High Resolution Analysis and Phylogenetic Network Construction Using Complete mtDNA Sequences in Sardinian Genetic Isolates

Cristina Fraumene, Elise M. S. Belle,† Loredana Castrì,‡ Simona Sanna,§ Gianmaria Mancosu,* Massimiliano Cosso,* Francesca Marras,* Guido Barbujani,† Mario Pirastu,*§ and Andrea Angius*§*

*Shardna Life Sciences, Pula (Cagliari), Italy; †Dipartimento di Biologia, Università di Ferrara, Ferrara, Italy; ‡Dipartimento di Biologia Evoluzionistica e Sperimentale–Sezione di Antropologia, Università di Bologna, Bologna, Italy; and §Istituto di Genetica delle Popolazioni, CNR, Alghero (Sassari), Italy

For mitochondrial phylogenetic analysis, the best result comes from complete sequences. We therefore decided to sequence the entire mitochondrial DNA (mtDNA) (coding and D-loop regions) of 63 individuals selected in 3 small Ogliastra villages, an isolated area of eastern Sardinia: Talana, Urzulei, and Perdasdefogu. We studied at least one individual for each of the most frequent maternal genealogical lineages belonging to haplogroups H, V, J, K, T, U, and X. We found in our 63 samples, 172 and 69 sequence changes in the coding and in the D-loop region, respectively. Thirteen out of 172 sequence changes in the coding region are novel. It is our hypothesis that some of them are characteristic of the Ogliastra region and/or Sardinia. We reconstructed the phylogenetic network of the 63 complete mtDNA sequences for the 3 villages. We also drew a network including a large number of European sequences and calculated various indices of genetic diversity in Ogliastra. It appears that these small populations remained extremely isolated and genetically differentiated compared with other European populations. We also identified in our samples a never previously described subhaplogroup, U5b3, which seems peculiar to the Ogliastra region.

Introduction

Mitochondrial DNA (mtDNA) analysis is considered an essential tool for studying human population structure, origins, migration patterns, and demographic history, given its polymorphism, its matrilineal mode of descent, and its lack of recombination (Torroni et al. 1996). The first complete sequence of human mtDNA, the Cambridge Reference Sequence (CRS), was published in 1981 (Anderson et al. 1981) and was recently revised (rCRS, Andrews et al. 1999).

The fact that mutations accumulate sequentially along maternal lineages allows associating many of them with different geographical regions of the world (Ingman et al. 2000; Herrnstadt et al. 2002). mtDNA sequence variations can thus be used to construct phylogenetic networks (Bandelt et al. 1999), displaying the relationships among sequences and estimating the time of appearance of mutations associated with each haplotype.

Historically, the first mtDNA polymorphisms used in human phylogenetic studies were identified in the noncoding or D-loop region, containing the hypervariable segments (HVS). Accurate phylogenetic networks for European mtDNAs were constructed using HVS-1 and HVS-2 sequence data (Richards et al. 1998; Helgason et al. 2000), complemented with coding-region restriction fragment length polymorphisms (RFLP) used to define mtDNA haplogroups (Torroni et al. 1996; Macaulay et al. 1999). High mutation rates, variability in site substitution rates, homoplasic sites, parallel mutation events, and reversion make the D-loop evolution complex and the phylogenetic analysis subject to error (Bandelt et al. 2002; Dennis 2003; Forster 2003). Complete mtDNA sequences, only recently becoming available (Ingman et al. 2000; Finnila et al. 2001; Maca-Meyer et al. 2001; Torroni et al. 2001; Herrnstadt et al. 2002; Achilli et al. 2004; Howell et al. 2004; Palanichamy et al. 2004; Rajkumar et al. 2005), represent the best possible solution for phylogenetic analysis (Richards and Macaulay 2001; Kivisild et al. 2006). However, data for mtDNA polymorphisms in different human populations are still limited and only few haplogroups are based on mtDNAs complete sequence (Finnila and Majamaa 2001). Most important, several reports associate some diseases with specific mtDNA haplogroups (Brown et al. 1997; Kalman et al. 1999; Chinnery et al. 2000; Ruiz-Pesini et al. 2000; Moilanen et al. 2003; Herrnstadt and Howell 2004; Mancuso et al. 2004; Pyle et al. 2005), and it is therefore fundamental to increase our understanding of mtDNA haplogroup (Rose et al. 2001; Niemi et al. 2003, 2005).

Interpopulation comparisons and phylogenetic tree construction through mtDNA studies can be useful for the characterization of populations with unusual genetic features (Tolk et al. 2000; Finnila and Majamaa 2001; Larruga et al. 2001; Meinila et al. 2001). mtDNA analysis of different Sardinian populations revealed specific genetic characteristic and a variable degree of subpopulation homogeneity within the island (Workman et al. 1975; Piazza et al. 1988; Morelli et al. 2000). However, so far no studies were carried out in eastern Sardinia using complete mtDNA sequences.

In the present article, we present the analysis of 63 complete mtDNA sequences of samples coming from 3 distinct villages of a Sardinian region, Ogliastra, characterized by both genetic and environmental homogeneity. This area encompasses 23 small isolated villages, whose founders presumably derived from the same original gene pool, with scant genetic exchanges with the rest of Sardinian areas recorded during the last 400 years of parish and historical records. Indeed, it is thought that geographic and cultural barriers have exerted a strong isolating effect in this part of Sardinia. Geographical isolation, small population size, high endogamy, and inbreeding are expected to lead to increased genetic differentiation among subpopulations as a consequence of founder effect and genetic drift (Angius et al. 2001; Fraumene et al. 2003).

Key words: complete mtDNA sequences, genetic isolates, phylogenetic network, subhaplogroup, genetic drift, founder effect.

E-mail: angius@shardna.it.

Mol. Biol. Evol. 23(11):2101–2111. 2006
doi:10.1093/molbev/msl084
Advance Access publication August 10, 2006

Table 1
Maternal Genealogical Lineages of the 3 Villages. "Year" Indicates the Date of Introduction of Maternal Lineages in the Village."Progeny" Indicates the Total Number of Living Descendants. "Hapl." Designates the Mitochondrial Haplogroup

