# DNA of prehistorical fossils: preliminary results in the Eastern Alps 

GIULIETTA DI BENEDETTO, MICHELE STENICO, LOREDANA NIGRO, MICHELE LANZINGER \& GUIDO BARBUJANI


#### Abstract

The population of the Eastern Italian Alps has been shown to be extremely variable at the mitochondrial DNA level. High sequence diversity is apparent both within populations, as has been observed in other European samples, but also between populations, especially when they belong to different linguistic groups. In particular, mitochondrial lineages that are present at low frequencies across much of Europe represent more than two-thirds of the sequences identified in two samples of Ladin speakers. Those lineages have probably originated in Palaeolithic times, although this does not necessarily mean that they spread in the Alps in the same moment.


KEY WORDS: DNA; Eastern Alps; Mitochondrial analyses
PAROLE CHIAVE: DNA; Alpi orientali; Analisi mitocondriale

Giulietta Di Benedetto, Michele Stenico \& Loredana Nigro - Dipartimento di Biologia, Università di Padova, V.le Colombo 3, I-35100 Padova

Guido Barbujani - Dipartimento di Biologia, Università di Ferrara, Via Borsari 46, I-44100 Ferrara Michele Lanzinger - Museo Tridentino di Scienze Naturali, Via Calepina 14, I-38100 Trento

## 1. INTRODUCTION

When a population, because of geografical or cultural barriers, becomes subdivided in two or more populations, genetic differences start to accumulate between them, the more rapidly the more the new populations are isolated. The analysis of such differences, as they appear for example in a small region of mithocondrial DNA, can give an estimate of the genetic distance between the populations involved, and therefore allow inferences on the possible timing of their separation.

Mithocondria are small organelles, present in several copies in the cytoplasm, each containing several copies of a molecule of circular DNA, about 16000 base pairs (bp) long. They are transmitted maternally, and this makes the analysis simpler, since the lineage to be reconstructed is only the maternal one. Because of that, and since the mutation rate is much faster in mithocondrial than in nuclear DNA, mithocondria are especially suitable
for the analysis and the phylogenetic reconstruction of recent events like those concerning human populations.

In a recent study (Stenico et al., 1996, 1998) the genetic structure of Ladin speaking populations living in the Eastern italian Alps has been analysed, to check if linguistic relatedness is accompanied by genetic similarity (wich would suggest a common origin), or, alternatively, if the Ladin language is spoken by genetically heterogeneous populations.

A sequence of 360 bp (first hypervariable region of the D-loop) has been analysed in 63 La din speaking individuals sampled in four different places (Val Badia, Val di Fassa, Val Gardena, and Colle Santa Lucia), and in Italian and German speaking individuals belonging to some neighbouring communities. It clearly appeared that the Eastern Alps populations are extremely variable at the mithocondrial DNA level, showing high variability both within populations, as has been observed in other European samples, but also between populations, especially when they belong to different
linguistic groups. In particular, Ladin populations are characterized by an extremely high frequency of a haplogroup (group 2, according to Richards et al., 1996) that is present everywhere in Europe, but at a much lower frequency. Similar frequencies are found only in the Near East, where this haplogroup could have originated about 57000 years BP (Richards et al., 1996).

A possible development of this work concerns the genotype analysis of differently aged fossils remains, to identify a minimum date after which the group 2 is present in the region and we started a project in collaboration with Svante Paabo of the University of Munich

## 2. METHODS AND RESULTS

Studying ancient DNA became possible in the last ten years because of the introduction of the PCR (Polymerase Chain Reaction), that allows the amplification of even minimum amounts of DNA, making it analysable. Indeed, being apart from the methabolism and the repair process that occur in living organisms, DNA quickly decays, because of the accumulation of hydrolytic and oxydative damages (Hoss et al., 1996). Moreover, when some DNA can be extracted, the main problem is the experimental reproducibility; there is a high risk of contamination by non ancient, or anyway exogenous, DNA. Finally, if the endogenous DNA is present in the extract in minimum amounts, it is possible that even the errors in the first PCR cycles are amplified, so that every attempt to analyse the sample will be irreproducible. Therefore, in order to judge endogenous the sequence obtained, some strict criteria have to be satisfied in the laboratory work (HANDT et al., 1994, 1996).

For this study, first we analysed teeth belonging to five individuals with respect to the amino acids racemization. It has been shown that ancient samples with high racemization level generally do not contain enough DNA (Poinar et al., 1996). We removed from each sample 10 mg of powdered tissue, then hydrolysed it in acid conditions, and we analysed the released amino acids by HPLC (High Performance Liquid Chromatography).

