

# Ancient DNA and human evolution

**Giorgio Bertorelle<sup>1\*</sup>, David Caramelli<sup>2</sup>, Carles Lalueza-Fox<sup>3</sup>, Cristiano Vernesi<sup>1</sup> and Guido Barbujani<sup>1</sup>**

<sup>1</sup>*Dipartimento di Biologia, Università di Ferrara, via Borsari 46, 44100 Ferrara, Italy*

<sup>2</sup>*Dipartimento di Biologia Animale e Genetica, Università di Firenze, Via del Proconsolo 12, 50122, Firenze, Italy*

<sup>3</sup>*Departament de Biologia Animal, Universitat de Barcelona, Av. Diagonal 645, 08028 Barcelona, Spain*

(\*): Corresponding author. Email: [ggb@unife.it](mailto:ggb@unife.it)

Running title: Human aDNA

## **ABSTRACT.**

In this paper, we shall review some important contributions of aDNA studies to the reconstruction of human evolutionary history. We shall argue that, despite non-negligible problems mainly related to difficult-to-detect DNA damage, contamination by exogenous human DNA and lack of statistical power due to the necessarily limited sample sizes, ancient DNA provides otherwise impossible-to-obtain information on past evolutionary processes. The crucial events since the origin of our species happened in a time frame compatible with the analysis of ancient DNA. Therefore, we expect that aDNA from humans and from species manipulated by humans, especially when typed in several individuals who lived in different time intervals and in different geographic region, will become in the near future an important tool in anthropological, archaeological, and evolutionary studies.

## **INTRODUCTION**

In 2005 we celebrate the 20<sup>th</sup> anniversary of the first analysis of ancient DNA from a human remain (Pääbo, 1985). In the pre-PCR era, a DNA fragment from an Egyptian mummy was cloned and sequenced, and we are all grateful to that paper that started the field of human archeo-genetics shortly after the publication of the first aDNA study in an extinct species (the quagga, Higuchi et al, 1984). Both papers stimulated a formidable boost of evolutionary genetics studies: the genotype, the final target of all evolutionary processes, transmitted in a very simple way through generations, could be identified in ancient remains as had been done in the past for complex (only partially heritable) phenotypic traits. In parallel with, and in support of, evolutionary inference based on contemporary genetic data, ancient DNA study offered for the first time the possibility to obtain and analyse genetic information about past individuals and populations.

Unfortunately, 20 years later, we know that two positions in the first quagga sequence were incorrect, and we do not know if the DNA sequence attributed to the mummy was really from the mummy, or rather from some less ancient human being who touched the mummy (Pääbo et al, 2004). Technical problems with ancient DNA, significantly but not completely reduced with the advent of the PCR technique and more rigorously dealt with by the strict standards later defined (Cooper and Poinar, 2000), remain serious and should be considered seriously. Still, the results obtained in this field in only 20 years are remarkable, and we are very pleased that the referees of the first two ancient DNA papers did not ask the authors to be 100% certain of their results. Had they done that, and had those papers failed to appear in press, chances are that this crucial evolutionary field would still be undeveloped.

After a brief discussion of the main problems in the analysis of ancient human DNA, we shall focus on 3 examples where we believe that the addition of the temporal dimension to the genetic information in human and non-human species was absolutely crucial to clarify otherwise cryptic and challenging phases of our evolutionary history.

## A LOT OF WORK AND MONEY FOR A FEW SUSPICIOUS AND ORDINARY HUMAN SEQUENCES?

We will not discuss the technical details of problems and solutions related to DNA damages. These are already presented in other chapters of this book, and do not directly depend on the species analysed. Preliminary treatments and evaluation of the samples, independent PCRs, and multiple cloning and sequencing of fragments from overlapping regions, are costly and time-consuming, but efficiently reduce the risk of describing nucleotide substitutions in the consensus sequence that were not present in the ancient individual when this individual was alive.

On the other hand, contamination can be a major problem in the study of human remains. Clearly, the simple inspection of sequences allows the identification of a contamination event with modern human DNA when other species are analysed. But, even an ancient human sequence several thousand of years old is not expected to differ from modern sequences much more than random pairs of modern sequences.

Assume for example to sample now and  $\tau$  generations ago a human population with constant effective size of  $N$  females. The expected pairwise number of nucleotide differences between two contemporary neutral mtDNA sequences,  $\pi_c$ , is equal to  $2N\mu$  (where  $\mu$  is the mutation rate per generation). This is of course also the number of differences we expect when modern and ancient samples are compared but recent contamination occurred. When modern and ancient sequences are compared, and contamination did not occur, the expected number of nucleotide difference,  $\pi_{ma}$ , will be  $\tau\mu$  (in the unlikely case of direct ancestry between the two individuals) and  $2N\mu + \tau\mu$  otherwise. The possibility to detect contamination depends on the relative increase of  $\pi_{ma}$  with respect of  $\pi_c$ , i.e.  $(\pi_{ma} - \pi_c) / \pi_c$ , which cannot be larger than  $((2N\mu + \tau\mu - 2N\mu) / 2N\mu) = \tau / 2N$ .

This result, which is interestingly independent from the mutation rate, is telling us that, within a genealogy of a neutral non-recombining gene in a panmictic and stationary population, an ancient sequence can be “spotted” (and contamination excluded) only when its age is large compared to the effective population size. If the population is small, all modern sequences will coalesce early (back in the past), and the lineage of an ancient sequence will have good chances to be independent from (and not embedded in) the modern sequences genealogy. On the other hand, if the population is large, most coalescence would not have occurred before the sampling time of the ancient sequence. This sequence will be therefore embedded in the modern sequences genealogy, and, even if the mutation rate is high (or more nucleotides are typed) it will be very hard, if at all possible, to identify it as a truly ancient sequence or as a modern contaminant.

