species. The oldest fossil consists of a cranium and a jaw fragment from Chad (Africa) and it is called *Sahelanthropus tchadensis* (Brunet et al., 2002). Indirect dating of the fossil suggests a date of 6-7 MYA. Several fossil remains from two genera (*Orronin* and *Ardipithecus*) dating to 5-6 MYA have been found in Kenya and East Africa (Senut, 2001). A large number of fossils found in the African continent date after 4.2 MYA and have

been attibuted to the genus *Australopithecus.* The oldest fossil associated to the genus *Homo* is *Homo habilis* dated from 1.8-1.9 MYA, although this genera has been also associated to the genus *Australophitecus*. However, it is not until 1.8 MYA when we found hominids of the species *Homo erectus* out of Africa. Fossils from this species have been found in Europe, Middle East, Asia and Africa (but not in America) with dates between 1.9 and 0.5 MYA (Anton, 2003); being the oldest fossil from this genera found in Africa and date from 1.8-1.9 MYA (Jobling, 2004).

Several later *Homo* species are known, including *Homo heidelbergensis, Homo antecessor, Homo neanderthalensis* and *Homo sapiens* (see Figure 15). *H*omo *heidelbergensis* has been found in Africa, and Europe. *Homo antecessor* has been found only in Atapuerca site, Spain (Bermudez de Castro et al., 1997) and it is dated to 780 KYA (kilo years ago). Many scientists include *Homo antecessor* within *Homo heidelbergensis* species while others consider them as a separate species.

Homo neandertalensis are found only in Europe, Middle East and Western Asia dating between 250KYA and 28KYA. They were relatively robust with large brains and short extremities. Well-known examples include the first Neanderthal fossil found in the Neander valley in Germany and "the old man of La Chapelle-aux-Saints" found in France. Neandertals are thought to be descendents of *Homo heidelbergensis* and their relationships with modern humans are still a theme of debate (see (Hublin and Paabo, 2006).

Homo sapiens appear about 1.6 KYA in Africa and spread all over the word. Nowadays is the unique alive species of the genus Homo that inhabits our planet.

Since the classification of the fossils inside the genus Homo is not universally accepted for all the experts in the field, all species between *Homo erectus* and *Homo sapiens* are sometimes collectively classified as "archaic *sapiens*".

The discovery in 2004 of a fossil with sufficient distinctive features to be assigned to a new hominid species in the island of Flores, Indonesia (Brown et al., 2004, Morwood et al., 2004) put into alert to the scientific community. This new fossil was called *H.floresensis* and its most distinctive features were its diminutive body size (about a metre in height) and the smallest brain size of all the hominids described until now ($380cm^3$). It was dated on 18 KYA and Morwood et al conclude that it was a dwarfed descendant from an ancestral *Homo erectus* population that survived until at least 18KYA, cohabiting in time with *Homo sapiens* (see Figure 16). In isolation, these populations were released from predation pressure and constrained by restricted resources (for instance food). The effect of these forces could lead to

this dwarfism. Island dwarfism is well known among mammals (e.g. (Lalueza-Fox et al., 2005a). Marta Miranzon Lahr and Robert Foley claimed (Mirazon Lahr and Foley, 2004), "*Homo floresciencis* is a challenge-it is the most extreme hominid ever discovered. An archaic hominid at that date changes our understanding of late human evolutionary geography, biology and culture". However, recently, some authors (Jacob et al., 2006). argued that the fossil found in the island of Flores is, in fact, an early pygmy *Homo sapiens* population. The studied individual would show, in addition, some developmental abnormalities, including microcephaly. Their argument is supported by additional mandible and post cranial remains from the same site that share small body size but not microcephaly.

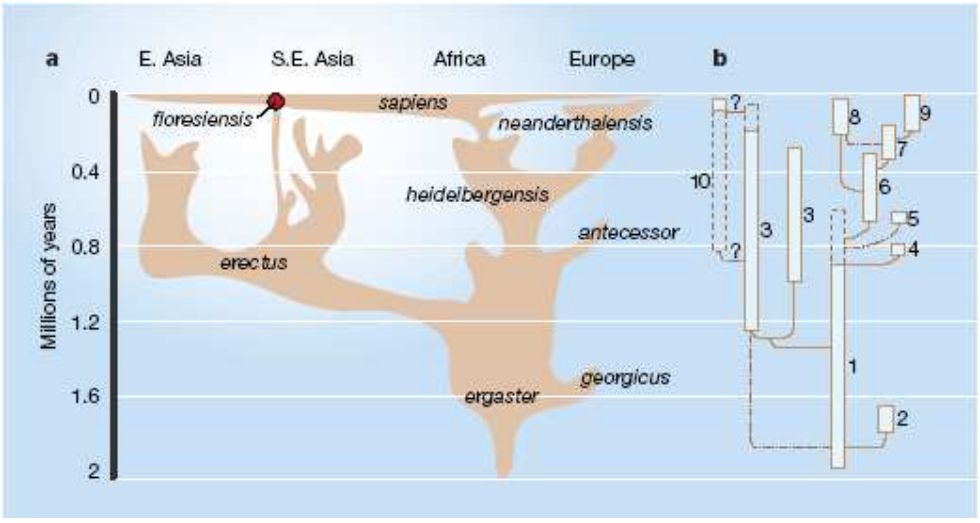


Figure 16: *Homo floresiensis* in the context of the evolution and dispersal of the genus Homo. . 1, *H. ergaster*/African *erectus*; 2, *georgicus*; 3, Javanese and Chinese *erectus*; 4, *antecessor*; 5, *cepranensis*; 6, *heidelbergensis*; 7, *helmei*; 8, *neanderthalensis*; 9, *sapiens*; 10, *floresiensis.* Solid lines show probable evolutionary relationships; dashed lines, possible alternatives. (adapted by (Mirazon Lahr and Foley, 2004)

1.7.3 THE ORIGINS OF MODERN HUMANS

The origin of anatomically modern humans (AMH) has been a matter of controversy between scientists during more than half a century. Historically, two extreme models, the **Multiregional model** and the **Out of Africa model,** have been proposed to explain the transition between *H.erectus* to *H.sapiens*. Although both models agree in the departure of *H.erectus* from Africa to Europe and Asia between 1.8 to 0.8 KYA, they disagree in the way of transition from one species to the other.

1.7.3.1 MULTIREGIONAL MODEL

The multiregional model proposes that there is not a unique geographic origin of anatomically modern humans but the transition from *H.erectus* to *H.sapiens* took place in a number of places all over the world (see Figure 17). Due that the isolation of archaic sapiens populations would have produced phenomena of speciation, it proposes the presence of long term local regional continuity with populations connected one each other by gene flow and therefore high population sizes which is unlikely. According to this model, current Europeans would be, in part, descendents from the Neandertals.

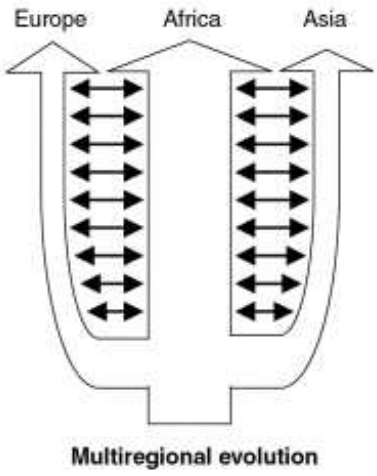


Figure 17: Multiregional Model. (Adapted from (Excoffier, 2002)

1.7.3.2 OUT OF AFRICA MODEL

The "Out of Africa" model proposes that the transition from *H.erectus* to *H.sapiens* took place recently (about 200 KYA) in the African continent (see Figure 18) and then around 100 KYA the anatomically modern humans left Africa to colonise the rest of the world. There are evidences of human population in the south of Asia and Australia around 60 KYA and of Europe around 40 KYA (Cavalli-Sforza and Feldman, 2003). According to this model, the anatomically modern humans evolved from a small African population that was able to colonize the whole world, replacing the hominid species already present in other continents (e.g. the *Homo neanderthalensis*). Reduced or inexistent inbreeding is assumed in this model.
A modified version of the "Out of Africa" model is the "Weak garden of Eden hypothesis". This model suggests that after the anatomically modern humans left Africa, the size of the population was reduced drastically due to a bottleneck phenomena and then the remained population expanded to the rest of the continents only 50 KYA ago (Harpending and Rogers, 2000).

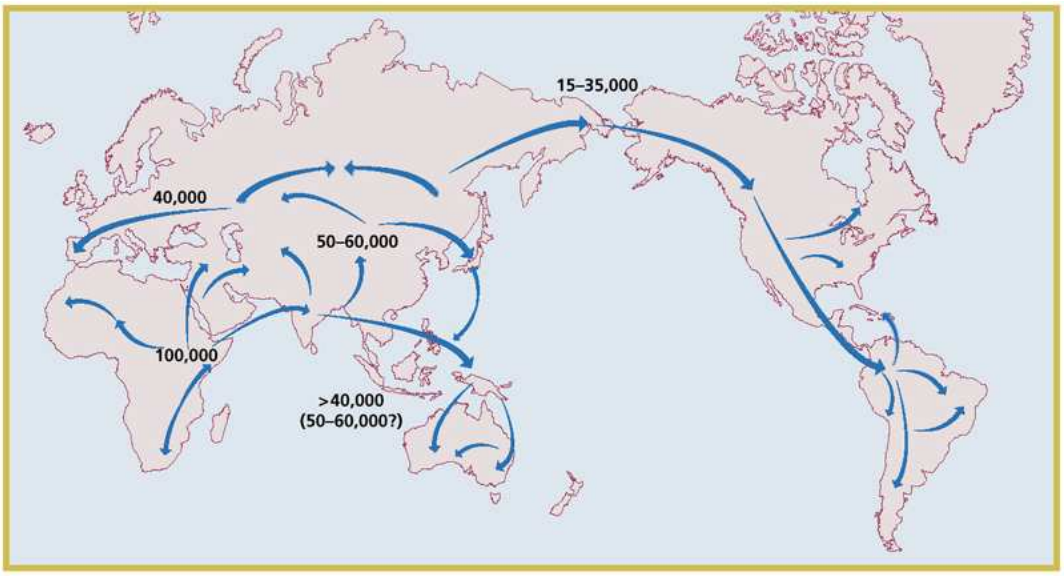


Figure 18: The migration of modern Homo sapiens (modified from(Cavalli-Sforza and Feldman, 2003)

Obviously, intermediate models are also possible (Templeton, 2002, Excoffier, 2002). Although the most direct prove of our past it is the fossil record, the information that we have got until now from the fossil record is scarce and very contradictory. It is here, where the genetic data can help us in disentangling the controversy that has surrounded the origin of the modern humans. Patterns of genetic diversity should contain information about the demographic history of *H.sapiens* due to the current genetic diversity, necessarily has to derive from an ancestral state (coalescence).

1.7.4 THE GENETIC FINGERPRINT IN THE HISTORY OF HUMAN POPULATIONS

However, population genetics is not a panacea. Interpreting the current genetic variability of the human populations is not an easily affordable task. As we have seen in chapter 1.5, there are different evolutionary forces that can shape the genetic diversity of a population. The same patterns of genetic variation can be obtained by the combination of different evolutionary forces acting at different times and strengths. Such number of different variables could potentially introduce a large degree of uncertainty when making inferences about the history of populations, and so population genetics has to be understood as another tool for studying the human history in addition to other fields such as the archaeology, anthropology or linguistics.

According to the multiregional model (see chapter 1.7.3.1), it is expected to find high levels of genetic diversity in human populations due to more than two million years of separate evolution between populations across several continents. Furthermore, there is no reason for one geographic region to show more genetic diversity than another. In contrast, according to the “Out of Africa model” (see chapter 1.7.3.2), the presence of a bottleneck in the original population that spread out of Africa and the short evolutionary time since this event (only 100 KYA) would have decreased the genetic diversity in the human populations out of the African continent; Africa should show greater genetic diversity that any other geographic region.

Recent human population studies using different genetics markers as mtDNA, Y-chromosome, Alu insertions, autosomic minisatellites or autosomic haplotypes point out in the same direction, that is, a recent African origin of our species. No genetic evidences supporting the multiregional model have been found so far. Some basic genetic findings that support the recent African origin of our species are: I) low human genetic diversity of human

species in comparison to other primate species, for instance chimpanzee diversity (reflecting a recent origin from a small population in contrast to ancient lineages and constant population size)(Gagneux et al., 1999) II) global human genetic diversity is a fraction of that found in Africa (reflecting that the origin of modern humans was in a population from Africa) (e.g. (Cann et al., 1987, Vigilant et al., 1991, Ingman et al., 2000) and III) Neandertals, the only extinct homo species from which genetic data is available, show their own evolutionary history with their mtDNA variability outside the mtDNA variability of modern humans (e.g. (Krings et al., 1997).

1.7.4.1 EVIDENCE FOR THE MITOCHONDRIAL GENOME: "The mitochondria Eve"

The mtDNA was the first genetic marker that supported the "Out of Africa" model. Brown et al (Brown, 1980) were the first in studying the human mtDNA variation by means of RFLPs (restriction fragment length polymorphism). They concluded that the diversity that actually exists in the human species started to accumulate 180,000 year ago.

The team of Allan Wilson in 1987 published an article in nature (Cann et al., 1987) where they postulated that all the genetic diversity of the mtDNA present in the current human population coalesced from a woman who lived 200,000 years ago, probably in Africa. This woman was called "**The mitochondrial Eve**". The study consisted in analysing the mtDNA of 147 individuals from five different geographic regions (Africa, Asia, Australia, Caucasus and New Guinea) by means of RFLPs of high resolution. A tree built by a parsimony method with the genetic diversity present in the mtDNA showed two branches. The deepest branch separates African mtDNAs while the other leads to mtDNAs from populations out of Africa. Based on these results, they suggested that Africa is a likely source of the human mitochondrial gene pool and that our common ancestor existed between 140,000 and 290,000 years ago.

This work was severely criticized by the scientific community due to technical problems when selecting the samples and to methodological problems (the method used to generate the tree was not the most parsimonious and the method used to generate the root was inaccurate (e.g. (Wills, 1992, Templeton, 2002). Despite these criticisms, subsequent studies of the variability of mtDNA in modern populations have supported the results of Cann et al. A subsequent study published by Vigilant and collaborators (Vigilant et al., 1991) was based on the mtDNA control region. They sequenced 189 individuals all over the world from which

121 individuals came from Africa. The phylogenetic tree (see Figure 19) of those sequences supported as well the “Out of Africa model”.

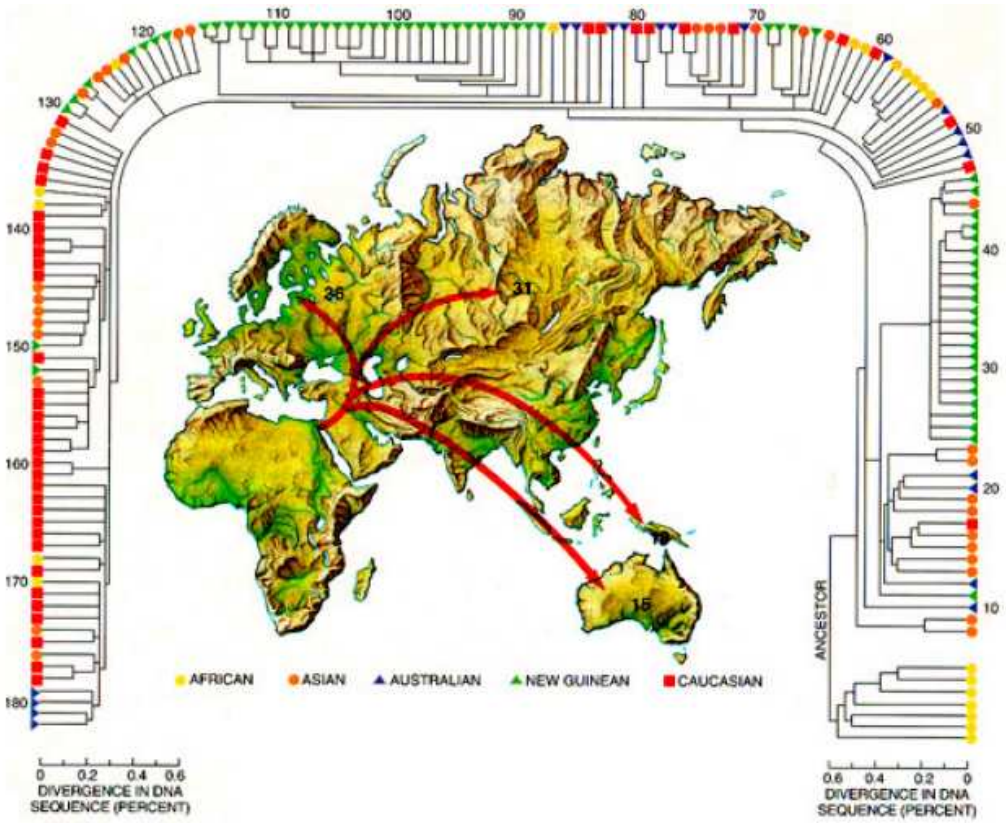


Figure 19: Genealogical tree of 182 mtDNA types (Adapted from (Vigilant et al., 1991)

A more recent study (Ingman et al., 2000) developed a mtDNA phylogeny based on the complete mtDNA sequences of 53 individuals of diverse geographical origins. All sequences were different. It was built a robust phylogeny without taking into account the variable positions in the control region. The results again showed a complete separation of African and non-African lineages with deep branches within African lineages and a starlike structure within non-African lineages.

1.7.4.1.1 DISTRIBUTION OF THE MtDNA LINAGES IN THE HUMAN POPULATIONS

However, the current genetic variability found in the mtDNA not only have been useful to study the origin of modern humans but also to get a better understanding of human migrations all over the world (Maca-Meyer et al., 2001). According to the Out of Africa model, the spread of *H. sapiens* out of Africa was not instantaneously and took a long period of generations (Cavalli-Sforza and Feldman, 2003); this allowed the possibility that new mutations appeared in the mtDNA genome of different populations, differentiating them. The posterior isolation by distance of the human populations would have geographically confined these genetic variants to the neighbour populations (but not to far away populations). In fact, genetic studies have shown that mtDNA substitutions that have accumulated along maternal lineages have diverged as human populations expanded through different geographic origins (Forster, 2004). An **Haplogroup** is defined as a group of sequences that share the same set of mutations; it can be defined by the presence of a particular SNP (Single Nucleotide Polymorphism) in the coding region of the mtDNA, that correlates to different haplotype sequences in the hypervariable control region (HVR) of the mitochondrial genome. Due to the isolation by distance, haplogroups tend to show some regional specificity.

Therefore the different geographic regions have their own characteristic mitochondrial haplogroups usually at the continental level (see Figure 20). Haplogroups L0, L1 and L2 (all of three subdivided in many subhaplogroups) belongs to the macrohaplogroup L and are specific from sub Saharan African populations (Salas et al., 2002). L3, M and N haplogroups, also belongs to the macrohaplogroup L. L3 is restricted to the African continent (Salas et al., 2002) whereas M and N haplogroups were originated in Eastern Africa but were posteriory dispersed into Europe and Asia when *H.sapiens* colonised these regions (Quintana-Murci et al., 1999, Maca-Meyer et al., 2001). Europe is characterized by a genetic homogeneity in the mtDNA variability (Simoni et al., 2000, Richards et al., 2002). Nevertheless, there are two main exceptions to this rule: (i) the Saami population, that is clearly differentiated from the rest of the European populations (probably due to genetic drift and geographic isolation (Simoni et al., 2000) and, (ii) the area around the Mediterranean sea, that shows a West to East clinal patterns (but not north to south) (Comas et al., 1997, Simoni et al., 2000). Haplogroups H, I, J, N1b, T, U, V and W are similarly frequent in European populations. All of them derived from the macrohaplogroup N (Torroni et al., 2006).

The genetic landscape of the mtDNA in Asia is more complex. Typical Asian haplogroups are derived from the macrohaplogroup N as C, D, E, G and Z or derived from the macrohaplogroup M as A, B, F and Y (Kivisild et al., 1999) but their frequencies in the different Asian regions strongly differs from one region to the other.

Haplogrups A, B, C and D are characteristic of Native Americans and together with the Western Asia X haplogroup comprise the 100% of the mtDNA linages (Forster, 2004). The haplogrups are represented in the following Figure 20:

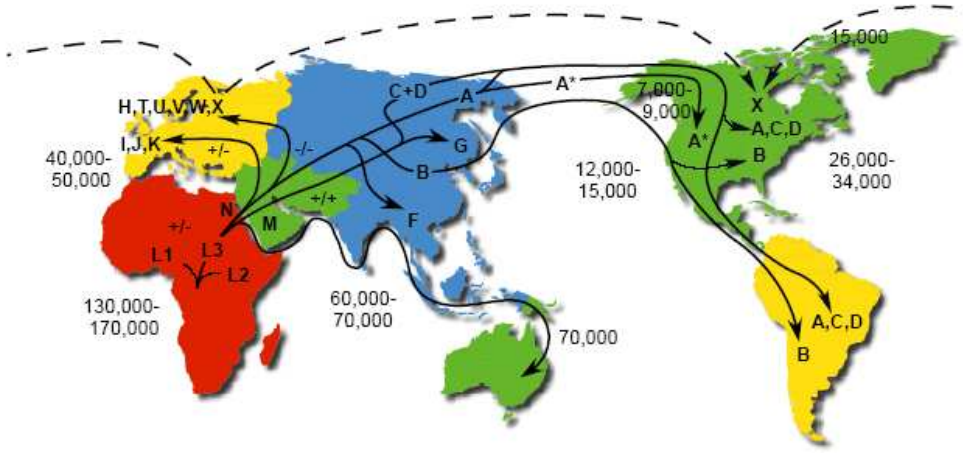


Figure 20: Human mtDNA migrations (Adapted from (MITOMAP: A Human Mitochondrial Genome Database. http://www.mitomap.org)

1.7.4.2 EVIDENCE FOR Y CHROMOSOME

The Y chromosome is, together with the X chromosome, a sexual chromosome. It is male specific (contain the SRY gene that codifies for the formation of testis), haploid, passes from fathers to sons (patrilineal inheritance) and only a small fraction recombines with its homologous part in the X chromosome. These properties make the Y chromosome an attractive marker for population genetics studies and, like in the case of the mtDNA, it has been widely used to study the history of modern human populations (e.g. (Perez-Lezaun et al., 1999, Bosch et al., 2001).

The Y-chromosomal phylogeny is well established (Underhill et al., 2000, Thomson et al., 2000) and its structure shows complete separation of African and non-African lineages with the root in Africa and supports as well the “Out of Africa” model.

1.7.4.3 EVIDENCE FROM AUTOSOMIC LOCUS

Together with the mtDNA and Y-chromosome genetic data, the analysis of the genetic variability in some autosomic locus reinforces the recent African origin of human populations. The stochastic nature of the neutral evolutionary forces shaping the neutral variability of our genome (see section 1.5) implies that conclusions about demographic processes can only be obtained by analysing more than few loci.

The (neutral) genetic variability present in different types of autosomic loci has been widely studied in human populations, including alu instertions (Stoneking et al., 1997), minisatellites (Armour et al., 1996), SNPs (HAPMAP, 2005), classical polymorphisms (Cavalli-Sforza, 1994), haplotype diversity (Tishkoff et al., 1996) and linkage disequilibrium patterns (Tishkoff et al., 1996). All these results tend to find higher degrees of genetic variation within the African populations and much lower in populations out of the African continent; furthermore, studies performed with a large number of widely spread populations (e.g.Tishkoff et al., 1996, Ramachandran et al., 2005) show that the genetic diversity of the different markers tend to be geographically distributed following clinal patterns from the African continent to out of Africa. Clinal patterns of genetic diversity can be obtained when successive population expansions in the space are performed from a single starting point (Barbujani and Goldstein, 2004). Thus, all these results support as well the “Out of Africa” model.

1.7.4.4 EVIDENCE FROM ANCIENT DNA

The possible analysis of the DNA from all the Homo species that have inhabited our planet by means of aDNA techniques should be the best way to disentangle the origins of modern humans. It would provide direct results, avoiding the uncertainty introduced by the evolutionary processes (i.e genetic drift) when using modern DNA data. If it were possible, the results would tell us directly whether there was a regional continuity of our species or a replacement of archaic sapiens by *H.sapiens* from Africa.

Unfortunately, as we have already seen in chapter 1.3, retrieving DNA from fossil remains it is not an easy task. Moreover, it is still more difficult due to the big problem of modern human contamination when working with "Homo" species. However, the improvement of aDNA techniques allowed investigators to get mtDNA sequences from the control region of thirteen *Homo Neanderthalensis* specimens spread all over Europe (Krings et al., 1997, Krings et al., 1999, Ovchinnikov et al., 2000, Krings et al., 2000, Schmitz et al., 2002, Serre et al., 2004, Lalueza-Fox et al., 2005b, Lalueza-Fox et al., 2006, Caramelli et al., 2006, Beauval et al., 2005, Orlando et al., 2006). When the genetic diversity of these Neandertal sequences was compared to that found in current human populations, it was shown that the mtDNA sequences retrieved from the thirteen Neandertal fossils have never been found in the pool of modern mitochondrial DNA sequences. However, concluding that there was no inbreeding based on these results could be tricky. It could be that these genetic variants have been lost in the actual human populations by genetic drift; thus, it would be better comparing the genetics affinities of the earliest modern humans of Europe and the Neandertals. However, such field is still a source of controversy (Hublin and Paabo, 2006). Caramelli and collaborators (Caramelli et al., 2003) extract and sequence mtDNA from two early anatomically modern *Homo sapiens sapiens* (Cro-Magnon) that had been dated at about 23 and 25 thousand years ago. Therefore they had a direct genetic comparison between individuals who lived at (ideally) the same time. Nevertheless, the obtained results have been severely criticised due to (again) the problem of modern human contamination (Serre et al., 2004, Gilbert et al., 2005). Following the most stringent criteria of validation aDNA sequences so far (Cooper and Poinar, 2000), they show that mtDNAs sequences of these individuals fall well within the range of variation of today's humans, but differed sharply from the available (at that moment only four specimens of Neandertal had been sequenced) sequences of contemporary Neandertals. This genetic discontinuity supported again the "Out of Africa model".

