Hum Genet (2007) 122:327–336
DOI 10.1007/s00439-007-0403-6

ORIGINAL INVESTIGATION

Genetic variation in prehistoric Sardinia

David Caramelli · Cristiano Vernesi · Simona Sanna · Lourdes Sampietro · Martina Lari · Loredana Castrì · Giuseppe Vona · Rosalba Floris · Paolo Francalacci · Robert Tykot · Antonella Casoli · Jaume Bertranpetit · Carles Lalueza-Fox · Giorgio Bertorelle · Guido Barbujani

Received: 4 June 2007 / Accepted: 29 June 2007 / Published online: 13 July 2007

Abstract We sampled teeth from 53 ancient Sardinian (Nuragic) individuals who lived in the Late Bronze Age and Iron Age, between 3,430 and 2,700 years ago. After eliminating the samples that, in preliminary biochemical tests, did not show a high probability to yield reproducible results, we obtained 23 sequences of the mitochondrial DNA control region, which were associated to haplogroups by comparison with a dataset of modern sequences. The Nuragic samples show a remarkably low genetic diversity, comparable to that observed in ancient Iberians, but much lower than among the Etruscans. Most of these sequences have exact matches in two modern Sardinian populations, supporting a clear genealogical continuity from the Late Bronze Age up to current times. The Nuragic populations appear to be part of a large and geographically unstructured cluster of modern European populations, thus making it difficult to infer their evolutionary relationships. However, the low levels of genetic diversity, both within and among ancient samples, as opposed to the sharp differences among modern Sardinian samples, support the hypothesis of the expansion of a small group of maternally related individuals, and of comparatively recent differentiation of the Sardinian gene pools.

Electronic supplementary material The online version of this article (doi:10.1007/s00439-007-0403-6) contains supplementary material, which is available to authorized users.

D. Caramelli · S. Sanna · M. Lari
Dipartimento di Biologia Animale e Genetica,
Laboratorio di Antropologia,
Università di Firenze, Firenze, Italy

C. Vernesi · G. Bertorelle · G. Barbujani (✉)
Dipartimento di Biologia ed Evoluzione,
via Borsari 46, 44100 Ferrara, Italy
e-mail: g.barbujani@unife.it

L. Sampietro · J. Bertranpetit
Dept. Ciències de la Salut i de la Vida,
Unitat de Biologia Evolutiva,
Universitat Pompeu Fabra, Barcelona, Spain

L. Castrì
Dipartimento di Biologia Evoluzionistica e Sperimentale,
Università di Bologna, Bologna, Italy

G. Vona · R. Floris
Dipartimento di Biologia Sperimentale,
Università di Cagliari, Cagliari, Italy

P. Francalacci
Dipartimento di Zoologia e Antropologia Biologica,
Università di Sassari, Sassari, Italy

R. Tykot
Department of Anthroopology,
University of South Florida, Tampa, FL, USA

A. Casoli
Dipartimento di Chimica generale ed Inorganica,
Chimica Analitica e Chimica Fisica,
Università di Parma, Parma, Italy

C. Lalueza-Fox
Unitat d'Antropologia, Dept. Biologia Animal,
Universitat de Barcelona,
Barcelona, Spain

Introduction

The population of Sardinia is one of the main European genetic outliers (Cavalli-Sforza and Piazza 1993). When compared with populations from all over the world, Sardinians are clearly part of a European genetic cluster (Rosenberg et al. 2002). However, they differ sharply from their European (Barbujani and Sokal 1990) and Italian (Barbujani and Sokal 1991; Barbujani et al. 1995) neighbours, so much so that they are often excluded from multivariate analyses, lest all other samples appear identical in comparison (Piazza et al. 1988; Semino et al. 2000). Mitochondrial (Morelli et al. 2000) and Y-chromosome (Semino et al. 2000; Quintana-Murci et al. 2003) haplotypes that are rare elsewhere in Europe occur at higher frequencies in Sardinia, and an extensive linkage disequilibrium has been described for autosomal markers (Tenesa et al. 2004). In addition, unusually strong genetic differences are observed among Sardinian communities, both for allele-frequency (Barbujani and Sokal 1991) and DNA (Fraumene et al. 2003) polymorphisms.

These peculiar features are the likely evolutionary product of the interaction between reproductive isolation and small population sizes, the former historically documented and the latter a consequence of the fragmented habitat. Starting perhaps 10,000 years ago, people of different provenance reached the island (Webster 1996), but strong isolating factors, such as the Mediterranean Sea and mountain ranges, as well as cultural barriers (Barbujani and Sokal 1991), have probably enhanced the evolutionary role of genetic drift, both within the island and between it and the rest of Europe. However the details of these processes are not well understood, and it is well known that ages of particular genealogical lineages present in a geographical region do not contain information on the arrival of the population in that region (Barbujani and Goldstein 2004). Therefore, to gain insight into the biological features of past populations, and into the relationships between modern populations and their ancient counterparts, the direct study of the DNA of ancient individuals is a crucial research priority.

Here we present the first analysis of the hypervariable region I (HVR-I) of mitochondrial DNA (mtDNA) from samples of Bronze and Iron Age inhabitants of Sardinia, who are called Nuragic people from the towers (*nuraghi*) that they built. Using a strict set of methodological criteria that give a high probability of reproducible results (which forced us to exclude more than half of the initial samples from the analysis) we obtained sequences that were compared with those of modern European populations, and with the sequences of the two pre-classical European populations whose DNA has been described so far, the Etruscans (Vernesi et al. 2004) and the Iberians (Sampietro et al. 2005). In particular, prior to sequencing, the samples were treated with Uracil-*N*-glicosidase, so as to avoid, as far as possible, artefacts due to post-mortem damage of DNA (Hofreiter et al. 2001; Gilbert et al. 2003).

