

## ARTICLE

# The population history of the Croatian linguistic minority of Molise (southern Italy): a maternal view

Carla Babalini<sup>1,5</sup>, Cristina Martínez-Labarga<sup>1,5</sup>, Helle-Viivi Tolkk<sup>2</sup>, Toomas Kivisild<sup>2</sup>, Rita Giampaolo<sup>1</sup>, Tiziana Tarsi<sup>1</sup>, Irene Contini<sup>1</sup>, Lovorka Barac<sup>3</sup>, Branka Janićijević<sup>3</sup>, Irena Martinović Klarić<sup>3</sup>, Marijana Peričić<sup>3</sup>, Anita Sujoldžić<sup>3</sup>, Richard Villems<sup>2</sup>, Gianfranco Biondi<sup>4</sup>, Pavao Rudan<sup>3</sup> and Olga Rickards<sup>\*1</sup>

<sup>1</sup>Department of Biology, University of Rome 'Tor Vergata', 00133 Rome, Italy; <sup>2</sup>Department of Evolutionary Biology, University of Tartu, and Estonian Biocentre, 51010 Tartu, Estonia; <sup>3</sup>Institute for Anthropological Research, 10000 Zagreb, Croatia; <sup>4</sup>Department of Environmental Sciences, University of L'Aquila, 67010 L'Aquila, Italy

This study examines the mitochondrial DNA (mtDNA) diversity of the Croatian-speaking minority of Molise and evaluates its potential genetic relatedness to the neighbouring Italian groups and the Croatian parental population. Intermatch, genetic distance, and admixture analyses highlighted the genetic similarity between the Croatians of Molise and the neighbouring Italian populations and demonstrated that the Croatian-Italian ethnic minority presents features lying between Croatians and Italians. This finding was confirmed by a phylogeographic approach, which revealed both the prevalence of Croatian and the penetrance of Italian maternal lineages in the Croatian community of Molise. These results suggest that there was no reproductive isolation between the two geographically proximate, yet culturally distinct populations living in Italy. The gene flow between the Croatian-Italians and the surrounding Italian populations indicate, therefore, that ethnic consciousness has not created reproductive barriers and that the Croatian-speaking minority of Molise does not represent a reproductively isolated entity.

*European Journal of Human Genetics* (2005) 13, 902–912. doi:10.1038/sj.ejhg.5201439;  
published online 11 May 2005

**Keywords:** mtDNA; haplogroups; Croatian-Italians; Croatians; Italians

## Introduction

People constituting an ethnic community share characteristic cultural traits of language, religion, patterns of social interaction and a feeling of unity and solidarity. While community members also believe that they share a common origin and history, very often this is only a myth. Many scholars maintain that language is an important if not the strongest feature of ethnic identity. Language represents the natural way of thinking, communicating,

and associating people with each other, thus providing the main social tie<sup>1</sup> that could influence the genetic structure of an ethnic group through marriage and reproduction. The spoken language can, in fact, act as a genetic and cultural barrier, increasing the genetic variation and divergence between populations. However, the construction of ethnic identity does not necessarily involve biological contributions.

Among the various groups speaking non-Italian languages who have slowly penetrated Italy since prehistoric times, the Croatian-speaking community of Molise represents the smallest ethnolinguistic minority.<sup>2</sup>

## Historic background

Population contacts between the two coasts of the Adriatic have existed since prehistoric times. According to historic

\*Correspondence: Professor O Rickards, Dipartimento di Biologia, Università di Roma 'Tor Vergata', Via della Ricerca Scientifica n. 1, 00133 Roma, Italy. Tel: +39 06 72594347; Fax: +39 06 2023500; E-mail: rickards@uniroma2.it

<sup>5</sup>These two authors contributed equally to this work  
Received 11 June 2004; revised 24 March 2005; accepted 12 April 2005;  
published online 11 May 2005

data, the first migrations of Croatians into central-southern Italy began as early as the 10th century and continued up to the 15th century. Small groups and families crossed the Adriatic and formed their own small colonies.<sup>3</sup> In contrast with these sporadic migrations, more lasting visible traces were left by Croatian immigrants to southern Italy during the 15th and 16th centuries. At that time, the Ottomans were advancing through the Balkan Peninsula to the Adriatic, forcing many population groups to flee from the continental interior toward the Adriatic coast where they resettled on the Middle and Southern Adriatic (Dalmatian) islands, the Istrian peninsula to the northwest and in the southern regions of Italy across the Adriatic. According to historic sources, 15 Croatian communities resettled in the region of Molise, with a total of 7000 to 8000 inhabitants.<sup>4</sup> Their number grew to over 15 000 and then gradually declined partially due to assimilation with the Italian population. Today, their descendents (total: 2081 individuals<sup>5</sup>) live in the mountainous inland villages of Acquaviva Collecroce, San Felice del Molise and Montemitro. Although Croatian is still spoken in all three villages, the number of individuals and families speaking it has greatly decreased through economically motivated migrations to other regions of Italy and overseas countries.<sup>2</sup>

### Linguistic evidence of the origin of Croatian populations

Throughout the last century, several different hypotheses as to the origin of these Croatian immigrants were put forth, along with a range of possible locations of origin including Istria and northern Dalmatia or Middle Dalmatia. However, a thorough linguistic analysis comparing the Molise speech with the local vernaculars of 52 villages in Istria and Dalmatia has shown that the speech of the inhabitants of Molise is undoubtedly a *štokavian-ikavian* dialect of the Croatian spoken in Middle Dalmatia before the Turkish invasion.<sup>6</sup> The linguistic analysis clearly supports the hypothesis that the ancestral homeland of the immigrants into the region of Molise is to be found in this area, specifically between the Cetina and Neretva rivers.

The present study examines the genetic effects of geographic and linguistic subdivision in Molise, a small region of Italy. In particular, direct sequencing of the two most variable segments of the mitochondrial DNA (mtDNA) control region and high-resolution RFLP analysis of the whole molecule were performed to describe mtDNA variation in the Croatian minority of Molise, the Croatian parental population and the neighbouring Italian groups in order to assess whether the spoken language has contributed to the preservation of the genetic identity of the Croatians living in Italy.

### Materials and methods

The study was carried out on three groups of unrelated and apparently healthy people with attested maternal genealogy. The first group consisted of 41 schoolchildren (age range, 11–15 years) from the three villages belonging to the Croatian ethnolinguistic minority of Molise (Acquaviva Collecroce, Montemitro and San Felice del Molise); the second comprised 199 individuals from central-southern Italy (52 from Lazio, 48 from Campania, 73 from Abruzzo/Molise, and 26 from Puglia); the third group consisted of 96 individuals from four villages located on the Makarska coast near the islands of Brač and Hvar, from north-west to south-east: Krilo Jesenice (35), Mimice (24), Zaostrog (13) and Živogošće (24). In addition, 311 individuals chosen randomly from the parish registries of three Croatian islands in the eastern Adriatic (Middle Dalmatian islands) (105 from Brač, 108 from Hvar, and 98 from Korčula<sup>7,8</sup>), for which HVS-I complete sequences and RFLP analyses were available, were also included. The Krk sample was not used for comparison since the ethnohistory of this island differs considerably from that of the other Croatian islands.<sup>7</sup> Figure 1 shows all sampling localities in Italy and Croatia and the reconstruction of historic migration directions.