If  $N = 10,000$  (an effective population size of this order of magnitude has been suggested several times for our species, see e.g. Jobling et al, 2004) and the ancient sample is 2,000 generations old (about 40,000 years), the average divergence expected when modern sequences are compared with genuine ancient sequences ( $\tau / 2N$ ) is only 10% larger than the divergence between modern sequences. If  $\tau = 5,000$  (about 100,000 years), the increase becomes only 25%. The variance of these values is very large, meaning that, in practice, the authentication of a single ancient sequence based on the comparison with modern sequences is not possible even when the age of the sample is as ancient as the colonization of Europe by anatomically modern humans.

Two questions arise from this simple computation: i) how can we exclude contamination? ii) why should we analyse ancient DNA in humans, if not very much occurred in the sequences during most of the short evolutionary history of this species?

Several *criteria of authenticity*, based on a combination of precautions, molecular and biochemical assay, and common sense, have been proposed. The most recent list can be found in Pääbo et al. (2004) and Jobling et al. (2004), and all these criteria are described and discussed in detail elsewhere in this book. Some authors suggest that the criteria of authenticity should not be used, when fulfilled, to favour the hypothesis that the DNA sequence is authentic, but only, when not fulfilled, to exclude this hypothesis (see Abbot, 2003). In other words, at least in human DNA studies, they suggest that even when all the criteria are met, it is impossible to establish the authenticity of a sequence. Beside the name of the criteria, which should be changed to the less positive *criteria to detect contamination*, they are right. Our ancestors had sequences very similar to ours (see above), and therefore it is simply impossible to reject for good the hypothesis that the sequence from an ancient sample is modern. For this reason, for example, Serre et al. (2004) did not provide the DNA sequence of several ancient human samples, but only verified that no Neandertal sequence could be amplified from them. This line of reasoning, which implies that all published human sequences older than a specific (and very difficult to establish) age should not be trusted, has no solution. The only way to be 100% sure that modern humans do not contaminate ancient human remains would be to appoint educated chimps (their contamination can be easily detected) or sophisticated machines for the manipulation of the samples from the excavation to the molecular laboratory.

We still believe that the fulfilment of all the criteria of authenticity of ancient DNA should be used also in human studies to support the endogenous origin of DNA sequences (Barbujani and Bertorelle, 2003). As in almost all scientific areas, absolute confidence in the results cannot be reached. However, the criteria available for ancient DNA studies are many and can be strict, including several levels of cross controls within and between laboratories. As usually happens in science, it is only through successive independent studies that the degree of reliability of the results and the likelihood of the subsequent inferences will increase or decrease until a general consensus can be reached across the scientific community.

So, we know that under a simple demographic model, ancient human sequences within most of the evolutionary history of our species are expected to be very similar to modern sequences, and we also know that excluding contamination with a reasonable degree of confidence is possible, but quite hard. It is therefore obvious to ask:

## WHEN IS IT WORTHWHILE TO STUDY ANCIENT DNA TO UNDERSTAND HUMAN EVOLUTION?

We envisage at least 3 general situations in which the study of ancient DNA is very important, so much so that the problems outlined in the previous section appear minor with respect to the information that would be missed if one decided not to type ancient samples. Each of them will be briefly presented and described with an example in the following sections.

1. Understanding human evolution implies of course studying also distinct, but closely related, taxonomic units. In this case, when the age of the sample is compatible with genetic typing, contamination can be easily detected, and even single ancient sequences can be very important. The analyses of the Neanderthal DNA sequences (Krings et al, 1997; Ovchinnikov et al, 2000; Krings et al, 2000; Schmitz et al, 2002; Serre et al, 2004; Lalueza-Fox et al, 2005), and their comparison with Cro-Magnon sequences (Caramelli et al, 2003; Serre et al, 2004), are paradigmatic of these types of studies and of the huge amount of information they provide. Going to a very recent finding in anthropology, the genetic analysis of the Flores man (Brown et al, 2004), if possible, would doubtless have a large impact on our understanding of human evolution.

2. Human populations share a vast majority of the genetic variation, but still significant differences between them can be found also at short geographic distances (Lewontin, 1972; Cavalli-Sforza et al, 1994; Jobling et al, 2004). Reduced gene flow and strong drift effects during past bottlenecks produced some level of geographic and temporal structure, with the consequent deviation from a simple demographic model whereby very little is changing through time. For example, ancient samples can be different from modern samples in the same area (in terms of variation and/or divergence), pointing to recent migrations, admixture, extinction and colonization events, or demographic crashes. Therefore, when the sample size of the analysed ancient group is not limited to just a few individuals, even the limited degree of differentiation between populations (present, past, and extinct) can be useful for micro-evolutionary inference. Clearly, the size of the ancient sample is crucial in these studies, if one wants to reach a reasonable statistical power. The study of the colonization of the Americas and the Pacific Island, for example, has been approached by ancient DNA analyses (see the review by Kaestle and Horsburgh, 2002), and the number of these type of analyses also in specific human groups is rapidly growing (e.g. Endicott et al, 2003; Vernesi et al, 2004; Maca-Meyer et al, 2004; Lalueza-Fox et al, 2004)

3. Aspects of human cultural habits of the past, their evolution through time and divergence through geographical space, and their influence on the global biodiversity patterns, can be reconstructed using nonhuman ancient DNA. The molecular classification of biological remains found in a site or isolated from coprolites, for example, has been frequently used to infer human diet in specific time-periods (Kaestle and Horsburgh, 2002). But, on top of that, the analysis of nonhuman ancient DNA can be very important to understand innovations in human evolution, such as the domestication of plants and animals (e.g., Jones and Brown, 2000; Bruford et al, 2003). Domestication was not only a cultural process, but also affected the genetic composition of humans (for example, promoting mass migration processes, Cavalli Sforza et al, 1994, Bellwood 2004, or selecting lactose tolerance, Bersaglieri et al, 2004), and drastically modified the biodiversity patterns and distributions of several species. Of course human DNA

contamination can be easily detected in these studies, reducing the costs to reach confidence in the results.