Talana					Urzulei					Perdasdefogu				
Founder	Year	Progeny	Sample	Hapl.	Founder	Year	Progeny	Sample	Hapl.	Founder	Year	Progeny	Sample	Hapl.
1	1600	97	2545T	H	1	1780	91	00136U	H	1	1750	76	0002P	H
2	1600	104	2175T	H1	2	1570	35	00207U	H1	2	1680	73	0218P	H1
3	1623	108	2554T	H1	3	1805	40	00097U	H1	3	1680	64	0305P	H1
4	1684	95	2798T	H1	4	1572	409	00004U	H3	4	1685	36	0695P	H1
5	1720	33	2472T	H1	5	1585	78	00070U	H3	5	1690	198	0399P	H1
6	1800	221	2579T	H1	6	1790	19	00307U	H3	6	1879	45	0157P	H1
7	1572	104	2702T	H3	7	1865	21	00120U	H3	7	1640	60	0437P	H3
8	1585	94	2060T	H3	8	1890	23	00271U	H3	8	1665	27	0019P	H3
9	1600	34	2477T	H3	9	1763	140	00062U	H4	9	1695	182	0009P	H3
10	1640	13	2465T	H3	10	1600	22	00257U	J1c	10	1700	35	0054P	H3
11	1760	170	2538T	H3	11	1610	217	00134U	J2b	11	1851	23	0066P	H3
12	1770	45	2662T	H3	12	1590	64	00100U	T2b	12	1650	101	0006P	PreV
13	1790	33	2008T	H3	13	1770	46	00260U	T2b	13	1870	34	0540P	J2a
14	1895	16	2282T	J2b	14	1776	63	00105U	T2b	14	1680	212	0349P	K1a
15	1740	109	2038T	T2	15	1660	66	00050U	U5b1	15	1884	32	0208P	K1a
16	1662	128	2096T	U5b3	16	1780	33	00191U	U5b2	16	1610	95	0259P	T2
17	1550	130	2010T	V	17	1785	183	00102U	U5b2	17	1854	38	0185P	T2
					18	1590	45	00368U	U5b3	18	1660	34	0288P	U1a
					19	1614	106	00032U	U5b3	19	1725	141	0008P	U1a
					20	1615	63	00187U	U5b3	20	1650	47	0044P	U5b3
					21	1650	104	00213U	U5b3	21	1831	34	0176P	U5b3
					22	1904	24	00343U	U6	22	1800	48	1074P	X2b
										23	1710	218	214P-358P	X2

Genealogical records systematically kept since the 17th century allow the careful and accurate reconstruction of genealogies for each village. We created a relational database and developed specific tools to access these data easily and reconstruct complete genealogical trees for up to 16 generations in a rapid manner (Mancosu et al. 2003, 2005). It is, therefore, possible to also reconstruct accurately all the maternal lineages present in the 3 villages all the way back to the 17th century. We analyzed one or more individuals from each founder maternal genealogical lineage.

Our study allowed us to monitor segregation and selection in the evolution of these lineages. Complete mtDNA sequences permitted us to evaluate genetic differentiation of each village in comparison to other European sequences.

Materials and Methods

Samples Selection

We sequenced complete mtDNAs of 63 individuals from 3 small villages within Ogliastra region (Eastern Sardinia). Talana (1,161 inhabitants) and Urzulei (1,419 inhabitants) are located in the northwestern part of Ogliastra, whereas Perdasdefogu (2,400 inhabitants) is located in the south. Samples were selected on the basis of genealogical information (Fraumene et al. 2003) aiming at examining one or more individuals for each founder maternal lineage.

All villages have a limited number of maternal lineages coalescing into few different haplogroups (table 1). Maternal lineages with less than 10 descendents were not taken into consideration as not being sufficiently representative of the community. For Talana, we estimate that 17 genealogical lineages account for about 87% of the present-day living inhabitants (1,765 residents and nonresidents), for Urzulei, 22 genealogical lineages account for about 90% present-day living inhabitants (2,091 residents and nonresidents), and for Perdasdefogu, 23 genealogical lineages account for about 79% present-day living inhabitants (2,340 residents and nonresidents). During the last 50 years, Perdasdefogu has undergone substantial immigration because of a military base nearby, but, by our sampling method, the maternal lineages immigrated in the village after 1950 were not considered.

Informed consent was obtained from each individual and all samples were collected in accordance with the Declaration of Helsinki.

DNA and Sequence Analysis

Genomic DNA was isolated from 5 ml of peripheral blood as previously described (Ciulla et al. 1988). Whole mtDNA was amplified using 24 partially overlapping fragments, each ≈800 bp in length (Rieder et al. 1998). Polymerase chain reaction (PCR) primers were designed to provide ≈200 bp overlap between neighboring fragments. Template DNA was amplified in 25-μl volume. PCR conditions were as follows: initial denaturation 2 min at 95 °C, 35 subsequent cycles of 30 s denaturation at 95 °C, 30 s annealing at each primer-specific temperature, 45 s extension at 72 °C, and a final 7 min extension at 72 °C. PCR products were purified with ExoSap-IT Kit (USB Corporation, Cleveland, OH). Sequencing reactions were performed with 7 μl purified PCR product, forward or reverse primers used in PCR reactions, and ABI BigDye Terminator Cycle Sequencing kits (Applied Biosystems, Foster City, CA). Sequencing reactions were purified using Millipore MultiScreen Assay System (Millipore, Billerica, MA), and fluorescent-labeled extension products were loaded onto an ABI

3730 DNA analyzer (Applied Biosystems, Foster City, CA). To avoid errors or artifacts, each sample was sequenced for both mtDNA strands. Moreover, ambiguous results were always confirmed by independent PCR and sequencing.

Comparisons with rCRS were performed using the BioEdit Sequence Alignment software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Sequences Evolutionary Analysis

All sequences were assigned to Western Eurasian haplogroups according to the nomenclature of Reidla et al. (2003), Maca-Meyer et al. (2003), Palanichamy et al. (2004), Achilli et al. (2004, 2005), Behar et al. (2006), and references therein.

The phylogenetic networks based on coding and D-loop region sequences were constructed by use of a median-joining algorithm (Bandelt et al. 1995, 1999) as implemented in the Network 4.1 program (http://www.fluxus-technology.com). By inspecting the network, one can identify homoplasic sites, that is, sites that are subjected to recurrent mutation (Bandelt et al. 1995, 1999).

Median-joining networks, based on nucleotide variation in the whole mtDNA, were generated with an ε value of 0. Because of mutation-rate heterogeneity between D-loop and coding region, we chose to give a smaller weight to nucleotide positions in the D-loop region. We assigned a weight = 2 to all coding-region variants and a weight = 1 for D-loop substitutions.

Analyses of Genetic Variation

Statistical analyses of genetic variation within each population of the 3 villages of Ogliastra (Talana, Urzulei, and Perdasdefogu) were conducted using DnaSP ver 4.0 software (Rozas et al. 2003).

Measures of genetic variation between populations were computed using Arlequin ver 2.0 software (Schneider et al. 2000). F_{ST} and ϕ_{ST} values, 2 indices of population differentiation, based, respectively, on allele frequencies and on both allele frequency and sequence differences, were computed between each pair of populations, and an analysis of molecular variation (AMOVA) was carried out. To assess the significance of the genetic variances thus estimated, individual genotypes were randomly reassigned to populations and populations to groups 1,000 times (Schneider et al. 2000). In this way, the significance of the estimated variances was tested by a nonparametric permutational procedure in which the distribution of the random variances thus obtained was compared with the observed values (Excoffier et al. 1992). Several tests of mutation-drift equilibrium were conducted using Tajima's *D* (Tajima 1989), Fu and Li's *D* and *F* (Fu and Li 1993), and Fu's F_s (Fu 1997).

Results

Sequence Variation in mtDNA from 63 Subjects of Talana, Urzulei, and Perdasdefogu

Complete mtDNA sequence was determined in 63 maternally unrelated healthy individuals from Talana, Urzulei, and Perdasdefogu belonging to haplogroups H, pre-V, V, J, K, T, U, and X.