Serre et al (Serre et al., 2004) trying to avoid de problem of modern human contamination tried to amplify Neandertal sequences in well-preserved fossils of early modern humans. They did not get any Neandertal sequence. From this study they concluded that it is unlikely that Neandertals lead a large genetic contribution in the early modern humans, but they do not rule out the possibility of a small contribution. It should be recalled that male contribution cannot be recorded in the mtDNA sequences and that we are only analysing one locus from all the genome. Cooper et al (Cooper et al., 2004) argued that future methodological developments, such us the retrieval of nuclear sequences from Neandertal fossils, are needed in order to

arrive a conclusive results about the possibility of inbreeding between Neandertals and anatomically modern humans. In July of this year (2006) the Max Plank institute for Evolutionary Anthropology and private company 454 Life Science announced that they are going to sequence the Neandertal genome over the next two years. It is thought that a comparison of the Neandertal genome and the human genome will help us in understanding the evolutionary history of this extinct homo species.

1.8 PEOPLING OF EUROPE

The peopling of Europe is a complex phenomenon full of migratory events. We can study this phenomenon through the fossil evidence, through the archaeological evidence and through the genetic evidence (current and ancient).

The temperatures that characterise Eurasia during the last 250 KY were very instable (see Figure 21), mostly colder and drier with lower sea level. Such environmental changes had a deep influence in the migration of human populations in the past, by changing the ecosystem of the Eurasian continent and opening (and closing) new migratory paths.

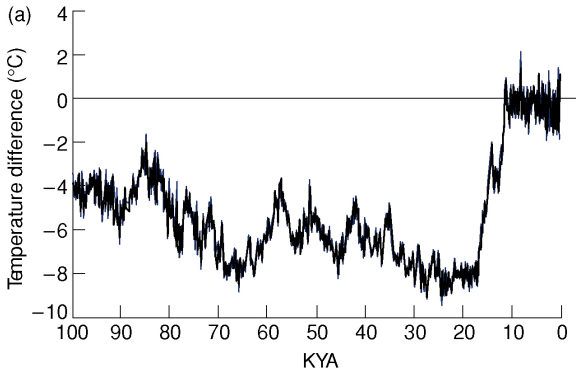


Figure 21: Temperature variation over the last 100 KY (Adapted from (Jobling, 2004).

1.8.1 UPPER PALEOLITHIC: THE NEANDERTALS

As we have already seen in chapter 1.7.2, the fossil record tells us that the first presence of the *Homo* genus in Europe was around 1 MYA after the early expansion from Africa of *Homo erectus*. Later descendents of this first diaspora of *Homo erectus* in Europe lead to the *Homo neanderthalensis*. Based on the fossil record Neandertals occupied Europe and Western Asia between 250 KYA and 30 KYA when they became extinct (Mellars, 2004).The Iberian peninsula is the place where the most recent fossils belonging to *Homo neanderthalensis* have been found so far (Hublin, 1995). This *Homo* species evolved physical adaptations to the cold climate present in Europe during this geological period (see previously) including large brain, short-but-robust bodies (weighting around 80kg) and large noses, those body proportions can be found nowadays in populations that inhabit in cold climates as artic populations. Other morphological features of this archaic form of *Homo* genus are receding foreheads and

supraorbital buttresses. Their faces were long and very projecting in the mid-part, with no chins. This morphometric features show that Neandertals were in fact different to modern humans (Hublin and Paabo, 2006). The first Neandertal skeleton was found in 1856 in the Neander Valley in Germany. Nowadays around 400 fossils of Neandertals are known, and several almost complete skeletons have been discovered, including immature individuals (Hublin and Paabo, 2006).

Traditionally they are related to **Mousterian culture**. This culture is characterized by flakes described as side-scrapes and points (Jobling, 2004).

Both archaeological and genetic evidences suggest that the first anatomically modern humans entered Europe from the Middle East 35 KYA to 40 KYA (Barbujani and Goldstein, 2004). Therefore, Neandertals and the incoming anatomically modern humans coexisted in Europe during approximately 10,000 years. The incoming anatomically modern humans were hunter gathers and they were related to a more sophisticated culture called **Aurignacian culture.** This was the first culture that has been associated to *Homo sapiens sapiens* and spread through central and Western Europe between 40 KYA and 25 KYA. It is characterized by the use of blade flint technology and bone tools (Jobling, 2004).

The cohabitation of Neandertals and modern humans in Europe during more than 10,000 years has stimulated considerable debate regarding hypothetical admixture between both *Homo* species. Paleoanthropological evidence indicates that Neandertals were mostly replaced, with very rare, if any, interbreeding. There is no direct evidence that the two groups lived in close contact (Hublin and Paabo, 2006). As it has been argued above (see chapter 1.7.4.4), molecular data from Neanderthal remains and from the early anatomically modern humans that inhabited Europe at the same time suggests that there were not genetic flow between these two homo species or, if it took place, it was irrelevant (Krings et al., 1997, Krings et al., 1999, Ovchinnikov et al., 2000, Krings et al., 2000, Schmitz et al., 2002, Caramelli et al., 2003, Serre et al., 2004, Lalueza-Fox et al., 2005b, Lalueza-Fox et al., 2006, Caramelli et al., 2006, Beauval et al., 2005, Orlando et al., 2006).; that is, anatomically modern humans would have replaced the Neandertals populations without presumable inbreeding with them. Furthermore, Krings et al (Krings et al., 1997) calculated the time when the most recent ancestral sequence common to the Neandertals and current modern human mtDNA sequences existed. They obtained a date of 550 KYA to 690 KYA for the divergence of the Neandertal mtDNA and contemporary human mtDNA whereas coalescence times for contemporary humans are between 120 KYA and 150 KYA from the same mitochondrial data. Therefore, the divergence between the Neandertal linage and the *Homo sapiens* linage

was between the first diaspora of *Homo erectus* outside Africa 2 MYA and the migration of modern humans out of Africa around 100 KYA. As we have already seen, the retrieval and the analysis of the thirteen Neandertal mtDNA sequences available till now, confirm that Neandertals differ sharply from modern humans. Also, none of the many thousands of humans living today have been found to carry such divergent mtDNA sequences. Given these facts, the contribution of the Neandertals to the gene pool of modern humans can only be minor, if any.

When considering both hypervariable mitochondrial regions, their average differences from European and Africans are 35.3 ±2.1 and 33.9 ±2.1 substitution respectively (Krings et al., 2000). In addition, a low genetic diversity has been observed among Neandertals, similar to that observed among modern humans (Krings et al., 2000). Lalueza-Fox et al (Lalueza-Fox et al., 2006) calculated the most recent common ancestor in the Neandertal linage with the seven mtDNA sequences that were longer than 300 bp from Neandertal specimens and they obtained an estimation of 250 KYA to 65 KYA by coalescent methods.

However, no one knows for sure why Neandertals become extinct. Presumable many factors concurred to cause that extinction, but two may have been very important, namely climatic instability (Neandertals were adapted for a cold environment, and it could be possible that they were unable to adapt to the warming climate at the end of the most recent ice age) and competition with the incoming anatomically modern humans that presumably carried a more sophisticated culture. At the end, modern humans went on towards a complex social organization whereas Neandertals become extinct.

The climatic stabilisation in Europe around 30 KYA lead local *Homo sapiens* populations grew strongly, as can be evidenced by the presence of starlike mtDNA clusters (Forster, 2004). One of the consequences of this growth was the settlement of the Americas trough Beringia (now submerged). According to mtDNA founder analysis (Forster et al., 1996),a small group of Asian from northern Siberia moved into Americas, whose characteristic mtDNA types A, B, C and D are found today in tribes across the Americas.

However, after anatomically modern humans settled in Europe, the average temperatures strongly decreased, leading to the last glacial maximum around ~18,000 years before present (BP)(see Figure 21). In that period, many species were extinct, whereas others and probably including humans, migrated to southern areas of milder climate, the glacial refuges, from which they re expanded when the climate improved (Barbujani and Goldstein, 2004). Based on the paleobiological evidence, three refugia have been recognized in Europe: Iberia peninsula, Italy and the south of the Balkanian range (Willis and Whittaker, 2000). The end of

the last maximum glacial in Europe was followed by a relative climatic stability in which the human populations prospered and grew in size, re-settling of the northern latitudes (Forster, 2004). According to certain authors, such re-expansion would have leaded its fingerprint in the current genetic variation present in the northern European populations. The resulting reduced diversity in the north of the continent is evidenced today by the predominance of a mtDNA haplogroup in Eskimo-Aleut and Na Dene speakers or by the very high percentage of H and V mtDNA haplogroups in north-western Europeans which appear to have arisen from a founder effect from Iberia peninsula or southern France (Torroni et al., 1998, Torroni et al., 2001). Furthermore, by dissecting haplogroups H (the most common in Europe) and U5 into subhaplogroups, it has been shown that in the Late Upper Palaeolithic there was a repopulation of much of Western and Northern Europe from the Franco-Cantabrian glacial refuge (Iberian refuge). This expansion must be dramatic and it is testified by the star-like structure of sub-haplogroups H1 and H3 (Achilli et al., 2004).

1.8.2 THE NEOLITHIC PERIOD

An important cultural event for the European population history took place about 10,000 years ago in the Near East. Neolithic revolution mainly consisted in the discovery of agriculture (plant and animal domestication), and the development of new technologies as the manufacture of pottery or the building of human settlements (Jobling, 2004). Neolithic revolution started between the current Syria and Israel, in a region called the “fertile crescent”. The human population of the “Fertile Crescent” changed its lifestyle from a hunter-gatherer society to a farmer society. This process was not unique in the history of human populations but such a technological improvement took place independently in time and space in China and Mesoamerica as well. In fact, all these areas share in common the presence of native species suitable for domestication (Diamond, 2002).

The capacity of the farmer societies to produce their own food lead to a crucial human population growth. Several reasons have been proposed for this crucial change in the demography of human populations (Diamond, 2002). One of the main reasons was that increasing the surface of the designated land to harvest edible food could support higher population densities. Moreover, better nutrition could lead to a longer period of fertility in women. Despite these apparent advantages, changes in early farmers were not as beneficial as a priori we could think. Skeletal remains show lower bone porosity among farmers than

among hunter-gathers, a sign indicative of anaemia (Jobling, 2004). This and other skeletal indicators suggest that malnutrition was common among the early farmers. Furthermore, skeletal remains suggest the presence of more lesions derived from pathogenic infections in the early farmers than in the hunter-gathers. In fact, the new environment developed by the farmers should be particularly favourable for the presence of infectious diseases: pathogens could be transferred from an animal host (zoonoses) and high population densities could help to the rapid transmission of infections from one individual to another (Jobling, 2004).

However, as farming was a very successful cultural innovation, it spread rapidly all over the world. Once the "agricultural package" had been learned in one region, it could be exported to other regions of similar climate. The arrival of agriculture to Europe is still a matter of debate among scientists. Two main theories have been proposed to explain the spread of the agriculture through Europe: the **cultural diffusion model** and **the demic diffusion model**. The cultural diffusion model supports the idea that the farmers did not move but the agricultural knowledge was transmitted from the Near East to Europe through the movement of technology and ideas. In contrast, the demic diffusion model (most commonly known as the **wave of advance**) supports the idea that the farmers from the Near East spread to Europe taking agricultural knowledge with them; thus, this model involves gene flow between the hunter-gatherers that inhabited Europe at that moment and the farmers that arrived. Other intermediate models have also been proposed (Ammerman, 1984).

There are two main sources of evidences used to disentangle between both hypotheses: the archaeology and the genetics.

1.8.2.1 EVIDENCE FROM THE ARCHAEOLOGY

Luigi Luca Cavalli-Sforza together with the archaeologist Albert Ammerman were the pioneers in using archaeological sciences to study the impact of the Neolithic revolution in the peopling of the Europe continent (see *The Neolithic transition and the Genetics of Population in Europe*, 1984). Ammerman and Cavalli-Sforza accepted the central role of sedentism and population growth in the early farming communities. They observed that there was a complex of elements at Neolithic sites in Europe (it was called the Neolithic package). This included cereal crops (especially wheat and barley), domestic animals, pottery, ground and polished stone tools and houses. They developed the use of **isochrones maps**, plotting similarly dated sites (radiocarbon dating) on a map of Europe as a series of isochrones lines. The result was a

remarkably uniform rate of expansion of the “Neolithic package” of about one kilometre per year (see Figure 22). They observed that there is a time gradient: the oldest sites, that indicate the origins of agriculture, date to almost 10,000 years ago and they are in the “Fertile Crescent”. Sites become consistently younger towards the northwest of Europe, with agricultural practices arriving at the Baltic and the British Isles between 5,500 and 4,200 years ago. Farming communities appeared in the Aegean area and Greece around 9,000 years ago and they followed two streams of movement, one into southeast Europe, and the other along the Mediterranean coast. From the Mediterranean there were expansions into Italy, France and the Iberian Peninsula.

Ammerman and Cavalli-Sforza introduced the term “*Wave of advance*” in order to explain the outcome of population growth together with range expansion as a radial expanding population wave, in which the culture spreads with the expansion of people. Farmers would have displaced the less numerous hunter-gatherer populations and as a consequence they predicted that the major component of the modern European gene pool is derived from Near Eastern farmers rather than the indigenous Mesolithic populations. A recent analysis of prehistoric population in densities inferred from archaeological data also showed that the likely demographic impact of Neolithic dispersal was greatest in the south eastern Europe, and decrease towards Iberia and the north (Lahr, 2001).

Although there is a debate about interpretation of individual pieces of archaeological evidences, this overall picture from archaeology has provided a framework for the disputes of genetics.

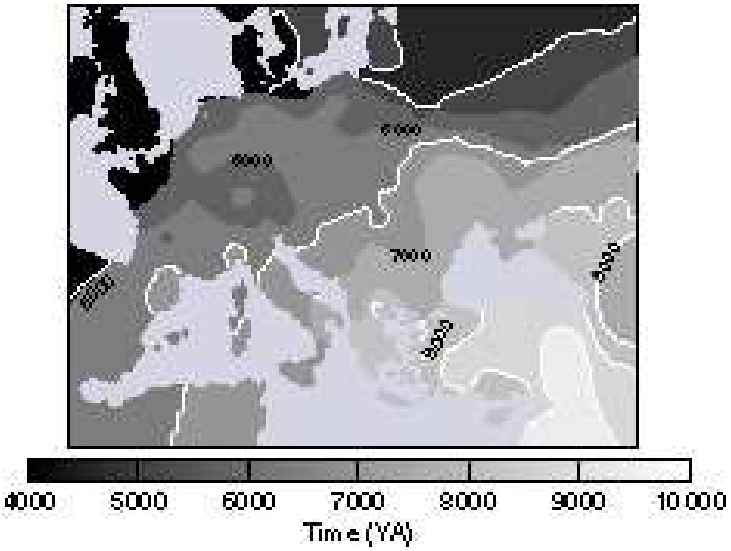


Figure 22: Map of the distribution of the earliest archaeological sites in Europe and the Middle East showing evidence of agriculture. Constructed from data in Ammerman and Cavalli-Sforza (from(Jobling, 2004).

1.8.2.2 EVIDENCE FROM THE GENETICS

Human population history in Europe resembles a palimpsest, where the same peace of paper is written on, scraped off, and used again. Similar migratory routes were followed by the first settlement of Europe by early anatomically modern humans and the later spread of agriculture, that is, from South-East Europe to North-West Europe (Barbujani and Goldstein, 2004) (see Figure 23). Furthermore, the hunter-gatherers that occupied Europe and the incoming farmers are likely to share the same genetic diversity, since the divergence took place only some hundreds generations ago, a time relatively small in the evolutionary timescale. Moreover, the genetic composition of the indigenous Europeans before the arrival of the agricultural is unlikely to have been uniform; the original occupation in the Palaeolithic was followed by a later re-expansion after the last glacial maximum from the southerly glacial refugia in the Iberia peninsula, Italy and the Balkans (see chapter 1.8.1). The sum of these population movements is very likely to have produced heterogeneous patterns of gene frequencies among hunter-gatherers populations before the arrival of agriculture (Barbujani and Bertorelle, 2001, Richards et al., 2000, Richards et al., 2002, Barbujani and Goldstein, 2004).

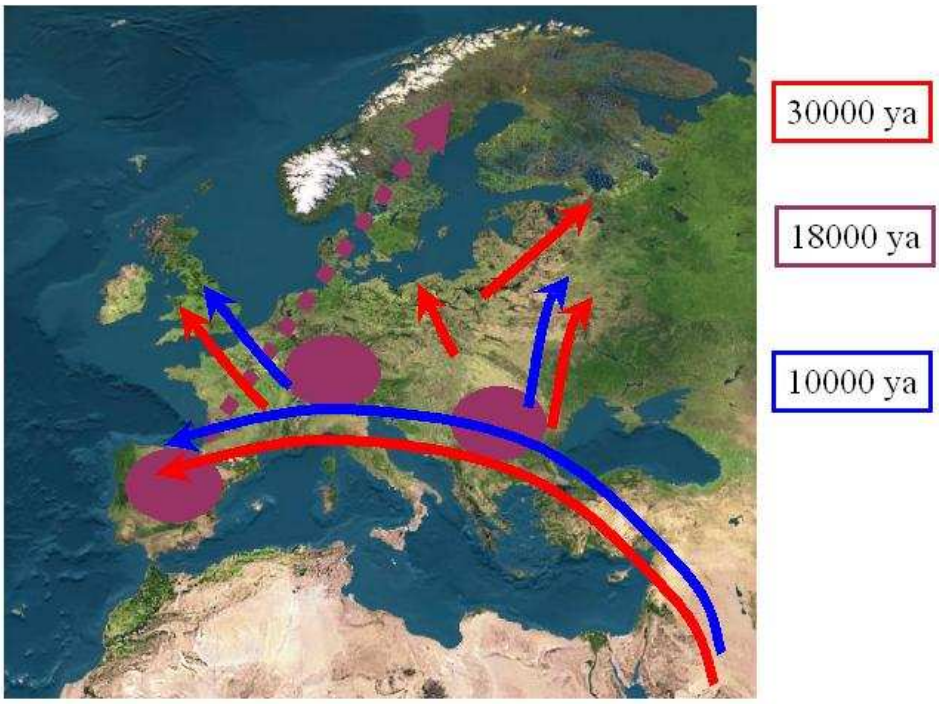


Figure 23: Major population movement in the European prehistory (Adapted from (Simoni et al., 2000).

Therefore, the interpretation of the available genetic data (current and ancient) is not an easy task; a lot of factors have to be taken into account in order to get a definitive conclusion about how was the impact of the Neolithic revolution, that took place in the Near East 10,000 years ago, in the current European genetic diversity.

1.8.2.2.1 EVIDENCE FROM AUTOSOMIC LOCUS

Menozzi et al (Menozzi et al., 1978) were the first in using genetics in order to study how big was the impact of the "Neolithic revolution" in the patterns of human genetic diversity that we observe today in Europe. They analysed the frequency of 38 alleles at 10 different loci in several European populations and, some years later, the same authors (Cavalli-Sforza, 1994) extended the study to 94 alleles at 34 loci. Since patterns for single locus should be treated with caution (e.g.: some might be under natural selection, some might be uninformative) they used **Principal Components analysis** (PCA) to analyse the data from all alleles simultaneously. These authors developed a method to present the output of PCA in the form

of **synthetic geographical maps** of individual principal components (PC). The synthetic map of the first principal component that summarises around the 30% of the total variance has a strong focus in the Near East and shows the same clines from the southeast to the northwest (see Figure 24) as in the case of the radiocarbon dating map (see Figure 22 and chapter1.8.2.1). This result was interpreted as supporting the demic diffusion model of the Neolithic expansion in Europe with a deep impact in the current genetic diversity of the European populations.

These conclusions were supported by posterior studies (Sokal et al., 1989, Sokal et al., 1991) using a different analytical approach based on the study of the spatial autocorrelation of the gene frequencies; other studies using multiple microsatellites loci and estimating the percentage of admixture between south-east and north-west populations have been interpreted also as an indication of the demic diffusion model (Chikhi et al., 1998, Dupanloup et al., 2004, Belle et al., 2006).

Nevertheless, the interpretation of the outcome of the PCA has been severely criticised by several authors. Firstly, (Zvelebil, 1988) pointed out that there was no strong reason for identifying the first PC solely with a Neolithic expansion given that it is very likely that Europe has suffered several migratory movements in the prehistory. So, the observed gradients might be the results of many dispersals, each one overlap the last. Later, Richards et al (Richards et al., 1996) pointed out that the spread of the first modern humans into Europe was about 45,000 years ago, following very similar routes to the later spread of the Neolithic, so the same routes into Europe may have been used time and time again and so it must be difficult to disentangle how was the effect of each dispersal in the current European genetic diversity.

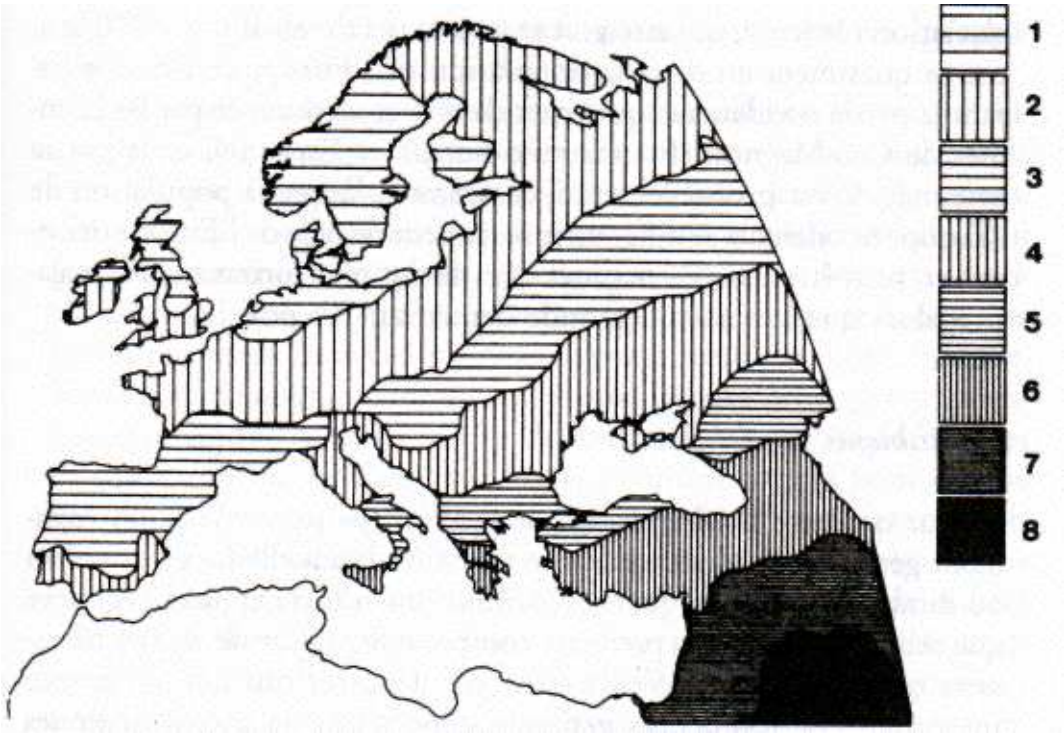


Figure 24:Synthetic map of Europe and western Asia obtained using the first principal component of classical genetic date (Adapted from (Cavalli-Sforza, 1994).

1.8.2.2.2 EVIDENCE FROM mtDNA

The first studies of the mitochondrial control region diversity found little geographical structure in Europe, with most European populations showing similar mtDNA haplotypes (Richards et al., 1996, Richards et al., 1998). The inferred evolutionary networks showed that mtDNA haplotypes form clusters, which were called haplogroups (see chapter 1.7.4.1.1). Richards et al in three different papers where they studied mtDNA diversity in population from Europe and the Near East (Richards et al., 1998, Richards et al., 2000, Richards et al., 2002) defined these haplogroups, calculated their ages based on estimates of mitochondrial mutation rates (founder analysis), and observed that for most haplogroups these ages indicated common molecular ancestors in Palaeolithic times (see Figure 25). The 95% confidence intervals for haplogroup ages proposed by Richards et al (Richards et al., 2000) ranged from 53,600-58,900 for haplogroup U, to 6,100-12,800 for T1. Only a minority of lineages (haplogroups J and T1) were a result of a Neolithic immigration, whereas the remaining lineages, dating back to between 15000 and 50000 years ago, seem to have Early, Middle or Late Upper Palaeolithic origin within Europe. Having these figures, they concluded

that the most European mtDNA sequences descended from local Upper Palaeolithic ancestors, and that only a small fraction of the mitochondrial European gene pool, less than 25% entered Europe in the Neolithic.

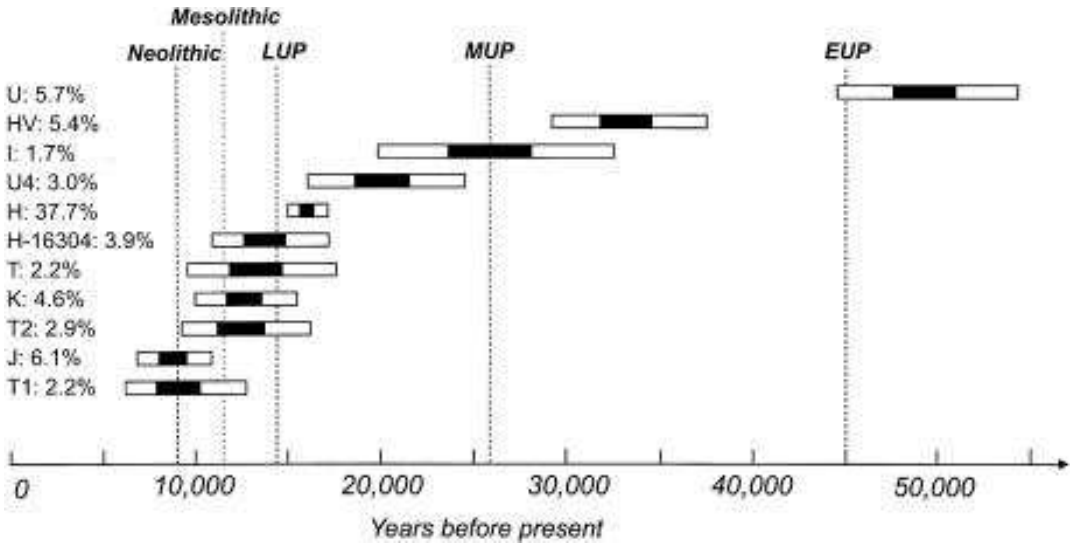


Figure 25: Estimated ages of the major European mtDNA haplogroups founders. The proportion of lineages in each cluster is indicated. The 95% (50%) CRs for the age estimates of each cluster are shown by white (black) bars. (Adapted from (Richards et al., 2000)

Nevertheless, these works have been widely criticized from a traditional population-genetic perspective by a number of authors (e.g (Barbujani and Bertorelle, 2001, Barbujani and Goldstein, 2004) focusing mainly on the way of dating the haplogroups. However, there is a general agreement that the current European gene pool is mainly derived from Paleolithic hunting-gathering and Neolithic farming ancestors, but different studies disagree on the relative weight of these contributions.