Materials and methods

DNA molecules are often scarce and damaged in ancient samples. As a consequence, experimental artefacts are possible or even likely, depending on the state of preservation of the material, during the analysis of DNA from fossil remains. The risk to mistake modern contaminating DNA for endogenous genetic material is higher when dealing with relatively recent human samples, whose DNA sequence may be similar to that of the archaeologists, museum personnel and geneticists who manipulated them. For all these reasons, in order to obtain reproducible results, the most stringent available standards were followed in this study, based on criteria listed by Cooper and Poinar (2000) and Hofreiter et al. (2001).

We collected 106 teeth, 2 from each of 53 different individuals, coming from excavations at six archaeological sites of Sardinia (Fig. 1) and dated between 4,300 and 3,000 years before the present (Table 1). The selected teeth had not been washed, which meant they had not been touched by anthropologists for morphological analysis. By analysing dental pulp from inside the tooth we reduced the risk of contamination from the archaeologists' DNA, and care was taken to choose integer teeth without fractures. As a rule, to avoid the risk of sampling twice the same individ-

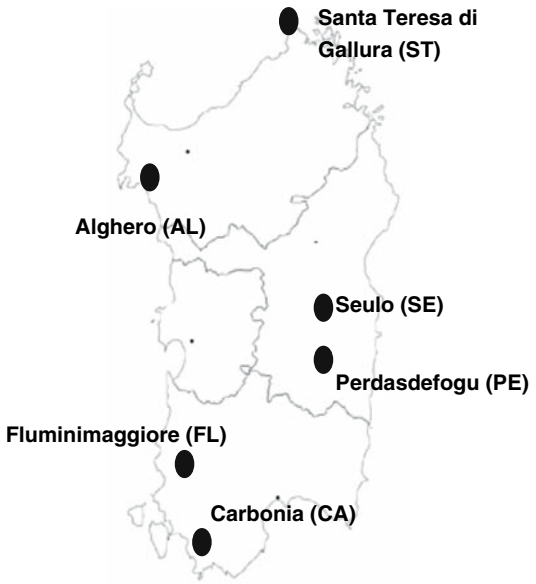

Fig. 1 A map of Sardinia showing the six Nuragic sampling sites. The two-letter codes are used to label the sequences in the text and in Fig. 2

Table 1 Samples analysed

Site	Series	Time	^{14}C Dating	Burial	N1	N2
Alghero	Lu Maccioni	Iron Age	700 ± 70 BC	Natural Cave	6	1
Carbonia	Su cungiareddu 'e Serafini	Middle Bronze Age	1430 ± 70 BC	Natural Cave	7	2
Fluminimaggiore	Capo Pecora	Late Bronze Age	1265 BC	Natural Cave	3	1
Perdasdefogu	Perdasdefogu	Late Bronze Age	930 ± 60 BC	Natural Cave	10	5
Seulo	Stampu Erdi e Cannisoni e Gastea	Middle Bronze Age	1300 ± 60 BC	Natural Cave	10	6
S.Teresa di Gallura	S.Teresa di Gallura	Late Bronze Age	1200–1300 BC	Tomb of Giants	17	8

N1 is the number of specimens initially available, N2 the number of specimens whose mitochondrial sequences could be determined

ual, whenever the teeth had not been found in place in the skull or mandible, we took only left lower canines and left lower molars. Because of the limited amount of amplifiable DNA present in most specimens, we decided to concentrate our efforts on the typing of the non-coding HVRI region.

Criteria of authenticity: an overview

Tooth specimens were handled (using mask, gloves and laboratory coats) in a space where no modern DNA had ever been extracted or analysed. We used only disposable sterile tubes, filtered tips, sterile reagents and solutions, exclusively dedicated to ancient DNA studies. Different sets of pipettes were used for DNA extraction, PCR amplification and analysis of the PCR products. DNA was extracted and PCR-amplified in separate rooms and under hoods, constantly irradiating with UV rays (254 nm). PCR products from ancient DNA templates were analysed in a third room. In each set of extractions or purifications we included a negative control, represented by all the reagents except the bone powder, and these negative controls, together with blanks (all amplification reagents minus DNA), were regularly analysed in every PCR experiment to control for presence of exogenous DNA.

All phases of the analysis were replicated at least twice. Four samples for which relatively large amounts of tooth powder were available were sent to the Barcelona laboratory, where the whole analysis was independently replicated. To test for preservation of other macromolecules as an indirect evidence for DNA survival (Poinar et al. 1996, Poinar and Stankiewicz 1999) we estimated the degree of aminoacid racemisation in each sample, using approximately 5 mg of tooth powder (Poinar et al. 1996). We also quantified the amount of target DNA by competitive PCR. PCR products were cloned, an average 25 clones were sequenced for each individual, and the sequences thus obtained were aligned and compared across clones. The consensus sequences were finally compared with a database of European mitochondrial sequences to test whether the ancient sequences obtained make phylogenetic sense.

DNA was also extracted from cattle (*Bos taurus*) remains retrieved in one burial (PE25), and we tried to amplify it using both human and cattle primers. The presence of human DNA sequences in extracts from non-human bones would suggest contamination in other bones in the burial, and possibly in other burials as well. All persons who worked in this study were mtDNA genotyped and none of their sequences matched those obtained.

DNA extraction

After brushing and irradiating each tooth surface (1 h under UV light), the root tip of each tooth was removed and the crown was repositioned into its mandibular alveolus. The root was powdered and DNA was extracted by means of a silica-based protocol (modified from Caramelli et al. 2003). For each individual we obtained two independent extracts from different teeth. A negative control was included in each extraction.