In all samples the ratio between the two enantiomeres D and L of the aspartic acid resulted compatible with the possibility to recover some DNA in good conditions (D/L Asp $£ 0.1$ ) (Poinar et al., 1996; Krings et al., 1997). Aspartic acid is the most indicative among the aminoacids in such kind of analysis, because its racemization speed is compa-
rable to the DNA depurination speed.
In a second phase DNA was extracted twice indipendently from each sample, and for every extract, we amplified the first hypervariable region (HVR I) of the mithocondrial D-loop. This region is 360 bp long, but, as it was known from previous experiences that ancient DNA is so damaged that it does not allow amplification products longer than 100-200 bp (PaAbo et al., 1989), the whole region was subdivided in five overlapping amplifications. Moreover, the number of template molecules in the extract was evaluated by a quantitative PCR. Indeed, amplifications starting from approximatively a thousand molecules normally give reproducible results, whereas amplifications starting from a few hundred molecules, or even less, often produce variable results among different experiments; this is due to sporadic contamination and, mostly, to errors in the first PCR cycles.

From each sample analysed it was possible to recover some DNA, with one exception (Fiavè sample) which, despite the HPLC result, did not give any amplifiable DNA. The DNA amount of the remaining four samples was quantified (what has been determined is the approximative number of molecules in each amplification, that is in 5 ml of extract), and the results are the following:

|  | Vatte | Mezzocorona, Villabruna, Borgo Nuovo |
| :--- | :---: | :---: |
| I extract | undetermined | $4000<$ \# of molecules < 20000 |
| II extract | some hundred | $4000<$ \# of molecules < 20000 |
|  | molecules |  |

Then the amplified regions were cloned in plasmids and about ten clones for every amplification were sequenced, to determine with high confidence the true sequence of each sample. In spite of the quantitation discouraging results, we sequenced even the amplification products obtained from the Vatte sample, which further confirmed the impossibility to obtain unambigous and reproducible results if the number of starting molecules is too low. An example of the ambiguity of the sequences obtained is shown in Fig.1. On the contrary, the sample called Mezzocorona, that seemed to contain many molecules in both extracts, gave unambiguous and reproducible sequences. In particular, we sequenced two fragments of the HVR I of the mithocondrial D-loop containing the "Ladin motif" (that is, the mutations characteristic of the haplotypes with high frequency among Ladinspeaking people), and the Mezzocorona sample appeared to belong to the subgroup 2B (Richards et al., 1996). The sequences are presented in Fig.2.

The DNA obtained from the samples called Villabruna and Borgo Nuovo has been amplified and cloned, and its sequencing is now in progress.

## 3. DISCUSSION

These are certainly preliminary results; indeed, for some samples the sequencing is still in
progress, and for others we have not yet started the analysis. It is anyway interesting that in the only sample of which, by now, the sequence is known, dated to more than 5600 years BP, we found mutations characteristic of the haplogroup 2. Since this sample dates back to a time preceding the Neolithic migration wave in the Alps, we can say that in the Ladin-speaking populations of the Alps are persisting ancestral characteristics, probably originated in Palaeolithic times.

SUMMARY - Analysis of genetic variation, both in present and past populations, is a way to reconstruct otherwise cryptic aspects of human history. The population of the Eastern Italian Alps has been shown to be extremely variable at the mitochondrial DNA level. High sequence diversity is apparent both within populations, as has been observed in other European samples, but also between populations, especially when they belong to different linguistic groups. In particular, mitochondrial lineages that are present at low frequencies across much of Europe represent more than two-thirds of the sequences identified in two samples of Ladin speakers (Stenico et al., 1996). Those lineages have probably originated in Palaeolithic times (Richards et al., 1996), although this does not necessarily mean that they spread in the Alps in the same moment. To identify a probable, minimum date for the presence of such lineages in the Eastern Alps we tried the extraction and typization of a short segment of the mitochondrial hypervariable region in five human remains from the Alps. The estimated age of the specimens varies between 11000 and 4000 years BP. From some of these specimens it was impossible to obtain any results, either because we did not retrieve enough DNA, or because of contamination, is one of the biggest problems in ancient DNA studies. Anyway, from three samples we were able to recover enough endogenous DNA, and for one of them we sequenced a 360 bp segment. Our results suggest that this individual belongs to one of the lineages present at high frequency among the Ladin speakers. Sequencing of the two others samples is in progress. A comparative analysis of modern and ancient mitochondrial lineages may allow inferences on the evolutionary and demographic processes occurring in the course of the human peopling of the Alps.