In the context of this debate, the results of the geographical dissection of the European haplogroup V have been interpreted as an evidence that late glacial expansion of Palaeolithic populations from refuge areas in Southern Europe could have had a major impact in the repopulation of the continent before the arrival of the farmers into Europe (Torroni et al., 1998, Torroni et al., 2001, Torroni et al., 2006).

1.8.2.2.3 EVIDENCE FROM Y-CHROMOSOME

Y-chromosome genetic analysis provides contradictory evidence about how big was the Neolithic contribution in the current European genetic diversity and contrary to the mtDNA variability point out towards a demic diffusion model in the transition from the Palaeolithic to the Neolithic period in Europe.

Several studies show (e.g. (Semino et al., 1996, Rosser et al., 2000) that the most common diagnostic Near Eastern haplogroups are J and E3b. Both J and E3 display declining gradients moving from the Near East to Europe, and both may have potentially been spread with the Neolithic so that these clines are compatible with the demic diffusion model. Semino et al (Semino et al., 2000) published that the Near Eastern Neolithic contribution to Europe would be around 20%-25%. However, this work was severely criticised by Chikhi et al (Chikhi et al., 2002) that reanalysing the date set of Semino et al (Semino et al., 2000) suggested that the Neolithic contribution is much higher, in fact greater than 50% in some parts of Europe.

1.8.2.2.4 EVIDENCE FROM ANCIENT DNA

As previously mentioned, Europe has an ancient and complex history where modern humans have been around for the last 40 KY and so that, a lot of population movement could have happened in response, for instance to climatic changes. Thus, disentangling the population history from the actual content of genetic variation in the European populations is not an easy task. Indeed, it would be preferable analysing the genetic diversity of the farmers and the hunter-gatherers rather than studying the genetic diversity of their (putatively) descendents. This could be achieved by means of the direct analysis of ancient DNA from ancient remains. The difficulties when working with ancient DNA of anatomically modern humans are now well known (see chapter 1.3) and reliable results could only been expected under special circumstances (Gilbert et al., 2005). However, Haak et al (Haak et al., 2005) successfully extracted and sequenced the HVR I of the mtDNA from 24 out of 57 Neolithic skeletons from various locations in Germany, Austria and Hungary. All human remains were dated to the LBK or AVK period (7000 to 7500 years ago). They found that 25% (6 out of 24) of the samples are of a distinctive and rare N1a lineage of the mtDNA well-known phylogeny. Furthermore, five of these six individuals display different N1a haplotypes and they were widespread in the LBK area. Europeans today have a 150-times lower frequency (0.2%) of

this mtDNA type, revealing that these first Neolithic did not have a strong impact in the genetic background of the modern European female lineages. In addition, they addressed the question of whether the 150-times lower frequency of N1a in modern Europeans might be due to a simple genetic drift scenario over the past 7500 years. To resolve that problem they used a simulation approach finding that the results of the simulations reject the simple hypothesis in which modern Europeans are direct descendants of these first farmers and have lost N1a mainly by genetic drift. They proposed that small pioneer farming groups carried farming into new areas of Europe, and that once the technique had taken root; the surrounding hunter-gatherers adopted the new culture and then outnumbered the original farmers, diluting their N1a frequency to the current residual figure.

Thus, this result supports the cultural diffusion model, where the farming culture itself spread without the people originally carrying these ideas. They proposed that within the current debate on whether Europeans are genetically of Palaeolithic or Neolithic origin, and leaving aside the possibility of significant post-Neolithic migration, their data lend weight to the arguments for a Palaeolithic origin of Europeans.

In summary, it is difficult to quantify with absolute confidence the contribution of the first Palaeolithic settlers, or of later Neolithic immigrants, to the current Europeans' genome.

1.8.3 POST-NEOLITHIC PERIOD

On the basis of archaeological data, only the following three large-scale demographical events that occurred in Europe since Palaeolithic times until Neolithic times has been documented to have a deep impact in the European demography: the continent colonization by anatomically modern humans coming from the Near East in the Palaeolithic (Mellars, 1992); the re-expansion of the population from glacial South refugia towards Northern Europe after the last maximum glacial (Otte, 1990) and the introduction of agriculture from the Near East towards western and Eastern Europe (Ammerman, 1984). The different genetic studies using different types of genetic markers explain the current geographic structure of the European population's genetic diversity within the context of whatever of those three main events (Barbujani and Bertorelle, 2001). Post-Neolithic migrations are probably not important to understand the origin of the continental genetic clines, (Barbujani and Goldstein, 2004). None of the post-Neolithic migration processes documented in the archaeological and historical records seems to have had a sufficient geographical and demographic scope to determine

spatial structuring of genes across Europe. In addition, these migrations occurred in all directions and hence they are not expected to generate a continent –wide pattern (Sokal et al., 1996). However, although small in comparison with the main migrations, such expansions could have left its genetic fingerprint at a micro geographic scale. Semino et al (Semino et al., 2004) explained the variability of the Y-chromosome for certain markers with regard to more recent migrations. In addition, Xiao et al (Xiao et al., 2004) when analysing the variability in a marker in the X-chromosome proposed a recent a high rate of recent post-Neolithic female migration in Europe.

Within this framework, the genetic analysis of European ancient human remains belonging to the post-Neolithic period allow us to understand better how was the impact of post-Neolithic migration in the history of particular, well delimited, geographical regions. Nevertheless, up till now, analyses based on the genetic of ancient population from post-Neolithic period in Europe are very scarce (Izagirre and de la Rua, 1999, Alzualde et al., 2005, Vernesi et al., 2004). Vernesi et al (Vernesi et al., 2004) analysed 28 individuals (following all authenticity criteria for the validation of ancient DNA sequences) belonging to the Etruscan population that populated Italy between the VIII and III BC. They found that genetic distances and sequences comparison at mtDNA level showed closer evolutionary relationship with the easter Mediterranean shores from the Etruscans than for modern Italian population. The Etruscan sites appear to have rather homogeneous genetic characteristics. In addition, Etruscan mitochondrial haplotypes were very similar, but rarely identical, to those commonly observed in contemporary Italy and they suggested that the links between the Etruscan and Eastern Mediterranean region were in part associated with genetic, and not only to culture exchanges. A year later, Alzualde et al (Alzualde et al., 2005), when analysing the mtDNA variability of the historical population of Aldaieta (VI-VII AD, Basque country) found that the variability of the mtDNA haplogroups fall within the range of the present-day population of Europe´s Atlantic fringe, whereas the prehistoric population of the Basque country display clear differentiation in relation to present day Basque population. These two genetic studies regarding to prehistoric population from post-Neolithic times, reveal a discontinuity between prehistoric and present day population, suggesting a post-Neolithic structuration of the populations between 5,000-1,500 years before present.

It may be that this phenomenon is more widespread and had occurred in other regions of Western Europe. Data on more prehistoric populations are required in order to confirm this phenomenon. It should also be taken into account that these differences have only been detected by means of studying the mtDNA that is known to be quite homogeneous in Europe

((Simoni et al., 2000, Richards et al., 2000); analysing the nuclear genome of prehistoric European populations could provide further evidences of post Neolithic micro migrations that have occurred in the European populations

2OBJETIVES

In the current thesis we were interested in different technological aspects of the aDNA research field and in the history of human populations in the European (mainly the Iberian Peninsula region) continent:

(i) The problematic associated to pre-laboratory modern human DNA contamination when working with ancient human remains (see chapter 4.1chapter). Several studies have deal the problems associated with laboratory derived contamination when working with ancient human remains and a number of guidelines has been suggested to help deal with this issue, but little is known regarding to sample contamination prior the genetic analysis. In the first chapter we address this issue. To do that, Neolithic remains from the Iberia peninsula were subjected to a genetic analysis adopting all the authenticity criteria proposed so far, but with the advantage that we have been able to monitor all the persons involved in the manipulation of the remains before and after their genetic analysis.

(ii) Development of non-invasive techniques that allow studying ancient samples but preserving it from the destruction due to the current methods of DNA extraction. In particular, we attempted to extract aDNA from Neandertals from soil sediments (in particular, clay sediments) from “El Sidrón Cave” (see chapter 0). The chemical characteristics of clay make this substrate suitable for binding and preserving aDNA.

(iii) The study of three different historical periods of the prehistoric peopling of Western Europe by using ancient DNA techniques: Palaeolithic, Neolithic and Post-Neolithic. The first historical period was covered by retrieving and comparing with existing data aDNA from two samples of a Neanderthal from “El Sidron cave” (Asturias, North of Spain) dated 43,000 years before present (see chapter 4.3). The Iberian Peninsula represents both the Western and the Southern European edge of the Neanderthal distribution. It is furthermore the place where Neandertals coexisted longest with modern humans and where it has been suggested that hybridisation between these species may have taken place. Consequently, the retrieval of mtDNA sequences from an Iberian Neandertal represents an important step in our understanding of the evolutionary history of this species and its past interaction with *Homo sapiens*. Next we studied a Neolithic population from the Iberian Peninsula, which allowed us to share light between the different hypotheses of cultural versus demic diffusion of the agriculture

(see chapter 4.5). Finally, we attempt to study a Post-Neolithic Iberian population (see chapter 4.6). This population is of particular interest because they spoke and wrote a non-Indo-European language and their origins and relationships with other population (like Etruscans or Basques) remain unclear. We tried to resolve this issue thought the genetic analysis of the scarce skeletal remains that have been preserved (the Iberians cremated the bodies).

3MATERIALS AND METHODS

3.1 DNA EXTRACTION

Extraction procedures were carried out in an isolated pre-PCR area exclusively dedicated to ancient DNA studies and where no previous DNA amplification had been undertaken, physically separated from the main laboratory (molecular biology laboratory) with positive air pressure, overnight UV light and continually bleach cleaning of the bench surfaces. Furthermore, all samples and reagent manipulation were performed in a laminar flow cabinet routinely irradiated with UV light and we used coverall, gloves, facemasks and sterile tips in order to help avoid intra-laboratory derived contamination.

The techniques that were employed to extract DNA from ancient specimens varied according to the tissue. Nevertheless, such methods rely on an initial digestion of the tissue to release DNA, with a subsequent purification step using either organic solvents (mainly phenol-chloroform protocol) or the DNA binding properties of silica. It is usual that before starting the extraction procedure and specially when working with ancient human samples (the main aim of this thesis), the specimen is 'decontaminated' using various techniques that are aimed at removing any surface contaminants such bleach surface cleaning or UV light surface irradiation. Following these steps, bones, teeth, and other ancient samples, are powdered using a drilling machine. Grinding has the effect of increasing surface area, thus aiding the posterior digestion of the ancient sample.

3.1.1 DNA ISOLATION FROM TEETH

All ancient teeth DNA extractions made in the current thesis (chapter 1,2 and 3) were performed by a **phenol-chloroform** extraction protocol. The surface of each sample were first scraped with a scalpel; cleaned with bleach and then ground to powder. We used from 0.1 to 1 g of teeth powder in each DNA extraction.

First, the sample was subjected to a decalcification step with ten millilitres of EDTA (ethylenediaminetetra-acetic acid; pH:8; 0.5M) overnight at 37ºC in order to remove mineral salts (mainly Ca^{++} and $PO_4^{—}$ ions). Next and after centrifugation at 5000x for 5min, the EDTA was carefully poured off and the powder was incubated overnight at 50ºC in a **lysis solution** (1ml SDS 5%, 0.5ml TRIS 1M, 8.5ml H_2O and loads of

proteinase K) in order to release DNA from the cells. Tris-HCI breaks up the cell walls by creating an hypotonic medium, proteinase K digests the protein fraction, and a nonionic detergent such as SDS (sodium dodecyl sulphate) emulsifies the lipids. The day after, after centrifugation at 5000x for 5min, the aqueous face (the DNA is now in the solution) were extracted three times, first with 10 ml phenol, second with 10 ml phenol-chloroform (1:1) and third with 10 ml chloroform-isoamilic alcohol (24:1). Since DNA is polar, it remains dissolved in the aqueous phase, whereas many other compounds commonly found in the tissue remain in the organic phase. Between each extraction step the sample is centrifuged at 5000x for 5 minutes, take the top layer (aqueous layer) with a Pasteur pipette, put in a fresh falcon tube and continues the following extraction step. Once the third extraction is made (chloroform-isoamilic alcohol), the resulting aqueous phase is carefully removed (as carryover of organic solvents to subsequent extraction and amplification stages may inhibit PCR), and desalted and concentrated using centricons (Millipore). To do that, we take 2 ml of the sample, put into a centricon with filter (removes macromolecules less than 30,000 molecular weight (MW) that may inhibit subsequent PCR reactions and keep the rest, including DNA) and centrifuge at 4000x for 30 min. The solution that come through the filter is throwing it away and we put another 2 ml of the sample in the centricon and repeat again the steps until the sample is finished. The DNA is now in the filter. Then we clean the filter with water (biological water). We add 2 ml of water and centrifuge at 4000x for 30 min. We repeat it twice. In the last step, we turn around the centricon to receive the DNA that is in the filter. Now we centrifuge at 4000x for 5min and keep the solution in the frozen.

3.1.2 DNA ISOLATION FROM SOIL SEDIMENTS

The extraction protocol followed in order to extract DNA from soil sediments from El Sidrón Cave (Asturias, Spain) was the following:

1. Add 0.25g soil to each FAST PREB soil tube in Flow Hood. Clean work area between loading samples.
2. Suspend pellet in 600µl enzymatic solution.

12.5ml enzymatic solution

2.5 ml of 10% Sarcosyl
0.625ml 1M Tris ph 7.8
0.5ml 0.5M EDTA
0.375ml 5M NaCl
Make up to 11.925ml (add water)

Filter with 3000 MWCO

Just before use:

0.43ml of Beta mercaptoethanol (50mM)
10mg of proteinase K (640 µl of an stock 14-22mg/ml)
0.625 ml 1M DTT
0.25ml PTB (100mM)

3. FAST PREP tubes (level 6 for 45 sec. x 4 sessions) and put the tubes on ice for 1-2 min. between each session.
4. Leave the enzymatic solution for agitation at 55ºC for overnight.
5. Spin briefly (5000rpm).
6. Add 150µl 5M NaCl to enzymatic solution
7. Add 375µl Chloroform/Octanol solution (24:1).
8. FAST PREP tubes briefly to mix
9. Rotate tubes at room temperature and leave for 30 minutes (or overnight).
10. Centrifuge the solution at 12,000g for 2 min. Taking care not to disturb the interface, combine the samples by transferring the aqueous (top) phase from the pairs of tubes to a new 1.5ml Eppendorf tube and incubate it at 2-3ºC for at least 1h allowing sediments to settle.
11. Centrifuge the solution at 12,000g for 2 min and move the supernatant to a 15ml tube.
12. Add Quiagen PB buffer (5 x the vol. of the supernatant) and agitate by hand.
13. Move remainder of the solution, (700µl a time) to a QIAquick spin column and centrifuge 1 min at 10,000g. Discard filtrate. Remember, samples can be split into multiple spin-columns at this stage.
14. Sub-aliquot buffers – don't use stocks. Remember, check Salton wash 1 hasn't precipitated.
15. Add 500µl Salton wash 1 buffer and centrifuge 1 min at 10,000g.

16. Add 500µl Salton wash 2 buffer and centrifuge 1 min at 10,000g.
17. Add 500µl AW 1 buffer and centrifuge 1 min at 10,000g.
18. Add 500µl AW 2 buffer and centrifuge 3 min at 15,000g. Discard filtrate and spin again to remove residual ethanol from base of spin column. *Ethanol in the final solution will inhibit PCR*.
19. Place QIAquick column in a clean 1.5 ml Eppendorf safe lock tube with the lids removed.
20. To elute the DNA, add 100µl buffer EB to the centre of the QIAquick membrane and leave at room temperature for 10 min. Then centrifuge the column for 1 min at 10,000g and the extract is ready for use.
21. Transfer and combine samples to clean 0.5ml Eppendorf tube.

3.1.3 DNA ISOLATION FROM HAIR

2-3 cm of hair were mixed with 200 µl chelex 5% ("chelating resin"), with 5µl of proteinase k 10ng/µl and with 7µl of DTT 1M. Then, the mix was heated one hour at 56ºC. After vortexing the mix, it was boiled during ten minutes. Finally, we vortex again the sample and spin it during 3min at 10.000-15000rpm (rounds per minute). At the end, we retrieved the supernatant where the DNA was.

3.2 PCR AMPLIFICATIONS

PCR reactions set up in this thesis were designed to amplify fragments from the mtDNA genome (basically fragments from the control region, although we have type some diagnostic SNPs of the coding region (see **Table 4**) to classify the sequences in haplogroups) either of ancient humans or Neandertals specimens. Since the DNA was fragmented, we designed (or used designed primers from previous studies) primers that covered the entire mtDNA control region both in modern humans (see **Table 3**) and in Neandertals (see **Table 5**)with a very long range of sizes.

L primer	Sequence (5'-->3')	H primer	Sequence (5'-->3')
L16022	ctaatttaaactattctct	H16142	atgtactacaggtggtcaag
L 16055	gaagcagatttgggtaccac	H16158	tgtggattgggtttttatgt
L16081	ttgactcacccatcaacaa	H16211	tagttgagggttgattgctg
L16122	cattactgccagccaccatgaata	H16218	tgtgtgatagttgagggttg
L16131	caccatgaatattgtacggt	H16247	caactatcacacatcaactgcaa
L16185	aacccaatccacatcaaaacc	H16281	ttaagggtgggtaggtttgt
L16209	ccccatgcttacaagcaagt	H16356	gtcatccatggggacgagaa
L16247	caactatcacacatcaactgcaa	H16378	caagggacccctatctgagg
L16223	agcaagtacagcaatcaac	H16385	ggtggtcaagggacccctat
L16261	caactatcacacatcaactgcaa	H16401	tgatttcacggaggatggtg
L16347	cgtacatagcacattacagt		

Table 3: HVRI Mitochondrial primer sequence information for PCR amplification in human extractions.

L primer	Sequence (5'-->3')	H primer	Sequence (5'-->3')	haplogroup
L12227	gaaagctcacaagaactgc	H12341	ggttatagtagtgtgcatgg	U (12208)
L 6999	caaactcatcactagacatcg	H 7066	gaatgaagcctcctatgatgg	H (7028)
L13669	cacccttactaacattaacg	H13725	tagtaatgagaaatcctgcg	J (13708)
L13257	aatcgtagccttctccacttca	H13372	ttgttaaggttgtggatgat	T (13368)
L56	gagctctccatgcatttggt	H131	ggatgaggcaggaatcaag	Pre-HV (073)
L10014	ttttagtataaatagtaccg	L10088	gtagtaaggctaggagggtg	I
L8215	acagtttcatgcccatcgtc	H8297	atgctaagttagctttacag	W

Table 4: Primers used to type SNPs in the codifying mtDNA human genome

Lprimer	Sequence (5'-->3')	H primer	Sequence (5'-->3')
NL00189	tacgttcaatattacaggcgag	NH00247	gcagctgtgcagacattt
NL16129	agccaccatgaatattgtaca	NH16169	gggggttgatgtggatta
NL16183	aaaacctaatccacatcaacc	NH16223	gtagttgatgtatgacagttgaa
NL16230*	cagcaatcaaccttcaactg	NH16262	gttgatatcctagtgggtgtaa
None	NL16230	NH16262G	gttgatatcctagtgggtgtag
NL16232	gcaatcaaccttcaactgtc	NH16260	gatatcctagtgggtgtaagg
NL16256**	atacatcaactacaactccaaaga	NH16278	aagggtgggtaggtttgttga
NL16258	catcaactacaactccaaagaca	NH16276	gggtgggtaggtttgttgata
NL16278	ccttacacccactaggatat	NH16311	gtacggtaaatgactttatgtg
L16034	actattctctgttctttcatgg	H16079	acgaaatacatagcggttg
NL16068C	gtaccacccaagtattgacc	H16097	gtggctggcagtaatgta
NL16068A	gtaccacccaagtattgaca	None	H16097
NL16078	agtattgactcacccatcag	H16112	ccgtacaatattcatggtg
L16168	ccacctgtagtacataaaaacc	NH16205	gttgattgctgtgcttg
L16203	ctccccatgcttacaa	NH16234	gctggagttgtagttgatgta
NL16244	cttcaactgtcatacatcaacta	H16268	taggtttgttggtatcctagtg

Table 5: Mitochondrial primer sequence information for PCR amplification in Neanderthal soil extractions.

PCR amplifications were performed in 25 μl reactions and were setting up in the pre-PCR area and then carried to the main laboratory where the PCR machines are. Material interchange between the aDNA laboratory and the main laboratory were strictly forbidden in order to avoid contamination due to the large amount of amplicons that are generated in the molecular biology laboratory. All the precautions that are described in the extraction procedure were fulfilled here as well. We have to distinguish to different PCR reactions: those set up to amplify DNA fragments from ancient teeth and those set up to amplify fragments from soil sediments (work made in Copenhagen University):

3.2.1 ANCIENT TEETH PCR

The standard 25ul PCR reaction conditions were the following:

COMPONENTS	STOCK CONCENTRATION	FINAL CONCENTRATION	ul for each sample
10x Ecotaq buffer	10X	1X	2,5
dNTP´s	2.5mM	0,2mM	0,5
$MgCl_2$	50mM	2,5mM	0,8
Taq DNA polymerase (Ecogen)	5 units/ul	1,2 unit	0,2
ddH2O			
BSA	10 mg/ml	1.4mg/ml	3,5
Primer L	10 uM	1uM	0,4
Primer H	10uM	1uM	0,4
DNA extract			
total volumen			25

The amount of DNA extract and water varies from one PCR to another depending on the extract quality (for example amount of inhibitors.).

The PCR reactions were subjected to 40 amplification cycles (1 min step at 94◦C, 1 min step at 50◦ and 1 min step at 72◦C) with an initial denaturing step at 94◦C for 5 min and a last elongation step at 72◦C for 7 min. Products were electrophoresed in 1.6% low-melting point agarose gels (Invitrogen) stained with ethidium bromide. When needed, PCR products were excised from the gel and subjected to another 35 cycles of PCR with limiting reagents.

3.2.2 SOIL SEDIMENT PCR

The standard 25ul PCR reaction conditions were the following:

COMPONENTS	STOCK CONCENTRATION	FINAL CONCENTRATION	ul for each sample
10x Hi-Fi buffer	10X	1X	2,5
dNTP´s	10mM	0,4mM	1
MgSO4	50mM	3,5mM	1,75
Platinum Hi-Fi	5 units/ul	1 unit	0,2
ddH2O			15,55
Primer stock mix	25uM	1uM	1
DNA extract			3
total volumen			25

The amount of DNA extract varies from 1 to 5ul depending on the extract. The PCR reactions were subjected to 40-45 amplification cycles (30 sec step at 94ºC, 30sec step at 50º-57ºC and 30sec step at 68ºC) with an initial denaturing step at 94ºC for 2 min and a last elongation step at 68◦C for 6 min. Products (12,5µl) were electrophoresed in 3% low-melting point agarose gel (100V for 2 hours). Whenever I obtained a band with the expected size or even a bit bigger I cut it and t purified it.

3.3 DNA PURIFICATION

Before sequencing the PCR products, it is essential purified them. This step is based on eliminating the remaining dNTP´s, the primers that have not been used, the PCR products partially amplified in the elongation step and all those compounds that could interfere in the sequencing reaction. There are many ways to purify PCR products based either on enzymatic reactions (Exo-sap reaction) or in the DNA binding properties of silica (geneclean protocol or column separation). In this thesis, we have used mainly silica purification protocols. We have to distinguish between purify directly PCR products or purify from gel band cut.

3.3.1 PCR PRODUCTS PURIFICATION

We use this kind of purification mainly to purify PCR cloning products. We did it throughout the use of available commercial kits (Amersham biosicence; Omega Bio-Tek (E-Z 96 Cycle-Pure Kits)). These methods use columns (single columns or 96 well-plates) that are pre-packed with a glass fibre matrix that is able to "capture" the DNA. The sample, treated previously with a chaotropic agent ("capture buffer") that promotes the binding of double-stranded DNA, is passed by centrifugation throughout the column, which captures the DNA onto the glass fibre matrix. Once the DNA is "captured" in the matrix, proteins and salt contaminants are washed away with a "wash buffer solution" (which is basically 70% ethanol plus Tris and EDTA). At the end, the purified DNA is eluted in the desired volume of sterile water or TE.
A standard protocol is the following:

- Add 125 µl of capture buffer to each 25µl of PCR product and mix.
- Transfer the DNA solution to the column and centrifuge 1min.
- Add 500 µl of wash buffer to the column and centrifuge 1 min.
- Discard the collection tube.
- Apply 30 µl of water. Incubate 1min and centrifuge.

An alternative method when there are a lot of samples to purify is using ExoSAP method because that requires a minimum of "hand-on" time. **ExoSAP** employs two hydrolytic enzymes, exonuclease (degrades residual single stranded primers) and phosphatase (hydrolyzes remaining dNTPs from the PCR mixture), to remove the unwanted dNTPs and primers.
Protocol:

- Add 2µl of exosap for each 5µl of PCR product.
- Mix and incubate at 37ºC for 15min (thermal cycler)
- Inactivate exosap by heating to 80ºC for 15min (thermal cycler)

3.3.2 GEL BAND PURIFICATION

Two kinds of protocols were used in order to purify bands cutting from low-melting point agarose gels: QIAquick Gel Extraction Kit protocol (Amersham bioscience) or **gene clean** protocol. The former is similar to those described in chapter 3.3.1 with the only difference that before using "the capture buffer" another buffer is needed (QC buffer) in order to dissolve the agarose that now is present together with the PCR product. Gene clean protocols is a more tedious protocol but is very efficient and highly recommended when the length of the amplified fragment is very small in order to avoid the product can get lost (the recovery threshold of the commercial kits are around 70bp).

Gene clean protocol:

- Mix in an eppendorf: 60µl NaI + 8µl Silica + 20µl PCR's products
- Vortex
- Incubate 5-10min at room temperature and vortex every 2min
- Centrifuge 10 sec at 12000rpm
- Discard the supernatant tipping the liquid on an absorbent paper
- Wash the pellet (three times)
    - -Add 300µl of new wash solution
    - -Vortex
    - -Centrifuge at 12000rpm 10sec
    - -Discard the supernatant
- Re suspend the DNA (twice)
    - -Add 8µl ddH_2O to dissolve the pellet
    - -Incubate in a water bath 5min at 55ºC
    - -Centrifuge and stop the centrifuge machine just before reaching 12000rpm
    - -Transfer the DNA solution to an eppendorf tube avoiding to take silica.
    - -Discard the pellet
- Keep at –20ºC (before using it, centrifuge shortly to avoid pippeting remains of silica).