## NEANDERTALS AMONG US?

In 1997, Pääbo's group sequenced for the first time a fragment of the mtDNA hypervariable region of a Neandertal individual (Krings et al, 1997) found in the Feldhofer Cave in the Neander Valley (Germany), and dated to approximately at 40,000 years ago (Schmitz et al, 2002). This study, which established novel strict standards for the analysis of ancient DNA, provided a fundamental contribution to the debate about the origin of anatomically modern humans (amh), and stimulated enormously the supporters of different schools to find additional evidence in favour of their theories or against the opposed views.

Neandertals are anatomically archaic hominids documented in Europe and Western Asia from about 300,000 to 30,000 years ago. They coexisted with anatomically modern humans for several thousand of years, and clearly what occurred between the two groups during this period, and the genealogical relationships between the two human forms, is of great interest and importance to reconstruct our origins.

The multiregional and the Out-of-Africa are two extreme models to explain the origin of amh. In its original formulation, the multiregional model proposed that Neandertals and modern humans are just one population observed at different times (Wolpoff, 1984). Under this theory, modern humans evolved in the last two million of years as a single polytypic species, and the transition from *erectus* to *sapiens* took place in different area at different times. Conversely, the Out-of-Africa model (Stringer and Andrews, 1988) suggests that amh first arose in Africa less than 200,000 years ago, and then dispersed replacing archaic forms such as Neandertals. Intermediate models are of course possible, but it would be unfair to consider any non-African contribution larger than 0.0 % to the modern gene pool as supporting the multiregional model. Small non African contributions (say, <5-10%) should be still regarded as supporting the Out-of-Africa model, whereby almost all the human genetic variation originated in Africa. For the sake of clarity, the fundamental opposition of the two classical models is usually maintained.

Neanderthal DNA is obviously very important to reject either of these hypotheses. Contrary to the Out-of-Africa model, the multiregional hypothesis predicts in fact that Neanderthals, being a step within a single evolving species, contributed to our genetic composition and some of their DNA lineages should be found in modern samples. At present, the only testable prediction concerns the presence of Neanderthal mitochondrial sequences in modern humans.

The Feldhofer specimen had a mtDNA sequence never found in modern samples. On the average, about 30 mutations separate this sequence from modern human sequences (the average among humans is about 8). No special relationship was apparent between Neandertals and modern Europeans who, under the multiregional model, would be their direct descendants. The divergence between Neandertals and modern humans lineages was estimated by Krings et al. (1997) between 550,000 and 690,000 years (only slightly increased or decreased in successive studies, Krings et al, 2000; Beerli and Edwards, 2002), about three to four times older than the estimated most recent common ancestors of modern mtDNA lineages. This result suggested that Neandertals are not among us, or,

as stated by Krings et al. (1997), “Neandertals went extinct without contributing mtDNA to modern humans”.

As usually for influential paper, Neandertal DNA data significantly amplified the discussion around the different theories. Taken by the Out-of-Africa supporters as another nail in the coffin of the multiregional model, the Krings et al. (1997) paper stimulated theoretical work to show how the Neandertal DNA sequence can be accommodated within the multiregional model (e.g, Nordborg, 1998; Gutierrez et al, 2002; Relethford, 2001, Templeton, 2002). For example, Nordborg (1998) argued that the level of mtDNA divergence between Neandertal and modern humans is unlikely for a single population genealogy (though the tree topology is not), but interbreeding between two previously differentiated groups cannot be excluded. On the basis of the available data, Nordborg (1998) concluded that Neandertal mtDNA lineages could have been lost by random drift or selection after the Neandertal-modern humans mixing, implying also a very high probability that Neandertal contributed to our nuclear genome.

Eight additional mtDNA sequences typed in Neandertal individuals (Ovchinnikov et al, 2000; Krings et al, 2000; Schmitz et al, 2002; Serre et al, 2004; Lalueza-Fox et al, 2005) were found to cluster with the first Feldhofer sequence, definitely excluding the single population topology and failing to detect any evidence of female gene flow from amh to Neandertals. However, if we exclude the still problematic analysis of nuclear markers, the only way to test directly the more relevant hypotheses that Neandertal did contribute or not to our genetic make-up seemed to be the mtDNA analysis of early amh who lived in the same time period as Neandertals (Stringer and Davies, 2001; Gibbons, 2001). If this introgression occurred 30,000 years ago, the chances to detect it are higher when drift or selections did not have much time to delete its signature, i.e. studying individuals chronologically close to the Neandertals.

The relevant study was published two years ago by Caramelli et al. (2003). Two anatomically modern individuals of the Cro-Magnon type, recovered in Southern Italy and dated at about 23 and 25 thousand years ago, were typed, and all the criteria of authenticity suggested that contamination by modern DNA did not occur. The mtDNA sequences of these individuals appeared fully compatible with the variation observed in modern samples, whereas it was as distant as modern ones from those of the almost contemporary Neanderthals (see Figure 1). Last year, five additional early modern humans of similar age were analyzed, and none of them yielded “Neandertal-like” sequences (Serre et al, 2004). All these results, together with the recent suggestion, based on a simulation study that the maximum interbreeding rate between amh and Neandertals in Europe should have been smaller than 0.1% (Currat and Excoffier, 2004), seems to add the last-but-one nail in the multiregional theory coffin. The last one is expected from the analysis of nuclear loci.