Competitive PCR

A competitor was used containing a 94 bp deletion (from position 16,106 to 16,189, nt position according to Caramelli et al. 2003), in the mitochondrial HVR-I. PCR components were the same as described below for the sequencing of mitochondrial HVR-1, and the primers were those used for the second fragment amplification. Thermal cycler conditions consisted of an initial 10-min incubation at 95°C followed by 45 cycles of 50 s at 94°C, 50 s at 48°C, and 50 s at 72°C, with a final extension step at 72°C for 5 min.

Real-time PCR

Real-time PCR amplification was performed to confirm the results of competitive PCR. We used Brilliant® SYBR® Green QPCR Master Mix (Stratagene) in MX3000P (Stratagene), using 0.5 μM of appropriate primers (forward primer at H 16,107 and reverse primer at L 16,261). Thermal cycling conditions were 95°C for 10 min, 40 cycles at

95°C for 30 s, 53°C for 1 min and 72°C for 30 s, followed by SYBR® Green dissociation curve steep. Tenfold serial dilutions of the purified and quantified standard were included in the experiment to create the standard curve, thus estimating the number of initial DNA molecules in the samples.

UNG treatment

Uracil-*N*-Glycosylase (UNG) reduces sequence artefacts caused by post-mortem damage of DNA, resulting in apparent C to T and G to A mutations and subsequent errors in the sequence (Hofreiter et al. 2001). For each specimen considered, 10 µl of DNA extract were treated with 1 U of UNG for 30 min at 37°C to excise uracil bases caused by the hydrolytic deamination of cytosines.

Amplification

Two microliters of DNA extracted from the bone and treated with UNG was amplified with this profile: 94°C for 10 min (Taq polymerase activation), followed by 50 cycles of PCR (denaturation , 94°C for 45 s, annealing, 53°C for 1 min and extension, 72°C for 1 min) and final step at 72°C for 10 min. The 50-µl reaction mix contained 2 U of AmpliTaq Gold (Applied Biosystems), 200 µM of each dNTP and 1 µM of each primer. The 360-bp long HVR-I was subdivided in three overlapping fragments using the following primer pairs: L15,995/H16,132; L16,107/H16,261; L16,247/H16,402 (Vernesi et al. 2004). Each extract was amplified at least twice. Since overlapping primers were used throughout the PCR amplifications, it is highly unlikely that we amplified a nuclear insertion rather than the organellar mtDNA.

Cloning and sequencing

PCR products were cloned using TOPO TA Cloning Kit (Invitrogen) according to the manufacturer's instructions. White recombinant colonies were screened by PCR, transferring the colonies into a 30-µl reaction mix [67 mM Tris–HCl (pH 8.8), 2 mM $MgCl_2$, 1 µM of each primer, 0.125 mM of each dNTP, 0.75 U of Taq Polymerase] containing M13 forward and reverse universal primers. After 5 min at 92°C, 30 cycles of PCR (30 s at 90°C, 1 min at 50°C, 1 min at 72°C) were carried out and clones with insert of the expected size were identified by agarose gel electrophoresis. After purification of these PCR products with Microcon PCR devices (Amicon), a volume of 1.5 µl was cycle-sequenced following the BigDye Terminator kit (Applied Biosystems) supplier's instructions. The sequence was determined using an Applied BioSystems 3100 DNA sequencer. Finally, to test for possible contamination within the laboratory, four teeth were subjected to DNA extraction, amplification, cloning and sequencing in Barcelona. In this lab the following primer pairs were used: L16,022/H16,095; L16,055/H16,218; L16,209/H16,401.

Amplification of associated animal remains

We tried to amplify the DNA extracted from cattle bones using both human-specific and bovine-specific primers, namely primers for a 152-bp fragment of the *Bos taurus* mtDNA D-loop, and for a fragment of the human D-Loop (respectively: L16,030/H16,137, and L16,107/H16,261). The PCRs were performed with 2 µl of DNA, 1 µM of each primer, 200 µM of each dNTP, 1× reaction buffer (Applied Biosystem), 1.5 mM $MgCl_2$, and 2 U of AmpliTaq Gold (Applied Biosystem) in a total volume of 50 µl. The cattle D-Loop was amplified using the following thermal cycle: initial denaturation at 94°C for 10 min, followed by 40 cycles of 94°C for 1 min, 57°C for 1 min, 72°C for 1 min, and final extension at 72°C for 5 min. The conditions described above for human mitochondrial DNA amplification were used in the amplification of bovine DNA with human-specific primer. After a run on a 1.5% agarose gel, bands of the appropriate size were excised from the gel and purified with Ultra Free DNA (Amicon, Beverly, Ma). Cloning and sequencing of the PCR products were as described above.

Detection of long amplificates

Appropriate molecular behaviour was also tested by amplification of longer mtDNA fragments (443 and 724 bp), which have been reported to be very unusual for ancient DNA. PCR conditions were those described for mtDNA analysis above, the primers used were L15,995 and H16,401 (for the 443 bp fragment), and L16,247 and H00,360 (for the 724 bp fragment).

Data analysis

We estimated three measures of intrapopulation diversity in the Nuragic sample. Nucleotide diversity (π) is the average number of nucleotide differences per site (Nei and Kumar 2000), over the 360 sites screened [from position 16,024 to 16,383 of the Cambridge reference sequence (CRS)]. The haplotype diversity (h, often referred to as heterozygosity) is defined as $h = 1 - \Sigma q_i^2$ where q_i is the frequency of each haplotype and summation is over all haplotypes observed (Nei and Kumar 2000). The average pairwise sequence difference or average mismatch (k) is the average difference between sequences in the sample.