### Sample preparation and DNA extraction

**Croatian ethnolinguistic minority of Italy** Total genome DNA was extracted from a small amount of blood samples according to the method described in Budowle *et al.*<sup>9</sup> The blood sample collection strategy is reported elsewhere.<sup>10</sup>

**Populations of Central Italy** After obtaining informed consent, blood or mouth swab samples were collected at the Blood Transfusion Centre of the Hospital of Rome 'Umberto I'. DNA was extracted following the procedures of Miller *et al.*<sup>11</sup> and Budowle *et al.*<sup>9</sup>

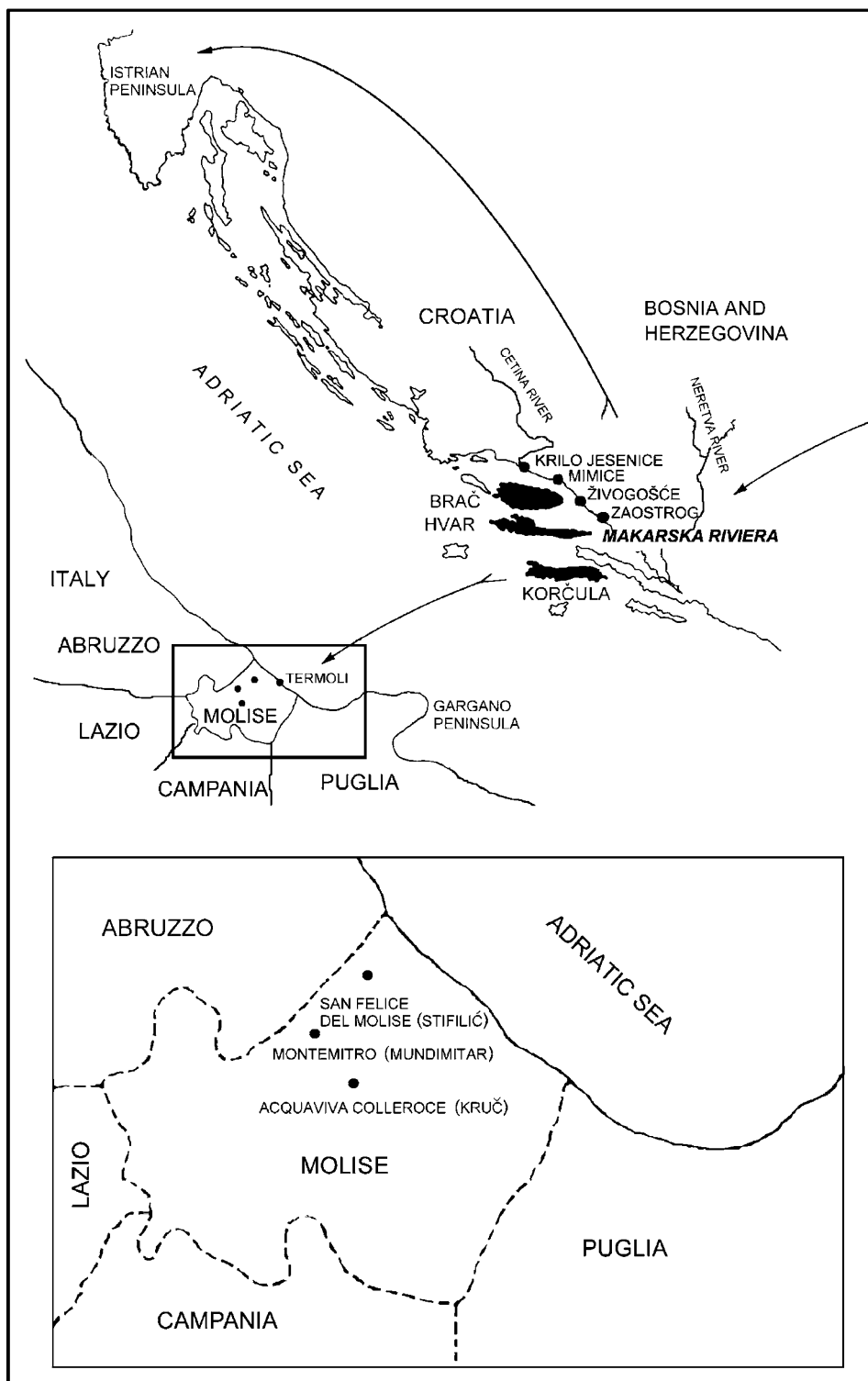
**Populations of Croatia** DNA extraction from blood samples and mtDNA analyses were performed following the procedures reported in Tolk *et al.*<sup>7</sup>

### mtDNA analyses

**D-loop PCR amplification and direct sequence analysis** The mtDNA HVS-I and HVS-II of the Croatian-Italian and the Italian samples were amplified simultaneously by applying the multiplex amplification strategy. For a detailed description of PCR and automated sequencing conditions, see Rickards *et al.*<sup>12,13</sup>

The HVS-I was amplified and sequenced for the southern Croatian samples between the region at nps 16024 and 16400.<sup>14,15</sup>

**RFLP screening** RFLP screening was performed in a hierarchical manner. Purified DNA of selected Croatian-Italian, Croatian and Italian samples was used to amplify



**Figure 1** Geographical map showing the location of the study samples and the migration routes from Croatia to Italy.

the polymorphic sites defining Eurasian and African mtDNA lineages.<sup>16</sup> Primers and conditions are reported in the Supplementary Information.

**Haplogroup classification strategy** For all the samples, 384 (376 for the coastal Croatians) bps of the HVS-I (from positions 16016/16024 to 16400) sequence and RFLPs were

determined to allocate a sequence to its haplogroup;<sup>16</sup> for the Italian and Croatian-Italian samples, information on HVS-II (from position 030–408) was also considered. Length variations were not considered in phylogenetic and statistical analyses. To avoid controversies with phylogenetic relations revealed by complete mtDNA sequences,<sup>17</sup> finer classification of control region sequences of haplogroups U, T, and J at subhaplogroup level (*sensu*<sup>18</sup>) was avoided.