3.4 CLONING OF PCR PRODUCTS

The easiest way to do it is using commercial kits. There are a lot of kits available in the market. All of them are based on the same steps: ligation of the PCR products into a vector, transformation into bacteria (*E.coli*) and then screening the colonies and pick up those that carry the insert. In the current thesis we have used three different kits: pMOS*Blue* blunt ended cloning kit (Amersham Biosciences), Topo TA cloning kit (Invitrogene) and Zero Blunt Topo PCR cloning kit (Invitrogene) and we have followed manufacturer's instructions. In the laboratory from Barcelona I have used mainly pMOS*Blue* whereas in the laboratory from Copenhagen I have used the two version of Topo cloning kit depending on the sequence of the fragment that I was cloning.

pMOS*Blue* blunt ended cloning kit protocol:

- Treat seven microliters of PCR product with 1µl of pK enzyme mix
- Incubate at 22ºC for 40 minutes
- Ligate into pMOS*Blue* vector overnight
- Transform 2µl of ligation product into MOSBlue competent cells (40-45 seconds at 42º C, the timing of this step is crucial)
- Grown in eppendorfs with 160µl of SOC medium at 37ºC during one hour
- Plate on IPTG/X-gal agar plates and allow to grow during 16 hours.
- Pick up white colonies with a sterile tip
- Transfer to 50 µl of water. Let it boil during 5min
- Transfer 10µl of the supernatant to a 40µl of PCR reaction (35cycles) using T7 and U-19 universal primers.
- Make an agarose gel, load the samples and run an electrophoresis
- Inserts that yield the correct size are identified, purified and sequenced.

Topo Protocol:

- 0.5 ul salt + 0.5ul vector + 2ul PCR product
- 5min at room temperature (ligation step)
- 13ul competent cells

- 10min on ice
- 30sec at 42 ºC (transformation step)
- 125ul SOC medium
- 1 hour at 37ºC. shaking
- Plate the entire mix in LB agar plates with ampicillin (x-gal in the topo TA cloning kit)
- 37ºC overnight
- Pick up the colonies. PCR using T7 and M13 universal primers.

3.5 QUANTIFICATION OF mtDNA BY MEANS OF RT-PCR

We performed Real Time-PCR experiments to find out if the amount of mtDNA templates was large enough to allow us to obtain reproducible results in the Iberian and Neolithic samples. To do that, firstly, we needed a mtDNA standard. The mtDNA standard was obtained by PCR-amplification of the HVR1 region of the mtDNA genome (we used L 15997 and H 017 primers; 628 bp in length) from an anonymous laboratory donor; the PCR product was purified and then quantified by a Nanodrop Spectrophotometer instrument. After making a dilution of 1/10,000 of the PCR product, we obtained a concentration of 22,48 ng/µl. As in each RT-PCR reaction we used 1µl of this standard, which means that we added 22.48ng of the amplified product in each reaction. Since we were interested in quantified the number of initial templates that we had in each Iberian or Neolithic sample, first, we needed to transform grams in number of templates. To do that, we used the Avogadro´s number (N_A); the number of molecules that are in one mol of whatever substance is 6,023 x 10^{23}. Thus, we had to calculate the weight of 628 bp in length amplify fragment (double strand fragment).

As we know that:

Adenine = C:5; N:5; H:4 =104g/mol

Thiamine = C:5; N:2; H:5;O:2 = 125 g/mol

Cytosine= C:4; N:3; H:4;O:1 = 110 g/mol

Guanine= C:5; N:5; H:4;O:1 = 150 g/mol

Deoxiribose + phosphate = 181 g/mol

The number of adenine, thiamine, guanine and cytosine were counted in each strand and it was calculated the total molecular weight, then it was added the molecular weight of the phosphate backbone. Therefore, the molecular weight of the 628bp amplified was 359.916 g/mol. With this value we could calculate the number of templates in each standard dilution.

As we did 10fold serial dilutions, our standard curve was composed by:

3.761.965 molecules
376.196 molecules
3.7619 molecules
3.761 molecules
376 molecules
37 molecules

Two different sized fragments (107bp and 278bp) within the HVR1 mtDNA region were assayed to estimate the mtDNA preservation as well. The design of primers and probes was performed with Primer Express 2.0 software (Applied Biosystem). The primer sequences for the small fragment were: L16001 ACCATTAGCACCCAAAGCTAAGA and H16065 GCGGTTGTTGATGGGTGAGT, and for the larger one: L16088 TCACCCATCAACAACCGCTAT and H16344 GGGACGAGAAGGGATTTGACT. The probe oligonucleotide sequence for the small fragment was FAM-CAAGCAAGTACAGCAA-MGB and for the large was VIC GAAGCAGATTTGGGTAC-MGB (Alonso *et al*. 2004). Real-time PCR amplification was performed in a 20μl reaction with 1x reaction TaqMan Universal PCR Master Mix (Applied Biosystem), 0,5 μM each primer, 50nM probes, 1mg/ml BSA and 1 μl DNA extract. Ten-fold serial dilutions of the purified and quantified standard were included in the experiment to create a standard curve, in order to quantify the number of initial mitochondrial DNA molecules of each size in the Iberian and Neolithic samples (see Figure 26).

The cycle threshold parameter (Ct) was determined by the SDS software as the fractional cycle number at which the fluorescence increases exponentially. The number of molecules (95% confidence interval) in each fragment is calculated by the standard curve.

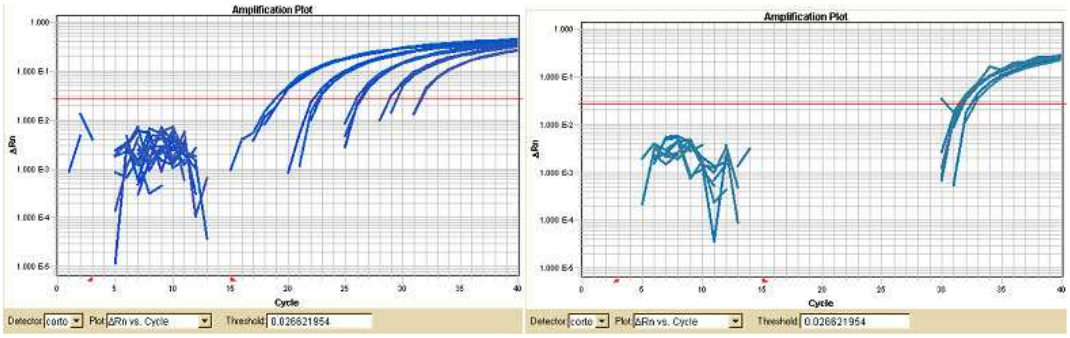


Figure 26: Real Time PCR result in the Iberian samples. The graphic on the right hand is the standard curve and on the left hand is the quantification result on the Iberian samples.

3.6 SEQUENCING

All the sequences analysed in the current thesis were produced using Big dye Terminator sequencing kit (3.0 version; Applied Biosystems) that uses ddNTP´s labelled with flourochromes of different absorbance. Since, almost all the sequences analyses belong to cloning PCR product, T7 universal primer were used in the reaction. The resulting products were loaded in the sequencer ABI prism 3100 (Appied Biosystems).

Sequencing PCR reaction:

REAGENTS	VOLUME
Big Dye terminator	2 µl
Primer 1µM (T7)	3.2 µl
DNA template	5.8 µl

The PCR reactions were subjected to 35 amplification cycles (10 sec step at 96ºC, 5sec step at 50ºC and 4 min step at 60ºC) with an initial denaturing step at 94ºC for 3 min. Then the samples were purified by a alcoholic precipitation:

- Mix in an eppendorf each PCR reaction with:

62,5 µl Ethanol 95%+ 3,0 µl 3M NaAc pH 4.6+23.5 µl ddH_20

-Vortex. Incubation 15min at room temperature.

-Centrifuge 20min. 14000 rpm

-Remove Ethanol solution

-Add 250 µl Ethanol 70%

-Centrifuge 5min.14000 rpm

-Remove Ethanol solution

-Dry the pellet (let open tubes air-dry or use a speed vacuum)

3.7 SEQUENCING ANALYSIS

Two different software were used in order to analyse the sequences: Bioedit (version 7.0) and DNASTART software package (version 3.1).

It is sometimes difficult to distinguish between sequence variation associated to post-mortem damage and sequence heterogeneity attributable to mixed contamination. It can be concluded that a sequence comes from a single source –and therefore, it is potentially endogenous- if all clones share the same pattern of substitutions, but this is not always the case. Occasionally, the majority of the clones harbour some motifs, while a minority –sometimes less than 5% of the clones- show a different haplotype; it seems more parsimonious to attribute the sequence displayed by the majority of the clones to the endogenous DNA (specially if we have evidence of a good biochemical preservation) and the sequence displayed by the minority of the clones to residual contaminants (basically for handling the specimen) than to the reversal (the majority being the contaminant). However, at least in some Neandertal specimens, it is obvious that the majority of the sequences retrieved (sometimes up to 95%) are modern contaminants.

In addition, quite often there are singletons (substitutions not shared by other sequences) in the clones; while it can be argued that each of these clones with singletons may come from different contaminant sequences, it is again more parsimonious to attribute it to cloning artefacts (substitutions generated during the growing of a particular bacterial colony) than to a huge diversity of contaminants (specially in modern humans, where the sequence variation is low).

4RESULTS

4.1 CHAPTER 1: Traking down human contaminations in ancient human teeth

María Lourdes Sampietro, M. Thomas P. Gilbert , Oscar Lao, David Caramelli, Martina Lari, Jaume Bertranpetit and Carles Lalueza-Fox.

Molecular Biology and Evolution. 2006.June;23(9):1801-07

4.2 CHAPTER 2: Dirty Neandertal DNA

María Lourdes Sampietro, Juan J. Sanchez, M. Thomas P. Gilbert, Anders Götherstrom, Kasper M Terkelsen, Rasmus Nielsen, Jonas Binladen, James Haile, Marco de la Rasilla, Javier Fortea, Antonio Rosas, Jaume Bertranpetit, Carles Lalueza-Fox, Eske Willerslev

(Submitted)

Dirty Neandertal DNA

María Lourdes Sampietro[1,3], Juan J. Sanchez[2], M. Thomas P. Gilbert[3], Anders Götherstrom[4,5], Kasper M Terkelsen[3, 4], Rasmus Nielsen[3, 6, 7], Jonas Binladen[3], James Haile[8], Marco de la Rasilla[9], Javier Fortea[9], Antonio Rosas[10], Jaume Bertranpetit[1], Carles Lalueza-Fox[11], Eske Willerslev[3*]

[1]Unitat de Biologia Evolutiva, CEXS, Universitat Pompeu Fabra, Barcelona, Spain.

[2]Department of Forensic Genetics, University of Copenhagen, Denmark.

[3]Centre for Ancient Genetics, University of Copenhagen, Denmark.

[4]Department of Evolutionary Biology, Uppsala University, Sweden.

[5]Centro Mixto UCM-ISCIII de Evolución y Comportamiento Humanos, Madrid, Spain.

[6]Centre for Bioinformatics, University of Copenhagen, Denmark.

[7]Department of Biology, University of Copenhagen, Denmark.

[8]Ancient Biomolecule Centre, University of Oxford, UK.

[9]Área de Prehistoria, Departamento de Historia, Universidad de Oviedo, Oviedo, Spain

[10]Departamento de Paleobiología, Museo Nacional de Ciencias Naturales, Madrid, Spain.

[11]Secció Antropologia, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain.

Author of correspondence: ewillerslev@bi.ku.dk

RESULTS

Online Abstract

By extracting DNA from clay sediment samples taken from the El Sidrón cave (Asturias, Spain), dated to around 43,000 years ago, we have been able to amplify mitochondrial sequences that are identical to previously retrieved Neandertal sequences. Using a Bayesian approach to assign the Neandertal-like sequences to any possible taxonomic group, we demonstrate that the probability that contemporary human contaminants could generate these sequences by chance is essentially zero. These results suggest, therefore, that Neandertal DNA derived from body decomposition is preserved in Pleistocene sediments. This finding could have wide applications in forensics and paleoenvironment reconstructions.

Text

To date, putative mitochondrial DNA (mtDNA) sequences have been reportedly recovered from teeth and bone samples of thirteen Neandertal individuals (*1* and references therein). The restricted number of Neandertal fossils (~400) and the destructive nature of ancient DNA (aDNA) sampling places limits on the skeletal remains available for analysis. However, so far no other alternative sources of Neandertal genetic material have been successfully explored. Here, we report the retrieval of what appears to be Neandertal mtDNA sequences preserved in clay sediments taken from the El Sidrón cave in Asturias (Spain) (*2*), a site dated to around 43,000 years ago (*1*).

Previous studies have shown the presence of aDNA from animals and plants in permafrost and cave sediments (*3, 4*). As clay is known to have a high capacity for binding free DNA, and protects DNA against degradation (*5*), it is plausible that DNA originating from the *in situ* decomposition of the Neandertal bodies may remain absorbed to the clay. Based on this evidence, clay samples were taken under controlled conditions (*6*) directly from stratigraphic layers of the El Sidrón cave containing Neandertal remains. DNA was extracted

from one clay sample (#5) obtained immediately next to an occipital fragment in a dedicated aDNA laboratory where no previous work with Neandertal DNA has been conducted. Polymerase chain reaction (PCR) was used to amplify two fragments of the mtDNA D-loop hypervariable region 1 (HVR1) in two different laboratories, using Neandertal specific primers: NL16,230/NH16,262 and NL16,263b/NH16,300 (*6*). The amplification products were cloned and 960 clones sequenced.

Of the clones, more than 22% (209 clones) were found to be of bacterial origin while less than 1% (6 clones) was found identical to contemporary humans and is likely the result of contamination (assignments based on simple BLAST search). Approximately 77% (741 clones) were likely PCR artifacts or sequences with no match in GenBank, although 28 of them contained fragmentary sequences identical to Neandertal haplotypes (Fig. S2). Finally, 5 clones (<1%) obtained in two independent amplifications with the NL16,263b/NH16,300 primer pair display the 16,278[T]-16,299[G] haplotype (Fig. S2) found in all Neandertal specimens investigated except that of Monte Lessini (*1*).

There are several lines of evidence suggesting that the sequences are truly of Neandertal origin. First, the 16,278[T]-16,299[G] haplotype is not known among >10,000 Caucasians (database updated from *7*), and has not been reported among modern humans in GenBank. The 16,278[T]-16,299[G] motifs are absent in all the excavators and laboratory researchers involved. Second, using a phylogenetic criterion to assign the Neandertal-like sequences to any possible taxonomic level for organisms represented in GenBank (*6*), we found the 16,278[T]-16,299[G] haplotype sequences forming a monophyletic group together with the GenBank Neandertal sequences with a posterior probability of 91% (Fig. 1A). Additionally, based on an alignment with the 50 most similar modern human sequences from GenBank, the Bayes factor in favour of Neandertal rather than *Homo sapiens sapiens* origin is 131. The probability that a human contaminant should generate a similarly high Bayes factor

by chance is extremely low ($p < 0.01$) (Fig. 1B). It is therefore highly unlikely that the 16,278[T]-16,299[G] haplotype sequence is a contemporary human contaminant or another contaminant present in GenBank.

Based on our results we cannot discard the possibility that an un-described soil contaminant, not represented in Genbank, could display sequences identical to Neandertal mtDNA haplotypes. However, should this be the case, then other short DNA sequences previously amplified from other Neandertal remains must be viewed with caution. Other explanations, such as DNA damage and recurrent chimaeric sequences from contaminations of unknown origin are less plausible, since the Neandertal-like fragment show a coherent and reproducible Neandertal haplotype. Thus, the most likely explanation is that short fragments of Neandertal mtDNA have been preserved in the sediment of El Sidrón, maybe due to the stable environmental conditions and fairly low annual mean temperature (10-13ºC on average) that exists within this karstic system.

The retrieval of Neandertal DNA sequences directly from sediments have important implications. If such sequences can be readily obtained from cave sediments it would provide a unique opportunity to explore Neandertal intragroup genetic variation through time and across occupational sites. Additionally, this new “Dirt DNA” approach could have wide applications in forensics and palaeoenvironment reconstructions.

References

1. C. Lalueza-Fox *et al., Curr. Biol.* (in press).
2. J. Fortea *et al., Estud. Geol.* **59,** 159 (2003).
3. E. Willerslev *et al., Science* **300,** 791 (2003).
4. M. Hofreiter *et al., Curr. Biol.* **13,** R693 (2003).

5. M. G. Lorenz, W. Wackernagel, *Microbiol. Rev.* **58,** 563 (1994).

6. Supporting Online Material.

7. M. Richards, *et al., Am. J. Hum. Genet.* **67,**1251 (2000).

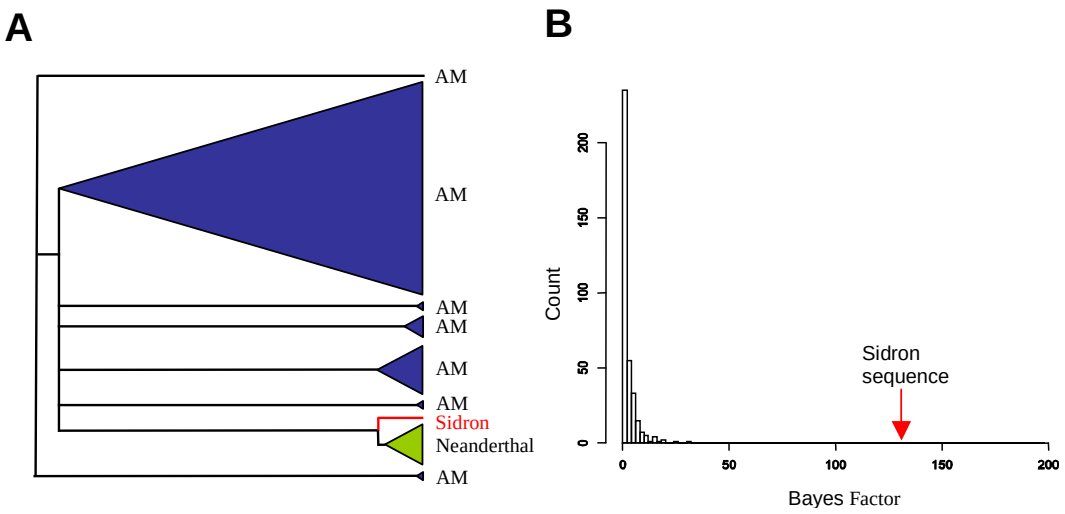


Fig. 1. Analyses of the 16,278[T]-16,299[G] haplotype sequences. **(A)** Strict majority rule consensus tree for the query sequence, Neanderthal sequences in GenBank, and the 50 best other Genbank hits (anatomically modern humans, AMH). **(B)** The empirical distribution of Bayes factors in favour of the Neanderthal hypothesis calculated for AMH sequences from the HvrBase++ database (*6*). The observed value for query sequence falls outside the distribution expected for AMHs, providing strong statistical evidence against the hypothesis that it could be of contemporary human origin.

Supporting Online Material

Dirty Neandertal DNA

María Lourdes Sampietro[1,3], Juan J. Sanchez[2], M. Thomas P. Gilbert[3], Anders Götherstrom[4,5], Kasper M Terkelsen[3, 6], Rasmus Nielsen[3, 6, 7], Jonas Binladen[3], James Haile[8], Marco de la Rasilla[9], Javier Fortea[9], Antonio Rosas[10], Jaume Bertranpetit[1], Carles Lalueza-Fox[11], Eske Willerslev[3*]

[1]Unitat de Biologia Evolutiva, CEXS, Universitat Pompeu Fabra, Barcelona, Spain.

[2]Department of Forensic Genetics, University of Copenhagen, Denmark.

[3]Centre for Ancient Genetics, University of Copenhagen, Denmark.

[4]Department of Evolutionary Biology, Uppsala University, Sweden.

[5]Centro Mixto UCM-ISCIII de Evolución y Comportamiento Humanos, Madrid, Spain.

[6]Centre for Bioinformatics, University of Copenhagen, Denmark.

[7]Department of Biology, University of Copenhagen, Denmark.

[8]Ancient Biomolecule Centre, University of Oxford, UK.

[9]Área de Prehistoria, Departamento de Historia, Universidad de Oviedo, Oviedo, Spain

[10]Departamento de Paleobiología, Museo Nacional de Ciencias Naturales, Madrid, Spain.

[11]Secció Antropologia, Facultat de Biologia, Universitat de Barcelona, Barcelona, Spain.

Author of correspondence: ewillerslev@bi.ku.dk

RESULTS

Material and Methods

Archaeological Site

El Sidrón is a deep and narrow karstic system close to the Cantabrian mountain range in Asturias (north of Spain) (*S1*). Numerous Neandertal skeletal fragments are continually being retrieved in a small pit lateral to the main cave gallery, around 250 meters from the current entrance. The geological evidence suggests that the bones of at least eight Neandertal individuals accumulated there, after the collapse of a doline at the surface (*S1*). The sedimentary deposits suggest that little or no depositional movements followed, which explains why some fragile skeletal structures, like foot bones and a partial thoracic cage, are still in anatomical connection.

DNA Extraction

During the 2004 excavation season, five soil samples were taken by one of us (C.L.-F.) at different levels in the Neandertal remains' layers (Fig. S1). The samples were obtained with sterile gloves and facemasks by pushing 50 ml Falcon tubes into the wet clay; the sub-samples analyzed correspond to the soil placed deeper into the archaeological section and therefore, not accessible to archaeologists' manipulations.

The sample processing and DNA extraction were carried out in dedicated ancient DNA facilities in Copenhagen with the standard precautions used on ancient DNA research (*S2, S3*). Importantly no work on Neandertals has previously been conducted in Copenhagen. A total of 40 DNA extractions were performed on five sections of the clay sediment # 5 (Fig. S1) using two different methods (both successful): one set of extractions followed the procedure described in (*S4*), except for the addition of PTB (N-phenacylthiazolium bromide) in the lysis buffer as in (*S5*). The other DNA extraction protocol is not previously described: About 300mg sediment (wet weight) was added to 0.75ml TE buffer also containing with proteinase K (15µg) and Urea (1M) and incubated at 50°C for 4 hours. 0.75ml phosphate-

buffer (0.2 M K_2HPO_4, set to pH 7.0 with phosphoric acid) was added and the reaction incubated at 50°C for additional 20 hours. The reaction was centrifuged down at 12.000 rpm for 5 min, the supernatant transferred to a 30.000 cutoff amicon filter (supplied by Millipore) and centrifuged for 10 min at 4000 rpm. Prior to additional centrifugation for 10 min at 4000 rpm, 1ml ddH2O was added to the filter. The filtrate (50-100µL) was moved to a silica-filter (Qiagen), already added 600µL PB-buffer (Qiagen) and incubated for 30-60 min at room temperature. The filter was then spun for 1 min at 12.000pm and washed twice with PE-buffer (Qiagen). A final spin was applied to dry out the filter. The filter was moved to a new collection tubes and100µL TE-buffer added. After 5 min the extract was collected by centrifugation at 12.000 rpm for 1 min.

PCR Amplification

The NL16,230-NH16,262 (*S6*) and NL16263b (5´CAACTCCAAAGACGCCCTTA3´)-NH16,300 (5´ACTTTATGTGCTATGTACTGT3´) primer pairs were used to amplify two mtDNA control region fragments of 71 bp and 76 bp, in Copenhagen and Uppsala, respectively. Blank controls were incorporated at a ratio of 1 blank to every four extractions or PCR reactions and processed together with the sediment samples.

The 25 or 50µl PCR reactions contained: 1x PCR High Fidelity PCR Buffer or 1*Qiagen PCR buffer, 3.5 mM Magnesium Sulfate solution (Invitrogen) or 2.5 mM $MgCl_2$, 0.4 mM dNTP or 0.2 mM dNTP Mix, 1 U Platinum® Taq DNA Polymerase High Fidelity (Invitrogen) or 3 units of HotStarTaq DNA polymerase (Qiagen, Valencia, CA), 1 µM each primer and 1-5 µL DNA extracts. Cycling was performed in an Eppendorf Mastercycler gradient (Eppendorf) thermal cycler or GenAmp PCR system 9700 (Applied Biosystems) with the following cycle program: denaturation at 94 or 95^{o}C for 2 or 10 min followed by 40-

45 cycles of 94^{o}C for 30 s, 50-59^{o}C for 30 s and 68 or 72^{o}C for 30 s, followed by 6 or 7 min at 68 or72^{o}C.

Approximately 12.5 µl or 5 µl of the reactions was electrophoresed on 3% or 2% GTC agarose gels, stained with ethidium bromide, and visualized by UV transillumination. The bands of the correct size were cut out and purified using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's instructions eluting the product in 30µl of EB buffer.

Cloning and Sequencing

Cloning and sequencing of PCR products was performed by using the TOPO TA, Zero Blunt PCR Cloning Kits (Invitrogen), JM109 competent cells, and pGEM\-T Vector System II cloning kit (Promega). Selected clones were picked directly into in 100 µl sterile water and 2µl were used for the colony PCR (25 µl final volume) containing 1x PCR Buffer, 2.5 mM $MgCl_2$ solution, 0.2 mM dNTP Mix, 1 U Taq DNA Polymerase or AmpliTaq Gold (Applied Biosystems), 1 µM each of universal M13R and T7 primers. Cycle conditions consisted of an initial denaturation at 94 °C for 2 min or 10 min followed by 30 cycles of 94 ^{o}C for 20 or 30 s, 53 ^{o}C or 53 ^{o}C for 20 or 30 s and 72 ^{o}C for 20 or 30 s, followed by 5 min or 7 min at 72 ^{o}C. Colony PCR products carrying the expected insert length were screened by 2% agarose gel electrophoresis. The PCR products were purified using the GFX-96 PCR (Amersham), the Omega-Bio-teck (E-Z 96 cycle-pure kits) purification kits, or with ExoSAP-IT^{TM} (USB Corporation). Cycle sequencing was carried out using the BigDye Terminator Kit (Applied Biosystems) or with the DYEnamic^{TM} cycle sequencing kit (Amersham Biosciences) as recommended by the manufacturer. The sequencing products were analyzed by capillary electrophoresis on an ABI PRISM™ 3130 Genetic Analyzer (Applied Biosystems), or on a MegaBACE 1000^{TM} (Amersham Biosciences). Further sequence analyses were carried out

using the programs Seqman II™ and MegAlign™ from the DNA Star Software package (version 4.05). The Neandertal-like sequences obtained from the sediment extracts were never recorded in the blank controls. Each of the reported Neandertal-like sequences (Fig. S2) was reproducibly obtained from independent reactions.