RIASSUNTO - L'analisi della variabilità genetica delle popolazioni attuale e del passato permette di ricostruire aspetti della storia umana che altrimenti rimarrebbero sconosciuti. La popolazione italiana dell'arco alpino orientale è risultata essere estremamente variabile a livello di DNA mitocondriale. L'elevata diversità appare sia all'interno delle popolazioni, come osservato in altri esempi europei, sia tra le diverse popolazioni specialmente quando queste appartengono a differenti gruppi linguistici. Le relazioni mitocondriali, presenti in basse frequenze in gran parte d'Europa, rappresentano più dei due terzi delle sequenze identificate nei due campioni di lingua Ladina (Stenico et al., 1996). Queste relazioni si sono originate probabilmente nel Paleolitico (Richards et al., 1996) sebbene questo non significhi necessariamente che si diffusero nelle Alpi nello stesso momento cronologico. Al fine di riscontrare prove della presenza di queste relazioni nelle Alpi orientali, abbiamo cercato la estrazione e tipizzazione di un corto segmento della regione mitocondriale in cinque resti umani delle Alpi. La probabile età dei campioni varia da 11000 a 4000 anni BP. Da alcuni di questi campioni, non fu possibile ottenere risultati sia per la scarsa presenza di DNA sia per la contaminazione, uno dei maggiori problemi nello studio di DNA antichi. Solo da tre campioni fu possibile ottenere sufficiente DNA endogeno e, in un solo campione, fu possibile ottenere un segmento di 360 bp . I risultati indicano che questo individuo appartiene ad una delle linee geneologiche maggiormente presente tra le genti di lingua Ladina. Lo studio degli altri due campioni è ancora in corso. Il confronto tra linee mitocondriali attuali e del passato permetterà di ottenere ulteriori informazioni sui processi demografici e evolutivi che hanno caratterizzato il popolamento delle Alpi.

## REFERENCES

Handt et al., 1994 - Ancient DNA: methodological challenges. Experientia 50: 524-529
Handt et al., 1996 - The retrevial of ancient human DNA sequences. Am. J. Hum. Genet. 59: 368-376.
Hoss et al., 1996 - DNA damage and DNA sequence retrevial from ancient tissues. Nucleic Acids Res. 24: 13041307
Krings et al., 1997 - Neandertal DNA sequences and the origin of modern humans. Cell 90: 19-30
PaAbo et al., 1989 - Ancient DNA and the Polymerase Chain Reaction. J. Biol. Chem. 264: 9709-9712

Poinar et al., 1996 - Amino acid racemization and the preservation of ancient DNA. Science 272: 864-866
Richards et al., 1996 - Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet. 59: 185-203
Stenico et al., 1996 - High mitochondrial sequence diversity in linguistic isolates of the Alps.Am. J. Hum. Genet. 59: 1363-1375.
Stenico et al, 1998 - Mitochondrial lineages in Ladin-speaking communities of the Eastern Alps. Proc. Royal Soc. B. 265: 555-561.

$2^{\circ}$ estratto
-- АСАGСААТСААСССТСААСТАТСАСАСАТСААСТGСААСТССАААGССА-ССССТСАСССАСТАGGATAССААСАААССТАСССАСССТтААСАGTAСАТАGTACATAAAGCCATTTACCGT-


Fig. 1 - It is shown a sequence fragment from two indipendent extractions of the Vatte sample. The sequences of different clones are in disagreement and therefore the real sequence is not understandable.
$1^{\circ}$ estratto

- ACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGT- -
т.т.
T.T.
.T.T.
.T.T.
T.T.
.T.T.
T.T.
T.T.
T.T.
T.T.
$2^{\circ}$ estratto
- ACAGCAATCAACCCTCAACTATCACACATCAACTGCAACTCCAAAGCCACCCCTCACCCACTAGGATACCAACAAACCTACCCACCCTTAACAGT--
TTT
.T.T.
T.T.
.T.T.
.T.T.
.T.T.
т. . . . . . . . . . . . . . . . . . . т.т. т
.T.T.
.т.T.

Fig. 2 - A fragment of the Mezzocorona sample's sequence: the different clones are in agreement; therefore we can hold the dominant sequence as the real one.