These sequences were defined using haplogroup definitions as described in literature (Torroni et al. 1996; Finnila et al. 2000, 2001; Herrnstadt et al. 2002; Maca-Meyer et al. 2003; Reidla et al. 2003; Achilli et al. 2004, 2005; Palanichamy et al. 2004; Behar et al. 2006). Although haplogroup distributions were different in each village (table 1), total frequencies were as follows: haplogroup H 52.4%; haplogroup T 9.5%; haplogroup V 1.6%; haplogroup pre-V 1.6%; haplogroup J 6.4%; haplogroup U 20.6%; haplogroup K 3.2%, and haplogroup X 4.7%. No mtDNAs from haplogroups I and W were found. These frequencies differ from those of Herrnstadt et al. (2002), where haplogroup U has a lower frequency, but our data are similar to Kivisild et al. (2006). Moreover, we found that Urzulei has the highest haplogroup U frequency. We found an underrepresentation of haplogroup K compared with the literature probably due to genetic drift: haplogroup K was present only in Perdasdefogu.

We found that our 63 mtDNA sequences contained 172 sequence changes in the coding-region and 69 sequence changes in the D-loop region. In the coding region, 46 positions were nonsynonymous, 84 positions were synonymous, 20 substitutions were in 12s and 16s rRNA genes, 16 substitutions were in tRNA genes, 2 substitutions were in the origin of replication, and 4 in the noncoding region (see supplementary table, Supplementary Material online). Thirteen out of 172 sequence changes in the coding region are novel as they have not been previously reported in literature or described in the Mitomap Web site (table 2).

Polymorphisms A263G, A750G, A1438G, A4769G, A8860G, and A15326G were shared by all mtDNA sequences. The 2706G and 7028T polymorphisms were not present in most haplogroup H mtDNA sequences. The 11719A polymorphism was present in all T, J, U, X, and K haplogroups. The C14766T polymorphism was also present in all T, J, U, X, and K haplogroups, with one exception: Talana, subhaplogroup T2.

Topology of Phylogenetic Networks and Comparison with Europeans

The median network of 63 complete mtDNA sequences of the Ogliastra region is shown in figure 1. Our analysis revealed 241 segregating sites characterizing 51 haplotypes. The topology of our network shows 4 mtDNA haplogroups clusters: HV, UK, TJ, and X according to a recent mtDNA tree based on complete mtDNA coding-region sequences (Kivisild et al. 2006).

Haplogroups H and V form a cluster sharing 73A, 14766C, and 11719G polymorphisms, with 4 exceptions: 6P and 2P samples from Perdasdefogu sharing 73G, 2798T, and 2579T samples from Talana sharing 14766T. The H haplogroup network is highly starlike, and at least 2 subclusters emerge: H1, defined by 3010A encompassing 34% of the HV cluster, and H3, defined by 6776C encompassing 48% of the HV cluster. The starlike nature of H3 and H1 subhaplogroups suggests that these lineages have undergone a recent expansion. Sample 62U could be assigned to H4 and is characterized by 3992T, 4024G,

Table 2
Novel Mutations Detected in the Coding Region of mtDNA of the Samples of the Ogliastra Region

			Haplogroup							
Nucleotide	Position	Amino Acid Substitution	H	Pre V	V	T	J	U	K	X
A3615G	ND1	Synonymous	4	−	−	−	−	−	−	−
T4314C	$tRNA^{Ile}$	—	−	−	+	−	−	−	−	−
C6491T	CO1	Synonymous	−	−	−	−	+	−	−	−
A7571C	$tRNA^{Asp}$	—	4	−	−	−	−	−	−	C
T7679C	CO2	Phe → Leu	−	−	−	3	−	−	−	−
T11116C	ND4	Synonymous	−	−	−	−	−	2	−	−
A11200G	ND4	Synonymous	+	−	−	−	−	−	−	−
A12059C	ND4L	Asn → His	2	−	−	−	−	−	−	−
A12662G	ND5	Asn → Ser	−	+	−	−	−	−	−	−
C13807T	ND5	Leu → Phe	−	−	−	−	−	−	−	2
A13907G	ND5	Asn → Ser	2	−	−	−	−	−	−	−
C14329T	ND6	Synonymous	3	−	−	−	−	−	−	−
C15318A	Cyt b	Ala → Asp	−	−	−	−	−	+	−	−

5004C, 7356A, 7521A, 9123A, 14365T, and 14582G sequence changes. There are also typical H haplotypes pertaining to 136U (186A, 2789T, 4823C, and 16192T), 2P (73G, 152C, 8557A, 12358G, and 16145A), and 2545T (16167T and 16261T) samples. Sample 2010T was identified as V and is characterized by sequence changes 72C, 3745A, 4314C, 4580A, 7852A, 15904T, 15905C, 16188T, 16294T, and 16298C. Sample 6P was identified as pre-V and is characterized by 72C, 73G, 12662G, and 15904T sequence changes. The 62U, 136U, 2P, 2545T, 2010T, and 6P samples, present at equal frequency within the HV cluster, represent 18% of samples of the cluster.

Within the HV cluster, some sequences differ from the previously published mtDNA phylogeny. Sample 6P was assigned to the HV cluster. It has 2 substitutions typical of V haplogroup (72C and 15904T) but lacks 73A, 4580A, and 16298C. Based on these findings and on its network position, this sample could be assigned to the pre-V haplogroup (Achilli et al. 2004). Sample 2P is similar to sample number 28 of Achilli et al. (2004), except for the substitution 9368G. Samples 2465T and 2477T present the 3010A, although they are H3. Moreover, sample 2554T has 3010G, although its network position is H1. All these cases could be explained by back mutation or homoplasy within the same haplogroup because such sites are known to be prone to recurrent mutations (Achilli et al. 2004; Loogvali et al. 2004; Palanichamy et al. 2004).

The TJ cluster is characterized by polymorphisms 4216C, 11251G, 15452A, and 16126C. In our population, the haplogroup T could be subdivided into 2 subhaplogroups, T2 and T2b. The haplogroup J could be subdivided into 3 subhaplogroups: J1c, J2a, and J2b.

Polymorphisms 12372A, 12308G, and 11467G characterize the UK cluster. In our network, 5 subhaplogroups are present within haplogroup U: U6, U1a, U5b1, U5b2, and U5b3. U6 and K1a are distinct from the other subhaplogroups by the presence of 16311C.

Haplogroup K is a subhaplogroup within U (Achilli et al. 2005; Behar et al. 2006; Kivisild et al. 2006). Subhaplogroup U6 is characterized by 3348G, 4336T, 4454C, 5147A, 6575G, 12501A, 14470C, 14518G, 16172C, 16219G, and 16261T sequence changes.

Subhaplogroup U1a is characterized by 195C, 285T, 663G, 2218T, 4991A, 6026A, 7403G, 7581C, 11116C, 12879C, 13104G, 13656C, 14070G, 14364A, 15115C, 15148A, 15217A, 15954C, 16145A, 16189C, and 16249C sequence changes.

Subhaplogroups U5b1, U5b2, and U5b3 are defined by sequence changes 150T, 3197C, 7768G, 9477A, 13617C, 14182C, and 16270T. A new, well-defined cluster, characterized by 6 transitions (373G, 7226A, 11177T, 16192T, 16235G, and 16519C) and 1 transversion at position 16169A, is identified; in accordance with the nomenclature system of Richards et al. (1998), it could be tentatively named U5b3. This subhaplogroup, never previously reported in the literature, seems to be restricted to Sardinia.