The sequences obtained with the NL16,230-NH16,262 primer pair from three extractions and four independent amplifications was found to be similar to mtDNA of Neandertals displaying the Neandertal-specific haplotype 16,234[T]-16,244[A] (Fig. S2). This DNA haplotype is described in all Neandertal specimens to date, including that from El Sidrón. Intriguingly, the sequences are putatively truncated at position 16,252 by a PCR event, forming a chimaeric sequence of undetermined origin.

The sequences obtained with the NL16,263b-NH16,300 primer pair display the Neandertal-specific haplotype 16,278[T]-16,299[G] (Fig. S2).

Sequence identification and statistical analysis

To assign the Neandertal-like DNA sequences to phylogenetic groups we used a Bayesian approach. First, using database searches we identified sets of sequences closely related to each sequence (query sequence). The 50 best hits defined by E-scores of a BLAST search were chosen, eliminating identical copies, but including additional homologues if the initial set did not include at least three families and two orders, while allowing no homologues with an E-value over 10. Sequences with <5 BLAST hits with E-value < 0.1 were discarded as possible PCR artifacts or species from taxonomic groups that are not sufficiently well represented in the database. An alignment of the 50 non-redundant sequences was then produced using ClustalW and MrBayes (*S7*) was applied to the sequences using two independent runs of 1,000,000 updates, heated with four simultaneous chains using a GTR+G model. $k = 2000$ trees were sampled from the results and based on the taxonomic assignments of sequences in

RESULTS

Genbank, the probability of each query sequence forming a monophyletic group with sequences from a particular phylogenetic group was assessed based on the sampled trees. For example, the probability that a query sequence (Q) belongs to family F would be assessed as:

$$\Pr(Q \in F \mid X) = \int_{G \in \Omega} I(Q, F \text{ monophyletic in } G)\, p(G \mid X)\, dG \approx \sum_{i=1}^{k} I(Q, F \text{ monophyletic in } G_i)$$

where Ω is the set of all possible trees (G) with branch lengths and G_i is the ith tree sampled from the Markov chain simulated in MrBayes, X is all of the sequence data, and $I(Q, F \text{ monophyletic in } G)$ is an indicator returning 1 if Q and F are monophyletic in tree G and 0 otherwise. In general, we would not require all sequences belonging to F to form a monophyletic group as long as Q and some sequences in F form a monophyletic group. This inference procedure would be performed for each sequence independently for all possible taxonomic assignments from the level of order to the level of species of sub-species. There are several caveats to this method, the most important being that assignment can only be done to taxonomic groups represented in the database. If only one relevant species or sub-species is represented in the database, the query sequence will be assigned to this species with probability one.

To test if the 16,278[T]-16,299[G] sequence could be a contemporary human contaminant, we use a combination of the preciously described Bayesian approach for assignment in conjunction with a frequentist procedure based on extant human HVR sequences using the Bayes factor as a statistic. The Bayes factor in favour of the hypothesis of Neanderthal origin compared to human origin is calculated as the ratio of posterior to prior probabilities in favour of Neanderthal origin divided by the ratio of posterior to prior probabilities in favour of a human origin. Based on an alignment with 50 human sequences and 4 Neanderthal sequences, the prior for human origin is 99/105 and the prior for Neanderthal origin is 6/105. The posterior probability is 88% and 12% in favour of Neanderthal and human origin,

respectively, resulting in a Bayes factor of 131. Using the HvrBase++ database (*S8*), we then constructed 365 sequences of same length and from the same region as the new sequence. For each of these sequences, we also calculated Bayes factor in favour of Neanderthal origin using the exact same procedure as used for the Sidron sequence (Figure 1). The maximal value observed among the 365 replicates was 40, providing strong statistical evidence against the hypothesis that the new sequence is of human origin.

In contrast, the posterior probability that the truncated 16,234[T]-16,244[A] haplotype sequence (*S2*) groups with the Neandertal sequences is only 48%, corresponding to a Bayes factor of 15, due to other GenBank sequences carrying similar motifs. Thus, it cannot be discarded that this fragmentary sequence could derive from non-Neandertal sources.

References

1. J. Fortea *et al., Estud. Geol.* **59,** 159 (2003).
2. A. Cooper H. N. Poinar, *Science* **289,** 1139 (2001).
3. E. Willerslev A. Cooper, *Proc. Royl. Soc. Lon. B.* **272,** 3 (2005).
4. E. Willerslev *et al., Science* **300,** 791 (2003).
5. H. N. Poinar *et al., Science* **281,** 402 (1998).
6. D. Serre *et al., PLOS Biol.* **2**, 313 (2004)
7. J. P. Huelsenbeck, F. Ronquist. *Bioinformatics* **17,** 754 (2001).
8. J. Kohl, *et al., Nucl. Acid. Res.* **34**, D700 (2006).

Fig S1. Stratigraphic cut of El Sidrón cave in September 2004. Several soil samples (numbered 1 to 5) were obtained at different layers. Sample #5 was selected for DNA analysis because it was obtained next to an occipital fragment (green dots, *Occ.*).

RESULTS

Fig S2: Neandertal-like DNA sequences obtained directly from El Sidrón cave sediments (Sidrón) using the NL16,230-NH16,262 (1) and NL16,263b-NH16,300 (2) primer pairs. The clone sequences are aligned to the Cambridge Reference Sequence and to Neandertal sequences obtained from bones and teeth (Engis 2, Feldhofer 1,2, La Chapelle-aux-Saints, Mezmaiskaya, Monte Lessini, Rochers de Villeneuve, Scladina, Vindija 75,77, 80, Sidrón bone and Sidrón 441). The number of DNA extraction E, and PCR amplifications are shown prior to each of the sediment sequences.

1) NL16,230-NH16,262

```
                                  2         2           2 2
                                  3         4           5 5
                                  4         4           6 8
Cambridge Reference Sequence      TCACACATCAACTGCAACTCCAAAGCCACCC
Sidrón E3,PCR1,clone 1-7          ...T.........A........
Sidrón E3,PCR1,clone 8            ...G.........A........
Sidrón E3,PCR2,clone 1-2          ...T.........A........
Sidrón E4,PCR1,clone 1-18         ...T.........A........
Sidrón E1,PCR1,clone 1-6          ...T.........A........
Sidrón bone, Sidrón 441           ...T.........A...........A.G...
Vindija 75,77,80                  ...T.........A...........A.G...
Rochers de Villeneuve             ...T.........A...........A.G...
Feldhofer 1                       ...T.........A...........A.G...
Monte Lessini                     ...T.........A...........A.G...
Feldhofer 2                       ...T.........A...........A.....
Mezmaiskaya                       ...T.........A...........A.....
Engis 2                           ...T.........A...........A.....
La Chapelle-aux-Saints            ...T.........A...........A.....
Scladina                          ...T........CA...........A.....
```

2) NL16,263b-H16,300

```
                                              2                   2
                                              7                   9
                                              8                   9
Cambridge Reference Sequence      CACCCACTAGGATACCAACAAACCTACCCACCCTTA
Sidrón E3, PCR1, clone 1-3        ..............T...................G
Sidrón E3, PCR1, clone 4          T.............T...................G
Sidrón E3, PCR2, clone 1          ..............T...................G
Sidrón bone                       ..............T...................G
Vindija 75,80                     ..............T...................G
Feldhofer 1                       ..............T...................G
Monte Lessini                     ...................................
Feldhofer 2                       ..............T...................G
Mezmaiskaya                       ..............T...................G
Scladina                          ..T...........T............T.......G
```

RESULTS

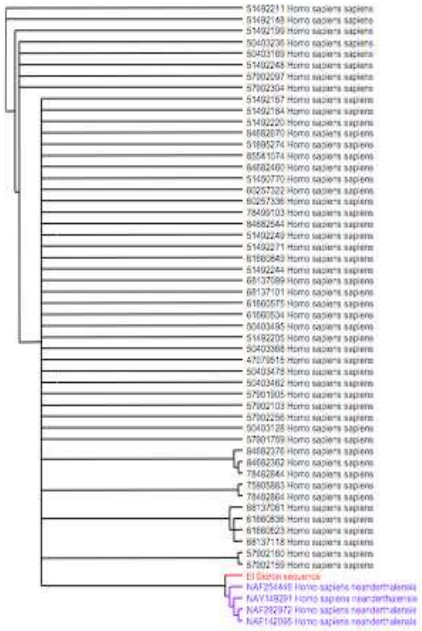


Fig. S4. Original strict majority rule consensus tree for the query sequence, Neanderthal sequences in GenBank, and the 50 best other Genbank hits (Modified version see Fig. 1).

4.3 CHAPTER 3: Neandertal evolutionary genetics: mitochondrial DNA data from the iberian peninsula

Lalueza-Fox, C., **Sampietro, M.L**., Caramelli, D., Puder, Y., Lari, M, Calafell, F., Martinez-Maza, C., Bastir, M., Fortea, J., de la rasilla, M., Bertranpetit, J. and Rosas, A.

Molecular Biology and Evolution. 2005 Apr;22(4):1077-81

4.4 CHAPTER 4:The mitochondrial hypervariable region I of an Iberian Neandertal suggests a population affinity with other European Neandertals

Carles Lalueza-Fox, Johannes Krause, David Caramelli, Giulio Catalano, Lucio Milani, María Lourdes Sampietro, Francesc Calafell, Cayetana Martínez-Maza, Markus Bastir, Antonio García-Tabernero, Marco de la Rasilla, Javier Fortea, Svante Pääbo, Jaume Bertranpetit, Antonio Rosas

Current Biology. 2006 Aug 22;16(16):R629-630

Supplemental data

Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neandertals

Carles Lalueza-Fox, Johannes Krause, David Caramelli, Giulio Catalano, Lucio Milani, María Lourdes Sampietro, Francesc Calafell, Cayetana Martínez-Maza, Markus Bastir, Antonio García-Tabernero, Marco de la Rasilla, Javier Fortea, Svante Pääbo, Jaume Bertranpetit, Antonio Rosas

Supplemental Experimental procedures

After excavation the bone sample from the El Sidron site was kept frozen at -20º and sent to the Max Planck Institute (Leipzig, Germany), where it was cleaned with sterile water in a laboratory dedicated to ancient DNA work. Three subsamples each around 0.5 g, were removed and used for genetic analyses in three different ancient DNA laboratories, at the University Pompeu Fabra (Barcelona, Spain), at the University of Florence (Florence, Italy) and the Max Planck Institute itself.

The stereoisomeric D/L ratios observed for three amino acids are: Asp (0.02±0.002), Ala (0.0036±0.0003) and Glu (0.008±0.003). The aspartic values are well within the proposed limit of 0.10 compatible with DNA preservation [S1].

DNA was extracted from about 0.5 g of bone in the ancient DNA laboratories of Barcelona, Leipzig and Florence. With specific variations, the sample was powdered, incubated with a proteinase K lysis buffer and extracted with phenol chlorophorm (Barcelona)[S2] or silica extraction (Leipzig and Florence)[S3] previous to concentrate it by column centrifugation. To generate a consensus HVR1 mtDNA sequence a two-step multiplex PCR protocol [S3] was used in Barcelona and Leipzig. Both steps

included 2 U AmpliTaq Gold (ABI, USA), 1X AmpliTaq Gold buffer (ABI, USA), 4 mM $MgCl_2$ (ABI, USA), 500 µM for each dNTP and 150 µM of each primer in the first multiplex step and 1.5 µM of each primer in the second step in a final volume of 20 µl. In the multiplex step up to 6 primer pairs were used in one reaction. Primary amplification consisted in a 10 min activation step at 94°C, followed by 27 cycles at 94°C for 20 s, 50º-55°C for 30 s, and 72°C for 30 s. In the second simplex PCR step just one primer pair was used. Conditions were as described for the multiplex step, except that the primer concentration was increased to 1.5 µM for each primer and that 33 cycles were performed. Five µl of a 1 to 10 (1:40) dilution of the primary amplification product were used as a template for the simplex PCR. After visualization of the products on 1% agarose gels, amplification products of the correct size were excised from the gel. DNA was purified with a gene clean silica method and cloned using the Topo TA cloning kit (Invitrogen, The Netherlands). Colonies were subjected to PCR with M13 universal primers; inserts with the right size were sequenced with an Applied BioSystems 3100 DNA sequencer. In Florence, the PCR conditions were based on a single amplification of 60 cycles, as described previously [S2].

Preliminary tests with Neandertal specific primers showed that the endogenous DNA was degraded to around <80 bp; after this, primers were designed to amplify 70 to 80 bp fragments. The specific fragments (those with primers that matched Neandertal specific positions, such as L16,135-H16,169; L16,161-H16,191; L16,182-H16,223; L16,220-H16,246; L16,244-H16,278; L16263b-H16,301 and L16,299-H16,320) yielded a high ratio of Neandertal versus contaminant sequences up to 100%. However, the figures were substantially lower in less specific fragments, especially those that had a complete match to modern human sequences; in these fragments, the ratio of Neandertal versus contaminant sequences ranged from only 5% in L16,310-H16,350 and 10% in L16,076-

H16,110 to 33% in L16,109-H16,140, 42% in L16,347-H16,378 and 60% in 16,319-H16,350. Overall, in can be concluded that the precautions undertaken during the excavation helped in the retrieval of genetic data from this specimen.

In total fourteen overlapping fragments were amplified to generate the HVR1 mtDNA El Sidrón consensus sequence (Figure S1). When possible, primer pairs were designed to obtain Neandertal specific haplotypes or Neandertal specific substitutions (such as the G in position 16,078, not described in modern Europeans). In the few cases were this was not possible (L16,299-H16,320, L16,319-H16,350, and L16,347-H16,378 fragments), the substitutions found are concordant with previously described Neandertal sequences. This, along with the low contamination ratio, supports the authenticity of the sequences, even in such short fragments.

Supplemental Date Estimates

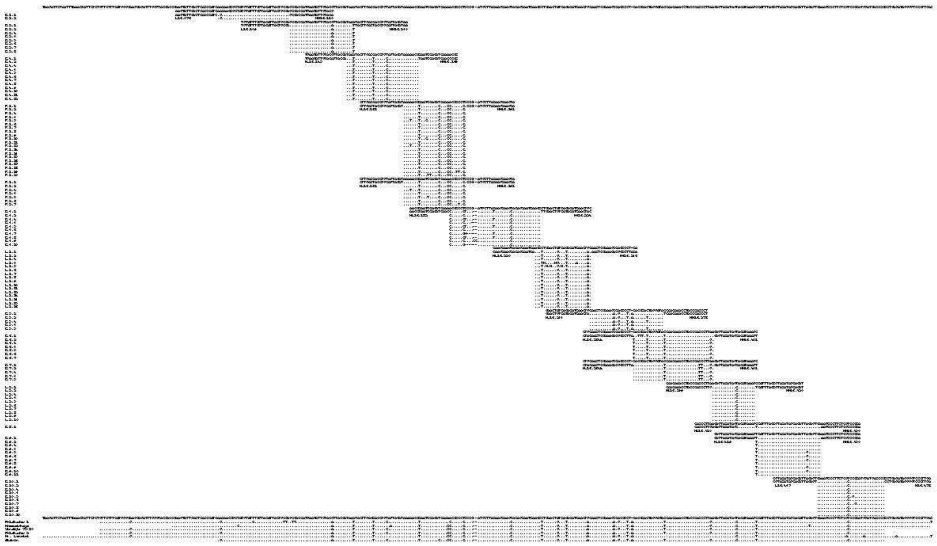
The coalescence time of the mtDNA Neandertal variation and the specific Neandertal subclade (G16,078-C16,154 haplotype) was calculated using the approach implemented in GeneTree[S4] for the mtDNA HVR1 between positions 16,076 and 16,378 (N = 7 Neandertals). Calculations were performed assuming constant population size, a figure of 10,000 for *Ne* and 20 year generation time, with 100,000,000 iterations used to estimate the $\theta = N_e\mu$ parameter.

Supplemental References

S1. Poinar, H.N., Hoss, M., Bada, J.L., Pääbo, S. (1996). Amino acid racemization and the preservation of ancient DNA. Science 272, 864-866.

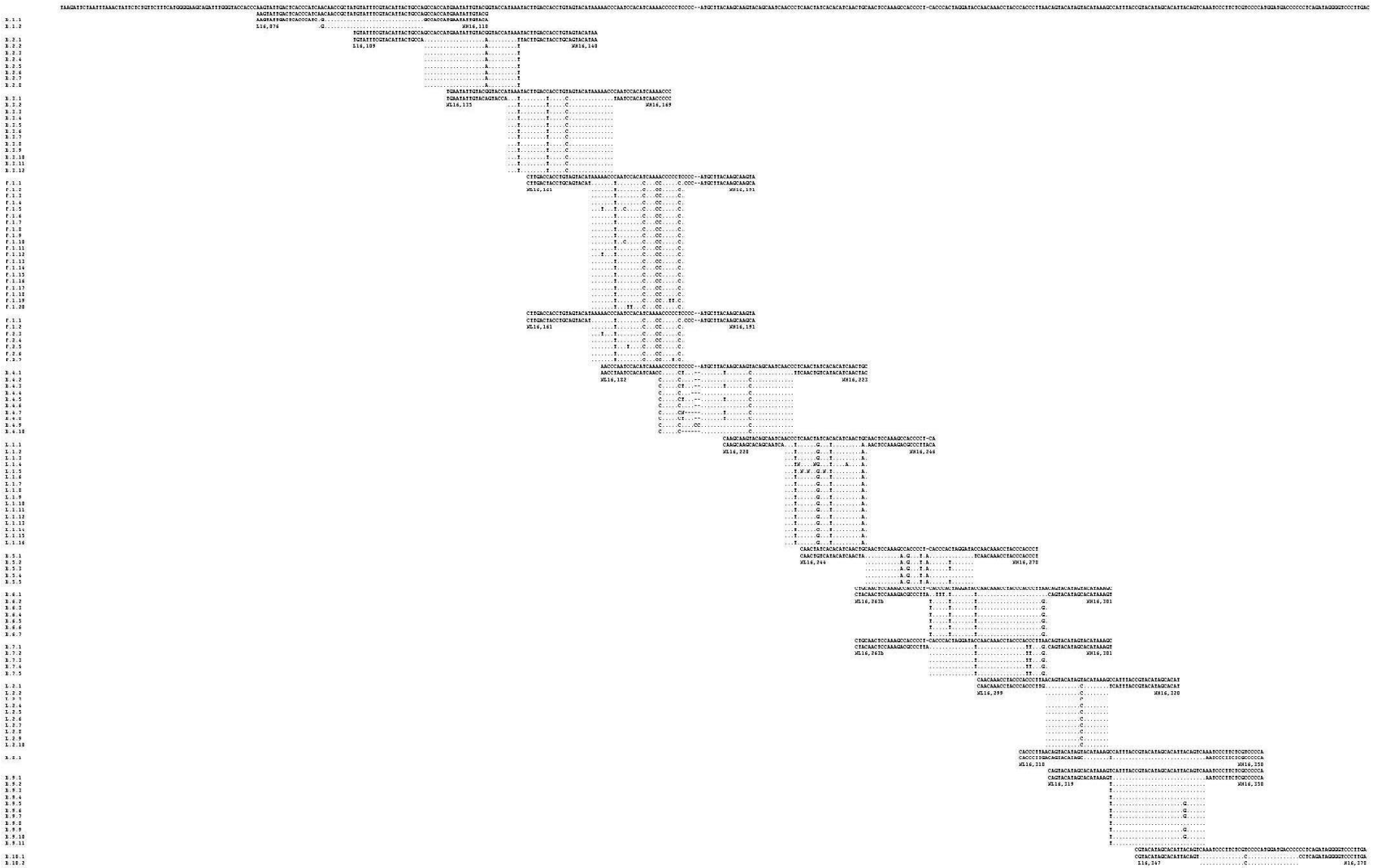
S2. Caramelli, D., Lalueza-Fox, C., Condemi, S., Longo, L., Milani, L., Manfredini, A., Saint Pierre, M. de., Adoni, F., Lari, M., Giunti, P., et al. (2006). A highly divergent mtDNA sequence in a Neandertal individual from Italy. Curr. Biol. 16, .xxx-xxx.

S3. Krause, J., Dear, P.H., Pollack, J.L., Slatkin, M., Spriggs, H., Barnes, I., Lister, A.M., Ebergsberger, I., Pääbo, S., and Hofreiter, M. (2005). Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. Nature 439,724-727.
Griffiths, R. C., and S. Tavaré. 1994. Ancestral inference in population genetics. Stat. Sci. 9, 307–319.



Supplemental Figure 1: Clones used to generate the HVR1 mtDNA El Sidrón consensus sequence as compared to the Cambridge Reference Sequence. B: means Barcelona, F: Florence, L: Leipzig. The second digit corresponds to the number of amplification; the third to the number of clone. For clarity reasons, only <15 clones are included for some fragments.

RESULTS



4.5 CHAPTER 5: Mitochondrial DNA from a late Neolithic site supports a long term genetic continuity in the Iberian Peninsula.

Sampietro, M.L., Lao, O., Caramelli, D., Lari, M., Pou, R., Martí, M., Bertranpetit, J., and Lalueza-Fox, C.

(Manuscript in preparation)

Mitochondrial DNA from a late Neolithic site supports a long-term genetic continuity in the Iberian Peninsula

Sampietro, M.L., Lao, O., Caramelli, D., Lari, M., Pou, R., Martí, M., Bertranpetit, J., and Lalueza-Fox, C.

Departament de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Dr. Aiguader 80, 08003 Barcelona, Spain

Laboratory of Anthropology,Department of Animal Biology and Genetics, University of Florence, via del Proconsolo 12, 50122 Florence, Italy

Unitat d'Antropologia, Departament de Biologia Animal, Facultat de Biologia, Universitat de Barcelona, Avda. Diagonal 645, 08028 Barcelona, Spain

INTRODUCTION

The main feature of the European genetic diversity is a clinal pattern that must reflect a population movement from the Southeast to the Northwest, with a significant demographic impact (Cavalli-Sforza). There are currently two processes in the demographic and evolutionary history of Europe, as documented from the archaeological record, that can account for such a cline: the Paleolithic colonisation of Europe (starting around 40,000 YBP) and the Neolithic agricultural diffusion (starting around 10,000 YBP) (Barbujani and Goldstein, 2004). Unfortunately, clines do not have dates associated, and as both population movements followed the same axis along Europe, the attribution of the genetic cline to either process is not straightforward. Different authors have tried to distinguish between these two hypothesis from the analysis of genetic data in current European populations but the conclusions obtained are contradictory (Barbujani, Simoni, Torroni, Richards,,,).

The arrival of agriculture to Europe is also a matter of debate among scientists (Diamond and Bellwood 2003). Two main hypotheses have been proposed to explain the spread of the agriculture through Europe: the *cultural diffusion model* and *the demic diffusion model*. The cultural diffusion model supports the idea that the farmers did not move and the agricultural knowledge was transmitted from the Near East to Europe through the movement of technology and ideas (Whittle 1996). In contrast, the demic diffusion model (most commonly known as the *wave of advance*) supports the idea that the farmers from the Near East spread to Europe taking agricultural knowledge with them; thus, this model involves gene flow between the hunter-gatherers that inhabited Europe at that moment and the farmers that arrived (Ammerman and Cavalli-Sforza 1984).

RESULTS

Interestingly, ancient DNA data could potentially be of great interest to resolve between both hypothesis since it allows us to directly study the ancient populations that were undergoing these evolutionary processes and not their descendent populations. Recently, Haak et al (2005) successfully extracted and sequenced the HVR I of the mtDNA from 24 out of 57 Neolithic skeletons from various locations in Germany, Austria and Hungary. All human remains were dated to the LBK or AVK period (7000 to 7500 years ago). They found that 25% (6 out of 24) of the samples are of a distinctive and rare N1a lineage of the mtDNA well-known phylogeny. Furthermore, five of these six individuals display different N1a haplotypes and they were widespread in the LBK area. Europeans today have a 150-times lower frequency (0.2%) of this mtDNA type, revealing that these first Neolithic did not have a strong impact in the genetic background of the modern European female lineages. They proposed that small pioneer farming groups carried farming into new areas of Europe, and that once the technique had taken root; the surrounding hunter-gatherers adopted the new culture and then outnumbered the original farmers, diluting their N1a frequency to the low modern value. Thus, this result supports the cultural diffusion model, where the farming culture itself spread without the people originally carrying these ideas. They proposed that within the current debate on whether Europeans are genetically of Palaeolithic or Neolithic origin, and leaving aside the possibility of significant post-Neolithic migration, their data lend weight to the arguments for a Palaeolithic origin of Europeans.

The results of Haak et al. (2005) have been critisized by Ammerman et al. (2006) because, among other things, of the limited sample size, the uniparental mode of inheritance of the mitochondrial (mt) DNA and the generalisation of the results to the whole of Europe. Moreover, ancient DNA studies on human samples, specially ancient Europeans, have also been put into question because of the impossibility of distinguishing between potential contaminants and endogenous sequences. Nevertheless, such a problem could be overwhelmed if information of all the putative contaminants present in a particular sample set is available. In this paper a Neolithic population from Southern Europe (Granollers, Catalunya, Northeast of Spain) has been subjected to aDNA genetic analysis, previous to the typing of all people involved in the manipulation of the samples. The putative endogenous sequences obtained do not match those found by Haak et al. (2005) in a sample from Central Europe and raise new questions on the heterogeneity of the Neolithic dispersal.

MATERIALS AND METHODS

The site “Camí de Can Grau” (Granollers, Barcelona, Spain) is a necropolis excavated in 1994, that comprised 23 tombs dated by C14 between 3,500-3,000 cal years B.C. There were two different funerary typologies with separate geographic locations that corresponded to different periods, spanning several hundreds of years; the older tombs formed squared sepulchral chambers, while the younger ones were hypogean tombs with an access through a vertical well.

A tooth sample was removed from 23 adult individuals for DNA analysis, with the exception of a toothless specimen, from which a bone fragment was obtained. In some specimens, a second tooth was removed for independent replication in Florence.

Standard methodological precautions were followed, to provide as much as possible support for the authenticity of the results. The contamination by handling has been recognized as a major problem facing ancient DNA studies that focuss on ancient human remains, specially when researchers and remains are from the same geogaphic area. The authentication criteria proposed by different authors (e.g. Cooper and Poinar 2000) can help in preventing putative intralaboratory contamination, but it is impossible to directly monitor pre-laboratory contamination.