**Population – genetic and phylogeographic analyses** - Gene and nucleotide diversities were calculated according to Nei.<sup>19</sup> The average number of nucleotide pairwise differences, the number of segregating sites, and Tajima's D statistics<sup>20</sup> were estimated. All these analyses were based on HVS-I (bps 16024–16400) and performed using Arlequin package software.<sup>21</sup> Distance matrices between populations were generated using two methods: Reynolds' distance<sup>22</sup> and Slatkin's linearized  $F_{st}$ 's.<sup>23</sup> Multidimensional scaling analysis (MDS<sup>24</sup>) based on the above reported distance matrices and principal component (PC) analysis based on haplogroup frequencies were performed with STATISTICA package software (StatSoft). The admixture proportions of the Italian and Croatian populations in the Croatian minority of Molise were quantified using the ADMIX 2.0 program.<sup>25</sup> Estimates of the admixture coefficients and their SEs were calculated using 1000 bootstrap replicates. A median network based on the algorithm by Bandelt *et al.*<sup>26</sup> was manually created to show the phylogenetic relationship of the combined RFLP and HVS-I data.

## Results and discussion

### Genetic diversity

The number of haplotypes, their frequency, the diversity indexes, and the mean number of pairwise differences according to HVS-I (Table 1) were calculated to assess the mtDNA genetic diversity of the ethnolinguistic minority of Molise and their neighbouring Italian and Croatian parental populations. Molecular information obtained from the three Croatian-Italian samples was pooled together since frequency analysis of the identified haplogroups did not show any significant differences between the three villages. In fact, when 95% credible regions were calculated, according to Richards *et al.*,<sup>16</sup> the estimated frequencies overlapped. Moreover, many haplotypes were identical in the three Croatian villages. Other haplotypes that were found were closely related to these. Also  $F_{st}$  values between the three villages did not show any significant differences (data not shown).

These findings are consistent with the homogeneity revealed by previous genetic studies on nuclear loci.<sup>10</sup>

The highest gene diversity in the mtDNA HVS-I region was observed among the southern Italians (Campania and Puglia); this is in line with the nonhomogenous genetic origin of these populations.<sup>27</sup> The Croatian-Italians, and the Croats in particular, displayed somewhat lower values. The reduced pattern of gene diversity and the high frequency of repeated haplotypes displayed by the Middle-Dalmatian sample is very likely a result of the effects of inbreeding, genetic drift and founder effect due to a certain degree of isolation of these populations, which is consistent with ethnohistoric records.<sup>6,28</sup> The nucleotide diversity index of the three groups displayed values that were

**Table 1** Diversity indexes for HVS-I in the Croatian-Italian, Italian and Croatian populations

Population	n	Haplotype number	Number of single haplotypes	Gene diversity	Nucleotide diversity	MPSD	Tajima's D	P
Croatian-Italians	41	29	22	0.972±0.015	0.015±0.008	5.311±2.617	-1.865	0.020
<i>Italians</i>								
Abruzzo-Molise	73	51	46	0.953±0.018	0.013±0.007	4.522±2.251	-2.171	0.005
Campania	48	41	39	0.981±0.014	0.015±0.008	5.371±2.636	-2.160	0.006
Lazio	52	37	32	0.959±0.020	0.013±0.007	4.503±2.259	-2.005	0.011
Puglia	26	24	23	0.991±0.015	0.016±0.009	5.834±2.881	-2.006	0.011
<i>Croatsians (coast)</i>								
Krilo Jesenice	35	21	13	0.960±0.017	0.009±0.005	3.422±1.792	-1.331	0.091
Mimice	24	17	11	0.971±0.019	0.010±0.006	3.993±2.068	-1.591	0.049
Zaostrog	13	9	5	0.949±0.042	0.012±0.007	4.615±2.421	-0.867	0.215
Zivogošće	24	14	9	0.924±0.038	0.008±0.005	3.373±1.791	-1.573	0.051
<i>Croatsians (islands)<sup>a</sup></i>								
Brač <sup>a</sup>	105	54	39	0.968±0.008	0.011±0.006	4.265±2.131	-1.942	0.014
Hvar <sup>a</sup>	108	51	31	0.973±0.006	0.014±0.008	5.642±2.727	-1.429	0.069
Korčula <sup>a</sup>	98	44	31	0.944±0.013	0.008±0.005	3.335±1.727	-2.113	0.006

<sup>a</sup>Data from Tolk *et al.*<sup>7,8</sup>

comparable with those of other European populations (range of variation: 0.010–0.019<sup>29</sup>). The average number of pairwise nucleotide differences (MPSD) was similar for the Italians and the Croatian-Italians, while the Croatians showed lower values. The Croatians and the Italians showed a bell-shaped distribution of HVS-I pairwise sequence differences, and negative and mostly significant Tajima's *D* values (see Table 1), signalling likely a late Pleistocene demographic expansion affecting the majority of European populations.<sup>30</sup> In contrast, the pairwise distributions in the Croatian ethnolinguistic minority of Molise showed two major peaks, one at six and another at three substitutional differences (data not shown), which is very likely compatible with a scenario of admixture events with Italian females, since heterogeneity among the three villages was excluded. In fact, Brakez *et al*,<sup>31</sup> analysing the mtDNA sequence variation in the Moroccan population, showed that admixture can create a bimodal pairwise difference distribution.

### Haplogroup frequency distribution

The samples were classified according to the existing mtDNA haplogroup nomenclature<sup>32,33</sup> on the basis of control region sequence data and diagnostic RFLP sites from coding region data (sequence and RFLP data are reported in Tables 2 and 3 of the Supplementary Information). The majority of the sequence types could be classified as belonging to typical European and Western Asian haplogroups (Table 2). Only one Italian sequence, from Lazio, could not be attributed to any specific haplogroup. It displayed only two transitions in HVS-I with respect to the Cambridge Reference Sequence (CRS),<sup>14</sup> namely at bps 16224 and 16273, and the RFLP analysis excluded its inclusion in either haplogroup H or K. The quantity of DNA for further analysis was too small to allow its proper classification in the phylogeny.