In our network, the haplogroup X is characterized by polymorphisms 153G, 195C, 225A, 1719A, 6221C, 6371T, 12705T, 13966G, 14470C, 16189C, 16223 T, 16278T, and 16519C. Additional sequence changes subdivide the main branch into subhaplogroups, X2b and X2 (Reidla et al. 2003).

We found homoplasy cases in different haplogroups: 709A (for haplogroups T–U), 5147A (for haplogroups T–U), 5656G (for haplogroups T–U), 10398G (for haplogroups J–K), 14798C (for haplogroups J–K), 14470C (for haplogroups U–X), 11914A (for haplogroups H–U), 11950G (for haplogroups H–U), and 15607G (for haplogroups H–T). We detected in all 18 homoplasic positions in the D-loop region and some even occurring at subhaplogroup-defining sites.

Figure 2 shows a network based on European sequences described by Herrnstadt et al. (2002) together with 76 European samples from Kivisild et al. (2006). Taking into account the ethnic origin (United States and United Kingdom) of Herrnstadt sequences, we added the Kivisild samples covering numerous European countries in order to produce a more complete portrait of the European network. The Kivisild samples include 5 Northern European, 12 Italian, 1 Greek, 2 Finn, 2 Ashkenazi, 1 Georgian, 17 Hungarian, 3 Icelander, 3 Czech, 1 Sardinian, 5 Basque, 1 Iberian, and 23 Dutch.

We inserted our samples in this European network to show the position of Ogliastra in comparison with

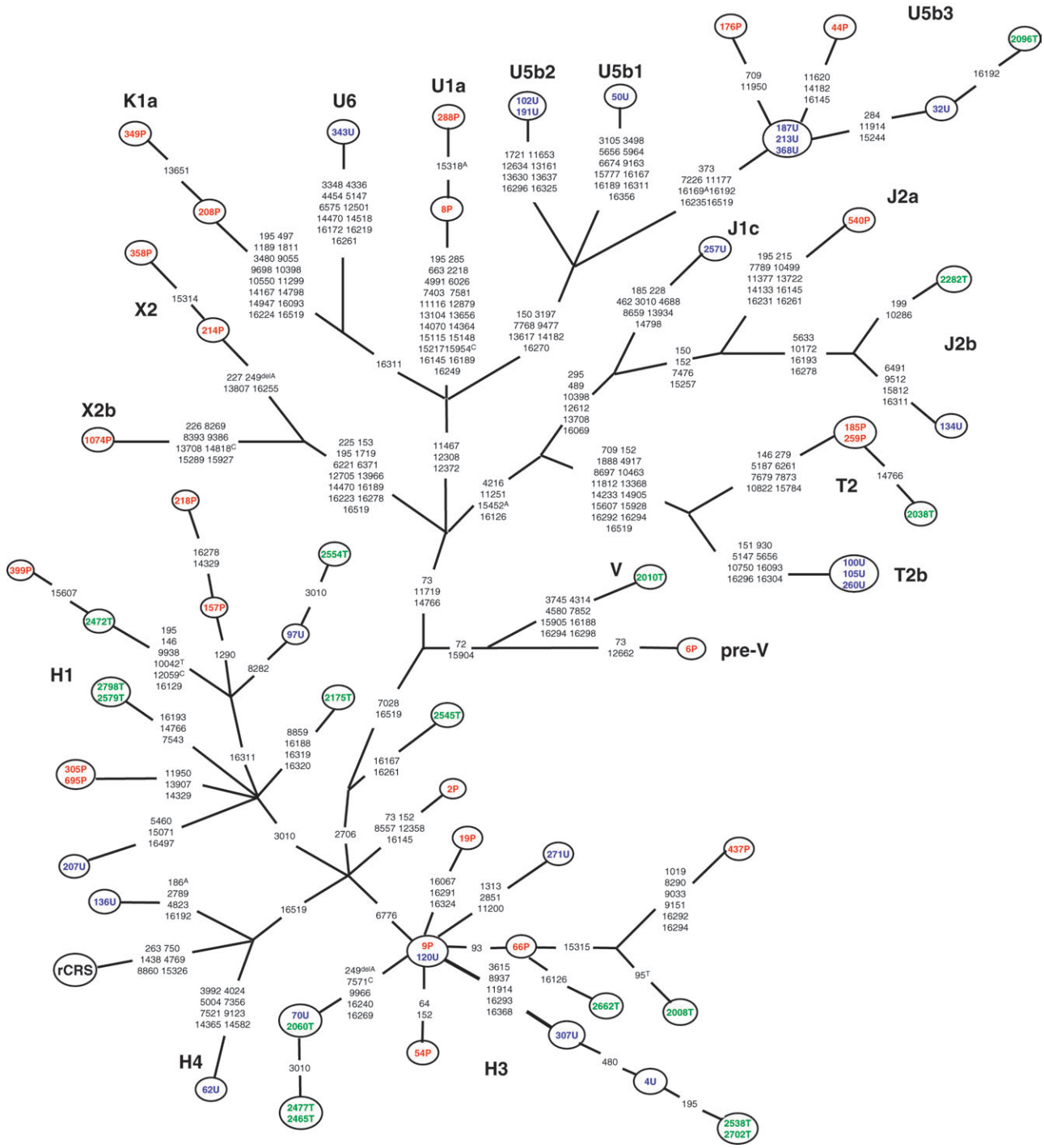

FIG. 1.—Phylogenetic network of 63 complete mtDNA sequences of the Ogliastra region relative to the recent rCRS. Green, blue, and red numbers at the nodes indicate the samples of villages of Talana, Urzulei, and Perdasdefogu, respectively. All nucleotide substitutions are transitions unless otherwise indicated by suffixes, which denote transversions. Some branch lengths have been distorted to increase legibility.

European lineages. Our samples are placed externally within each haplogroup, reflecting influence of factors such as founder effect and genetic drift. There is general agreement regarding the formation of H, H1, H3, V, T2, T2b, J1c, J2a, J2b, U5b1, U5b2, X2b, and X2 subhaplogroups, but for subhaplogroup K1a, we changed Herrnstadt et al. (2002) network topology according to Behar et al. (2006).