Phylogenetic trees, i.e. reduced median networks (Bandelt et al. 1995), summarising mtDNA variation were esti-

mated by means of Network 4.1, using a reduction threshold = 2 and the same weights for all loci. To put Nuragic DNA diversity in the proper evolutionary context we used a database of HVR-I sequences from 60 populations, namely 57 modern populations from Europe, Central Asia and the Southern Mediterranean shores, plus the two pre-classic European populations typed so far at the DNA level, the Etruscans (Vernesi et al. 2004) and Iberians (Sampietro et al. 2005), and a medieval sample from Spain (Casas et al. 2006). The modern database includes a sample generically labelled as Sardinia (Di Rienzo and Wilson 1991), and two samples from two regions of Sardinia, namely Ogliastra in the Southeast (Fraumene et al. 2003) and Gallura in the Northeast (Morelli et al. 2000). Complete references are available at this URL: http://web.unife.it/progetti/genetica/pdata.htm.

Genetic distances between populations, F_{st} distances, were calculated by the Arlequin ver 3.01 software (Excoffier et al. 2005), based on Kimura's two-parameter model and using $\alpha = 0.26$. A two-dimensional representation of population relationships was obtained by multidimensional scaling (MDS), using the software Statistica 5.5 (Statsoft Inc.).

Results

For 10 of the 53 samples the D enantiomer of Asp represented more than 10% of the total amino acid (Table 2), implying that there was little hope to retrieve well-preserved DNA molecules from those samples (Poinar et al. 1996; Serre et al. 2004). Therefore, we extracted the DNA from the remaining 43 samples. Sporadic contamination is considered unlikely when the number of molecules that PCR will use as template (*target DNA*) is greater than 1,000 (Poinar et al. 1996); in 7 samples quantitative PCR showed that the DNA available was less than that, whereas 36 samples yielded sufficient DNA for amplification, cloning (about 30 clones for each sample) and sequencing. Real Time PCR confirmed that these 36 samples were suitable for further analysis. In comparing sequences across clones, nucleotide substitutions observed in one clone only were considered due to Taq-polymerase errors. Conversely, in six samples the repeated occurrence of the same substitution in several (but not all) clones made it impossible to infer a consensus sequence without ambiguity (Handt et al. 1996).

Globally, a low rate of *Taq* misincorporation was estimated (0.21 substitutions every 1,000 bp within the HVRI), with at least 95% of the clones showing the consensus nucleotide at each position. This result suggests that the DNA templates were not damaged. On the other hand, seven samples yielded multiple sequences and were excluded from the analysis at this stage, thus bringing down the number of nuragic HVR-I sequences to 23. Additional teeth from four of these 23 individuals (PE23, PE20, SE14, SE2) were available and were sent to the Barcelona laboratory; the sequences determined there were identical to those obtained in the Florence laboratory. Amplification of large DNA fragments, unusual for ancient DNA, was not observed, and the final consensus sequences (Table 3) make phylogenetic sense, i.e. do not appear to be combinations of different known sequences. Finally, in the only case in which cattle remains were available, PCR amplification

Table 2 Elimination of the specimens that did not comply with quality standards

Initial sample size (individuals)	Test	Eliminated	Kept
53	AA racemisation	10	43
43	Quantitative PCR	7	36
36	Cloning, sequencing	6	30
30	Sequence comparisons	7	23
23	Replication in a second lab	0[a]	23
23	Amplification of non-human associated remains	0[b]	23

[a] This test was carried out for four specimens only

[b] This test was carried out for one specimen only

Table 3 DNA sequences of 23 Nuragic specimens

	1111111111	
	6666666666	
	0011112223	
	5622892791	
	1969933881	
CRS	ACTGTCCCTT	
(AL07)	....C.....	(Alghero)
(CA02)	....	(Carbonia)
(CA14)	G.........	(Carbonia)
(FL04)	...C......	(Fluminimaggiore)
(PE11)	....	(Perdasdefogu)
(PE15)	...C....C.	(Perdasdefogu)
(PE20)	....	(Perdasdefogu)
(PE23)	....	(Perdasdefogu)
(PE25)	......T...	(Perdasdefogu)
(SE01)	..C.......	(Seulo)
(SE02)	.TC..T....	(Seulo)
(SE13)	G.........	(Seulo)
(SE60)	.......T..	(Seulo)
(SE81)	....	(Seulo)
(SE84)	....	(Seulo)
(ST08)	......T...	(S. Teresa di Gallura)
(ST10)	....	(S. Teresa di Gallura)
(ST15)	...C......	(S. Teresa di Gallura)
(ST16)	C	(S. Teresa di Gallura)
(ST30)	..C.......	(S. Teresa di Gallura)
(ST38)	....	(S. Teresa di Gallura)
(ST47)	....	(S. Teresa di Gallura)
(ST54)	...C......	(S. Teresa di Gallura)

CRS Cambridge reference sequence

was successful only using primers specific for *Bos taurus*, and not human primers.