## **PREHISTORIC POPULATION GENETICS: THE ETRUSCANS**

In the previous section we have seen that even single, or a few, ancient DNA sequences can provide fundamental information when the question we are asking implies the comparison between highly divergent groups or sequences. But we have also remarked that additional typing of Neandertals and early modern humans allows the shift from a phylogenetic to a population genetics approach, with the consequent increase of statistical power.

When ancient DNA is used to infer the structure and the evolution of past human populations, the sample size becomes crucial. Single mtDNA sequences can be used sometimes in individual, paternity, family, and population assignment (or, better, exclusion) problems (Kaestle and Horsburgh, 2002), but useful information to compare human populations (present and past) can be obtained only after the assessment of the genetic variation both within and between groups. Despite the enormous effort necessary to describe genetic variation in ancient human populations, several studies are starting to use this approach with reasonable sample sizes to analyse significant processes occurred since the Neolithic. Lalueza-Fox et al. (2004), for example, elucidated some aspects of the migration patterns in Central Asia studying 36 individuals dated between 1,500 and 3,500 years ago, and Yao et al (2003), reanalyzing 58 sequences from samples dated between 2,000 and 2,500 years ago, were able to propose a demographic scenario to explain the changes in the matrilineal genetic structure in a Chinese region. Here we will briefly describe the results we obtained (Vernesi et al, 2004) typing the remains from a pre-historical population dated at about 2,500 years ago: the Etruscans.

Etruscans were a non-Indo-European-speaking population of pre-classical Italy. Their culture developed about 3,000 years ago, and their organization in independent cities established a political and cultural leadership over an approximately triangular area spanning several hundreds of kilometres in Central Italy. The decline begun during, and probably as a consequence of, the Roman expansion. The Etruscan language disappeared about 2,000 years ago.

Two important questions regarding this population could be addressed using ancient DNA. First, were the Etruscans a panmictic group over their distribution range, or were they an assemblage of genetically different groups sharing a language and a culture? Second, was the frequently suggested Eastern Mediterranean influence on the Etruscans a cultural process or did it involve also some level of gene flow?

Starting from 80 well preserved skeletons from 10 necropoleis, the application of the criteria of authenticity reduced the final number of validated mtDNA sequences to 27, from 6 necropoleis. Some of these sequences have been questioned and suspected of DNA damage (Bandelt, 2004; Malyarchuk and Rogozin, 2004), but we believe that the current knowledge of the mutational process at mtDNA and the strict application of the standard criteria in our study do not justify these criticisms (Barbujani et al, 2004). In addition, even a few potentially non-genuine substitutions in a few sequences would not modify the general conclusions of this study.

As commonly observed in mtDNA samples, most of the Etruscan individuals had different sequences (the haplotype network of the different sequences is reported in Figure 2). However, even if some geographic structuring seems to emerge in relation to the 14766 *MseI* restriction site, individuals from the same archaeological site were not statistically closer to individuals from different sites (about 5 substitutions on the average for both comparisons). In other words, the molecular variation between necropoleis was not statistically larger than the variation within them, suggesting that migration between Etruscan cities was sufficient to prevent genetic divergence. When the Etruscan population was compared to 19 modern populations from Europe and the Near East (almost 2,000 individuals), and the genetic distance was measured combining with equal weight two different indexes of population divergence (molecular  $F_{st}$  and net nucleotide divergence), the first three closest groups resulted, in order, Tuscans, Bulgarians, and Turks.

Tuscans live now where Etruscans did 2,500 years ago, and the fact that their resemblance to Etruscans is larger than any other population was therefore expected. Unexpected, on the other hand, was the fact that Tuscans and Etruscans, geographically overlapping and temporally separated by “only” about 100 generations, share only 2 sequences. The high rate of extinction of the Etruscan sequences could be due to their displacement by related lineages, possibly during the decline of this civilization. Ongoing typing of other ancient Italian populations (unpublished results) seems to indicate that such deviation from the simple model of genealogical continuity from prehistoric times is not an isolated case, making the genetic study of ancient populations even more exciting. Finally, the affinity between Etruscans and Eastern populations (especially Turks) suggests that the documented trading routes of the Etruscans across the Mediterranean Sea were possibly accompanied by interbreeding that left a detectable signature in the Etruscans’ genomes.

## **NONHUMAN DNA: STUDYING CATTLE DOMESTICATION TO UNDERSTAND NEOLITHIC BREEDERS**

The shift from hunting-gathering lifestyle to agricultural means of food production, the Neolithic transition, had a dramatic impact on the history of our, as well as many other, species. We can consider it as one of the most momentous and influential innovations occurred since the emergence of *Homo sapiens*.

The process of Neolithic transition was an authentic demographic and cultural revolution, accompanied by a dramatic set of genetic changes, in humans but also in nonhuman species, which can by no means be overlooked. Demographic expansions in farming groups and bottlenecks in others, mass migrations to colonize new fields to cultivate, fragmentation related to settling and/or new cultural (for example, language) barriers, diet modification, increase in infectious disease prevalence, selective breeding and trades of animal and plants, are just some events which were prompted by the Neolithic changes in subsistence technologies (Bellwood, 2004) and are expected to have genetic consequences. Genetic data can be used, in turn, to reconstruct aspects of these processes that are not sufficiently documented in the archaeological records. Several genetic analyses of modern samples confirm that levels, patterns, and geographic distributions of both neutral and adaptive genetic variation in humans and domesticated species were affected by the Neolithic revolution (Cavalli-Sforza et al, 1994; Jobling et al, 2004). The application of ancient DNA techniques to understand some aspects of this process seems therefore very promising in domesticated species, also because less complex controls to exclude human contamination are required. Here we shall briefly discuss the results obtained typing the wild ancestor of the cattle, the aurochs.

The process of cattle domestication from aurochs (*Bos primigenius*) ancestors started about 10,000 years ago. The deep genetic divergence (>100,000 years ago) between taurine (*B.taurus*) and zebu (*B. indicus*) completely interfertile cattle breeds points to at least two independent domestication events from two distinct aurochs groups (see Figure 3).