Analyses of Genetic Variation

The results of the statistical analysis of genetic diversity within each population are presented in table 3. The gene diversity for the whole sequence within populations ranges from 0.970 to 0.993 depending on the locality, thus showing a relatively high genetic variability within each population. Those high values could be explained by the

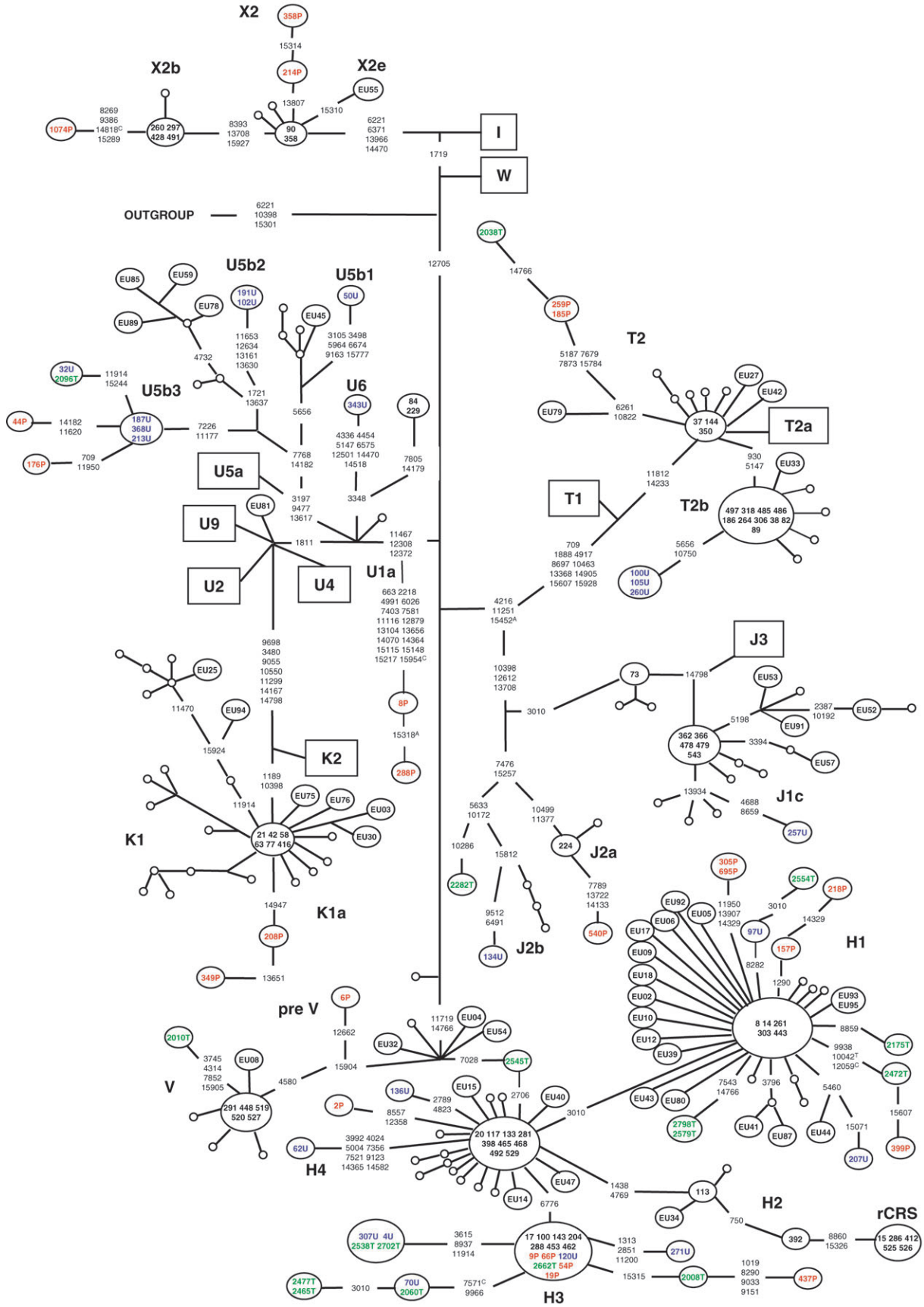
X2
358P
15314
214P
X2b
X2e
EU55
13807
15310
8269
9386
14818C
15289
1074P
260 297
428 491
8393
13708
15927
90
358
6221
6371
13966
14470
I
1719
W
OUTGROUP
6221
10398
15301
12705
2038T
14766
259P
185P
T2
5187 7679
7873 15784
EU79
6261
10822
37 144
350
EU27
EU42
T2a
930
5147
EU33
T2b
497 318 485 486
186 264 306 38 82
89
11812
14233
T1
709
1888 4917
8697 10463
13368 14905
15607 15928
5656
10750
100U
105U
260U
U5b2
U5b1
EU85
EU59
EU78
EU89
191U
102U
50U
EU45
3105 3498
5964 6674
9163 15777
11653
12634
13161
13630
4732
1721
13637
5656
U6
343U
84
229
4336 4454
5147 6575
12501 14470
14518
7805
14179
32U
2096T
11914
15244
U5b3
44P
14182
11620
187U
368U
213U
7226
11177
176P
709
11950
7768
14182
U5a
3197
9477
13617
3348
EU81
U9
U2
U4
1811
11467
12308
12372
U1a
663 2218
4991 6026
7403 7581
11116 12879
13104 13656
14070 14364
15115 15148
15217 15954C
4216
11251
15452A
J3
10398
12612
13708
3010
73
14798
EU53
2387
10192
EU52
5198
EU91
362 366
478 479
543
3394
EU57
13934
4688
8659
J1c
257U
9698
3480
9055
10550
11299
14167
14798
EU25
11470
EU94
15924
K2
8P
15318A
288P
7476
15257
5633
10172
10499
11377
224
J2a
10286
15812
2282T
7789
13722
14133
540P
9512
6491
134U
J2b
1189
10398
11914
EU75
EU76
EU03
EU30
21 42 58
63 77 416
K1
14947
K1a
208P
349P
13651
305P
695P
2554T
218P
3010
11950
13907
14329
14329
EU92
EU06
EU05
97U
EU17
EU09
157P
H1
8282
EU18
1290
EU02
EU93
EU95
EU10
8 14 261
303 443
8859
EU12
2175T
9938
10042T
12059C
EU39
2472T
7543
14766
3796
5460
EU43
EU80
EU41
EU87
EU44
15071
15607
2798T
2579T
207U
399P
pre V
6P
12662
11719
14766
EU04
EU54
EU32
7028
2545T
15904
2010T
3745
4314
7852
15905
EU08
4580
V
291 448 519
520 527
136U
2789
4823
EU15
2706
EU40
3010
2P
8557
12358
20 117 133 281
398 465 468
492 529
62U
3992 4024
5004 7356
7521 9123
14365 14582
H4
EU47
1438
4769
113
H2
EU34
750
392
8860
15326
rCRS
15 286 412
525 526
EU14
6776
307U 4U
2538T 2702T
3615
8937
11914
17 100 143 204
288 453 462
9P 66P 120U
2662T 54P
19P
1313
2851
11200
271U
15315
2008T
1019
8290
9033
9151
437P
2477T
2465T
3010
70U
2060T
7571C
9966
H3

Table 3
Genetic Diversity within Each Population: Number of Segregating Sites, Gene Diversity, Nucleotide Diversity, and Average Number of Pairwise Difference or Mismatch. In Bold Characters are the Values Relative to the Whole Mitochondrial Sequence and in Normal Character, the Ones Relative to the D-Loop Region Only

	Talana	Urzulei	Perdasdefogu	All
Segregating sites	**101**, 38	**129**, 40	**148**, 43	**234**, 69
Gene diversity	**0.978**, 0.963	**0.970**, 0.965	**0.993**, 0.975	**0.992**, 0.980
Nucleotide div Pi	**0.0010**, 0.0060	**0.0016**, 0.0075	**0.0016**, 0.0072	**0.0015**, 0.0072
Mismatch media	**17.10**, 6.75	**26.40**, 8.41	**25.95**, 8.09	**24.31**, 8.06

biased sampling strategy. Indeed, individuals were not sampled randomly but one individual from each lineage was selected, thereby increasing the genetic diversity within each population. However, for the D-loop, the values fall within the range of values reported for Italian populations (Handt et al. 1998).