Ten sites appeared polymorphic in the Nuragic samples, with nine transitions and one transversion, leading to identification of ten distinct haplotypes. Nucleotide diversity was 0.0041, and haplotype diversity a low 0.83, close to the value observed in Ogliastra, 0.78. By contrast, haplotype diversity values of 0.96 (Di Rienzo and Wilson 1991) and 0.97 (Morelli et al. 2000) were observed in the other modern Sardinian samples and ≥0.93 in all other European samples with the exception of Saami. The low haplotype diversity among the Nuragic people is due to the occurrence at relatively high frequencies of five haplotypes, including the CRS (9 individuals, 39% of the total). The other five haplotypes, AL07, PE15, SE02, SE60 and ST16, are observed once in the Nuragic sample, and so, despite the small sample sizes, different Nuragic populations share one or more haplotypes (Table 4). The average difference between pairs of sequences, or average mismatch, was a low 1.43 ± 0.90 compared to the Etruscans (3.90; Vernesi et al. 2004) and the modern samples (average 4.68), but close to the Iberians' 2.12 (Sampietro et al. 2005) and to the modern Ogliastra 2.63 (Fraumene et al. 2003).

The phylogenetic relationships of the Nuragic sequences were summarised in a network comparing them with the modern sample from Ogliastra (Fraumene et al. 2003), for which the RFLPs data necessary for haplogroup identification were available (Fig. 2). Eight Nuragic sequences, representing 21 individuals, appear to fall into the haplogroup H cluster, whereas the remaining two appear to belong to haplogroups V and J. However, in the FL04-ST15-ST54 sequence a C is observed in position 129, a substitution commonly observed in subhaplogroup U2 (Achilli et al. 2005). Therefore, in the absence of RFLP data, it is impossible to attribute with confidence this sequence to either haplogroup H or U.

The comparison with the database including sequences from 57 modern populations, ancient Etruscans and Iberians, and a medieval sample from Spain, is clearly affected by the different sample sizes. However, 8 of the sequences representing 19 individuals (83%) found an exact match in modern European samples, to the exclusion of PE15 (with substitutions at 16,129 and 16,298) and FL04-ST15-ST54 (with a substitution at 16,129). These two sequences find no match in comparisons with 92 African samples either (data not given). Six haplotypes are shared between modern and ancient Sardinians, representing 61% of the ancient individuals. Four and two haplotypes, representing respectively 52 and 48% of the total Nuragic sequences, is the level of sharing between the Nuragic people and both Etruscans and ancient Iberians. The minimum allele sharing was one for the Sardinian sample from Gallura (and for a few European, Caucasian, Near Eastern and North African populations, including Tuscany and the Ladin linguistic isolate of Northern Italy). Two modern samples from Sardinia (Di Rienzo and Wilson 1991) and Ogliastra (Fraumene et al. 2003) share four haplotypes each with the Nuragic sample, and, in particular, the Ogliastra sample shares three haplotypes with the ancient samples from the same area (Seulo and Perdasdefogu). The CRS, which represents 39.1% of the

Table 4 Number of haplotypes shared between the population of Nuragic Sardinia and other regions of Eurasia (relative frequencies between parentheses)

Haplotype	Frequency	North Africa	Near East	Caucasus	Central Asia	Europe	Ancient Iberians	Etruscans	Sardinia	Ogliastra	Gallura
Sample size		456	261	397	205	3,379	17	27	73	175	27
CA02, PE11, PE20, PE23, SE81, SE84, ST10, ST38, ST47	39.1	44 (9.6)	18 (6.9)	56 (14.1)	11 (5.4)	451 (13.3)	3 (17.6)	2 (7.4)	15 (20.5)	78 (44.5)	5 (18.5)
ST16	4.4	11 (2.4)	6 (2.2)	4 (1.0)	2 (1.0)	89 (2.6)		1 (3.7)	3 (4.1)	12 (7.1)	
AL07	4.4	2 (0.4)		3 (0.8)	2 (1.0)	46 (1.4)			1 (1.4)		
SE02	4.4		1 (0.4)	1 (0.3)		5 (0.1)				2 (1.2)	
SE60	4.4		1 (0.4)			11 (0.3)					
PE25, ST08	8.7		1 (0.4)	1 (0.3)		1 (0.0)		1 (3.7)			
SE01, ST30	8.7		1 (0.4)	2 (0.3)		22 (0.6)	2 (11.8)	2 (7.4)		3 (1.8)	
CA14, SE13	8.7	7 (1.5)		1 (0.3)		1 (0.0)			1 (1.4)		
PE15	4.4										
FL04, ST15, ST54	13.0										

The three modern Sardinian samples are from Di Rienzo and Wilson (1991) (Sardinia, SR in the legend to Fig. 3), Fraumene et al. (2003) (Ogliastra, SO), and Morelli et al. (2000) (Gallura, GR)