Sequences belonging to haplogroups that are uncommon to European populations, like L1b, L1c, M1, R\*, U6a, U7, D and F were found sporadically. The presence of three

**Table 2** Haplogroup frequency distribution in the Croatian-Italian, Italian and Croatian populations

Populations	Croatian-Italians (n=41)	Abruzzo/Molise (n=73)	Campania (n=48)	Lazio (n=52)	Puglia (n=26)	Krilo (n=35)	Mimice (n=24)	Zaostrog (n=13)	Živogošće (n=24)	Brač <sup>7,8</sup> (n=105)	Hvar <sup>7,8</sup> (n=108)	Korčula <sup>7,8</sup> (n=98)
H	31.7	45.2	31.3	44.2	53.9	54.3	37.5	23.1	54.2	35.2	27.7	60.3
Pre-HV		2.7										
HV	7.3	4.1	8.3	1.9		2.9			4.2	10.5	4.6	4.1
Pre-V	2.4			3.9	3.8					1.9	1.9	1.0
V		1.4	2.1			2.9	8.3	15.4		5.7	4.6	8.3
K	7.3	12.3	6.3	9.6		14.3			4.2	9.5	3.7	4.1
U		1.4	2.1							1.0		
U1		1.4							4.2	1.0		2.0
U2	2.4			1.9	3.8		4.2	23.1			4.6	1.0
U3	2.4	2.7		1.9	3.8	2.9			4.2	2.8		
U4			2.1				8.3			1.0	0.9	1.0
U5a	4.9	5.5	2.1	3.9	3.8	11.4			16.7	6.7	6.5	1.0
U5b			2.1				4.2	7.7		1.9	8.3	1.0
U6a				5.8								
U7				1.9			4.2					
Total U	17.0	23.3	14.7	25.0	11.4	28.6	20.9	30.8	29.3	23.9	24.0	10.1
T1	7.3		4.2		3.8			15.4		1.9	1.9	
T2	9.8	5.5	10.4	13.5	7.7		12.5		4.2	3.8	12.0	5.1
T4		1.4										
Total T	17.1	6.9	14.6	13.5	11.5	0	12.5	15.4	4.2	5.7	13.9	5.1
J	12.2	11.0	10.4	1.9	3.8	2.9			8.3	8.6	9.3	6.0
F											8.3	
R*	2.4		2.1					7.7				
X	2.4		2.1		7.7						1.9	
W			4.2	1.9	3.8					2.8	1.9	3.1
I						2.9	8.3	7.7		1.0	1.9	1.0
N*			2.1				4.2					
N1a	4.9	1.4								1.9		
N1b		2.7	8.3	3.9						2.8		
N1c		1.4										
D							8.3					
M1	2.4			1.9								
L1b					3.8							
L1c												1.0
NC				1.9								

sequences belonging to the northwestern African subhaplogroup U6a in the Italians (Lazio) might be of interest since this haplogroup has been previously found in Europe in notable frequencies only in Iberian Peninsula, where U6 could have been introduced already through a prehistoric African occupation.<sup>34,35</sup> The occurrence of these haplotypes in Lazio could also be attributed to gene flow from North West Africa, hypothesizing a possibly wider range of contacts between the circum-Mediterranean populations in the past.

Although Italians, Croatian-Italians and Middle-Dalmatian Croatians share the same basic haplogroups at similar frequencies, some of the identified haplogroups were differentially distributed among the three studied groups, a possible indication of genetic affinities between the Croatian-Italians and the other two populations. It should be emphasized, however, that since the haplogroup frequencies in individual populations, as exemplified by the wide range of proportions demonstrated by individual haplogroups in the Croatian villages (ie 0–15% for T), is mainly a subject of drift, notwithstanding possible sampling errors, the frequency pattern could be misleading and may not be the best way to estimate the parental contributions. Therefore, molecular differences between mitochondrial sequences rather than frequency differences between haplogroups were used in the following analyses.

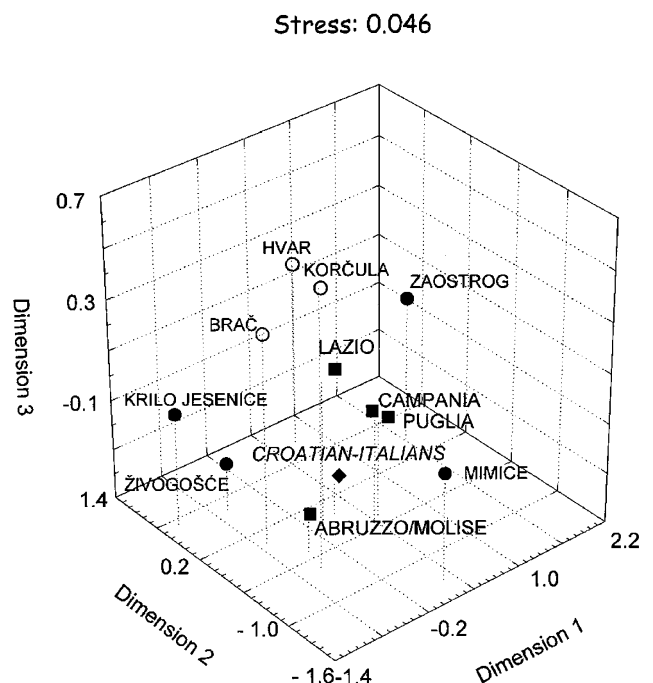
### Interpopulation differences

**Intermatch analysis** To investigate the affinities of the linguistic minority of Italy, the Croatian source populations and the Italians, the mean number of pairwise differences between the Croatian-Italian mtDNA sequences and those detected in the various Croatian and Italian villages (intermatch) was calculated, wherein the lower the pairwise difference, the closer genetically related the populations under comparison are. This analysis shows that the MPSD values are similar and low for both Croatia (coastal region: 4.633,  $P=0.004$ ; islands: 4.953,  $P=0.058$ ) and total Italy (5.057,  $P=0.238$ ; no differences in the MPSD values were found when the comparisons were made with each Italian region separately (data not shown)), indicating that the mtDNA sequences found in the three Croatian-speaking villages of Molise are closely related to both the present day Croatians, especially those living in the coastal area, and the surrounding Italian population. Such a pattern is expected from populations that may have interchanged a significant number of migrants over long periods of time.<sup>36</sup> In fact, when the intermatch analysis was extended to populations from different continents the obtained values were higher (from 7.902,  $P=0.691$  for Africa to 5.985,  $P=0.668$  for Asia).

**Genetic distances and migration rates between populations** To better define the maternal contribution of the Italian and the Croatian populations to the Croatian

ethnolinguistic minority living in Molise, genetic distance analysis was performed. The obtained distance matrices between the populations were visualized by MDS analysis. Figure 2 shows only the MDS plot based on Slatkin's linearized  $F_{st}$ , since the Reynolds' distances gave a similar plot (data not shown). The values reported by Kruskal<sup>24</sup> for the stress statistic, which measures the goodness of fit between distances in the configuration space and the monotonic function of the original distances, indicate that the level of stress of the present analysis is good, that is, the degree of distortion of the pairwise distances in the three-dimensional space is acceptable. Moreover, the not appreciably lower value of the stress provided by the four-dimensional plot is a further indication that the three-dimensional representation reproduces the distance matrix of the populations reasonably well.

A striking feature of the MDS representation is the intermediate position of the Croatian-Italians between the Italian and the Croatians. This pattern again seems to be compatible with a certain degree of admixture between the Croatian population of Molise and the surrounding Italian populations. The genetic relationships between populations obtained when PC analysis based on haplogroup frequencies is applied are similar to those based on Slatkin's linearized  $F_{st}$  distance matrix (data not shown).