On the other hand, the nucleotide diversity and the number of segregating sites is lower than that observed by Ingman et al. (2000) for non-Africans, using sequences from individuals from 32 different countries. This is also true for the D-loop sequence only. Similarly, the average number of pairwise differences is lower than the one found by Ingman et al. (2000), both for the complete sequence and for the D-loop, although, for the D-loop, it is slightly higher (4.22) than the value previously found in Sardinia by Di Rienzo et al. (1991). These values are similar to the ones found in other European populations, for instance, 8.4 in Central Italy (Handt et al. 1998; Tagliabracci et al. 2001).

In brief, most of our values of genetic diversity were lower than the ones found by Ingman et al. (2000) but higher than the ones previously estimated for Sardinia; this could be explained, respectively, by the facts that 1) our study was based on only 3 localities (instead of 32 countries in the study of Ingman et al.) and 2) the sampling was not random but specifically designed to explore intrapopulation diversity, thus increasing measures of internal diversity compared with other studies in Sardinia.

The values of genetic distance between populations, both F_{ST} and ϕ_{ST}, are relatively low ($F_{ST} = 0.0173$; $\phi_{ST} = 0.0398$). The percentage of variation based on haplotype frequencies among populations was 1.73 ($P < 0.01$) and within populations was 98.27, whereas the percentage of variation based on molecular distances was 3.98 ($P < 0.05$) and 96.02, respectively, among and within populations. However, AMOVA reveals that between 1.7% and 4.0% of the genetic variation is due to differences among populations (considering either haplotype frequencies or molecular distances), which is statistically significant. However, these values are consistent with the ones observed for mainland Italians and Sardinians by Barbujani et al. (1995), who suggested that less than 6% of the mitochondrial diversity could be attributed to differences between localities.

Tests of Mutation-Drift Equilibrium

Tajima's *D* is negative in all populations sampled but never significant, thus not showing any departure from neutrality, unless pooling all 3 populations together (table 4). This is in contrast with most previous studies on mtDNA, which have shown significant negative values, generally interpreted as a consequence of population expansions (Excoffier and Schneider 1999). For instance, based on the D-loop region, both Merriwether et al. (1991) and Barbujani et al. (1995) found significant Tajima's *D* values in Sardinians; and Verginelli et al. (2003) also found significant values in central-eastern Italy. Similarly, in a study of complete mitochondrial sequences (Ingman et al. 2000), significant negative values of Tajima's *D* were observed in the non-African samples. Fu and Li's *D* and *F* tests were not significant in our study, even after pooling the populations.

The insignificant values obtained in our tests of mutation-drift equilibrium do not seem due to the sampling scheme. Tajima's *D* is based on the differences between the number of segregating sites and the average number of nucleotide differences. Fu and Li's *D* test is based on the differences between the number of singletons (mutations appearing only once among the sequences) and the total number of mutations, whereas Fu and Li's *F* test is based on the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences.

Positive values are observed in stationary populations, in which a substantial number of mutations are shared by different lineages (Rogers and Harpending 1992). Conversely, in expanding populations, most mutations tend to be unique to a single lineage, resulting in negative values. By selecting and completely sequencing mitochondrial genomes of different haplogroups or by jointly analyzing members of different populations, one is overestimating the fraction of singletons (and segregating sites) and, hence, biasing downwards the estimates.

←

FIG. 2.—Phylogenetic network of 259 European mtDNA sequences from Herrnstadt et al. (2002), 76 European mtDNA sequences from Kivisild et al. (2006), and 63 mtDNA sequences of the Ogliastra region based on coding-region variations relative to the recent rCRS. Numbers at the nodes indicate samples; green, blue, and red numbers indicate the samples of villages of Talana (T), Urzulei (U), and Perdasdefogu (P), respectively. Empty nodes were extensively described by Herrnstadt et al. (2002). Samples indicated by European Union code come from Kivisild et al. (2006). All nucleotide substitutions are transitions unless otherwise indicated by suffixes, which denote transversions. Some branch lengths have been distorted to increase legibility. The branches I, W, T1, T2a, J3, U5a, U9, U2, U4, and K2 were not expanded because our samples did not lie in these mtDNA lineages.

Table 4
Tests of Selection of Tajima's *D*, Fu and Li's *D* and *F*, Fu's F_s; in Bold Characters are the Values Relative to the Complete Mitochondrial Sequence

	Talana	Urzulei	Perdasdefogu	All
Tajima's *D*	**−1.819**, −1.648	**−1.029**, −0.912	**−1.381**, −1.144	**−1.799***, −1.532
Fu and Li's *D*	**−2.037**, −1.259	**−1.126**, −0.822	**−0.119**, −0.143	**−1.456**, −0.734
Fu and Li's *F*	**−2.288**, −1.587	**−1.284**, −0.993	**−0.610**, −0.534	**−1.910**, −1.247
Fu's F_s	**−1.319**, −3.639*	**−0.167**, −3.809*	**−4.193***, −5.589**	**−15.080***, −20.213***

*$P < 0.05$; **$P < 0.01$; ***$P < 0.0001$.

On the other hand, Fu's F_s values are both significant for complete mtDNA sequences and for the D-loop, when considering all localities together. The F_s test statistic (Fu 1997) is based on the probability of having a number of alleles greater or equal to the observed number in a sample drawn from a stationary population. This statistic is particularly sensitive to population growth (Excoffier and Schneider 1999). Using the data of Di Rienzo et al. (1991) on Sardinia, Excoffier and Schneider also found significant negative values, indicating a period of population expansion. It is possible that we found significant values only with this last test because it is the most sensitive to population demographic processes, and there is evidence that population growth in Ogliastra might have been slower than in other European countries (Barbujani et al. 1995).

Discussion

Haplogroup definition and population genetic inference would greatly benefit from a better understanding of mtDNA variations in coding regions, but complete mtDNA sequences are rarely produced. Analysis of complete coding-region sequences of our Ogliastra samples revealed 13 novel substitutions. Our study ascertained that 26% of nonsynonymous substitutions involve threonine codon: most changes replace threonine with alanine. Comparing human, primate, and other mammalian mtDNA reveals that threonine/alanine substitutions are overrepresented in humans, whereas methionine/leucine changes are common in other mammalian species (Kivisild et al. 2006).

So far, only a few studies have carried out complete mitochondrial sequences, and we have analyzed the complete coding region of Sardinian samples for the first time. The number of never previously described substitutions discovered in this study is probably due to the low number of complete mtDNA sequences reported in the literature. We could hypothesize that some substitutions are typical of the Ogliastra region and/or Sardinia, whereas others will be found in the future in other populations. Moreover, we found some typical substitutions, that is, the U5b3 subhaplogroup, probably peculiar to the Ogliastra region.

In the present study, we constructed a network showing the general topology of coding and D-loop regions of Talana, Urzulei, and Perdasdefogu samples, and we inserted it in a broader network containing other European sequences described in literature.