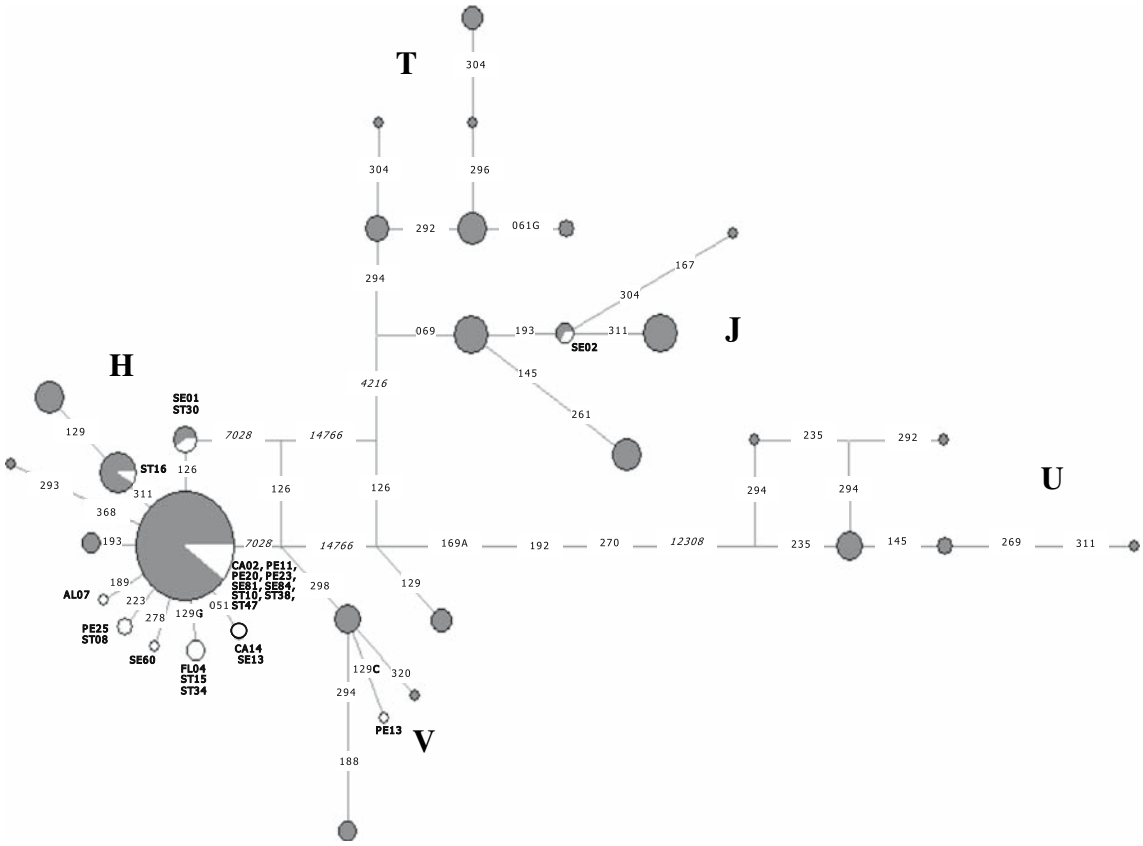

Fig. 2 Reduced median networks of Nuragic (*white*) and modern Ogliastra (*grey*) samples. The *size of the circles* represents the haplotype's absolute frequency. Positions of HVR-I mutations are indicated minus 16,000 and transversions by the nucleotide change; RFLP sites are indicated as follows: −7025AluI = 7,025; −14766MseI = 14,766; +12308HinfI = 12,308; +4216NlaIII = 4,216. *Capital letters* refer to the haplogroups inferred from sequence variation and RFLP status

Nuragic sequences and only 13.3% all over Europe, reaches similarly high frequencies only in the modern Sardinians (globally 35.5%), and especially in Ogliastra.

An analysis of molecular variance (Excoffier et al. 2005) showed no difference among the Perdasdefogu ($n = 5$), Seulo ($n = 6$) and Santa Teresa di Gallura ($n = 8$) samples, with a negative estimate of the among-population variance component (−0.60%). Among the Etruscans, the differences among populations, albeit insignificant, accounted for 1.11% of the overall variance.

The multidimensional scaling graph (Fig. 3) summarises the genetic relationships inferred from F_{st} distances between populations. Saami were eliminated from this graph, because their high level of differentiation drastically reduces the apparent differences among all other populations. A mild stress value was observed in the analysis, 0.19, meaning that the projection of the data on a plane distorted population relationships only to a limited extent. In this global representation, Nuragic Sardinia falls in a large European cluster, where little or no geographical structure is apparent. All outliers are either populations separated by large geographic distances from the other Europeans (mainly North Africans and Central Asians), or well-known genetic outliers, such as the modern population of Ogliastra (Fraumene et al. 2003) and the ancient, non-Indo-European speaking Etruscans.

Discussion

The first archaeological evidence for the Nuragic civilisation dates back to the Middle Bronze Age, 3,800–3,300 years ago. Populations previously scattered in small units integrated in larger communities, still having only sporadic contacts with non-Sardinian people (Webster 1996). In the Late Bronze Age pre-existing nuraghi were enlarged and metallurgy intensified, leading to increased trade across the Mediterranean Sea. Population growth and warfare can be inferred from the archaeological record of that period, the former being a possible cause of the latter.

The Late Bronze Age Sardinians whose mtDNA we could type show a remarkably low level of genetic differentiation. Despite sampling covering an area of approximately 24,000 km^2, and a 700-year time interval (up to the early Iron Age, 2,700 years BP), the same sequences were repeatedly found at different locations, and all belong to