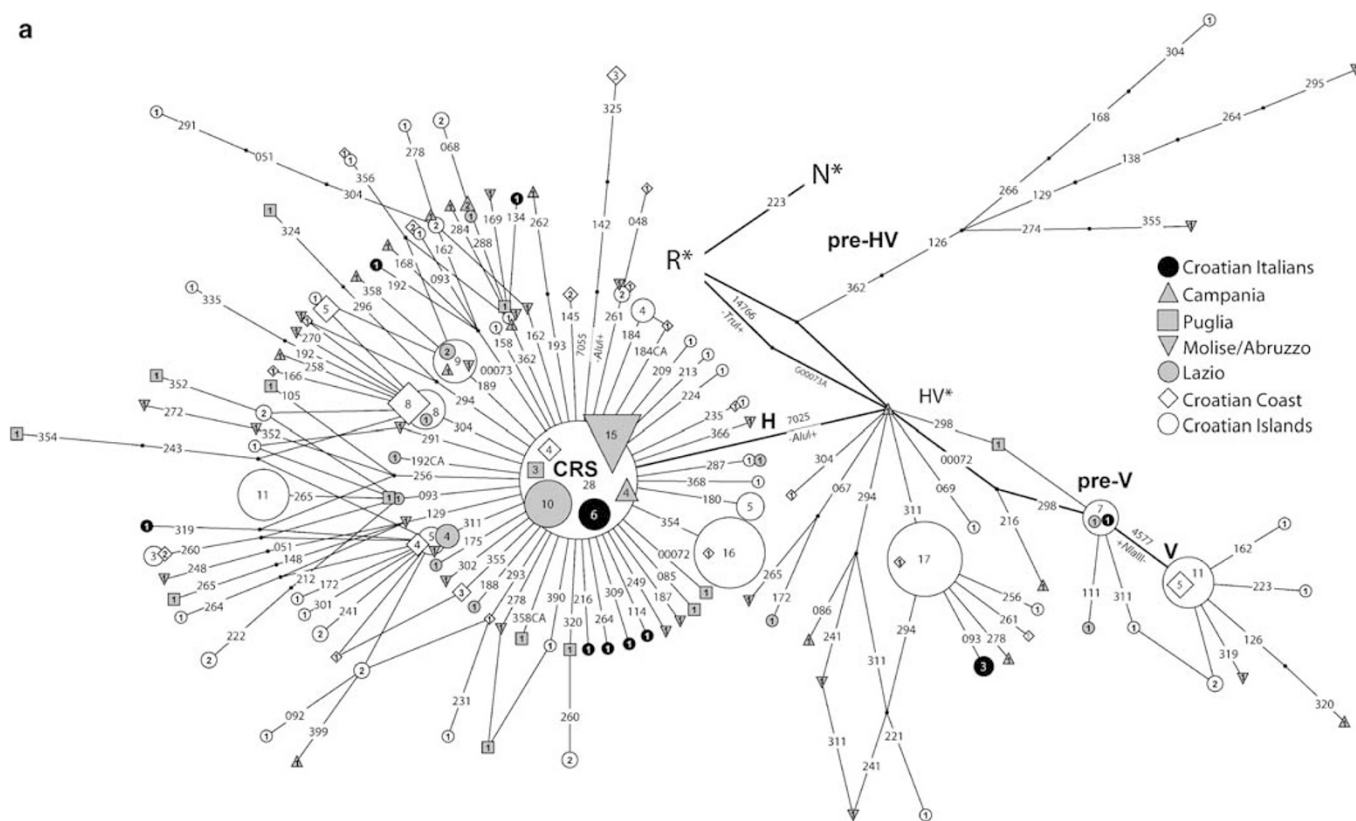


**Figure 2** Multidimensional scaling (MDS) plot of the genetic distances between the Croatian-Italian, Italian (full squares) and Croatian (empty circles: islands; full circles: coastal villages) populations on the basis of mtDNA HVS-I sequences in three-dimensional space.

To quantitatively estimate the degree of admixture in the ethnic minority of Molise, the method proposed by Dupanloup and Bertorelle<sup>25</sup> was applied. The admixture coefficients are 41% (SD 20%) for the coastal Croatian parental population and 59% for the Italian population from Molise. The estimated Croatian and Italian contributions to the Croatian-Italian gene pool did not change substantially considering as Croatian parental population either the coastal or the island villages, and as Italian parental population either the Molise sample or the total Italian sample (data not shown). However, this result must be taken with due caution. In fact, due to the large standard deviation associated with the admixture coefficients, the present data do not allow a reliable estimates of the parental contributions. To get a more reliable evaluation on these estimates, a phylogeographic approach was

performed, which allows as well to distinguish between a possible Croatian origin (from either the coast or the islands) and an Italian origin of each of the sequences found in the Croatian-Italians.

A median network was constructed to detect the exact matches and derivatives within the study populations (Figure 3a–c). The network shows three specific matches between the Croatian-Italians and the Italians that were not found in Croatian sample, equally large in size: K (16188–16218–16224–16311), U2 (16051–16189–16266–16294), T (16126–16294–16304), and four matches with the Croatians of Croatia that were not observed in Italian sample: T1 (16126–16163–16186–16189–16294), U5a (16256–16270–16399), J (16069–16126–16366). Only two of these matches (ie in K and U2) are specific, whereas other lineages occur with



**Figure 3** Median-network showing the phylogenetic relationship of mtDNA haplotypes detected among Croatian-Italians of Molise, Italians, populations of the Croatian coast and the southern Croatian islands. Diagnostic RFLP sites and nucleotide positions of the mutations in the HVSI (without the prefix 16 000) compared with the Cambridge Reference Sequence<sup>14,15</sup> are indicated along the branches connecting the haplotypes. The node sizes are proportional to haplotype frequencies specified by the number of individual sequences within each node. Origin of the samples: black nodes — Croatian-Italians of Molise; grey nodes — Italians; white nodes — populations of the Croatian coast and the southern Croatian islands. (a) The structure of mtDNA haplogroups H, HV, V, pre-V, and pre-HV. Proportion of individuals in these clades vs total sample size: Croatian-Italians of Molise (17/41), Italians from Campania (20/48), Puglia (15/26), Molise/Abruzzo (39/73), Lazio (26/52), Croatian Coast (51/96), Middle Dalmatian Islands (170/311). (b) The substructure of mtDNA superhaplogroups L3\* and N\*. Proportion of individuals in these clades vs total sample size: Croatian-Italians of Molise (4/41), Italians from Campania (8/48), Puglia (4/26), Molise/Abruzzo (3/73), Lazio (4/52), Croatian coast (9/96), Middle Dalmatian Islands (20/311). (c) The structure of mtDNA haplogroups J, T, U, F and R. Proportion of individuals in these clades versus total sample size: Croatian-Italians of Molise (20/41), Italians from Campania (20/48), Puglia (7/26), Molise/Abruzzo (31/73), Lazio (21/52), Croatian coast (36/96), Middle Dalmatian Islands (121/311).