Archaeological data indicate that Middle-East or Anatolia and Pakistan were the primary site of taurine and zebu cattle domestication, respectively (Clutton-Brook, 1999). This hypothesis implies that present day European breeds, which belong to the taurine type,

would be all descend from Fertile Crescent ancestors. However, the aurochs, widespread in Europe and Western Asia, got extinct no more than 400 years ago and hence modern (*Bos taurus*) and archaic (*Bos primigenius*) forms coexisted for millennia during which they had the opportunity to interbreed. A major question regarding the cattle domestication, as well as the behaviour and life-style of early cattle breeders, is what happened during the spread of selected breeds, pastoralism, and farmers from the major centre of origin.

The first genetic study on the aurochs DNA date back to 1996 (Bailey et al, 1996), one year before the publication of the first Neandertal DNA sequence. Interestingly, the results of the two studies were, in some sense, analogous. As Neandertal sequences were very different from all modern humans, the sequences of a mtDNA fragment from two British aurochs were very different from those observed in several hundreds of modern cattle from different breeds (see Figure 3). If we exclude the sometimes suggested possibility that initial DNA studies were biased towards divergent sequences because modern-like were considered contaminations (Trinkaus, 2001), and we assume that the British aurochs sequences represent well the genetic composition of this species in Europe, this result can be (and was) interpreted as the result of the analysis of Neandertal specimens: neither archaic European populations, aurochs and Neanderthals, contributed to present-day genetic variation in cattle and humans. In contrast with the human case, however, the zero contribution hypothesis is much more strongly supported by the data for cattle. In fact, the aurochs got extinct at a time when modern cattle lineages were numerous, which makes it highly unlikely that drift could have removed all aurochs sequences from the modern gene pool

In other words, ancient DNA, corroborated by the observed patterns of modern DNA genetic variation (Bailey et, 1996; Loftus et al, 1999; Troy et al, 2001), could be taken as a strong evidence for a single Near Eastern origin of the European cattle breeds, followed by their diffusion during the Neolithic without introgression from local wild animals. In turn, this would also imply that the domestication process was already very advanced at the time of the first Neolithization of Europe, and cattle breeders were therefore never interested, and were also able to avoid, introgressions into their selected herds of genomes from local aurochs commonly found in the wild.

However, all aurochs typed so far at the genetic level come from a single geographically restricted area in Northern Europe, and significant differences between regions of the same continent are usually observed in modern ungulates (e.g., Vernesi 2002). We therefore decided to type five Italian aurochs specimens dated between 7,000 and 17,000 years BP. Preliminary results (unpublished) suggest that all these sequences fall within the range of variation of modern cattle, with no resemblance with the British aurochs. In addition, Italian aurochs and Italian modern breeds have the same most common allele with very similar frequencies, about 60%, but the same allele has only a frequency of about 30% in Anatolian and Middle Eastern breeds.

As is, this result would suggest a direct ancestry of European breeds from Italian aurochs. However, Middle Eastern aurochs, unfortunately not typed yet, are expected to be very similar to European breeds as well, both because archeological data suggest that the major center of domestication for European *Bos taurus* breeds was the Fertile Crescent (Clutton-Brook, 1999), and also because the mtDNA sequence from a Syrian specimen attributed to a cattle and dated at 8,000-9,000 years ago (Edwards et al, 2004) (that is, to an individual chronologically very close to its wild ancestors), shows a typical European haplotype found both in modern breeds and the Italian aurochs. In

conclusion, the absence of British aurochsen alleles in modern European breed can no longer be taken as evidence for the absence of European aurochs introgression. On the contrary, some evidence seems to suggest that small to moderate levels of local gene flow from wild *Bos primigenius* females in selected breeds were either accepted or maybe enforced by Neolithic breeders.

## PERSPECTIVES

Cooper and Poinar (2000) suggested, when dealing with ancient DNA, to “do it right, or not at all”. This suggestion is clearly valid in all fields of science, since nobody can possibly be against care and strive for excellence. In this review we tried to make two points, namely: (a) that in ancient DNA studies, and particularly when the subjects are humans, 100 per cent confidence that the sequence is accurate and genuine is impossible to reach, at present and in the foreseeable future; (b) that ancient DNA data offer a wholly new perspective, and precious scientific evidence, on evolutionary episodes that in the past could not be studied directly and about which our knowledge was purely conjectural. Our conclusion is that one should be fully aware of both points. “Replicability” seems the key-word here. Given the importance of ancient DNA data, scientists should do as much as they can to obtain data that can be independently replicated. When that happens, and when data were generated following carefully all available standards, we see no good scientific reason to dismiss their importance. A “highly sceptical attitude to one’s own work” (Pääbo et al, 2004) is doubtless important, but, after an initial phase of unjustified optimism in this discipline (Cooper and Poinar, 2000), when published artefacts were not uncommon, it would make little sense to jump to the opposite extreme and embrace a radically pessimistic view, in the absence of evidence that that pessimism is justified. The advances in understanding the evolutionary history of several species have been enormous in only 20 years of aDNA studies, and we expect that more will come in the future.