However, what makes Can Grau an exceptional site is that its pre-laboratory history is perfectly recorded; the remains were excavated, handled and washed by the archaeologists R.P. and M.M.; once dried, they were reconstructed and studied by a physical anthropologist (E. Vives) and posteriorly stored in closed plastic boxes for about ten years in a local museum until the genetic study was attempted. By typing the mitochondrial DNA of all the people involved in the manipulation of the skeletal remains and the laboratory analysis (M.L.S., C.L.-F. and D.C.), we have been able to trace all contaminants present in our samples (Sampietro et al. 2006). As far as we know it is the first time in the recent history of aDNA research where it has been possible to control the putatively pre-laboratory derived contaminant DNA sequences and consequently to eliminate them of the cloning dataset. Thus and under such circumstances, the Neolithic remains that we subjected to the genetic analysis were unique.

DNA extraction

The surface of each sample was cleaned with bleach and then ground to powder. The extraction method has been described elsewhere (e.g. Sampietro et al 2006). One extraction blank was included every three Neolithic samples. In brief, ten millilitres of EDTA (pH:8; 0.5M) were added to the powder overnight at 37ºC to remove mineral salts; after centrifugation, the EDTA was carefully poured off and the powder was incubated overnight at 50ºC in a lysis solution (1ml SDS 5%, 0.5ml TRIS 1M, 8.5ml H_2O and loads of proteinase K). Then the samples were extracted three times with phenol, phenol-chloroform and chloroform-isoamilic alcohol and concentrated with centricons (Millipore) up to a 50-100µl volume.

Extraction procedures were carried out in an isolated pre-PCR area exclusively dedicated to ancient DNA studies, physically isolated from the main laboratory, with positive air pressure, overnight UV light and frequent bench cleaning with bleach. All samples and reagent manipulation were performed in a laminar flow cabinet routinely irradiated with UV light. To help avoid intralaboratory contaminations, aliquoted reagents, filter pipette tips, sterile gloves, sterile pipettes, facemasks and cover-all coats were used.

Amplification, Cloning and Sequencing

The mtDNA HVR 1 region (Anderson et al, 1981) was amplified in 21 Neolithic samples in different overlapping fragments with sizes ranging from 98 to 212 bp combining several primers pairs (Table I). In addition, some additional primers pairs were used to amplify mtDNA coding regions where diagnostic SNPs that define unequivocally an haplogroup in the mtDNA genealogy are located (Table I). PCR amplifications were performed in 25µl reactions with 1µl to 5 µl of extract (some

extracts were subjected to 1:3 dilution in order to overcome inhibitors), 1.2 U of taq polymerase (Ecogen), 1X reaction buffer (Ecogen), 1.4 mg/ml BSA, 2.1mM $MgCl_2$, 0.2mM dNTP´s and 1μM of each primer. The PCR reactions where subjected to 40 amplification cycles (1min step at 94ºC, 1min step at 50ºC and 1min step at 72ºC) with an initial denaturing step at 94ºC for 2min and a final elongating step for 7min at 72ºC.

PCR products were electrophoresed in 1.6% low-melting point agarose gels (Invitrogen) stained with ethidium bromide. Bands with the expected correct size were excised from the gel, diluted in 150μl of double destilled water, heated 1hour at 65ºC and subsequently subjected to another 35 cycles of PCR with limited reagents conditions. Resulting bands were purified with GFX columns (Amersham Biosciences) and routinatelly cloned using pMOS blue blunt ended cloning kit (Amersham Biosciences) following the manufacture´s instruction. In brief, seven microlitres of PCR product were treated with pK enzyme mix, incubated al 22ºC for 40 min and ligated into pMOS*Blue* vector overnight. 2 μl of the ligation product were transformed into 40 μl of competent cells, grown in 160 μl of SOC medium at 37ºC during one hour and plated on IPTG/X-gal agar plates. After 16 hours, white colonies were subjected to direct PCR screening using T7 and U-19 universal primers. Inserts that yielded the correct size were identified by agarose gel electrophoresis, purified and sequenced with an ABI 3100® DNA sequencer (Applied Biosystems), following the supplier's instructions.

Uracil-N-Glycosylase (UNG) treatment

Hydrolytic deamination of cytosines causes uracil residues that are incorrectly read by the polymerase, resulting in false C→T/ G→A changes in the clone sequences (Hofreiter et al, 2001); this is the most common form of post-mortem damage in ancient DNA sequences (Stiller et al 2006). In those clone sequences where these C→T/ G→A substitutions were observed, a UNG treatment was followed in order to eliminate possible miscoding lesions. Ten microliters of DNA-extract were treated with 1U of UNG during thirty minutes at 37ºC to excise uracil residues in the original template (Hofreiter et al, 2001). After this treatment, extracts were subjected to the same PCR´s amplifications described above and subsequently cloned.

Quantitation of the templates molecules

To quantify if the number of mtDNA molecules is large enough to obtain reproducible results a real time PCR experiment were performed in several samples. We used the same quantify standard DNA, primers and probes that we used in Sampietro et al. (2005). However, it is obvious that, in cases of human contaminated samples, the quantitation results cannot discriminate between endogenous and contaminant sequences.

RESULTS

23 Neolithic remains were analysed; two samples yielded no amplification products and subsequently were discarded, nine samples were as well discarded due to the irreproducible or fragmentary results, and four more samples could not be unambiguously attributed to one of the main European mitochondrial DNA lineages.

The remaining eleven sequences were considered to be endogenous and included in the posterior population analysis.

A total of 572 clones were sequenced, from which 98 (17.13%) could be identified as being from 1 of the only 6 people involved in the manipulation and laboratory analysis of the Neolithic remains (Sampietro et al. 2006). Since we were able to monitor all the persons who had ever had access to this set of samples, it was possible to track down the pre-laboratory derived contaminant sequences and consequently we could definitely eliminate them from the generated cloning data set.

However, we faced some problems in particular situations, related to the impossibility of working with long DNA fragments. For instance, two Neolithic samples display the haplotype 069T, 126C and two out of the six handlers also have that haplotype, albeit only in the fist part of the HVR1 sequence (M.L.S has the haplotype 069T, 126C, 185T, 189C and R.P. 069T, 126C, 278T, 366T). Therefore, 069T, 126C could potentially be a contaminant, but two things allow us to consider these sequences as endogenous: they reach frequencies up to 100%? Of the clones in the amplified 055-218? CHECK fragment, while up to 90%?? Of the clones for the second half of the HVR1 are CRS. Therefore, the alternative hypothesis that the first fragment was totally contaminated while the second one was almost free of contaminants seems less plausible. Another situation is to consider the problem that contaminant sequences that have no substitutions in particular fragments will result in a background of CRS sequences (for instance, M.M. has the 129A haplotype and accounts for 20.41% of detected contaminant sequences will undoubtedly result in CRS sequences in the second half of the HVR1 fragments, while E.V. has the 298C haplotype and will result in CRS sequences in the first HVR1 half). Therefore, some fragments display a rather high level of CRS sequences that are likely this unspecific contaminant background; however, the putative endogenous sequences share some characteristics, such as to be reproducible, to be in many cases exclusive of a particular sample and to be present in higher frequencies than the distinguishable contaminants. Moreover, three Neolithic samples have probably CRS as endogenous haplotype, since these sequences are overhelmingly majoritary and present in higher frequency than the detected contaminants in the other Neolithic samples.

Neolithic haplotype sequences

Neolithic sequences showed haplotypes widely distributed through Europe when comparing them to a haplotype dataset composed by more than 10,000 individuals from Europe and the Middle East (see table 1). Interestingly, one of them carries one haplotype (223T, 292T, 295T, 304C) that is only found in the Middle East while other two Neolithic samples display haplotypes that are only found in the Iberian peninsula (264T 270T 311C 319A) and Italy (126C, 140C, 189C, 294, 296T, 311C**)**. Nevertheless, there were two samples that show a particular haplotype never found in the dataset: one Neolithic individual is assigned to the haplogroup T but the haplotype(126C, 140C, 294, 296T, 311C)does not show all the mutations that are currently found within the genetic diversity in the haplogroup T2. Another Neolithic sample display and haplotype (134T) that although belongs to the haplogroup U that position has never found alone but always together with other positions.

RESULTS

Neolithic haplogroup

A correspondence analysis were performed taking into account modern population from the Middle East, from south west Europe and from the Iberian peninsula together with the ancient Iberian Neolithic population. The correspondence analysis shows that the Neolithic population clusters together with the modern southwest Europe populations and with the Iberian population but not with the Middle East population (see figure1).
This result shows that the haplogroup composition of the Iberian Neolithic population is quite similar to the actual population in the Iberian Peninsula. This genetic continuity between the ancient population and the current populations suggested by these results contrasts with these found by Haak et al (2005).

Discussion

Haak et al when analysing an older Neolithic population (7000-7,500 years BP) from Central Europe found that 25% (6 out of 24) of the samples were of a distinctive and rare N1a lineage (currently present at 0.2% in the European population) of the mtDNA well-known phylogeny.
Discarding the possibility of a possible lost of this mtDNA lineages by genetic drift over the last 7500 years by means of demographic models, they proposed that small pioneer farming groups carried farming into new areas of Europe, and therefore the dispersal of the agricultural techniques were through a cultural model; so that, at the end of the process, the frequency of the N1a haplogroup was diluted to the low modern value that its observed today.
The absence of sequences carrying the N1a haplogroup in the Iberian Neolithic population could be due to the difference of time (approximately 2000 years) and geographic distance (North versus South Europe) with these analysed by Haak et al. However, since the absence of N1a lineage cannot be explained by genetic drift, other hypotheses have to be invoked in order to explain such a discrepancy between their results and the observed in the current work.

Therefore, we hypothesise that the dispersal of agriculture involved both demographic and cultural diffusion, depending on the region where it took place. Whereas the dispersal of the agricultural in the North of Europe could be a cultural diffusion model, in the south our results suggest a demic diffusion model. But, if that hypothesis is true, we should observe a “barrera” in the genetic diversity between North and South of Europe.

To test this hypothesis we should have analysed a paleolithic population from the Iberian peninsula. If that Palaeolithic population is quite similar to the Neolithic population therefore, the continuity had had been since paleolithic times. This result would support haak result.
Thus, in order to totally clarify whether the actual European populations are direct descendents of the original settlers that replaced the Neandertals or the result of posterior immigrants from the Near East, the analysis of more ancient Neolithic remains from different European regions are definitively needed. In addition, the analysis of paleolithic population of different regions would be of great relevance.

bibliography

Albert J. Ammerman, Ron Pinhasi, Eszter Bánffy (2006) Comment on "Ancient DNA from the First European Farmers in 7500-Year-Old Neolithic Sites" *Science*:Vol. 312. no. 5782, p. 1875

G. Barbujani, D. B. Goldstein, *Annu. Rev. Genomics Hum. Genet.* **5**, 119 (2004).

L. Chikhi, G. Destro-Bisol, G. Bertorelle, V. Pascali, G. Barbujani, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9053 (1998).

L. Chikhi, R. A. Nichols, G. Barbujani, M. A. Beaumont, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11008 (2002).

M. Richards *et al., Am.J. Hum. Genet.* **67**, 1251 (2000).

I. Dupanloup, G. Bertorelle, L. Chikhi, G. Barbujani, *Mol. Biol. Evol.* **21**, 1361 (2004).

O. Semino *et al., Science* **290**, 1155 (2000).

L. Simoni, F. Calafell, D. Pettener, J. Bertranpetit, G. Barbujani, *Am. J. Hum. Genet.* **66**, 262 (2000).

A. Torroni *et al., Am. J. Hum. Genet.* **62**, 1137 (1998)

G. Barbujani, G. Bertorelle, L. Chikhi, *Am. J. Hum. Genet.* **62**, 488 (1998).

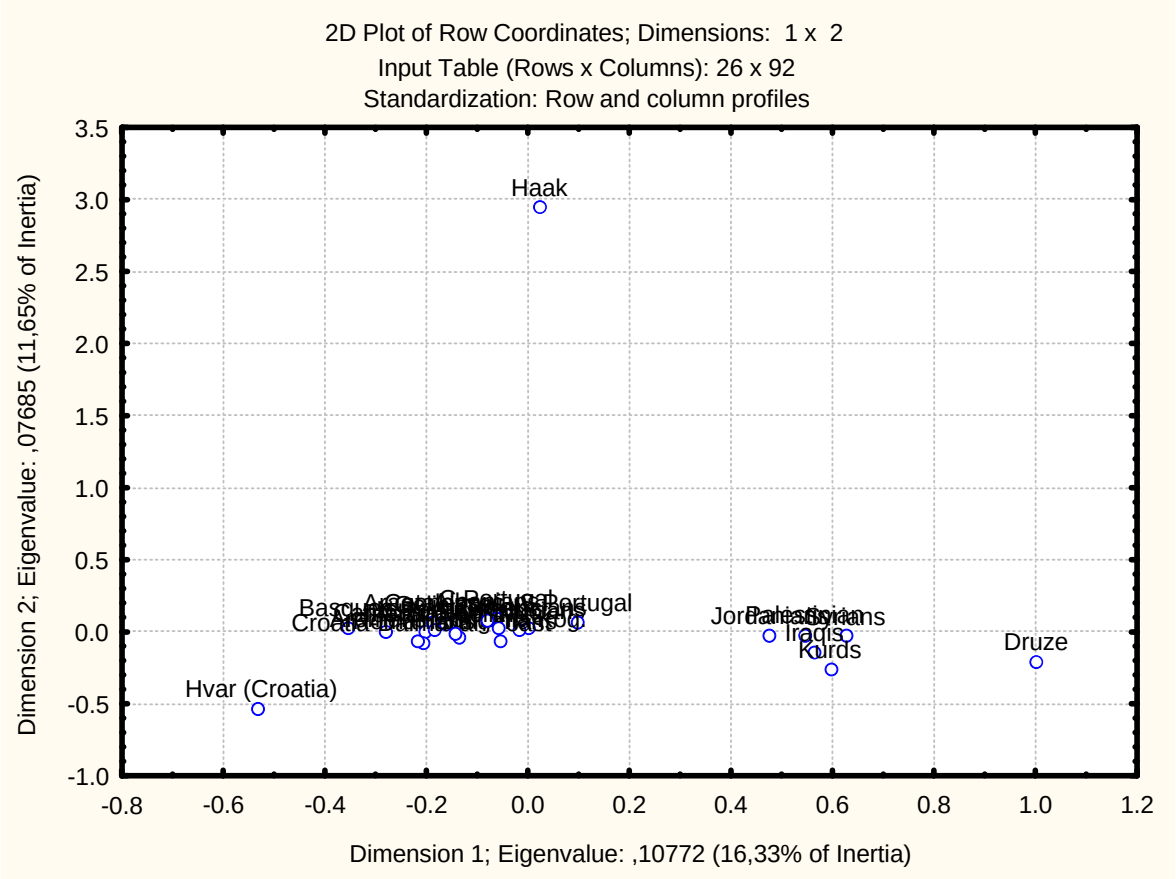
J. Diamond, P. Bellwood, *Science* **300**, 597 (2003).

A. J. Ammermann, L. Cavalli-Sforza, *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, Princeton, NJ, 1984).

Whittle, *Europe in the Neolithic. The Creation of New Worlds* (Cambridge Univ. Press, Cambridge, 1996).

RESULTS

A. Figure1 Corresponde analysis between populations



Neolithic	Haplotype	Coding region	haplogroup
Neolithic 1	CRS	6999-7066; 1 clon H and 3 clon no H	H
Neolithic 5	264T 270T 311C 319A	It didn't work the cloning	I1
Neolithic 6	CRS	I didn't clone	H
Neolithic 8	069T 126C	I didn't clone	J1c
Neolithic 10	223T, 292T,295T,304C	Only one clone but it isn't W	W1
Neolithic 11	288C 362C	I didn't clone yet	H
Neolithic 12	CRS	I didn't clone	H
Neolithic 14	069T 126C	I didn't clone	J1c
Neolithic 21	126C, 140C, 294T, 296T, 311C	13257-13372; 3clon T and 1 clon noT	T2
Neolithic 22	126C, 140C, 189C,294, 296T, 311C	13257-13372; 3clon T and 3 clon noT	T2
Neoltihic 23	134T	12227-12341; 2clon U and 4clon noU	U4

Table1: Haplotypes and haplogroup of each of the neolithic samples.

4.6 CHAPTER 6: The genetics of the pre-Roman Iberian Peninsula: a mtDNA study of ancient Iberians.

Sampietro, M.L., Caramelli, D., Lao, O., Calafell, F., Comas, D., Lari, M., Agustí, B., Bertranpetit, J., and Lalueza-Fox, C.

Annals of Human Genetics 2005. Sep;69(5):535-48.

5DISCUSSION

In the development of the current thesis we have addressed two different although very related topics. First, we studied the post-mortem mutation damage rate of contaminated sequences in ancient human remains and focusing on the development of strategies that can avoid pre-laboratory derived contaminations. Second, we analysed ancient human remains from different evolutionary times (from Paleolithic to post-Neolithic period) in order to make inferences about the peopling of Western Europe.

5.1 PRE-LABORATORY-DERIVED CONTAMINATIONS IN ANCIENT HUMAN TEETH

The importance of authenticating the obtained results when trying to retrieve aDNA from ancient tissues is well known for the aDNA community. At the beginning of the discipline, several revolutionary scientific discoveries were proven to be false due to the presence of contaminants (see chapter4.1). Since then, lots of efforts have been focused on proposing list of authentication criteria in order to validate the obtained results when working with ancient tissues. These criteria are addressed mainly to prevent sample contamination during the DNA extraction and amplification procedures but they do not contemplate the possibility that the ancient samples could have been impregnated with exogenous although very similar sources of DNA before the arrival to the genetic laboratory (pre-laboratory contaminants).

Particularly, human pre-laboratory contaminates has been reported in a number of studies (Richards, 1995, Hofreiter et al., 2001b, Malmstrom et al., 2005). In addition, in other situations, human pre-laboratory contaminant DNA sequences have been unambiguously identified in Neandertal remains to belong to some of the anthropologist or archaeologist that previously had studied the remains (Serre et al., 2004, Lalueza-Fox et al., 2005b). Furthermore, Gilbert et al (2005b, 2005a) demonstrated that both teeth and bone are readily contaminated presumably through handling and washing when excavating them and once contaminated in this way, both are difficult if not impossible to decontaminate when using routine decontamination protocols like surface bleach and UV-light bathing.

However, and despite these evidences, little is known about this complex (although fundamental) area, particularly when pre-laboratory derived contamination and the endogenous sequences could be undistinguishable. This is the case when working with ancient human remains that has been excavated by archaeologists under uncontrolled conditions. We have attempted to address this issue through the extraction, PCR amplification and cloning mtDNA HVR I sequences from twenty-three human Neolithic remains. These samples were of particular interest for such a kind of study, since we were able to monitor all the persons (archaeologists, anthropologist and geneticists) who had ever had access to this set of human samples. Therefore, we had the opportunity to explore deeper the phenomenon of pre-laboratory derived contamination in ancient humans teeth through the genetic analysis of 572 clones generated from twenty-three Neolithic remains. The contaminants track was monitored through typing all those persons. We followed all the recommended authentication criteria in aDNA studies (Cooper and Poinar, 2000, Paabo et al., 2004) to generate the clones in order to asses whether by their implementation we were able to avoid or detect pre-laboratory contaminations.

From the 572 clones sequences, 98 (17.13%) were definitively identified as being contaminant DNA sequences derived from one of the six handlers who had ever had access to the samples. Since, the HVR I was amplified through overlapping primers and the capability of distinguishing between an endogenous sequence and a contaminant sequence is dependent on the number of nucleotides that both sequences share, it was unavoidable that the detection of contaminant sequences will be differentially underestimated in some fragments depending on the studied contaminant. Based on the number of detected number of contaminants and the similarities between endogenous and contaminant sequences we computed the total expected number of contaminant sequences. Our results showed that this number derived from those people who were involved in the initial washing and cleaning of the remains are represented at a statistically higher frequency than the expected frequency of contaminant DNA frequency derived from the others participants. Consequently, the current decontamination protocols do not eliminate all the putatively exogenous DNA molecules and so that they could remain in somewhere inside the teeth and together with the endogenous DNA. This implies that it will be difficult if not impossible to differentiate between endogenous and contaminant sequences if human remains and archaeologists and/or anthropologists have a similar mtDNA genetic background.

Furthermore, it proves that human teeth remains are most susceptible to contamination at initial excavation and washing and that this is a critical step for contaminating the samples, rather than when performing the genetic analysis. Therefore these results add more proof to the inadequacy of current methods used to ensure the generation of authentic ancient DNA retrieved from human teeth. Naturally the specific results of this study will vary from those derived from alternative datasets, as no two datasets will have undergone exactly the same handling treatment. Nevertheless, although our results are dependent on factors such as sample preservation and the extent of sample manipulation at excavation and during subsequent anthropological analyses, it is clear that more research is needed in order to avoid pre-laboratory derived contamination. This research should be focused mainly on studying the exact mechanism of how the exogenous DNA can enter to the teeth by handling. Such knowledge could allow us to prevent these kinds of contamination, to develope new protocols in order to decontaminate better ancient human samples, or even to differentiate between contaminants and endogenous molecules and extracting only the endogenous DNA molecules by means of more efficient and specific DNA extraction protocols. In addition, ancient DNA researchers should pay more attention to the possibility that the human remains that they are analysing could be contaminated prior to the arrival to the genetic laboratory. This is especially important when European researchers excavate European human remains, since it could be impossible to differentiate between both endogenous and exogenous DNA sources.

The results of this work allow us to suggest a potential guideline to control pre-laboratory contaminants both when the samples have already been excavated under uncontrolled conditions (a very common situation since museums are in fact storehouses of fossil remains) as well as when the samples are freshly excavated (the best case scenario). In the former situation, the guideline consists in typing every single person involved on the manipulation of the remains, especially when the remains have not been excavated and washed under controlled conditions. In the latter situation, a possible way to study ancient human remains would be to excavate them under strictly controlled conditions, including: the use of sterile gloves, face masks and coveralls; the placement of excavated samples intended for later DNA analyses in sterile, sealed DNA-free containers; the avoidance of any sample washing, or if washing cannot be avoided, this should be undertaken with sterile water under controlled conditions.

Moreover, as the manipulations were made over a ten years period prior to the current genetic study, we were also able to test whether subsequent degradation of the contaminants has produced any sequence modifications, something that has previously been postulated (Willerslev and Cooper, 2005), although not definitively observed. In addition, since it has spent ten years since the samples were excavated till the genetic analysis were performed, we could compare whether both sources of contamination (we named “old contaminants” those sequences that belong to the archaeologists or to the anthropologist and “new contaminants” those sequences that belongs to geneticists) display the same amount of DNA damage (if any).

When studying the contaminant sequences and after classifying them into “old” and “new” contaminants we showed that these sequences can undergo observable levels of miscoding lesion damage post handling; in particular, we found high levels of transitions. Furthermore, we showed that the damage level found in “old” (approximately 10 years old) contaminant sequences was greater than that in “new” contaminants, thus demonstrating a time dependent occurrence. In addition, the level of damage found in the believed endogenous DNA sequences (in this case we considered only those sequences that we were sure that were endogenous) was indistinguishable from those in the “old” contaminant sequences. Therefore, we could correlate this higher damage level found in the “old” contaminants both with the time and with the temperature.

These findings are extremely important, as one commonly used argument in ancient DNA studies for data authenticity is the presence of such damage, following the reasoning that as damage accumulates roughly with time, authentic, thus old, sequences will be damaged, while new contaminants will remain undamaged. Clearly the results presented in this study demonstrates the above argument to be flawed and as such has to be used with caution due to not only the endogenous sequences can undergo damage but also those “exogenous” DNA contaminants that have been penetrated into the teeth several years ago can undergo damage as well.

Thus, it would be a very valuable and helpful information to know where and in which environmental conditions (temperature, for instance) the remains have been kept once excavated. This information could help us to decipher how likely is that those putatively modern contaminant sequences that are now inside the ancient teeth could display the same level of miscoding lesions than the endogenous ones.

5.2 HUMAN POPULATION HISTORY IN WESTERN EUROPE

Humans are a young species in evolutionary terms (Carroll, 2003, Paabo, 2003). Nevertheless, making inferences about our evolutionary history is quite difficult because it involves a large number of migratory events through all around the word since the first early anatomically modern humans left Africa (Cavalli-Sforza and Feldman, 2003).

The European continent is of special interest for understanding the evolutionary history of the *Homo sapiens* species. First of all, it is known that two different Homo species (neanderthalensis and sapiens) coexisted for more than 10,000 years in the European continent (Barbujani and Goldstein, 2004). However, little is known about the natural history of the *Homo Neanderthalensis*, the biological relationship between the two Homo species or the degree of inbreeding (if any) between both species. Second, one of the most dramatic demographic events, the agricultural revolution of the Neolithic, took place in the near east and then spread through the European continent. Nevertheless, how this process (either cultural or demographic) was performed is still a matter of debate (Barbujani and Goldstein, 2004). Third, it is still unclear the demographic history of European populations once the Neolithic process was accomplished. Although these demographic events can be traced from the current genetic variability in human populations, a better approach would be through the direct genetic analysis of these ancient populations. In the current thesis three different western European populations from different historical periods have been studied in order to address these issues. Specifically we have extracted and sequenced DNA from three timely different European populations, ranging from Upper Paleolithic times since the post-Neolithic period. We have analysed two Neandertal remains (a teeth sample and a clay sediment sample) from El Sidron Cave (Asturias, North of Spain); twenty-three Neolithic remains from Granollers (Catalunya, Northern Spain) and twenty-two Iberian remains from Girona (Catalunya, North Eastern Spain).

5.2.1 NATURAL HISTORY OF NEANDERTALS IN THE IBERIA PENINSULA

Neandertals became extinct about 28 KYA years ago after coexisting during more than 10,000 years with the anatomically modern humans (Hublin and Paabo, 2006). The last Neandertal populations have been found in the Iberian Peninsula (Hublin, 1995) and it is as well the Iberia peninsula the place where it has been suggested the possibility of hibridization between both species (Duarte et al., 1999). Consequently, the retrieval of mtDNA sequences of Iberian Neandertals specimens represents an important step in our understanding of the evolutionary history of Neandertal species and its past interaction with *Homo sapiens*. In the current thesis we have successfully retrieved Neandertal mtDNA sequences from two different sources of material. In the first case we explored the potential of preserving the Neandertal genetic material in clay sediments. In the second case we extracted and sequenced 47 bp of the mtDNA genome from one Neandertal upper left first incisor remain from el Sidron Cave (Sidron 441) dated from 43 KYA.