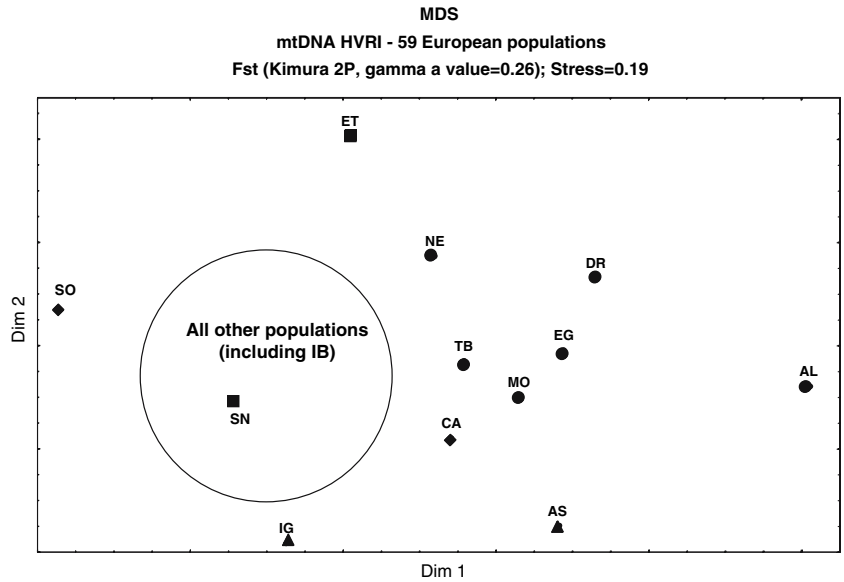

Fig. 3 Multi-dimensional scaling. *Squares* ancient populations, *circles* North Africa and the Near East, *diamonds* Europe, *triangles* Asia and the Caucasus. Population codes as follows: Algerians (*AL*); Central Asia (*AS*); Catalans (*CA*); Druze (*DR*); Egyptians (*EG*); Etruscans (*ET*); Ancient Iberians (*IB*); Ingushians (*IG*); Moroccan Arabs (*MO*); Near Easterners (*NE*); Nuragic Sardinians (*SN*); Sardinians Ogliastra (*SO*). The *circle* includes all other populations, namely: Abazinian; Adygei; Armenians; Austrians; Azerbaijan; Basques; Belgium; Belarus; British; Bulgarians; Chechenians; Cherkessians; Cornwell; Danes; Darginians; Estonians; Finns; French; Galicians; Germany; Georgians; Sardinians Gallura; Germans North; Germans South; Icelanders; Italians, Ladin and German speakers; Italians South; Italians, Tuscany; Karelians; Kurds; Moroccan Berbers; Dutch; Norwegians; Portuguese; Russians; Saami; Italians, Sicily; Sardinians; Swedes; Medieval Spaniards; Spaniards; Spaniards, South; Swiss; Sirians; Tunisians Berbers; Turks; Volga-Finns; Welsh. The total number of individuals considered is 4,875

only 3 or 4 of the 20 haplogroups documented in Western Mediterranean populations (Sampietro et al. 2005). Only another ancient population, the sixth to second century BC Iberians (Sampietro et al. 2005), showed a comparable mitochondrial diversity, much lower than in all modern samples and in the Etruscans.

Genetic variation in Sardinia appears limited across time as well. Six haplotypes observed in the Nuragic sample are still present in the modern Ogliastra and Sardinia samples, and they include not only the widespread CRS (whose frequency is higher in moderns as well as in ancient Sardinians compared to other samples), but also rarer sequences such as those that we labelled AL07, SE02, ST16, SE01-ST30, and CA14-SE13. Three such sequences, AL07, ST16, and CA14-SE13 also occur in North Africa; their presence in both modern and Nuragic Sardinia suggest the effects of common ancestry or ancient gene flow, rather than those of gene flow in historical times. In particular, modern people from Ogliastra share three haplotypes with Nuragic people from the nearby localities of Perdasdefogu and Seulo, i.e. half of both the modern (83/175) and the ancient (12/23) sequences (Table 4). On the contrary, the modern population of Gallura appears genetically distant from both ancient and modern Sardinian samples, in agreement with the effects of relatively recent immigration from continental Italy, known to have affected Gallura more than Central and Southern Sardinia (Morelli et al. 2000).

The other pre-Roman Italic population described so far at the genetic level, the Etruscans (Vernesi et al. 2004), showed very different patterns of mtDNA diversity. The Etruscans appear more closely related to modern Tuscans than to any other European or Near Eastern population, but only 2 of 23 different Etruscan sequences were found to have an exact match in the modern Tuscan populations, including some that were selected for their supposed Etruscan origins (Achilli et al. 2007). The difference with Nuragic Sardinia (where shared sequences are seven out of ten) is highly significant ($\chi^2 = 13.21$, 1 d.o.f., $P < 0.001$). Plausible explanations for the low resemblance between Etruscans and modern Tuscans include loss of haplotypes by genetic drift and/or extensive immigration, possibly but not necessarily related to the social changes that occurred after the Roman assimilation (Belle et al. 2006), or in modern times. By contrast, the results of the present study show a much clearer genealogical continuity in Sardinia, presumably reflecting the island's higher degree of isolation.

Despite this clear genealogical continuity, there are differences between ancient and modern Sardinians, notably in their levels of internal differentiation, much higher in

the latter, and very high at the allele-frequency level (Cappello et al. 1996). A previous genetic analysis excluded pre-Bronze-Age expansions similar to those that occurred in continental Italy and other European regions (Barbujani et al. 1995). Although mitochondrial and autosomal variation are not necessarily correlated, this finding suggests that Sardinian populations differentiated recently, probably because of local immigration and population subdivision. Consistent with a recent differentiation is the fact that zones of increased genetic change in Sardinia often correspond to linguistic boundaries, which cannot be more than 2,000 years old (Barbujani and Sokal 1991). The simplest way to reconcile these findings with the archaeological evidence of demographic growth in the Late Bronze Age is to envisage the expansion of a small group of geographically dispersed but maternally related individuals. The successive subdivision in geographically and linguistically differentiated isolates enhanced the evolutionary effects of genetic drift.

In the multidimensional scaling of Fig. 3, Nuragic Sardinians cluster with the majority of the European populations. Given the small sample size, inevitable in ancient DNA studies, it is at present impossible to infer their evolutionary relationships from mtDNA affinities. Nevertheless, in relation with ancient samples, Nuragic Sardinians appear more related to the Iberians than to the Etruscans, whose position in the graph is eccentric. Three data points are not enough for a robust generalisation. However, one can at least conclude that Sardinians and Iberians show a greater genealogical continuity with the Bronze-Age inhabitants of the same regions than the Tuscans. To better understand the processes leading to these differences it will be necessary to genetically characterise people who lived in those areas between 2,000 years ago and the present time. At present, a study of paternal genealogies in Holland shows that gene flow in the last 3 centuries resulted in a massive population displacement (Manni et al. 2005). We do not know to what extent this finding can be generalised to other European populations and to maternal genealogies, although female migration is known to be higher than male migration in most human societies (Seielstad et al. 1998; Dupanloup et al. 2003). However, the Dutch data raise the possibility that mildly isolated populations, such as those of Holland and Tuscany, have recently experienced drastic demographic changes, a process that has affected only marginally more isolated regions, such as Sardinia.