Our complete mtDNA sequences can be arranged into an unambiguous network, in agreement with previously published D-loop–based networks. Each haplogroup was defined by many coding-region sequence changes, whereas only a few D-loop substitutions appeared to be important in this respect (as reported for 6P sample). In Talana, Urzulei, and Perdasdefogu (Fraumene et al. 2003), all reticulations were resolved, demonstrating the importance of integrating the study of the 2 regions in order to define a correct and simplified haplogroup phylogeny. This study characterizes also with greater precision some haplogroups previously based only on control region data and RFLP analysis. We find that there are actually many haplogroup-associated, haplogroup-defining, and haplogroup-specific sequence changes and a large number of homoplasic sites. The network topology of our samples is congruent with previously published Western Eurasian basal haplogroup phylogenies (Herrnstadt et al. 2002; Reidla et al. 2003; Achilli et al. 2004; Palanichamy et al. 2004; Achilli et al. 2005), with some minor external branch differences probably due to retromutation and events of homoplasy. Homoplasy or back mutation presence within the coding region is already documented (Loogvali et al. 2004) and confirms some site variability in mtDNA evolution (Aris-Brosou and Excoffier 1996). Comparing with other European coding-region sequences, it appears that these 3 small populations remained extremely isolated and genetically differentiated in the European context. Different biodemographic histories determined by population expansions, immigration rates, endogamy, founder effect, and genetic drift shaped each one of these populations characterized by unusual genetic features.

Comparing statistics describing genetic diversity in Ogliastra with those of other populations is not straightforward because of the sampling scheme chosen. Here we observed relatively high values of gene diversity, with respect to other isolated populations. This is the expected consequence of the sampling strategy; indeed when specific mtDNA lineages are selected, one is artificially increasing diversity within populations. However, the number of segregating (or polymorphic) sites is less than that in other studies, both for the whole sequence and for the D-loop alone. Therefore, it is highly likely that variation within populations is actually reduced in the Ogliastra region, a result suggesting that the isolation documented in the historical record has actually affected the people's genetic features. Along with the observation that F_{ST} and Φ_{ST} values are high among the 3 villages sampled, our finding would suggest that the Ogliastra population is unlikely to show internal stratification. Only an analysis of genetic diversity for autosomal markers implementing a random sampling strategy could confirm that the 3 populations studied in this region are distinct and internally homogeneous in general. This could also clarify a possible departure from neutral expectations (for

Fu's F_s statistic) due to an excess of low-frequency haplotypes. A similar excess of rare haplotypes (and a related deficit of intermediate-frequency haplotypes) was observed in most previous mitochondrial studies, with some populations of hunters and gatherers being the main exception (see e.g., Excoffier and Schneider 1999). We assume chances are that the Ogliastra populations, despite the relative paucity of the resources available and with their isolation, did slowly expand, much like other food-producing populations. This hypothesis should be further investigated in the future. The results of the Tajima's *D*, Fu and Li's *D* and *F* tests, and Fu's *F*s suggest that because of the relative paucity of the resources available and their isolation, the Ogliastra populations increased in size too slowly, unlike other food-producing populations, to carry a clear mark of the expansion in their mitochondrial genome.

To conclude, complete networks could help to distinguish between a rare polymorphism and a pathogenic mutation in clinically affected people. Similar phylogenetic network studies should be carried out in other populations for furthering medical and population genetics.

Electronic Database Information

The URLs for data in this article are as follows:

Genome Data Base, http://www.gdb.org/.

Helsinki declaration, http://www.wma.net/e/policy/17-c_e.html.

Life Sciences and Engineering Technology Solutions, http://www.fluxus-engineering.com/ (for Network 4.1 software).

Supplementary Material

Complete coding-region sequence information for 63 individual samples that have been submitted to GenBank under accession nos. DQ523619–DQ523681 and supplementary table are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Giuseppe Ledda and Teresa Manias for genealogical analysis. We are very grateful to the populations of Talana, Perdasdefogu, and Urzulei for their collaboration and to *Parco Genetico dell'Ogliastra* and the municipal administrations for their economic and logistic support. We acknowledge intramural funding from Shardna Life Sciences and support by grant from the Italian Ministry of Education, University and Research n: 5571/DSPAR/2002.

Literature Cited

Achilli A, Rengo C, Battaglia V, et al. (13 co-authors). 2005. Saami and Berbers—an unexpected mitochondrial DNA link. Am J Hum Genet 76(5):883–6.

Achilli A, Rengo C, Magri C, et al. (21 co-authors). 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. Am J Hum Genet 75(5):910–8.

Anderson S, Bankier AT, Barrell BG, et al. (14 co-authors). 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–65.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23(2):147.

Angius A, Melis PM, Morelli L, Petretto E, Casu G, Maestrale GB, Fraumene C, Bebbere D, Forabosco P, Pirastu M. 2001. Archival, demographic and genetic studies define a Sardinian sub-isolate as a suitable model for mapping complex traits. Hum Genet 109:198–209.

Aris-Brosou S, Excoffier L. 1996. The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. Mol Biol Evol 13:494–504.

Bandelt HJ, Forster P, Rohl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48.

Bandelt HJ, Forster P, Sykes BC, Richards MB. 1995. Mitochondrial portraits of human populations using median networks. Genetics 141:743–53.

Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V. 2002. The fingerprint of phantom mutations in mitochondrial DNA data. Am J Hum Genet 71(5):1150–60.

Barbujani G, Bertorelle G, Capitani G, Scozzari R. 1995. Geographical structuring in the mtDNA of Italians. Proc Natl Acad Sci 92(20):9171–5.

Behar DM, Metspalu E, Kivisild T, et al. (20 co-authors). 2006. The matrilineal ancestry of ashkenazi jewry: portrait of a recent founder event. Am J Hum Genet 78(3):487–97.

Brown MD, Sun F, Wallace DC. 1997. Clustering of Caucasian Leber hereditary optic neuropathy patients containing the 11778 or 14484 mutations on an mtDNA lineage. Am J Hum Genet 60(2):381–7.

Chinnery PF, Taylor GA, Howell N, Andrews RM, Morris CM, Taylor RW, McKeith IG, Perry RH, Edwardson JA, Turnbull DM. 2000. Mitochondrial DNA haplogroups and susceptibility to AD and dementia with Lewy bodies. Neurology 55(2):302–4.

Ciulla TA, Sklar RM, Hauser SL. 1988. A simple method for DNA purification from peripheral blood. Anal Biochem 174(2):485–8.

Dennis. 2003. Error reports threaten to unravel databases of mitochondrial DNA. Nature 421(6925):773–4.

Di Rienzo A, Wilson AC. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci 88:1597–601.

Excoffier L, Schneider S. 1999. Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. Proc Natl Acad Sci 96:10597–602.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–91.

Finnila S, Hassinen IE, Ala-Kokko L, Majamaa K. 2000. Phylogenetic network of the mtDNA haplogroup U in Northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. Am J Hum Genet 66(3):1017–26.

Finnila S, Lehtonen MS, Majamaa K. 2001. Phylogenetic network for European mtDNA. Am J Hum Genet 68(6):1475–84.

Finnila S, Majamaa K. 2001. Phylogenetic analysis of mtDNA haplogroup TJ in a Finnish population. Am J Hum Genet 46:64–9.

Forster P. 2003. To err is human. Ann Hum Genet 67(Pt 1):2–4.

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):1–10.

Fu Y-X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915–25.

Fu Y-X, Li W-H. 1993. Statistical tests of neutrality of mutations. Genetics 133:693–709.