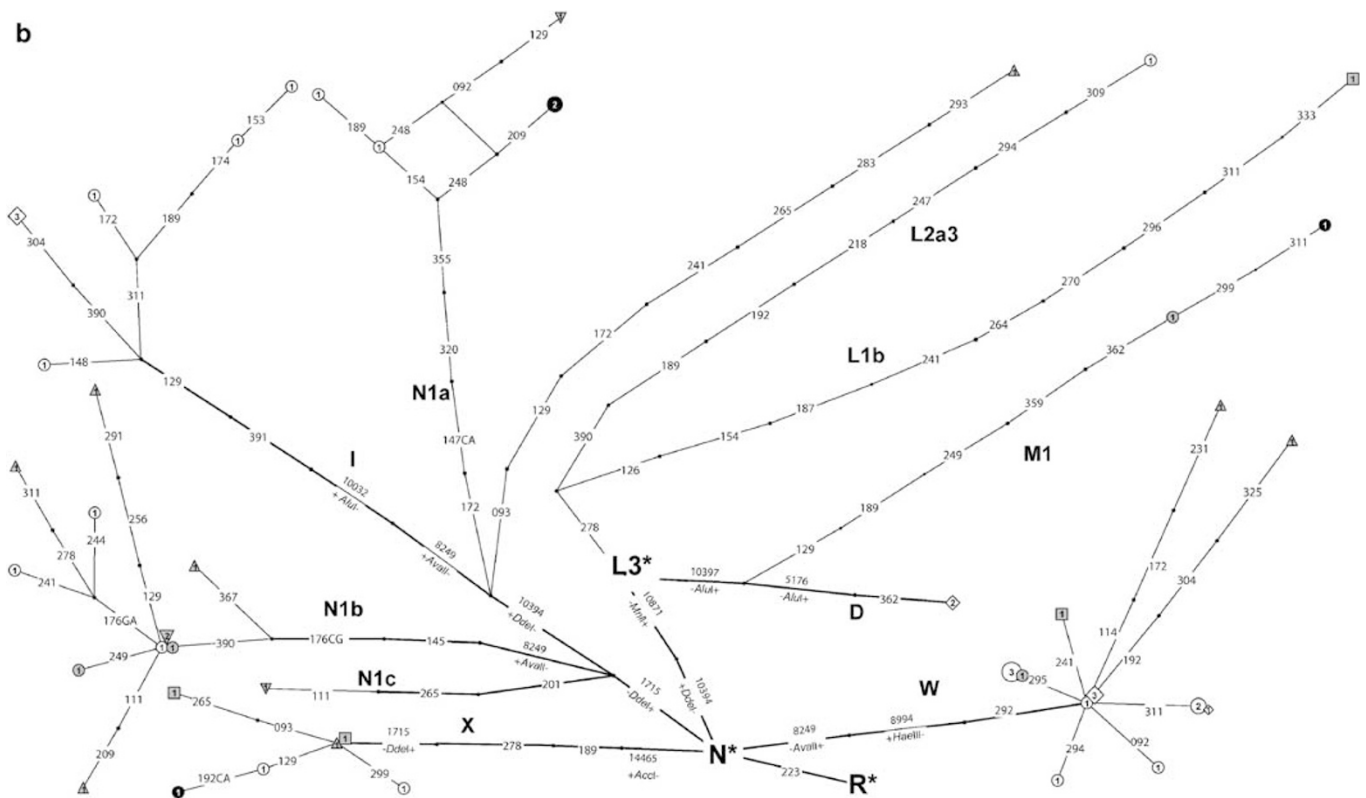


Figure 3 Continued

different frequencies among many European populations; the haplogroup J lineage (16069–16126–16366) seems to be frequent in the northwestern Balkan area but almost absent elsewhere (author's unpublished data).

Interestingly, the unusual haplogroup HV type characterized by transitions at 16093 and 16311, while quite frequent in the Croatian-Italians (7%), is missing in both Italy and Croatia. It has been observed so far only in the Romanians,<sup>16</sup> pointing to a possible eastern European origin of this haplotype. Otherwise, the unusual Croatian-Italian type could also have derived from an independent mutation at the hot-spot position 16093 in the frequent HV haplotype defined by the single mutation at 16311. The HV-16311 type is completely absent in continental Italy, whereas it is highly frequent in Croatia, where it has been observed in all local populations, with the highest frequency in Brač (9%). The Croatian-Italian unique N1a haplotype, shared by two individuals, derives from one transition at np 16209 from their common founder type sampled in Estonia, Lithuania, and Russia<sup>37</sup> (our unpublished data). So, a potential Slavic source could be argued or at least speculated for this lineage as well. The haplogroup H haplotype with 16134 and 16362 mutations has its closest relative among the Kabardinians.<sup>16</sup> This

motif derives from the likely paraphyletic subfounder lineage defined by 16362 mutation. H-16362 frequency in continental Italy is about 0.5%, whereas the Croatian average is five times higher (2.5%), reaching 6% (8/133) in the northwest island of Krk (our unpublished data). The USA founder haplotype 16114A–16192–16256–16270–16294, from which the Croatian-Italian sample CRO189 uniquely derives at np 16369, is absent in the total pool of continental Italy, but it is common in Slavic and north-eastern European populations<sup>38,16</sup> and it has been sampled twice also in our unpublished Croatian and Bosnian samples, indicating its likely Slavic origin. The same potential Slavic origin applies to the R\* lineage characterized by the recurrent 16343 mutation, according to the available distribution of R\* CRS haplotypes and their absence in continental Italy. The ancestral haplotype of X (CRO179) has been sampled from central Italy<sup>39</sup> and Turkey but not in Croatia. The H-CRS haplotype frequency in the Croatian-Italian sample (15%) is also more similar to the Italian average (16%) than to averages of either the Croatian mainland (4%) or the island (9%) populations. The bulk of evidence gained from the phylogeographic approach therefore points again to the neighborhood influence on the Croatian community of Italy.