Acknowledgments This paper was supported by funds from the Italian Ministry of the Universities (PRIN 2006) and by funds from the Universities of Ferrara and Florence. We thank Antonio Torroni for a thorough preliminary discussion of the results of this study, and three anonymous referees for several useful suggestions.

References

Achilli A, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, Bandelt HJ, Semino O, Torroni A (2005) Saami and Berbers—an unexpected mitochondrial DNA link. Am J Hum Genet 76:883–886

Achilli A, Olivieri A, Pala M, Metspalu E, Fornarino S, Battaglia V, Accetturo M, Kutuev I, Khusnutdinova E, Pennarun E, Cerutti N, Di Gaetano C, Crobu F, Palli D, Matullo G, Santachiara-Benerecetti S, Cavalli-Sforza LL, Semino O, Villems R, Bandelt HJ, Piazza A, Torroni A (2007) Mitochondrial DNA variation of modern Tuscans supports the Near Eastern origin of Etruscans. Am J Hum Genet 80:759–768

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Barbujani G, Bertorelle G, Capitani G, Scozzari R (1995) Geographical structuring in the mtDNA of Italians. Proc Natl Acad Sci USA 92:9171–9175

Barbujani G, Goldstein D (2004) Africans and Asians abroad: Genetic diversity in Europe. Annu Rev Genomics Hum Genet 5:119–150

Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci USA 87:1816–1819

Barbujani G, Sokal RR (1991) Genetic population structure of Italy. II. Physical and cultural barriers to gene flow. Am J Hum Genet 48:398–411

Belle EM, Ramakrishnan U, Mountain JL, Barbujani G (2006) Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. Proc Natl Acad Sci USA 103:8012–8017

Cappello N, Rendine S, Griffo R, Mameli GE, Succa V, Vona G, Piazza A (1996) Genetic analysis of Sardinia: I. Data on 12 polymorphisms in 21 linguistic domains. Ann Hum Genet 60:125–141

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 USA 100:6593–6597

Casas MJ, Hagelberg E, Fregel R, Larruga JM, Gonzalez AM (2006) Human mitochondrial DNA diversity in an archaeological site in al-Andalus: genetic impact of migrations from North Africa in medieval Spain. Am J Phys Anthropol 131:539–551

Cavalli-Sforza LL, Piazza A (1993) Human genomic diversity in Europe: a summary of recent research and prospects for the future. Eur J Hum Genet 1:3–18

Cooper AR, Poinar H (2000) Ancient DNA: do it right or not at all. Science 289:1139

Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci USA 88:1597–1601

Dupanloup I, Pereira L, Bertorelle G, Calafell F, Prata MJ, Amorim A, Barbujani G (2003) A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. J Mol Evol 57:85–97

Excoffier LG, Laval S, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50

Fraumene C, Petretto E, Angius A, Pirastu M (2003) Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. Hum Genet 114:1–10

Gilbert MTP, Hansen AJ, Willerslev E, Rudbeck L, Barnes I, Lynnerup N, Cooper A (2003) Characterization of genetic miscoding

lesions caused by postmortem damage. Am J Hum Genet 71:48–61

Handt O, Krings M, Ward RM, Pääbo S (1996) The retrieval of ancient human DNA sequences. Am J Hum Genet 59:368–376

Hofreiter M, Serre D, Poinar HN, Kuch M, Pääbo S (2001) Ancient DNA. Nat Rev Genet 2:353–359

Manni F, Toupance B, Sabbagh A, Heyer E (2005) New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling. Am J Phys Anthropol 126:214–228

Morelli L, Grosso MG, Vona G, Varesi L, Torroni A, Francalacci P (2000) Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. Hum Biol 72:585–595

Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, Oxford

Piazza A, Cappello N, Olivetti E, Rendine S (1988) A genetic history of Italy. Ann Hum Genet 52:203–213

Poinar HN, Hoss M, Bada JL, Pääbo S (1996) Amino acid racemization and the preservation of ancient DNA. Science 272:864–866

Poinar HN, Stankiewicz BA (1999) Protein preservation and DNA retrieval from ancient tissues. Proc Natl Acad Sci USA 96:8426–8431

Quintana-Murci L, Veitia R, Fellous M, Semino O, Poloni ES (2003) Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis. Am J Phys Anthropol 121:157–171

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Sampietro ML, Caramelli D, Lao O, Calafell F, Comas D, Lari M, Agusti B, Bertranpetit J, Lalueza-Fox C (2005) The genetics of the pre-Roman Iberian Peninsula: a mtDNA study of ancient Iberians. Ann Hum Genet 69:535–548

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–280

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290:1155–1159

Serre D, Langaney A, Chech M, Teschler-Nicola M, Paunovic M, Mennecier P, Hofreiter M, Possnert G, Paabo S (2004) No evidence of Neandertal mtDNA contribution to early modern humans. PLOS Biol 2:0001–0005

Tenesa A, Wright AF, Knott SA, Carothers AD, Hayward C, Angius A, Persico I, Maestrale G, Hastie ND, Pirastu M, Visscher PM (2004) Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. Hum Mol Genet 13:25–33

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

Webster GS (1996) A prehistory of Sardinia 2300–500 BC. Sheffield Academic Press, Sheffield