5.2.1.1 RECOVERING NEANDERTAL mtDNA FROM CLAY SEDIMENTS

Screening other sources of neandertal genetic materials (different from the normally used teeth and bone fossil remains) is of great importance due to two main reasons: (i) the restricted number of well preserved Neandertal fossils (~400) (Hublin and Paabo, 2006) that has been discovered and (ii) trying to avoid the destructive nature of the current ancient DNA methodology when retrieving DNA from a valuable and scarce Neandertal fossil remain.

The use of soil sediment has been proven to be an alternative source of aDNA in a previous study (Willerslev et al., 2003). By means of studying permafrost and cave sediments, Willerslev et al (2003) were able to retrieve aDNA sequences from plants and animals from the Miocene (~400,000 years ago). However, this study was based on soil from a cold environment, which tends to prevent DNA degradation (see introduction). Because of this, the applicability of such approach could be reduced in warmer areas such as those in the south of Europe.

Nevertheless, the highly natural capability of binding free DNA of clay (Lorenz and Wackernagel, 1994) makes this kind of soil sediment another plausible good candidate for preserving aDNA after the body decomposition. Because of this reason, we attempted to retrieve mtDNA sequences from clay sediments taking from El Sidron Cave immediately next to a Neandertal occipital fragment remain. Using the Neandertal specific primers: NL16,230/NH16,262 and NL16,263b/NH16,300 and after sequencing a large amount of clones, we could retrieve Neandertal like mtDNA HVR I sequences from clay sediments from El Sidron Cave (see chapter 0). The sequences obtained with the set of primers NL16,230/NH16,262 were truncated at the very end and although they displayed a Neandertal specific haplotype 16,234[T]-16,244[A] (Krings et al., 1997) we can not discard the possibility of that the truncated sequence could belong to a undescribed soil microorganism displayed in the Genbank dataset. In contrast, the sequences obtained with the NL16,263b-NH16,300 primer pair display the Neandertal-specific haplotype 16,278[T]-16,299[G] (Krings et al., 1997). Statistical analysis based on Bayesian probability showed that the probability that this haplotype was due to a contemporary human contaminant or any other putative contaminant present in the GeneBank was highly unlikely. Thus, these results proves that clay sediment is a suitable source of aDNA and open the possibility to use other sources of material to retrieve Neandertal aDNA rather than destroying the anthropological and archeological valuable tissue samples. In addition, this new approach could allow us to study the genetic diversity of an archeological area were several individuals (e.g. Neandertals) have been found or even scan if there were Neanderthals remains or not in a particular archeological area where a particular industry has been found.

Nevertheless, the efficiency of this process was extremely low. A large fraction of the retrieved sequences consisted of PCR artifacts (see chapter five), mainly due to the low amount of Neandertal sequences in the clay sediment. This allowed primers to hybridize between them instead of binding the Neandertal DNA template. Furthermore, another source of non Neandertal sequences was soil microorganisms. This strongly contrasts with the results observed in other aDNA Neandertal studies, where a significant fraction of the retrieved sequences are, in fact, modern human contaminants (Krings et al., 1997, Serre et al., 2004, Lalueza-Fox et al., 2005b). This result could be explained by the fact that these soil sediments were freshly taken under controlled conditions. In addition, only the deeper layers of the soil sediment were used, thus carefully reducing the pre

laboratory derived contamination, which is an extremely important factor when working with bone and teeth remains (see chapter 4.1).

In this particular study we have used clay sediments due to the interesting binding DNA properties of this material. However, more research is needed in order to improve the DNA retrieval efficiency of this technique and to explore other putative sources of long time DNA preservation. For example, it has been described that some microorganisms can introduce exogenous DNA inside their genome, thus preventing the degradation of the DNA (Chen and Dubnau, 2004, Gogarten and Townsend, 2005, Thomas and Nielsen, 2005) along the time. If that was the case, it could be possible that some microorganisms could be used as ancient genomic libraries.

In the next future, one possible way to increase the efficiency of this new extraction protocol from clay sediments (where a lot of sources of DNA are present) would be try to get rid of those sources of DNA that we are not interested in during the extraction procedure. One possible way to do that would be developing cleaning strategies in order to capture in the extract (and before setting up any PCR reaction) only those DNA that we are interested in. Specifically, it would be possible binding the DNA of interest to specific proves and then cleaning the remaining DNA that have not been linked. At the end, we will have increased the efficiency of the protocol because of after this "cleaning" step only the DNA of interest will remain.

5.2.1.2 DEMOGRAPHIC HISTORY OF THE NEANDERTAL LINEAGE IN EUROPE (Teeth Neandertal remain mtDNA extraction)

We successfully extracted and sequenced 47bp of the mtDNA control region from the upper left first incisor Neandertal remain El Sidron 441 in two overlapping fragments. with the following Neandertal-specific primers: NL16230/NH16262 and NL16256/NH16278. The 47bp sequenced range between positions 16231 and 16277 of the reference sequence (Anderson et al., 1981). Since the DNA was much degraded, it was impossible to amplify longer fragments than 80bp. In addition, only a very small fraction of all the sequenced clones were definitively identified as Neandertal origin

(~5%). The remaining sequences were clearly modern contaminants; some of them could be assigned to the researchers who had superficially handled the specimen. Consequently, the retrieval of the whole mtDNA HVRI region of the El Sidron 441 remain was technically impossible. This is not an unusual case when sequencing Neandertal sequences. In fact, from the thirteen Neandertal sequences retrieved so far, only seven Neandertal sequences(Krings et al., 1997, Krings et al., 1999, Ovchinnikov et al., 2000, Krings et al., 2000, Schmitz et al., 2002, Serre et al., 2004, Lalueza-Fox et al., 2006) cover a significant section (>300 bp) of the HVRI of the mtDNA whereas the remaining sequences only covered a small fraction of that mtDNA region (Krings et al., 2000, Serre et al., 2004, Beauval et al., 2005, Orlando et al., 2006).

The genetic diversity observed in the retrieved 47 bp mtDNA sequence from sidron 441 specimen has also been found in other mtDNA HVR I European Neandertal sequences (Krings et al., 1997, Krings et al., 1999, Ovchinnikov et al., 2000, Krings et al., 2000, Schmitz et al., 2002, Serre et al., 2004, Lalueza-Fox et al., 2006, Beauval et al., 2005, Orlando et al., 2006). Furthermore, the mtDNA sequence from Sidron 441 remain together with the reported mtDNA control region sequences of the other twelve Neandertal specimens have never been found in the pool of modern mitochondrial DNA sequences. In addition Serre et al (Serre et al., 2004) did not get any Neandertal sequence when trying to amplify Neandertal sequences in well-preserved fossils of early modern humans. Moreover, Caramelli et al (2003) when analysing two early anatomically modern *Homo sapiens sapiens* dated from 23-25 thousand years ago showed that mtDNAs sequences of these individuals fall well within the range of variation of today's humans, but differed sharply from the sequences of contemporary Neandertals. Consequently, mtDNA molecular data from Neanderthal remains and from the early anatomically modern humans that inhabited Europe at the same time suggests that there was not (or, if there were, it was irrelevant) genetic flow between both Homo species. That is, presumably both species of the homo genus (neanderthalensis and sapiens) had two independent and unrelated evolutionary histories (Krings et al., 1997) and, although they coexisted in the European continent, the anatomically modern humans replaced the Neandertal populations presumably without inbreeding with them. This conclusion is of special relevance in the case of the Iberian Peninsula, where the large period of coexistence of Neandertals and anatomically modern humans during more than 10,000 years could have suggested the presence of inbreeding. By the way,

this result supports the “out of Africa model” of human evolution in front of the multiregional model (see chapter 1.7.3).

The presence of other nine sequences at the time of this study allowed us to make deeper analyses on the population structure of the European Neandertal population. In particular we estimate the TMRCA (Time Most Recent Common Ancestor) and the effective population size (N_{fe}) of the Neandertal population at this time. The obtained values (TMRCA: 245,500 ±108,00 and N_{fe}: 5000-9000 individuals) indicated that the genetic history of Neandertals was not shaped by a dramatic population bottleneck associated with the 130,000 years ago glacial maximum (Jouzel et al., 1999). Furthermore, among the nine Neandertals sequences studied until that moment there is a highly polymorphic genetic marker, an A to G transition at position 16258 (4 out of 9 Neandertals mtDNA have an A and the rest a G). Estimates of the age of this highly polymorphic site (153,000 ± 81,000 years ago) together with the estimation of the TMRCA (245,500 ±108,00 years ago) suggests that the genetic variation at this position existed among European Neandertals prior to their retreat into the Southern refugia (Iberian peninsula and the Balkans). It was also noteworthy that this polymorphic position showed a different spatial distribution in Europe: whereas Neandertal sequences from the South of Europe carries the G nucleotide, Neandertals sequences retrieved from North Europe show the A nucleotide. This coincides with the full emergence of the “classical” Neandertal morphology and fits chronologically with a proposed speciation event of Homo Neandertalensis that it is supposed to took place 250,000-300,000 years ago (Rightmire, 2001).

However, the reliability of these results (both the natural history of the extinct *Homo Neandertalensis* species and their biological relationships with the anatomically modern humans) is conditioned to the fact that we have only analysed one marker (the mtDNA) that, in addition, is only of maternal inheritance (see chapter 1.6). Although the technical problems when working with very ancient degraded DNA templates make using mtDNA the best choice (see chapter 1.3), Cooper et al (2004) argued that future methodological developments, such us the retrieval of nuclear sequences from Neanderthal fossils, are needed in order to arrive a conclusive results. This is a challenging issue that would involve important technical and scientific improvement. However, it seems that getting that challenge is not so far away. In fact, in July of this year (2006) the Max Plank institute for Evolutionary Anthropology and the private company 454 Life Science announced the project of sequencing the Neandertal genome

over the next two years. It is thought that a comparison of the Neandertal genome and the human genome will help us in understanding the evolutionary history of this extinct homo species and their biological relationship with *Homo sapiens* species.

5.2.2 THE NEOLITHIC PERIOD IN EUROPE: aDNA FROM AN IBERIAN NEOLITHIC POPULATION.

It is still a matter of discussion whether the actual European populations are direct descendents of the original settlers that replaced the Neandertals in Europe or the result of posterior immigrants from the Near East (Barbujani and Goldstein, 2004). As we have already seen (see chapter 1.8), at least two other dramatic demographic events took place after the peopling of the European continent: the last maximum glacial event (~18,000 years BP) and the Neolithic expansion from the Fertile Crescent (~10,000 years BP). Whereas the first one was within the descendants from the original settlement that moved towards the glacial refugia in the south of Europe, the second one could involve a replacement of the original European populations. Two different hypotheses have been proposed to explain the agricultural diffusion in Europe: the demic and the cultural diffusion models; whereas the first one invokes the presence of large population replacements, the second proposes a cultural rather than demographic diffusion of the new technologies (see chapter 1.8.2). The overlapping nature of the peopling of the European continent and the agricultural dispersal in addition to the small evolutionary time period between both processes make hard to disentangle between both evolutionary scenarios (Barbujani and Bertorelle, 2001) when using current European population genetics. The reconstruction of the demography and evolutionary history of European populations has been based mainly on analysing neutral genetic markers in current European populations but the conclusions obtained when analysing that data are still a matter of debate among scientists. Interestingly, ancient DNA data could be of great interest to resolve this problem since it allows us to directly study the ancient populations that were undergoing these evolutionary processes and not their descendent populations. However, aDNA data is very scarce at population level in Europe(Izagirre and de la Rua, 1999, De Benedetto et al., 2000, Caramelli et al., 2003, Vernesi et al., 2004, Alzualde et al., 2005), mostly due to the big threat that represents modern human contamination when analysing ancient human samples (see chapter 1.3.5). Nevertheless, such a problem could be overwhelmed if there were information available of all the putative contaminants present in the sample. In the current thesis we have analysed a Neolithic population from the Granollers area (in Catalunya, Northeast of Spain) making a special effort in tracing all possible contaminants from the sample.

A total of twenty-three Neolithic remains dated by 5000-5500 years BC were subjected to aDNA analysis. Using different pairs of overlapping primers we attempt to amplify the whole HVR I of the mtDNA genome in each one. At the end, we generated a total of 572 clones. Since, we were able to monitor all the persons who had ever had access to this set of samples (see chapter 4.1), it was possible to track down the pre-laboratory derived contaminant sequences and consequently we could definitely eliminate them from the generated cloning data set. However, this task was not easy and in some situations we faced some problems basically related to the impossibility of working with longer DNA fragments. This was mainly due to the degraded nature of the DNA. Thereby, in some situations it was not possible to distinguish between contaminants and endogenous sequences basically because both sources of sequences were equal in a particular overlapping fragment and thereby impossible to differentiate. In these doubtful situations we discarded the Neolithic sample for posterior analysis. By the way, as far as we know it is the first time in the recent history of aDNA research where it has been possible to control the putatively pre-laboratory derived contaminant DNA sequences and consequently to eliminate them of the cloning dataset. Thus and under such circumstances, the Neolithic remains that we subjected to the genetic analysis were unique.

We found than in these particular set of samples, there were a lot of pre-laboratory derived contaminants and therefore we discard 10 out of 23 Neolithic remains. Preliminary results with the remaining samples show that the haplogroup composition of the Iberian Neolithic population is quite similar to the actual population that nowadays inhabits this area .This genetic continuity between the ancient population and the current populations suggested by these results contrasts with these found by Haak et al (2005). Haak et al when analysing a Neolithic population (7000-7,500 years BP) from Central Europe found that 25% (6 out of 24) of the samples were of a distinctive and rare N1a lineage (currently present at 0.2% in the European population) of the mtDNA well-known phylogeny. Discarding the possibility of a possible lost of this mtDNA lineages by genetic drift over the last 7500 years by means of demographic models, they proposed that small pioneer farming groups carried farming into new areas of Europe, and therefore the dispersal of the agricultural techniques were through a cultural model; so that, at the end of the process, the frequency of the N1a haplogroup was diluted to the low modern value that its observed today.

DISCUSSION

The absence of sequences carrying the N1a haplogroup in the current set of analysed sequences could be due to the difference of time (approximately 2000 years) and space (North versus South Europe) with these analysed by Haak et al. However, since the absence of N1a lineage cannot be explained by genetic drift, other hypotheses have to be invoked in order to explain such a discrepancy between their results and the observed in the current work. Therefore, we hypothesise that the dispersal of agriculture involved both demographic and cultural diffusion, depending on the region where it took place. Whereas the dispersal of the agricultural in the North of Europe could be a cultural diffusion model, in the south our results suggest a demic diffusion model.

Thus, in order to totally clarify whether the actual European populations are direct descendents of the original settlers that replaced the Neandertals or the result of posterior immigrants from the Near East, the analysis of more ancient Neolithic remains from different European regions are definitively needed.

5.2.3 POST-NEOLITHIC PERIOD IN THE IBERIAN PENINSULA: A mtDNA STUDY OF THE ANCIENT IBERIANS

Although archaeology does not document other large-scale demographical events after the Neolithic period, it is known that there have been continuous small migrations through the European continent (Sokal et al., 1996). Raisings and fallings of empires, such as the Roman Empire, have been associated to cultural and population movements on the conquest territories. Sometimes such processes have almost erased the previous culture thus making more difficult the study of these ancient populations. This is the case of the Iberian populations, that before the Roman conquest had a non Indo-European language which is still not understood. The current Basque language and the extinct Etruscan languages (in Italy) were also non-Indoeuropean languages; this could indicate the existence of a Palaeolithic substratum according to a model of a demic expansion associated with agriculture and Indo-European languages from de Middle East. Although the analysis of classical genetic markers in the Basque population has been interpreted as a result of long isolation period (Bertranpetit and Cavalli-Sforza, 1991), genetic analysis of these extinct populations is of great interest in order to clarify the present and past population structure in the Western Europe. We have analysed seventeen humans' remains belonging to the Iberian culture (2600 years BP). In contrast to other aDNA studies based on post-Neolithic populations (e.g (Vernesi et al., 2004, Alzualde et al., 2005), working with Iberian remains had an additional complication. Since Iberians incinerated their dead, the majority of the Iberian necropolises were unsuitable for aDNA analysis. However, some skeletal remains were occasionally found in the large town of Ullastret (Catalonia, North of Spain); these are related to a ritual practise of nailing skulls, in which the heads of the enemies were displayed in public places with a long nail going all the way through the skull from the forehead to the cranial base. In this study, we have analysed some of these nailed skulls plus other Iberian remains to obtain, for the first time, a genetic picture of the human pre-Roman substratum in the northeast of the Iberian Peninsula. Following all the authenticity aDNA criteria proposed so far (Cooper and Poinar, 2000), the mtDNA HVR was amplified in different overlapping fragments and some diagnostic SNPs in the control region were typed. It was possible to assign each remain to an unequivocally haplogroup in the mtDNA gene genealogy. When analysing the genetic composition of

the ancient Iberians at mtDNA level, we found that they were not significantly different from modern populations from the same region (roughly, North-East and East of the Iberian Peninsula) in haplogroup composition (see chapter 4.6). This finding pointed out towards a genetic continuity in the Iberia peninsula since Neolithic times. This result contrasts with those found in previous genetic studies when analysing populations from the same historical period but in different regions of Europe such as the Basque country (Alzualde et al., 2005) and the Italian peninsula (Vernesi et al., 2004). These studies gave support to possible post-neolithic migration events because of the presence of differences between modern populations from those regions and the ancient populations. Clearly more research is needed in other ancient post-Neolithic population in order to clarify this phenomenon.

Furthermore, the lack of African lineages in the ancient Iberians, although not conclusive due to the small number of individuals analyzed, does not provide support for an earlier gene flow between North Africa and the Iberian Peninsula before the arrival of the Roman Empire. But again, this cannot be discarded due to the small sample size. Moreover, the three non Indo-European populations, Basque, Etruscan and the enigmatic Iberian language, do not seem to be especially close to one another from a genetic point of view. Interestingly, when analyzing the genetic diversity of the ancient Iberians, it cannot be observed that in comparison to the modern population of the same area, the Iberians had a lower amount of genetic diversity. We hypothesized that the limited genetic diversity of the ancient Iberians could be due to the existence of small populations and endogamic processes related to the tribe-structured Iberian society. Moreover, the documented, posterior arrivals of groups from Europe and North Africa did not alter significantly the pre- Roman genetic background but they probably increased the relatively low genetic diversity of the Iberian groups.

In conclusion, there is a long term genetic continuity in the Iberia peninsula at least since the Neolithic. The only clear genetic discontinuity found is this involving two different human species, *H. sapiens* and *H. neanderthalensis*.

6BIBLIOGRAPHY

ACHILLI, A., RENGO, C., MAGRI, C., et al. (2004) The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet,* 75**,** 910-8.

ALONSO, A., MARTIN, P., ALBARRAN, C., et al. (2004) Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Sci Int,* 139**,** 141-9.

ALZUALDE, A., IZAGIRRE, N., ALONSO, S., et al. (2005) Temporal mitochondrial DNA variation in the Basque Country: influence of post-neolithic events. *Ann Hum Genet,* 69**,** 665-79.

AMMERMAN, A. J., CAVALLI-SFORZA L. L. (1984) *The Neolithic transition and the Genetics of Population in Europe*.

ANDERSON, S., BANKIER, A. T., BARRELL, B. G., et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature,* 290**,** 457-65.

ANDREWS, R. M., KUBACKA, I., CHINNERY, P. F., et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet,* 23**,** 147.

ANTON, S. C. (2003) Natural history of Homo erectus. *Am J Phys Anthropol,* Suppl 37**,** 126-70.

ARCTANDER, P. (1999) Mitochondrial recombination? *Science,* 284**,** 2090-1.

ARMOUR, J. A., ANTTINEN, T., MAY, C. A., et al. (1996) Minisatellite diversity supports a recent African origin for modern humans. *Nat Genet,* 13**,** 154-60.

AWADALLA, P., EYRE-WALKER, A. & SMITH, J. M. (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science,* 286**,** 2524-5.

BANDELT, H. J. (2005) Mosaics of ancient mitochondrial DNA: positive indicators of nonauthenticity. *Eur J Hum Genet,* 13**,** 1106-12.

BARBUJANI, G. & BERTORELLE, G. (2001) Genetics and the population history of Europe. *Proc Natl Acad Sci U S A,* 98**,** 22-5.

BARBUJANI, G. & GOLDSTEIN, D. B. (2004) Africans and Asians abroad: genetic diversity in Europe. *Annu Rev Genomics Hum Genet,* 5**,** 119-50.

BARBUJANI, G., MAGAGNI, A., MINCH, E., et al. (1997) An apportionment of human DNA diversity. *Proc Natl Acad Sci U S A,* 94**,** 4516-9.

BIBLIOGRAPHY

BEAUVAL, C., MAUREILLE, B., LACRAMPE-CUYAUBERE, F., et al. (2005) A late Neandertal femur from Les Rochers-de-Villeneuve, France. *Proc Natl Acad Sci U S A,* 102**,** 7085-90.

BEJA-PEREIRA, A., CARAMELLI, D., LALUEZA-FOX, C., et al. (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proc Natl Acad Sci U S A,* 103**,** 8113-8.

BELLE, E. M., LANDRY, P. A. & BARBUJANI, G. (2006) Origins and evolution of the Europeans' genome: evidence from multiple microsatellite loci. *Proc Biol Sci,* 273**,** 1595-602.

BENDALL, K. E. & SYKES, B. C. (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet,* 57**,** 248-56.

BENSASSON, D., FELDMAN, M. W. & PETROV, D. A. (2003) Rates of DNA duplication and mitochondrial DNA insertion in the human genome. *J Mol Evol,* 57**,** 343-54.

BERMUDEZ DE CASTRO, J. M., ARSUAGA, J. L., CARBONELL, E., et al. (1997) A hominid from the lower Pleistocene of Atapuerca, Spain: possible ancestor to Neandertals and modern humans. *Science,* 276**,** 1392-5.

BERTRANPETIT, J. & CAVALLI-SFORZA, L. L. (1991) A genetic reconstruction of the history of the population of the Iberian Peninsula. *Ann Hum Genet,* 55**,** 51-67.

BINLADEN, J., WIUF, C., GILBERT, M. T., et al. (2006) Assessing the fidelity of ancient DNA sequences amplified from nuclear genes. *Genetics,* 172**,** 733-41.

BOSCH, E., CALAFELL, F., COMAS, D., et al. (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet,* 68**,** 1019-29.

BOWER, M. A., SPENCER, M., MATSUMURA, S., et al. (2005) How many clones need to be sequenced from a single forensic or ancient DNA sample in order to determine a reliable consensus sequence? *Nucleic Acids Res,* 33**,** 2549-56.

BROWN, P., SUTIKNA, T., MORWOOD, M. J., et al. (2004) A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature,* 431**,** 1055-61.

BROWN, W. M. (1980) Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Acad Sci U S A,* 77**,** 3605-9.

BROWN, W. M., GEORGE, M., JR. & WILSON, A. C. (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A,* 76**,** 1967-71.

BRUNET, M., GUY, F., PILBEAM, D., et al. (2002) A new hominid from the Upper Miocene of Chad, Central Africa. *Nature,* 418**,** 145-51.

BUNCE, M., SZULKIN, M., LERNER, H. R., et al. (2005) Ancient DNA provides new insights into the evolutionary history of New Zealand's extinct giant eagle. *PLoS Biol,* 3**,** e9.

CANN, R. L., STONEKING, M. & WILSON, A. C. (1987) Mitochondrial DNA and human evolution. *Nature,* 325**,** 31-6.

CANO, R. J., POINAR, H. N., PIENIAZEK, N. J., et al. (1993) Amplification and sequencing of DNA from a 120-135-million-year-old weevil. *Nature,* 363**,** 536-8.

CARAMELLI, D., LALUEZA-FOX, C., CONDEMI, S., et al. (2006) A highly divergent mtDNA sequence in a Neandertal individual from Italy. *Curr Biol,* 16**,** R630-2.

CARAMELLI, D., LALUEZA-FOX, C., VERNESI, C., et al. (2003) Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc Natl Acad Sci U S A,* 100**,** 6593-7.

CARROLL, S. B. (2003) Genetics and the making of Homo sapiens. *Nature,* 422**,** 849-57.

CAVALLI-SFORZA, L. L. & FELDMAN, M. W. (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet,* 33 Suppl**,** 266-75.

CAVALLI-SFORZA, L. L., MENOZZI P, PIAZZA A (1994) *History and geography of human genes,* Princeton (NJ).

CHEN, I. & DUBNAU, D. (2004) DNA uptake during bacterial transformation. *Nat Rev Microbiol,* 2**,** 241-9.

CHIKHI, L., DESTRO-BISOL, G., BERTORELLE, G., et al. (1998) Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. *Proc Natl Acad Sci U S A,* 95**,** 9053-8.

CHIKHI, L., NICHOLS, R. A., BARBUJANI, G., et al. (2002) Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A,* 99**,** 11008-13.

COMAS, D., CALAFELL, F., MATEU, E., et al. (1997) Mitochondrial DNA variation and the origin of the Europeans. *Hum Genet,* 99**,** 443-9.

BIBLIOGRAPHY

COMAS, D., CALAFELL, F., MATEU, E., et al. (1998) Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. *Am J Hum Genet,* 63**,** 1824-38.

COMAS, D., PAABO, S. & BERTRANPETIT, J. (1995) Heteroplasmy in the control region of human mitochondrial DNA. *Genome Res,* 5**,** 89-90.

COOPER, A., DRUMMOND, A. J. & WILLERSLEV, E. (2004) Ancient DNA: would the real Neandertal please stand up? *Curr Biol,* 14**,** R431-3.

COOPER, A., LALUEZA-FOX, C., ANDERSON, S., et al. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. *Nature,* 409**,** 704-7.

COOPER, A. & POINAR, H. N. (2000) Ancient DNA: do it right or not at all. *Science,* 289**,** 1139.

COOPER, G. M. (1997) *The Cell: a molecular approach.*

CROW, J. F. (1997) The high spontaneous mutation rate: is it a health risk? *Proc Natl Acad Sci U S A,* 94**,** 8380-6.

DE BENEDETTO, G., NASIDZE, I. S., STENICO, M., et al. (2000) Mitochondrial DNA sequences in prehistoric human remains from the Alps. *Eur J Hum Genet,* 8**,** 669-77.

DESALLE, R., GATESY, J., WHEELER, W., et al. (1992) DNA sequences from a fossil termite in Oligo-Miocene amber and their phylogenetic implications. *Science,* 257**,** 1933-6.

DIAMOND, J. (2002) Evolution, consequences and future of plant and animal domestication. *Nature,* 418**,** 700-7.