The uncertainties regarding the analysis of nuclear DNA and the limited sample sizes are the two major limitations of present aDNA studies. There are many mitochondria in each cell, each containing several DNA copies. Especially in poorly preserved ancient samples, the chances to recover amplifiable DNA molecules is much higher for the mitochondrial than for the nuclear genome. Therefore, typing aDNA nuclear sequences is technically challenging, and at presents hardly feasible; however, there is no doubt that progress in this area would prompt a number of extremely important applications. Coding genes can be used to understand selective processes, and analysis of multiple independent non coding regions would dramatically increase the statistical power to reconstruct demographic and historical events. The possibility to analyse the sequence of a gene (and possibly the frequency of different alleles) before and after the occurrence of a selective event could clarify the role of that gene in adaptation. For example, lactose tolerance is associated with a mutation located 13.910 kb upstream the lactase gene, and understanding the dynamic through time of this mutation could be used to test directly the hypothesis of a strong selective effect associated with pastoralism development in some populations (Beja-Pereira et al, 2003; Bersaglieri et al 2004). Similarly, the temporal distribution of mutations associated to diseases, especially when heterozygotes are suspected to have, or have had, an advantage (for example at CFTR or G6PD loci), could be important to reconstruct the tempo and mode of the selection process. Nuclear DNA from extinct Neanderthals could provide invaluable data to identify the genes involved in the development of the cognitive abilities of modern humans (Bishop. 2002). Some

nuclear aDNA studies, mainly, but not only, in nonhuman species, have been published recently (Poinar et al, 2003; Bramanti et al., 2003; Huynen et al, 2003; Ricaut et al, 2004; Jaenicke-Despres et, 2004). The technical problems related with the difficulty to identify contamination, and with the high risk to generate experimental artifacts when there are but a few amplifiable molecules, are undeniable and, at present, there is no consensus on the reliability of ancient nuclear data on humans. Still, the potential importance of nuclear data is such that serious efforts in this area are a crucial research priority. In addition, both for nuclear and mtDNA analysis, sequences of single individuals will not suffice to understand population processes. Comparisons across time between modern and ancient populations will be often inconclusive if the genetic variation within and between ancient groups will not be, at least approximately, estimated. Therefore, whereas we stress that the available aDNA methodologies have already produced empirical evidence of enormous relevance, additional methodological advances and efforts are indispensable if we are to address the main emerging questions.

## BIBLIOGRAPHY

Abbot A (2003): Anthropologists cast doubt on human DNA evidence. *Nature* 423:468.

Bandelt H-J (2004): Etruscan artefacts *Am J Hum Genet* 75:919-920.

Barbujani G, Bertorelle G (2003): Were Cro-Magnons to like us for DNA to tell? *Nature* 424:127.

Barbujani G, Vernesi C, Caramelli D, Castri L, Lalueza-Fox C, Bertorelle G (2004): Etruscan artefacts: much ado about nothing. *Am J Hum Genet* 75:923-927.

Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, Bertorelle G, Chamberlain AT, Nunes TP, Metodiev S, Ferrand, N, Erhardt G (2003): Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat Genet* 35:

Bailey JF, Richards MB, Macaulay VA, Colson IB, James IT, Bradley DG, Hedges RE, Sykes, BC (1996): Ancient DNA suggests a recent expansion of European cattle from a diverse wild progenitor species. *Proc R Soc Lond B Biol Sci* 263,1467-1473.

Beerli P, Edwards SV (2002): When did Neanderthals and modern humans diverge? *Evol Anthropol Suppl* 1:60-63.

Bellwood P (2004): *First farmers. The origins of agricultural societies.* Blackwell Publishing (Oxford, UK)

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004): Genetic signatures of strong recent positive selection at the lactase gene *Am J Hum Genet.* 74:1111-20.

Bishop DVM (2002): Putting language genes in perspective. *Trends Genet* 18:57-59

Bramanti B, Hummel S, Chiarelli B, Herrmann B (2003): Ancient DNA analysis of the delta F508 mutation. *Hum Biol* 75:105-15.

Brown P, T. Sutikna M J, Morwood R P, Soejono, Jatmiko, Saptomo EW, Due RA. (2004): A new small-bodied hominid from the Late Pleistocene of Flores, Indonesia. *Nature* 431: 1055-1061.

Bruford MW, Bradley DG, Luikart G (2003): DNA markers reveal the complexity of livestock domestication. *Nat Rev Genet.* 4:900-10.

Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, Mallegni F, Chiarelli B, Dupanloup I, Bertranpetit J, Barbujani G, Bertorelle G (2003): Evidence for a genetic discontinuity between Neandertals and 24,000-year-old anatomically modern Europeans. *Proc Natl Acad Sci U S A.* 100:6593-6597.

Cavalli-Sforza LL, Piazza A, Menozzi P (1994): *The history and geography of human genes.* Princeton University Press (Princeton, NJ).

- Clutton-Brook J (1999): *Natural History of Domesticated Mammals*. Cambridge University Press. (Cambridge, UK).
- Cooper A, Poinar HN (2000): Ancient DNA: Do it right or not at all. *Science* 289:1139.
- Currat M, Excoffier L (2004): Modern humans did not admix with Neanderthals during their range expansion into Europe. *PLOS Biology* 2:1-10.
- Edwards CJ, MacHugh DE, Dobney KM, Martin L, Russell N, Horwitz LK, McIntosh SK, MacDonald KC, Helmer D, Tresset A, Vigne J-D, Bradley DG (2004): Ancient DNA analysis of 101 cattle remains: limits and prospects. *J Archaeol Sci* 6:695-710.
- Endicott P, Gilbert MT, Stringer C, Lalueza-Fox C, Willerslev E, Hansen AJ, Cooper A (2003): The genetic origins of the Andaman Islanders. *Am J Hum Genet* 72:178-184.
- Gibbons A (2001): The riddle of coexistence. *Science* 291:1725-9.
- Gutierrez G, Snachez D, Marin A (2002): A reanalysis of the ancient mitochondrial DNA sequences recovered from Neandertal bones. *Mol Biol Evol* 19:1359-1366.
- Higuchi R, Bowman B, Freiberger M, Ryder OA, Wilson AC (1984): DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312:282-284.
- Huynen L, Millar CD, Scofield RP, Lambert DM (2003) Nuclear DNA sequences detect species limits in ancient moa. *Nature* 425:175-8.
- Jaenicke-Despres V, Buckler ES, Smith BD, Gilbert MT, Cooper A, Doebley J, Paabo S (2004): Early allelic selection in maize as revealed by ancient DNA.
- Jobling MA, Hurler ME, Tyler-Smith C (2004): "Human Evolutionary Genetics". Garland Science (New York).
- Jones M, Brown T (2000): Agricultural origins: the evidence of modern and ancient DNA. *Holocene* 10:769-776.
- Kaestle FA, Horsburgh KA (2002): Ancient DNA in anthropology: methods, applications, and ethics. *Y Phys Anthropol* 45:92-130.
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S (1997): Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19-30.
- Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, von Haeseler A, Grossschmidt K, Possnert G, Paunovic M, Paabo S (2000): A view of Neandertal genetic diversity. *Nature Genetics* 26:144-146.
- Lalueza-Fox C, Sampietro ML, Gilbert MT, Castri L, Facchini F, Pettener D, Bertranpetit J (2004): Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc R Soc Lond B Biol Sci.* 271:941-7.
- Lalueza-Fox C, Sampietro ML, Caramelli D, Puder Y, Lari M, Calafell F, Martinez-Maza C, Bastir M, Fortea, J, de la Rasilla M, Bertranpetit, J, Rosas A (2005): Neandertal