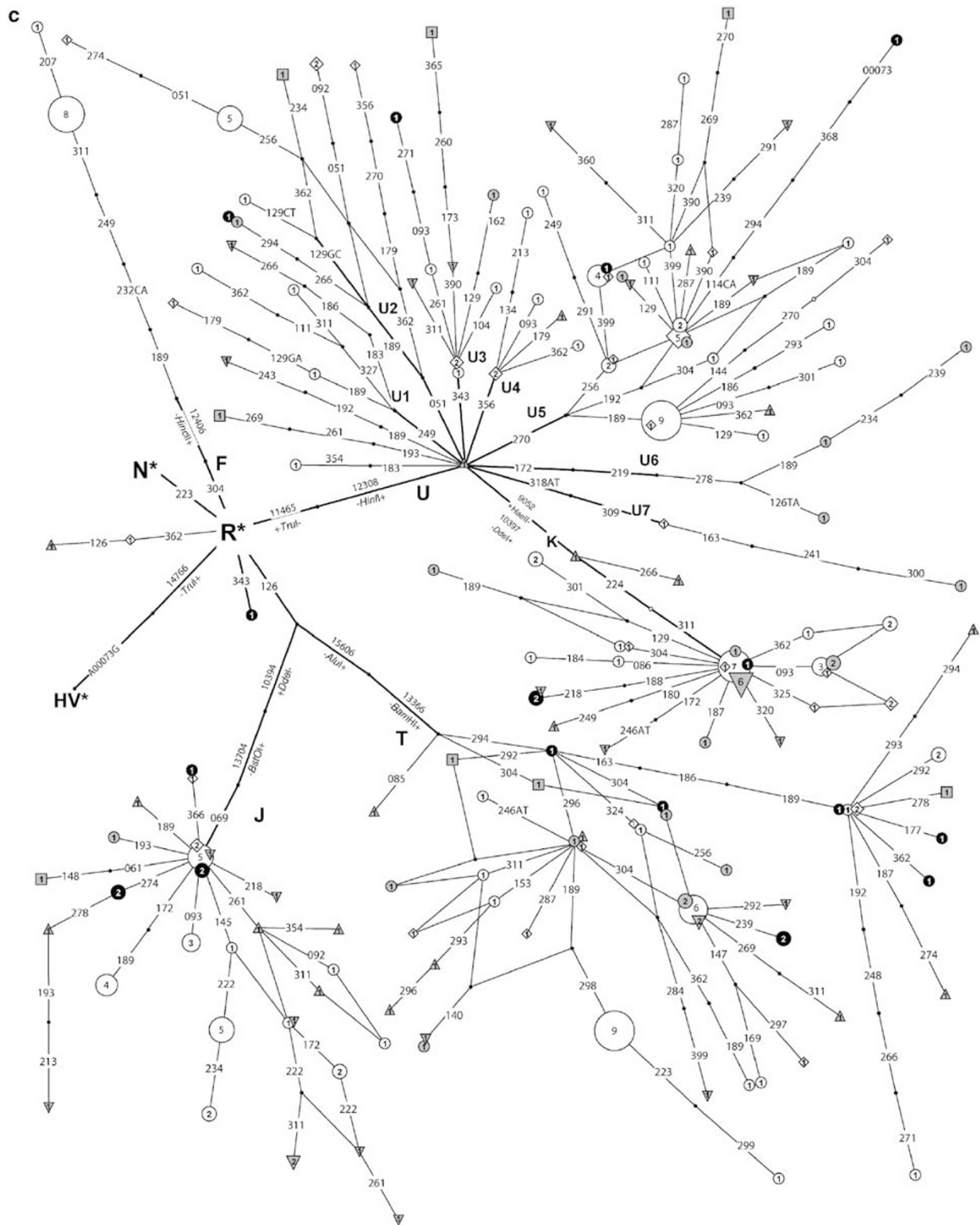


Figure 3 Continued

## Conclusions

The founder analysis by individual haplogroups based on HVS-I resolution level indicated that the fraction of the maternal gene pool of the three communities of Molise derives equally from the Croatian parental population and the Italian receptor population of Molise, highlighting the high penetrance of Italian maternal lineages into their gene pool. This result validated what inferred from the admixture analysis, which although associated with large standard errors converges in the picture obtained by the phylogeographic approach. Moreover, the genetic similarity between the Croatians of Molise and their neighbouring Italian populations, as highlighted by the MDS representation and intermatch analysis, indicates that there was no reproductive isolation between the two geographically proximate, yet culturally distinct populations. Therefore, the presence of appreciable levels of gene flow between the Croatian ethnic minority and their neighbouring Italian populations indicates that differences in language or other cultural traits did not create reproductive barriers in the maternal gene pool. Other examples of genetic admixture of culturally distinct groups have been reported elsewhere in Europe.<sup>40–42</sup> For example, very strong cultural differences exist between Romani and non-Romani populations but these groups are genetically indistinguishable.<sup>43</sup>

Overall, our results support the idea that the Croatian-speaking population living in Molise was not reproductively isolated. Of course, the fact that Croatian-Italians from Molise are not genetically distinct does not undermine the notion of ethnic and cultural distinctiveness.

Studies dealing with other DNA markers would augment the current mtDNA data and better delineate the scenario outlined in this study. For instance, it would be interesting to test whether analysis of Y-chromosome polymorphisms reveals a higher degree of genetic isolation of the Croatian minority of Molise from the neighbouring Italian populations due to asymmetric migration of males. While there are examples of a differential rate of migration between sexes among human populations,<sup>44–46</sup> recent analysis of the global distribution of human mtDNA and Y chromosome variation<sup>47</sup> shows a relatively symmetrical gene flow for females and males. It would be interesting to see if the Croatian-Italian genetic history conforms to this global model or is another interesting exception.

If this reconstruction will be proved to be correct, very likely further studies on Y chromosome polymorphisms would not modify the Croatian Italian genetic history inferred from mtDNA data.

## Acknowledgements

We wish to thank all the subjects, Professor Gabriella Girelli and Sandra Guglielmi for their cooperation in sample collection, Professor Giorgio Bertorelle for his advice on the admixture analysis, and Kenneth Britsch for the English revision of the manuscript. We also very much appreciated the critical comments of the anonymous

referees, which improved the quality of the paper. This work was supported by CNR (Contract No. C00A547-002 AG 2000), by the 'Ministero dell'Università e della Ricerca Scientifica 1 (MURST) through 60 and 40% grants allotted to OR, by the Croatian Ministry of Science, Education and Sports Grant No. 0196005 to PR and 0196002 to AS, and by Estonian Basic Research Grant No. 514 to RV and ETF Grant No. 5574 to TK.