DUARTE, C., MAURICIO, J., PETTITT, P. B., et al. (1999) The early Upper Paleolithic human skeleton from the Abrigo do Lagar Velho (Portugal) and modern human emergence in Iberia. *Proc Natl Acad Sci U S A,* 96**,** 7604-9.

DUPANLOUP, I., BERTORELLE, G., CHIKHI, L., et al. (2004) Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol Biol Evol,* 21**,** 1361-72.

EXCOFFIER, L. (2002) Human demographic history: refining the recent African origin model. *Curr Opin Genet Dev,* 12**,** 675-82.

EXCOFFIER, L., SMOUSE, P. E. & QUATTRO, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics,* 131**,** 479-91.

FATTORINI, P., CIOFULI, R., COSSUTTA, F., et al. (1999) Fidelity of polymerase chain reaction-direct sequencing analysis of damaged forensic samples. *Electrophoresis,* 20**,** 3349-57.

FORSTER, P. (2004) Ice Ages and the mitochondrial DNA chronology of human dispersals: a review. *Philos Trans R Soc Lond B Biol Sci,* 359**,** 255-64; discussion 264.

FORSTER, P., HARDING, R., TORRONI, A., et al. (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet,* 59**,** 935-45.

GAGNEUX, P., WILLS, C., GERLOFF, U., et al. (1999) Mitochondrial sequences show diverse evolutionary histories of African hominoids. *Proc Natl Acad Sci U S A,* 96**,** 5077-82.

GIBBONS, A. (1994) Possible dino DNA find is greeted with skepticism. *Science,* 266**,** 1159.

GILBERT, M. T., BANDELT, H. J., HOFREITER, M., et al. (2005) Assessing ancient DNA studies. *Trends Ecol Evol,* 20**,** 541-4.

GILBERT, M. T., BINLADEN, J., MILLER, W., et al. (2006) Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic Acids Res.*

GILBERT, M. T., CUCCUI, J., WHITE, W., et al. (2004) Absence of Yersinia pestis-specific DNA in human teeth from five European excavations of putative plague victims. *Microbiology,* 150**,** 341-54.

GILBERT, M. T., HANSEN, A. J., WILLERSLEV, E., et al. (2003a) Characterization of genetic miscoding lesions caused by postmortem damage. *Am J Hum Genet,* 72**,** 48-61.

GILBERT, M. T., WILLERSLEV, E., HANSEN, A. J., et al. (2003b) Distribution patterns of postmortem damage in human mitochondrial DNA. *Am J Hum Genet,* 72**,** 32-47.

GILBERT, M. T. P., A.J. HANSEN, E.WILLERSLEV, G. TURNER-WALKER AND M.COLLINS (2005a) Insights into the Processes behind the Contamination of Degraded Human teeth and Bone Samples with exogenous Sources of DNA. *International Journal of Osteoarchaeology,* 15**,** 1-9.

GILBERT, M. T. P., LARS RUDBECK, ESKE WILLERSLEV, ANDERS J. ANDERS, COLIN SMITH, KIRSTY E.H. PENKMAN, KURT PRANGENBERG, CHRISTINA M. NIELSEN-MARSH, MIRANDA E. JANS,

PAUL ARTHUR, NIELS LYNNERUP, GORDON TURNER-WALKER, MARTIN BIDDLE, BIRTHE KJOLBYE-BIDDLE, MATTHEW J. COLLINS (2005b) Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *Journal of Archaeological Science,* 32**,** 785-793.

GOGARTEN, J. P. & TOWNSEND, J. P. (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol,* 3**,** 679-87.

GOLENBERG, E. M., GIANNASI, D. E., CLEGG, M. T., et al. (1990) Chloroplast DNA sequence from a miocene Magnolia species. *Nature,* 344**,** 656-8.

GREENWOOD, A. D., CAPELLI, C., POSSNERT, G., et al. (1999) Nuclear DNA sequences from late Pleistocene megafauna. *Mol Biol Evol,* 16**,** 1466-73.

HAAK, W., FORSTER, P., BRAMANTI, B., et al. (2005) Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science,* 310**,** 1016-8.

HAGELBERG, E., GOLDMAN, N., LIO, P., et al. (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proc Biol Sci,* 266**,** 485-92.

HAGELBERG, E., SYKES, B. & HEDGES, R. (1989) Ancient bone DNA amplified. *Nature,* 342**,** 485.

HANDT, O., HOSS, M., KRINGS, M., et al. (1994a) Ancient DNA: methodological challenges. *Experientia,* 50**,** 524-9.

HANDT, O., KRINGS, M., WARD, R. H., et al. (1996) The retrieval of ancient human DNA sequences. *Am J Hum Genet,* 59**,** 368-76.

HANDT, O., RICHARDS, M., TROMMSDORFF, M., et al. (1994b) Molecular genetic analyses of the Tyrolean Ice Man. *Science,* 264**,** 1775-8.

HANNI, C., BROUSSEAU, T., LAUDET, V., et al. (1995) Isopropanol precipitation removes PCR inhibitors from ancient bone extracts. *Nucleic Acids Res,* 23**,** 881-2.

HANSEN, A., WILLERSLEV, E., WIUF, C., et al. (2001) Statistical evidence for miscoding lesions in ancient DNA templates. *Mol Biol Evol,* 18**,** 262-5.

HAPMAP (2005) A haplotype map of the human genome. *Nature,* 437**,** 1299-320.

HARDY, C., CALLOU, C., VIGNE, J. D., et al. (1995) Rabbit mitochondrial DNA diversity from prehistoric to modern times. *J Mol Evol,* 40**,** 227-37.

HARPENDING, H. & ROGERS, A. (2000) Genetic perspectives on human origins and differentiation. *Annu Rev Genomics Hum Genet,* 1**,** 361-85.

HEBSGAARD, M. B., PHILLIPS, M. J. & WILLERSLEV, E. (2005) Geologically ancient DNA: fact or artefact? *Trends Microbiol,* 13**,** 212-20.

HIGUCHI, R., BOWMAN, B., FREIBERGER, M., et al. (1984) DNA sequences from the quagga, an extinct member of the horse family. *Nature,* 312**,** 282-4.

HOFREITER, M., JAENICKE, V., SERRE, D., et al. (2001a) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Res,* 29**,** 4793-9.

HOFREITER, M., SERRE, D., POINAR, H. N., et al. (2001b) Ancient DNA. *Nat Rev Genet,* 2**,** 353-9.

HOSS, M., DILLING, A., CURRANT, A., et al. (1996a) Molecular phylogeny of the extinct ground sloth Mylodon darwinii. *Proc Natl Acad Sci U S A,* 93**,** 181-5.

HOSS, M., JARUGA, P., ZASTAWNY, T. H., et al. (1996b) DNA damage and DNA sequence retrieval from ancient tissues. *Nucleic Acids Res,* 24**,** 1304-7.

HUBLIN, J. J., C. BARROSO RUIZ, P. MEDINA LARA, M. FONTUGNE, AND J.-L. REYSS. (1995) The Mousterian site of Zafarraya (Andalucia,Spain): dating and implications on the palaeolithic peopling processes of Western Europe. *C. R. Acad. Sci. Paris IIa,* 321**,** 931–937.

HUBLIN, J. J. & PAABO, S. (2006) Neandertals. *Curr Biol,* 16**,** R113-4.

INGMAN, M., KAESSMANN, H., PAABO, S., et al. (2000) Mitochondrial genome variation and the origin of modern humans. *Nature,* 408**,** 708-13.

IZAGIRRE, N. & DE LA RUA, C. (1999) An mtDNA analysis in ancient Basque populations: implications for haplogroup V as a marker for a major paleolithic expansion from southwestern europe. *Am J Hum Genet,* 65**,** 199-207.

JACOB, T., INDRIATI, E., SOEJONO, R. P., et al. (2006) Pygmoid Australomelanesian Homo sapiens skeletal remains from Liang Bua, Flores: population affinities and pathological abnormalities. *Proc Natl Acad Sci U S A,* 103**,** 13421-6.

JAENICKE-DESPRES, V., BUCKLER, E. S., SMITH, B. D., et al. (2003) Early allelic selection in maize as revealed by ancient DNA. *Science,* 302**,** 1206-8.

JOBLING, M., HURLES ME TAYLER-SMITH C (2004) *Human Evolutionary Genetics. Origins. Peoples and Disease,* New York.

JORDE, L. B. & BAMSHAD, M. (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science,* 288**,** 1931.

BIBLIOGRAPHY

JORDE, L. B., WATKINS, W. S., BAMSHAD, M. J., et al. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet,* 66**,** 979-88.

JOUZEL, J., PETIT, J. R., SOUCHEZ, R., et al. (1999) More than 200 meters of lake ice above subglacial Lake Vostok, Antarctica. *Science,* 286**,** 2138-41.

KITTLES, R. A. & WEISS, K. M. V. (2003) Race, ancestry, and genes: implications for defining disease risk. *Annu Rev Genomics Hum Genet,* 4**,** 33-67.

KIVISILD, T., BAMSHAD, M. J., KALDMA, K., et al. (1999) Deep common ancestry of indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol,* 9**,** 1331-4.

KOLMAN, C. J. & TUROSS, N. (2000) Ancient DNA analysis of human populations. *Am J Phys Anthropol,* 111**,** 5-23.

KRAUSE, J., DEAR, P. H., POLLACK, J. L., et al. (2006) Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature,* 439**,** 724-7.

KRINGS, M., CAPELLI, C., TSCHENTSCHER, F., et al. (2000) A view of Neandertal genetic diversity. *Nat Genet,* 26**,** 144-6.

KRINGS, M., GEISERT, H., SCHMITZ, R. W., et al. (1999) DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. *Proc Natl Acad Sci U S A,* 96**,** 5581-5.

KRINGS, M., STONE, A., SCHMITZ, R. W., et al. (1997) Neandertal DNA sequences and the origin of modern humans. *Cell,* 90**,** 19-30.

KUMAR, S., HEDRICK, P., DOWLING, T., et al. (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science,* 288**,** 1931.

LAHR, M. M. R. A. F. (2001) *Genes Fossils and Behaviour: When and Where do they fit?,* Brussels: IOS Press.

LALUEZA-FOX, C., CALDERON, F. L., CALAFELL, F., et al. (2001) MtDNA from extinct Tainos and the peopling of the Caribbean. *Ann Hum Genet,* 65**,** 137-51.

LALUEZA-FOX, C., CASTRESANA, J., SAMPIETRO, L., et al. (2005a) Molecular dating of caprines using ancient DNA sequences of Myotragus balearicus, an extinct endemic Balearic mammal. *BMC Evol Biol,* 5**,** 70.

LALUEZA-FOX, C., GILBERT, M. T., MARTINEZ-FUENTES, A. J., et al. (2003) Mitochondrial DNA from pre-Columbian Ciboneys from Cuba and the prehistoric colonization of the Caribbean. *Am J Phys Anthropol,* 121**,** 97-108.

LALUEZA-FOX, C., KRAUSE, J., CARAMELLI, D., et al. (2006) Mitochondrial DNA of an Iberian Neandertal suggests a population affinity with other European Neandertals. *Curr Biol,* 16**,** R629-30.

LALUEZA-FOX, C., SAMPIETRO, M. L., CARAMELLI, D., et al. (2005b) Neandertal evolutionary genetics: mitochondrial DNA data from the iberian peninsula. *Mol Biol Evol,* 22**,** 1077-81.

LALUEZA-FOX, C., SAMPIETRO, M. L., GILBERT, M. T., et al. (2004) Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc Biol Sci,* 271**,** 941-7.

LAMBERT, D. M., RITCHIE, P. A., MILLAR, C. D., et al. (2002) Rates of evolution in ancient DNA from Adelie penguins. *Science,* 295**,** 2270-3.

LINDAHL, T. (1993a) Instability and decay of the primary structure of DNA. *Nature,* 362**,** 709-15.

LINDAHL, T. (1993b) Recovery of antediluvian DNA. *Nature,* 365**,** 700.

LORENZ, M. G. & WACKERNAGEL, W. (1994) Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev,* 58**,** 563-602.

MACA-MEYER, N., ARNAY, M., RANDO, J. C., et al. (2004) Ancient mtDNA analysis and the origin of the Guanches. *Eur J Hum Genet,* 12**,** 155-62.

MACA-MEYER, N., GONZALEZ, A. M., LARRUGA, J. M., et al. (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genet,* 2**,** 13.

MALMSTROM, H., STORA, J., DALEN, L., et al. (2005) Extensive human DNA contamination in extracts from ancient dog bones and teeth. *Mol Biol Evol,* 22**,** 2040-7.

MARGULIES, M., EGHOLM, M., ALTMAN, W. E., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature,* 437**,** 376-80.

MARGULIS, L., OLENDZENSKI, L. & AFZELIUS, B. A. (1990) Endospore-forming filamentous bacteria symbiotic in termites: ultrastructure and growth in culture of Arthromitus. *Symbiosis,* 8**,** 95-116.

MELLARS, P. (2004) Neanderthals and the modern human colonization of Europe. *Nature,* 432**,** 461-5.

MELLARS, P. A. (1992) Archaeology and the population-dispersal hypothesis of modern human origins in Europe. *Philos Trans R Soc Lond B Biol Sci,* 337**,** 225-34.

BIBLIOGRAPHY

MENOZZI, P., PIAZZA, A. & CAVALLI-SFORZA, L. (1978) Synthetic maps of human gene frequencies in Europeans. *Science,* 201**,** 786-92.

MIRAZON LAHR, M. & FOLEY, R. (2004) Palaeoanthropology: human evolution writ small. *Nature,* 431**,** 1043-4.

MISHMAR, D., RUIZ-PESINI, E., BRANDON, M., et al. (2004) Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. *Hum Mutat,* 23**,** 125-33.

MITOMAP: A HUMAN MITOCHONDRIAL GENOME DATABASE. HTTP://WWW.MITOMAP.ORG.

MORWOOD, M. J., SOEJONO, R. P., ROBERTS, R. G., et al. (2004) Archaeology and age of a new hominin from Flores in eastern Indonesia. *Nature,* 431**,** 1087-91.

MULLIS, K. B. & FALOONA, F. A. (1987) Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol,* 155**,** 335-50.

OOTA, H., SAITOU, N., MATSUSHITA, T., et al. (1995) A genetic study of 2,000-year-old human remains from Japan using mitochondrial DNA sequences. *Am J Phys Anthropol,* 98**,** 133-45.

ORLANDO, L., DARLU, P., TOUSSAINT, M., et al. (2006) Revisiting Neandertal diversity with a 100,000 year old mtDNA sequence. *Curr Biol,* 16**,** R400-2.

OTTE, M. (1990) The World at 18,000 BP, eds.Soffer, O. & Gamble, C. (Unwin, London). Vol I**,** 54–68.

OVCHINNIKOV, I. V., GOTHERSTROM, A., ROMANOVA, G. P., et al. (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature,* 404**,** 490-3.

PAABO, S. (1985) Molecular cloning of Ancient Egyptian mummy DNA. *Nature,* 314**,** 644-5.

PAABO, S. (1989) Ancient DNA: extraction, characterization, molecular cloning, and enzymatic amplification. *Proc Natl Acad Sci U S A,* 86**,** 1939-43.

PAABO, S. (2003) The mosaic that is our genome. *Nature,* 421**,** 409-12.

PAABO, S., GIFFORD, J. A. & WILSON, A. C. (1988) Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic Acids Res,* 16**,** 9775-87.

PAABO, S., HIGUCHI, R. G. & WILSON, A. C. (1989) Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. *J Biol Chem,* 264**,** 9709-12.

PAABO, S., IRWIN, D. M. & WILSON, A. C. (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem,* 265**,** 4718-21.

PAABO, S., POINAR, H., SERRE, D., et al. (2004) Genetic analyses from ancient DNA. *Annu Rev Genet,* 38**,** 645-79.

PAABO, S. & WILSON, A. C. (1991) Miocene DNA sequences - a dream come true? *Curr Biol,* 1**,** 45-6.

PAKENDORF, B. & STONEKING, M. (2005) Mitochondrial DNA and human evolution. *Annu Rev Genomics Hum Genet,* 6**,** 165-83.

PEREZ-LEZAUN, A., CALAFELL, F., COMAS, D., et al. (1999) Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am J Hum Genet,* 65**,** 208-19.

PERGAMS, O. R., BARNES, W. M. & NYBERG, D. (2003) Mammalian microevolution: Rapid change in mouse mitochondrial DNA. *Nature,* 423**,** 397.

PLAZA, S., CALAFELL, F., HELAL, A., et al. (2003) Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann Hum Genet,* 67**,** 312-28.

POINAR, H., KUCH, M., MCDONALD, G., et al. (2003) Nuclear gene sequences from a late pleistocene sloth coprolite. *Curr Biol,* 13**,** 1150-2.

POINAR, H. N. (2002) The genetic secrets some fossils hold. *Acc Chem Res,* 35**,** 676-84.

POINAR, H. N., HOSS, M., BADA, J. L., et al. (1996) Amino acid racemization and the preservation of ancient DNA. *Science,* 272**,** 864-6.

POINAR, H. N., SCHWARZ, C., QI, J., et al. (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science,* 311**,** 392-4.

POULTON, J. & MARCHINGTON, D. R. (2002) Segregation of mitochondrial DNA (mtDNA) in human oocytes and in animal models of mtDNA disease: clinical implications. *Reproduction,* 123**,** 751-5.

PUSCH, C. M. & BACHMANN, L. (2004) Spiking of contemporary human template DNA with ancient DNA extracts induces mutations under PCR and generates nonauthentic mitochondrial sequences. *Mol Biol Evol,* 21**,** 957-64.

QUINTANA-MURCI, L., SEMINO, O., BANDELT, H. J., et al. (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet,* 23**,** 437-41.

BIBLIOGRAPHY

RAMACHANDRAN, S., DESHPANDE, O., ROSEMAN, C. C., et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci U S A,* 102**,** 15942-7.

REICH, D. E., SCHAFFNER, S. F., DALY, M. J., et al. (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet,* 32**,** 135-42.

RICHARDS, M., B. SYKES, AND R. HEDGES. (1995) Authenticating DNA extracted from ancient skeletal remains. *J. Archeol. Sci.,* 22**,** 291-299.

RICHARDS, M., CORTE-REAL, H., FORSTER, P., et al. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet,* 59**,** 185-203.

RICHARDS, M., MACAULAY, V., HICKEY, E., et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet,* 67**,** 1251-76.

RICHARDS, M., MACAULAY, V., TORRONI, A., et al. (2002) In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet,* 71**,** 1168-74.

RICHARDS, M. B., MACAULAY, V. A., BANDELT, H. J., et al. (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet,* 62**,** 241-60.

RIGHTMIRE, G. P. (2001) Patterns of hominid evolution and dispersal in the Middle Pleistocene. *Quat. Int,* 75**,** 77–84.

ROBIN, E. D. & WONG, R. (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol,* 136**,** 507-13.

ROMPLER, H., ROHLAND, N., LALUEZA-FOX, C., et al. (2006) Nuclear gene indicates coat-color polymorphism in mammoths. *Science,* 313**,** 62.

ROMUALDI, C., BALDING, D., NASIDZE, I. S., et al. (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res,* 12**,** 602-12.

ROSSER, Z. H., ZERJAL, T., HURLES, M. E., et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet,* 67**,** 1526-43.

SALAS, A., RICHARDS, M., DE LA FE, T., et al. (2002) The making of the African mtDNA landscape. *Am J Hum Genet,* 71**,** 1082-111.

SCHMITZ, R. W., SERRE, D., BONANI, G., et al. (2002) The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc Natl Acad Sci U S A,* 99**,** 13342-7.

SCHWARTZ, M. & VISSING, J. (2002) Paternal inheritance of mitochondrial DNA. *N Engl J Med,* 347**,** 576-80.

SEMINO, O., MAGRI, C., BENUZZI, G., et al. (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet,* 74**,** 1023-34.

SEMINO, O., PASSARINO, G., BREGA, A., et al. (1996) A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet,* 59**,** 964-8.

SEMINO, O., PASSARINO, G., OEFNER, P. J., et al. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science,* 290**,** 1155-9.

SENUT, B., PICKFORD, M., GOMMERY, D., MEIN, P., CHEBOI, K. & COPPENS, Y. (2001) First hominid from the. Miocene (Lukeino formation,Kenya). *C. R. Acad. Sci. Ser. IIa,* 332**,** 137–144.

SERRE, D., LANGANEY, A., CHECH, M., et al. (2004) No evidence of Neandertal mtDNA contribution to early modern humans. *PLoS Biol,* 2**,** E57.

SHAPIRO, B., DRUMMOND, A. J., RAMBAUT, A., et al. (2004) Rise and fall of the Beringian steppe bison. *Science,* 306**,** 1561-5.

SIGURGARDOTTIR, S., HELGASON, A., GULCHER, J. R., et al. (2000) The mutation rate in the human mtDNA control region. *Am J Hum Genet,* 66**,** 1599-609.

SIMONI, L., CALAFELL, F., PETTENER, D., et al. (2000) Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet,* 66**,** 262-78.

SMITH, C. I., CHAMBERLAIN, A. T., RILEY, M. S., et al. (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. *J Hum Evol,* 45**,** 203-17.

SOKAL, R. R., HARDING, R. M. & ODEN, N. L. (1989) Spatial patterns of human gene frequencies in Europe. *Am J Phys Anthropol,* 80**,** 267-94.

BIBLIOGRAPHY

SOKAL, R. R., ODEN, N. L., WALKER, J., et al. (1996) Historical population movements in Europe influence genetic relationships in modern samples. *Hum Biol,* 68**,** 873-98.

SOKAL, R. R., ODEN, N. L. & WILSON, C. (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature,* 351**,** 143-5.

STILLER, M., GREEN, R. E., RONAN, M., et al. (2006) Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proc Natl Acad Sci U S A,* 103**,** 13578-84.

STONE, A. C. & STONEKING, M. (1993) Ancient DNA from a pre-Columbian Amerindian population. *Am J Phys Anthropol,* 92**,** 463-71.

STONE, A. C. & STONEKING, M. (1998) mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. *Am J Hum Genet,* 62**,** 1153-70.

STONEKING, M., FONTIUS, J. J., CLIFFORD, S. L., et al. (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res,* 7**,** 1061-71.

SUTOVSKY, P., MORENO, R. D., RAMALHO-SANTOS, J., et al. (1999) Ubiquitin tag for sperm mitochondria. *Nature,* 402**,** 371-2.

SUTOVSKY, P., MORENO, R. D., RAMALHO-SANTOS, J., et al. (2000) Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod,* 63**,** 582-90.

TAMURA, K. & NEI, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol,* 10**,** 512-26.

TEMPLETON, A. (2002) Out of Africa again and again. *Nature,* 416**,** 45-51.

THOMAS, C. M. & NIELSEN, K. M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol,* 3**,** 711-21.

THOMAS, R. H., SCHAFFNER, W., WILSON, A. C., et al. (1989) DNA phylogeny of the extinct marsupial wolf. *Nature,* 340**,** 465-7.

THOMSON, R., PRITCHARD, J. K., SHEN, P., et al. (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc Natl Acad Sci U S A,* 97**,** 7360-5.

TISHKOFF, S. A., DIETZSCH, E., SPEED, W., et al. (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science,* 271**,** 1380-7.

TORRONI, A., ACHILLI, A., MACAULAY, V., et al. (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet,* 22**,** 339-45.

TORRONI, A., BANDELT, H. J., D'URBANO, L., et al. (1998) mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet,* 62**,** 1137-52.

TORRONI, A., BANDELT, H. J., MACAULAY, V., et al. (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet,* 69**,** 844-52.

TULLY, L. A., PARSONS, T. J., STEIGHNER, R. J., et al. (2000) A sensitive denaturing gradient-Gel electrophoresis assay reveals a high frequency of heteroplasmy in hypervariable region 1 of the human mtDNA control region. *Am J Hum Genet,* 67**,** 432-43.

UNDERHILL, P. A., SHEN, P., LIN, A. A., et al. (2000) Y chromosome sequence variation and the history of human populations. *Nat Genet,* 26**,** 358-61.

VERNESI, C., CARAMELLI, D., DUPANLOUP, I., et al. (2004) The Etruscans: a population-genetic study. *Am J Hum Genet,* 74**,** 694-704.

VIGILANT, L., STONEKING, M., HARPENDING, H., et al. (1991) African populations and the evolution of human mitochondrial DNA. *Science,* 253**,** 1503-7.

WALLACE, D. C., BROWN, M. D. & LOTT, M. T. (1999) Mitochondrial DNA variation in human evolution and disease. *Gene,* 238**,** 211-30.

WALLACE, D. C., STUGARD, C., MURDOCK, D., et al. (1997) Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. *Proc Natl Acad Sci U S A,* 94**,** 14900-5.

WALLACE, D. C., YE, J. H., NECKELMANN, S. N., et al. (1987) Sequence analysis of cDNAs for the human and bovine ATP synthase beta subunit: mitochondrial DNA genes sustain seventeen times more mutations. *Curr Genet,* 12**,** 81-90.

WHELAN, J. A., RUSSELL, N. B. & WHELAN, M. A. (2003) A method for the absolute quantification of cDNA using real-time PCR. *J Immunol Methods,* 278**,** 261-9.

WILLERSLEV, E. & COOPER, A. (2005) Ancient DNA. *Proc Biol Sci,* 272**,** 3-16.

BIBLIOGRAPHY

WILLERSLEV, E., HANSEN, A. J., BINLADEN, J., et al. (2003) Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science,* 300**,** 791-5.

WILLIS, K. J. & WHITTAKER, R. J. (2000) Perspectives: paleoecology. The refugial debate. *Science,* 287**,** 1406-7.

WILLS, C. (1992) Human origins. *Nature,* 356**,** 389-90.

WOODWARD, S. R., KING, M. J., CHIU, N. M., et al. (1994a) Amplification of ancient nuclear DNA from teeth and soft tissues. *PCR Methods Appl,* 3**,** 244-7.

WOODWARD, S. R., WEYAND, N. J. & BUNNELL, M. (1994b) DNA sequence from Cretaceous period bone fragments. *Science,* 266**,** 1229-32.

XIAO, F. X., YOTOVA, V., ZIETKIEWICZ, E., et al. (2004) Human X-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts. *Eur J Hum Genet,* 12**,** 301-11.

ZINK, A., HAAS, C. J., REISCHL, U., et al. (2001) Molecular analysis of skeletal tuberculosis in an ancient Egyptian population. *J Med Microbiol,* 50**,** 355-66.

ZVELEBIL, M., ZVELEBIL, KV (1988) Agricultural transition and Indo-european dispersal. *Antiquity,* 62**,** 574-583.