Handt O, Meyer S, von Haeseler A. 1998. Compilation of human mtDNA control region sequences. Nucleic Acids Res 26(1):126–9.

Helgason A, Sigureth ardottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefansson K. 2000. Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 67:697–717.

Herrnstadt C, Elson JL, Fahy E, et al. (11 co-authors). 2002. Reduced-median-network analysis of complete mitochondrial DNA coding-region sequences for the major African, Asian, and European haplogroups. Am J Hum Genet 70(5):1152–71.

Herrnstadt C, Howell N. 2004. An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. Mitochondrion 4(5–6):791–8.

Howell N, Elson JL, Turnbull DM, Herrnstadt C. 2004. African Haplogroup L mtDNA sequences show violations of clock-like evolution. Mol Biol Evol 21(10):1843–54.

Ingman M, Kaessmann H, Paabo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408(6813):708–13.

Kalman B, Li S, Chatterjee D, O'Connor J, Voehl MR, Brown MD, Alder H. 1999. Large scale screening of the mitochondrial DNA reveals no pathogenic mutations but a haplotype associated with multiple sclerosis in Caucasians. Acta Neurol Scand 99(1):16–25.

Kivisild T, Shen P, Wall D, et al. (17 co-authors). 2006. The role of selection in the evolution of human mitochondrial genomes. Genetics 172(1):373–87.

Larruga JM, Diez F, Pinto FM, Flores C, Gonzalez AM. 2001. Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. Eur J Hum Genet 9:708–16.

Loogvali EL, Roostalu U, Malyarchuk BA, et al. (35 co-authors). 2004. Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. Mol Biol Evol 21(11):2012–21.

Maca-Meyer N, Gonzalez AM, Larruga JM, Flores C, Cabrera VM. 2001. Major genomic mitochondrial lineages delineate early human expansions. BMC Genet 2(1):13.

Maca-Meyer N, Gonzalez AM, Pestano J, Flores C, Larruga JM, Cabrera VM. 2003. Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. Genetics 16:4–15.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64: 232–49.

Mancosu G, Cosso M, Marras F, Borlino CC, Ledda G, Manias T, Adamo M, Serra D, Melis P, Pirastu M. 2005. Browsing isolated population data. BMC Bioinformatics 6 (Suppl 4):S17.

Mancosu G, Ledda G, Melis PM. 2003. PedNavigator: a pedigree drawing servlet for large and inbred populations. Bioinformatics 19:669–70.

Mancuso M, Conforti FL, Rocchi A, et al. (26 co-authors). 2004. Could mitochondrial haplogroups play a role in sporadic amyotrophic lateral sclerosis? Neurosci Lett 371(2–3):158–62.

Meinila M, Finnila S, Majamaa K. 2001. Evidence for mtDNA admixture between the Finns and the Saami. Hum Hered 52:160–70.

Merriwether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, Wallace DC. 1991. The structure of human mitochondrial DNA variation. J Mol Evol 33(6): 543–55.

Moilanen JS, Finnila S, Majamaa K. 2003. Lineage-specific selection in human mtDNA: lack of polymorphisms in a segment of MTND5 gene in haplogroup. J Mol Biol Evol 20(12):2132–42.

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 72:585–95.

Niemi AK, Hervonen A, Hurme M, Karhunen PJ, Jylha M, Majamaa K. 2003. Mitochondrial DNA polymorphisms associated with longevity in a Finnish population. Hum Genet 112(1):29–33.

Niemi AK, Moilanen JS, Tanaka M, Hervonen A, Hurme M, Lehtimaki T, Arai Y, Hirose N, Majamaa K. 2005. A combination of three common inherited mitochondrial DNA polymorphisms promotes longevity in Finnish and Japanese subjects. Eur J Hum Genet 13(2):166–70.

Palanichamy MG, Sun C, Agrawal S, Bandelt HJ, Kong QP, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP. 2004. Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. Am J Hum Genet 75(6):966–78.

Piazza A, Cappello N, Olivetti E, Rendine S. 1988. A genetic history of Italy. Ann Hum Genet 52:203–13.

Pyle A, Foltynie T, Tiangyou W, et al. (12 co-authors). 2005. Mitochondrial DNA haplogroup cluster UKJT reduces the risk of PD. Ann Neurol 57(4):564–7.

Rajkumar R, Banerjee J, Gunturi HB, Trivedi R, Kashyap VK. 2005. Phylogeny and antiquity of M macrohaplogroup inferred from complete mtDNA sequence of Indian specific lineages. BMC Evol Biol 5(1):26.

Reidla M, Kivisild T, Metspalu E, et al. (43 co-authors). 2003. Origin and diffusion of mtDNA haplogroup X. Am J Hum Genet 73(5):1178–90.

Richards M, Macaulay V. 2001. The mitochondrial gene tree comes of age. Am J Hum Genet 68(6):1315–20.

Richards MB, Macaulay VA, Bandelt HJ, Sykes BC. 1998. Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62(Pt 3):241–60.

Rieder MJ, Taylor SL, Tobe VO, Nickerson DA. 1998. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. Nucleic Acids Res 26(4):967–73.

Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–69.

Rose G, Passarino G, Carrieri G, Altomare K, Greco V, Bertolini S, Bonafe M, Franceschi C, De Benedictis G. 2001. Paradoxes in longevity: sequence analysis of mtDNA haplogroup J in centenarians. Eur J Hum Genet 9(9):701–7.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analysis by the coalescent and other methods. Bioinformatics 19(18):2496–7.

Ruiz-Pesini E, Lapena AC, Diez-Sanchez C, et al. (11 co-authors). 2000. Human mtDNA haplogroups associated with high or reduced spermatozoa motility. Am J Hum Genet 67(3):682–96.

Schneider S, Roessli D, Excoffier L. 2000. Arlequin ver. 2.000: a software for population genetics data analysis. Switzerland: Genetics and Biometry Laboratory, University of Geneva. Available from: http://lgb.unige.ch/arlequin/.

Tagliabracci A, Turchi C, Buscemi L, Sassaroli C. 2001. Polymorphism of the mitochondrial DNA control region in Italians. Int J Leg Med 114(4–5):224–8.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–95.

Tolk HV, Pericic M, Barac L, Klaric IM, Janicijevic B, Rudan I, Parik J, Villms R, Rudan P. 2000. MtDNA haplogroups in the populations of Croatian Adriatic Islands. Coll Antropol 24:267–80.

Torroni A, Bandelt HJ, Macaulay V, et al. (33 co-authors). 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. Am J Hum Genet 69:844–52.

Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus ML, Wallace DC. 1996. Classification of European mtDNAs from an analysis of three European populations. Genetics 144:1835–50.

Verginelli F, Donati F, Coia V, Boschi I, Palmirotta R, Battista P, Mariani Costantini R, Destro-Bisol G. 2003. Variation of the hypervariable region-1 of mitochondrial DNA in central-eastern Italy. J Forensic Sci 48(2):443–4.

Workman PL, Lucarelli P, Agostino R, Scarabino R, Scacchi R, Carapella E, Palmarino R, Bottini E. 1975. Genetic differentiation among Sardinian villages. Am J Phys Anthropol 43:165–76.

Sarah Tishkoff, Associate Editor

Accepted July 31, 2006