evolutionary genetics; mitochondrial DNA from the Iberian Peninsula. *Mol Biol Evol.* In press.

Lewontin RC (1972): The apportionment of human diversity. *Evol Biol* 6:381-398.

Loftus RT, Ertugrul O, Harba AH, El-Baroyd MAA, Machugh DE, Park SDE, Bradley DG (1999): A microsatellite survey of cattle from a centre of origin: the Near East. *Mol. Ecol.* 8, 2015-2022.

Maca-Meyer N, Arnay M, Rando JC, Flores C, Gonzalez AM, Cabrera VM, Larruga JM (2004): Ancient mtDNA analysis and the origin of the Guanches. *Eur J Hum Genet.* 12:155-62.

Malyachuk BA, Rogozin IB (2004): On the Etruscan mitochondrial DNA contribution to modern humans. *Am J Hum Genet* 75:920-923.

Nordborg M (1998): On the probability of Neanderthal ancestry. *Am J Hum Genet.* 63:1237-1240.

Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, Goodwin W (2000): Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490-493.

Pääbo S (1985): Molecular cloning of ancient Egyptian mummy DNA. *Nature* 314:644-645.

Pääbo S, Poiner H, Serre D, Jaenicke-Dupreés V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, Hofreiter M (2004): Genetic analyses from ancient DNA. *Ann Rev Genet* 38:645-679.

Poinar H, Kuch M, McDonald G, Martin P, Pääbo S (2003): Nuclear gene sequences from a Late Pleistocene sloth coprolites. *Curr Biol* 12-1150-52.

Relethford JH (2001): Absence of regional affinities of Neandertal DNA with living humans does not reject multiregional evolution. *Am J Phys Antropol* 115:95-98.

Ricaut FX, Fedoseeva A, Keyser-Tracqui C, Crubezy E, Ludes B (2004) : Ancient DNA analysis of human Neolithic remains found in northeastern Siberia. *Am J Phys Anthropol* [Epub ahead of print].

Schmitz RW, Serre D, Bonani G, Feine S, Hillgruber F, Krainitzki H, Paabo S, Smith FH (2002): The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc Natl Acad Sci U S A.* 99:13342-13347.

Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Menecier P, Hofreiter M, Possnert G, Pääbo S (2004): No evidence of Neandertal mtDNA contribution to early modern humans. *PLOS Biology* 2:313-317.

Stringer CB and Andrews, P. Genetic and fossil evidence for the origin of modern humans. *Science* 239, 1263-1268 (1988).

Stringer C, Davies W (2001): Archaeology. Those elusive Neanderthals. *Nature* 413:791-792.

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

Trinkaus E (2001): The Neanderthal paradox. In: Finlayson C (ed): "Neanderthals and Modern Humans in Late Pleistocene Eurasia" pp.73-74. Gibraltar Museum (Gibraltar)

Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlein AT, Sykes BC, Bradley DG (2001): Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410: 1088-1091.

Vernesi C, Pecchioli E, Caramelli D, Tiedemann R, Randi E, Bertorelle G (2002): The genetic structure of natural and reintroduced roe deer (*Capreolus capreolus*) populations in the Alps and central Italy, with reference to the mitochondrial DNA phylogeography of Europe. *Mol Ecol* 11, 1285-97.

Vernesi C, Caramelli D, Dupanloup I, Bertorelle G, Lari M, Cappellini E, Moggi-Cecchi J, Chiarelli B, Castri L, Casoli A, Mallegni F, Lalueza-Fox C, Barbujani G (2004): The Etruscans: a population-genetic study. *Am J Hum Genet* 74:694-704.

Wolpoff MH, Wu X, Thorne A (1984): Modern Homo sapiens origins: A general theory of hominid evolution involving the fossil evidence from East Asia. In: Spencer F (ed): "The origins of modern humans: A world survey of the fossil evidence" p. 411-483. Alan Liss (New York).

Yao YG, Kong QP, Man XY, Bandelt HJ, Zhang YP (2003): Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. *Mol Biol Evol* 20:214-219.

## Figure Legends

Figure 1. Multidimensional scaling analysis representing present day mtDNA sequences (filled squares, selected to represent the most divergent lineages from all around the world), two anatomically modern humans dated at about 25,000 years ago (open squares), and 4 different Neandertal sequences (observed among the 9 Neandertal individuals typed so far).

Figure 2. Reduced median network of 23 different sequences identified in Etruscan individuals. Mutations are numbered relative to the Cambridge Reference Sequence. Sites undergoing recurrent mutations are underlined. Different filling patterns refer to different necropoleis.