## References

- 1 Smith AT: *The ethnic revival*. Cambridge: Cambridge University Press, 1981.
- 2 Salvi S: *The lingue tagliate*. Milano: Rizzoli, 1975.
- 3 Rešetar M: *Die serbokroatischen Kolonien Südtaliens*. Wien: Kaiserliche Akademie der Wissenschaften - Schriften der Balkankommission-Linguistische Abteilung, 1911.
- 4 Scotti G: *Z one bane mora*. Rijeka: Otakar Kersovani, 1980.
- 5 ISTAT: *Censimento 2001. Primi risultati*. Roma, 2002.
- 6 Sujoldžić A: Continuity and change reflected in synchronic and diachronic linguistic variation of Middle Dalmatia. *Coll Antropol* 1997; **21**: 285–299.
- 7 Tolk H-V, Peričić M, Barać L. *et al*: MtDNA haplogroups in the populations of Croatian Adriatic Islands. *Coll Antropol* 2000; **24**: 267–279.
- 8 Tolk H-V, Barać L, Peričić M *et al*: The evidence of mtDNA haplogroup F in a European population and its ethnohistoric implications. *Eur J Hum Genet* 2001; **9**: 717–723.
- 9 Budowle B, Moretti T, Smith J, Dizinno J: *DNA typing protocols: molecular biology and forensic analysis*. Natick, MA: Books Publication, Eaton Publishing, 2000.
- 10 Biondi G, Rickards O, Fuciarelli M *et al*: Genetic polymorphisms in the Croatian ethno-linguistic minority of Italy. *Gene Geogr* 1990; **4**: 71–79.
- 11 Miller SA, Dykes KK, Plesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.
- 12 Rickards O, Martínez-Labarga C, Favaro M, Frezza D, Mallegni F: DNA analysis of the remains of the Prince Branciforte Barresi family. *Int J Legal Med* 2001; **114**: 141–146.
- 13 Rickards O, Martínez-Labarga C, Lum JK, De Stefano GF, Cann RL: mtDNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the Native Americans populations. *Am J Hum Genet* 1999; **65**: 519–530.
- 14 Anderson S, Bankier AT, Barrel BG *et al*: Sequence and organization of the human mitochondrial genome. *Nature* 1981; **290**: 457–465.
- 15 Andrews RM, Kubaka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999; **23**: 147.
- 16 Richards M, Macaulay V, Hickey E *et al*: Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000; **67**: 1251–1276.
- 17 Herrnstadt C, Elson JL, Fahy E *et al*: Reduced-median network analysis of complete mitochondrial DNA coding-sequences for the major African, Asian, and European haplogroups. *Am J Hum Genet* 2002; **70**: 1152–1171.
- 18 Helgason A, Hickey E, Goodacre S *et al*: mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet* 2001; **68**: 723–737.
- 19 Nei M: *Molecular evolutionary genetics*. New York: Columbia University Press, 1987.
- 20 Tajima F: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989; **123**: 585–595.
- 21 Schneider S, Roessli D, Excoffier L: *A software for population genetic data analysis. ARLEQUIN ver. 2.0*, Genetics and Biometry Laboratory Department of Anthropology and Ecology, University of Geneva, Switzerland, 2000.

- 22 Reynolds J, Weir BS, Cockerham CC: Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 1983; **105**: 767–779.
- 23 Slatkin M: A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 1995; **139**: 457–462.
- 24 Kruskal JB: Nonmetric multidimensional scaling and numerical method. *Psychometrika* 1964; **29**: 115–129.
- 25 Dupanloup I, Bertorelle G: Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol* 2001; **18**: 672–675.
- 26 Bandelt HJ, Forster P, Sykes BC, Richards HB: Mitochondrial portraits of human populations using median networks. *Genetics* 1995; **141**: 743–753.
- 27 Rickards O, Biondi G, De Stefano GF, Vecchi F, Walter H: Genetic structure of the population of Sicily. *Am J Phys Anthropol* 1992; **87**: 395–406.
- 28 Sujoldžić A, Šimunović P, Finka B, Bennett LA, Angel JL, Rudan P: Linguistic microdifferentiation on the island of Korčula, Yugoslavia. *Anthropol Linguistic* 1986; **28**: 405–432.
- 29 Belledi M, Poloni ES, Casalotti R *et al*: Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. *Eur J Hum Genet* 2000; **8**: 480–486.
- 30 Excoffier L, Schneider S: Why hunter-gatherer populations do not show signs of Pleistocene demographic expansions. *Proc Natl Acad Sci USA* 1999; **96**: 597–602.
- 31 Brakez Z, Bosch E, Izaabel H *et al*: Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol* 2001; **28**: 295–307.
- 32 Macaulay V, Richards M, Hickey E *et al*: The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 1999; **64**: 232–249.
- 33 Richards M, Macaulay V: Genetic Data and the Colonization of Europe: Genealogies and Founders; in: Renfrew C, Boyle K (eds): *Archaeogenetics: DNA and the population prehistory of Europe*. McDonald Institute Monographs. Cambridge: McDonald Institute for Archaeological Research, 2000, pp 139–151.
- 34 Maca-Meyer N, Arnay M, Rando JC *et al*: Ancient mtDNA analysis and the origin of the Guanches. *Eur J Hum Genet* 2004; **12**: 155–162.
- 35 Plaza S, Calafell F, Helal A *et al*: Joining the Pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann Hum Genet* 2003; **67**: 312–328.
- 36 Harpending HC, Sherry ST, Rogers AR, Stoneking M: The genetic structure of ancient human populations. *Curr Anthropol* 1993; **34**: 483–496.
- 37 Sajantila A, Lahermo P, Anttinen T *et al*: Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res* 1995; **5**: 42–52.
- 38 Malyarchuk BA, Derenko MV: Mitochondrial DNA variability in Russians and Ukrainians: implication to the origin of the Eastern Slavs. *Ann Hum Genet* 2001; **65**: 63–78.
- 39 Tagliabracci A, Turchi C, Buscemi L, Sassaroli C: Polymorphism of the mitochondrial DNA control region in Italians. *Int J Legal Med* 2001; **114**: 224–228.
- 40 Bertranpetit J, Sala J, Calafell F *et al*: Human mitochondrial DNA variation and the origin of the Basques. *Ann Hum Genet* 1995; **59**: 63–81.
- 41 Larruga JM, Díez F, Pinto FM *et al*: Mitochondrial DNA characterisation of European isolate: the Maragatos from Spain. *Eur J Hum Genet* 2001; **9**: 708–716.
- 42 Malyarchuk BA, Derenko MV: Mitochondrial DNA variability in Russians and Ukrainians: implication to the origin of the Eastern Slavs. *Ann Hum Genet* 2001; **65**: 63–78.
- 43 Gresham D, Morar B, Underhill PA *et al*: Origins and divergence of the Roma (Gypsies). *Am J Hum Genet* 2001; **69**: 1314–1331.
- 44 Seielstad MI, Minch E, Cavalli-Sforza LL: Genetic evidence for a higher female migration rate in humans. *Nat Genet* 1998; **20**: 278–280.
- 45 Stoneking M: Women on the move. *Nat Genet* 1998; **20**: 219–220.
- 46 Jorde LB, Watkins WS, Bamshad MJ *et al*: The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 2000; **66**: 979–988.
- 47 Wilder JA, Kingan SB, Mobasher Z *et al*: Global patterns of human mitochondrial DNA and Y-chromosome structure are not influenced by higher migration rates of females versus males. *Nat Genet* 2004; **36**: 1122–1125.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)