Figure 3. Multidimensional scaling analysis of mtDNA sequences selected to represent cattle and British aurochs variation.

Figure 1.

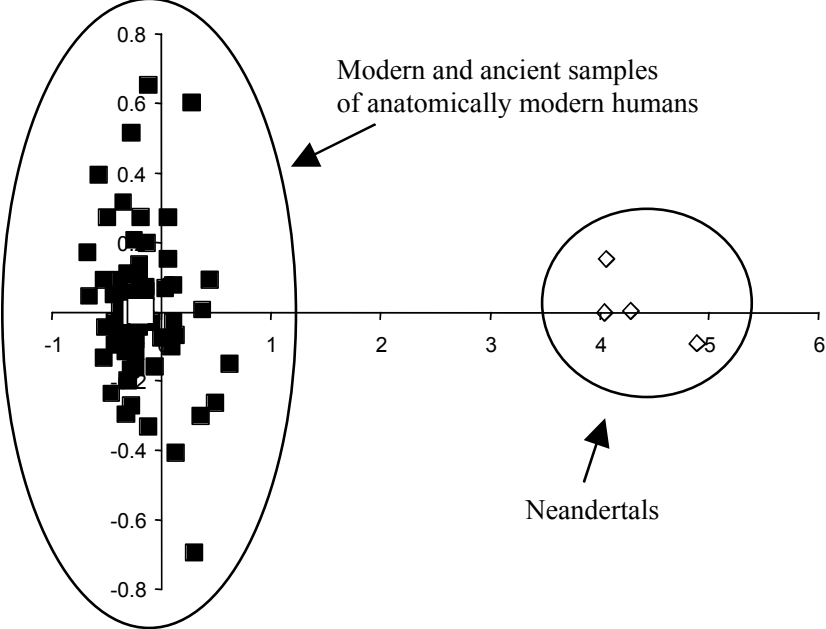


Figure 2.

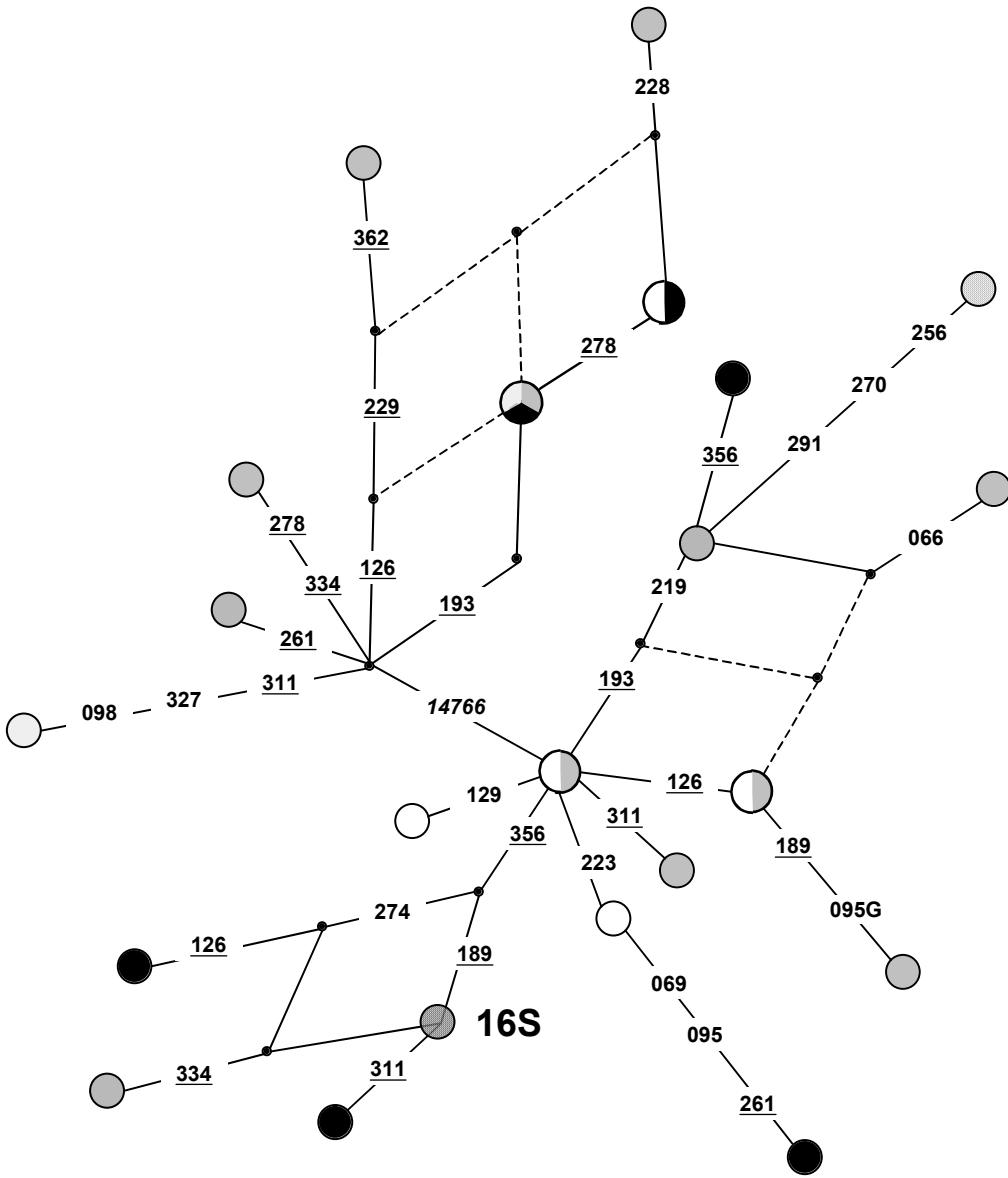


Figure 3

