

ARTICLE

# Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe

Vincenza Battaglia<sup>1</sup>, Simona Fornarino<sup>1,12</sup>, Nadia Al-Zahery<sup>1</sup>, Anna Olivieri<sup>1</sup>, Maria Pala<sup>1</sup>, Natalie M Myres<sup>2</sup>, Roy J King<sup>3</sup>, Siiri Rootsi<sup>4</sup>, Damir Marjanovic<sup>5,6</sup>, Dragan Primorac<sup>7,8</sup>, Rifat Hadziselimovic<sup>5</sup>, Stojko Vidovic<sup>9</sup>, Katia Drobnic<sup>10</sup>, Naser Durmishi<sup>11</sup>, Antonio Torroni<sup>1</sup>, A Silvana Santachiara-Benerecetti<sup>1</sup>, Peter A Underhill<sup>3</sup> and Ornella Semino<sup>\*,1</sup>

<sup>1</sup>Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia, Italy; <sup>2</sup>Sorenson Molecular Genealogy Foundation, Salt Lake City, UT, USA; <sup>3</sup>Department of Psychiatry and Behavioral Sciences, School of Medicine, Stanford University, Stanford, CA, USA; <sup>4</sup>Department of Evolutionary Biology, University of Tartu and Estonian Biocentre, Tartu, Estonia; <sup>5</sup>Institute for Genetic Engineering and Biotechnology, University of Sarajevo, Sarajevo, Bosnia and Herzegovina; <sup>6</sup>Genos doo, Zagreb, Croatia; <sup>7</sup>Medical School, Split University, Split, Croatia; <sup>8</sup>Medical School, Osijek University, Osijek, Croatia; <sup>9</sup>Faculty of Medicine, University of Banja Luka, Banja Luka, Bosnia and Herzegovina; <sup>10</sup>Forensic Laboratory and Research Center, Ministry of the Interior, Ljubljana, Slovenia; <sup>11</sup>Medical School, University of Tetovo, Tetovo, Former Yugoslavia Republic of Macedonia

The debate concerning the mechanisms underlying the prehistoric spread of farming to Southeast Europe is framed around the opposing roles of population movement and cultural diffusion. To investigate the possible involvement of local people during the transition of agriculture in the Balkans, we analysed patterns of Y-chromosome diversity in 1206 subjects from 17 population samples, mainly from Southeast Europe. Evidence from three Y-chromosome lineages, I-M423, E-V13 and J-M241, make it possible to distinguish between Holocene Mesolithic forager and subsequent Neolithic range expansions from the eastern Sahara and the Near East, respectively. In particular, whereas the Balkan microsatellite variation associated to J-M241 correlates with the Neolithic period, those related to E-V13 and I-M423 Balkan Y chromosomes are consistent with a late Mesolithic time frame. In addition, the low frequency and variance associated to I-M423 and E-V13 in Anatolia and the Middle East, support an European Mesolithic origin of these two clades. Thus, these Balkan Mesolithic foragers with their own autochthonous genetic signatures, were destined to become the earliest to adopt farming, when it was subsequently introduced by a cadre of migrating farmers from the Near East. These initial local converted farmers became the principal agents spreading this economy using maritime leapfrog colonization strategies in the Adriatic and transmitting the Neolithic cultural package to other adjacent Mesolithic populations. The ensuing range expansions of E-V13 and I-M423 parallel in space and time the diffusion of Neolithic Impressed Ware, thereby supporting a case of cultural diffusion using genetic evidence. *European Journal of Human Genetics* advance online publication, 24 December 2008; doi:10.1038/ejhg.2008.249

**Keywords:** Balkan Neolithic; farming transition; peopling of Europe; Y-chromosome haplogroups

\*Correspondence: Dr O Semino, Dipartimento di Genetica e Microbiologia, Università di Pavia, Pavia 27100, Italy.

Tel: + 39 0382 985543; Fax: + 39 0382 528496;

E-mail: semino@ipvgen.unipv.it

<sup>12</sup>Present address: Human Evolutionary Genetics, CNRS URA 3012, Institut Pasteur, Paris, France.

Received 2 September 2008; revised 17 November 2008; accepted 20 November 2008

## Introduction

Europe was colonized by modern humans about 40 000 years ago and underwent a second colonization wave during the Neolithic, with the spread of farming.<sup>1,2</sup> The relative Palaeolithic and Neolithic contribution to the current European gene pool has been widely debated and is still under discussion.<sup>3–6</sup> Two opposing models have been

cited to account for the spread of farming in Europe: the demic diffusion model, which implies a movement of people and therefore a significant Near Eastern genetic input,<sup>7</sup> and the cultural diffusion model, which, on the contrary, considers the transition to agriculture as a cultural phenomenon, and therefore without major changes at the genetic level.<sup>8</sup> Archaeological evidences suggest, however, that the spread of agriculture has been a complex process characterized by migrations and local admixture.<sup>9,10</sup>

Genetic studies have described allele frequency clines for different markers along the European colonization routes.<sup>11–13</sup> These have been interpreted in favour of demic diffusion and considered a strong indication of a Neolithic contribution to the modern European gene pool.<sup>14,15</sup> However, a number of simulation studies have demonstrated that allele frequency clines can also occur in range expansions where admixture is not present.<sup>16–19</sup> The impact of the Neolithic in the Balkans, as archaeological records show, is regarded to be considerable,<sup>20</sup> although Mesolithic hunter-gatherers were present in this region just before the first appearance of Neolithic pottery.<sup>21</sup> Thus, the question to address is the following: can we detect present-day signals of interactions between indigenous Mesolithic people and agricultural colonists in the southeast European gene pool? In this regard, particularly useful are Y-chromosome markers, whose distribution is often highly correlated with their phylogeny.<sup>22</sup>

Interestingly, the Y-chromosome gene pool of southeast European populations is characterized both by 'autochthonous' markers, such as haplogroup (Hg) I, present in the Balkans in pre-Neolithic times,<sup>23,24</sup> and by markers mainly belonging to Hgs E and J, which originated outside Europe, in Africa and the Middle East, respectively.<sup>12,13,25–28</sup> In addition, new Y-chromosome polymorphisms have added further sophistication to phylogenetic relationships,

especially within Hgs E, J and I,<sup>29–32</sup> providing the opportunity to more fully evaluate the above issue. This possibility induced us to carry on a deeper genetic structure characterization of the Balkan area through the analysis of 80 Y-chromosome bi-allelic markers and 12 linked STR loci in 1206 subjects from 17 population samples mainly from southeast Europe.

## Materials and methods

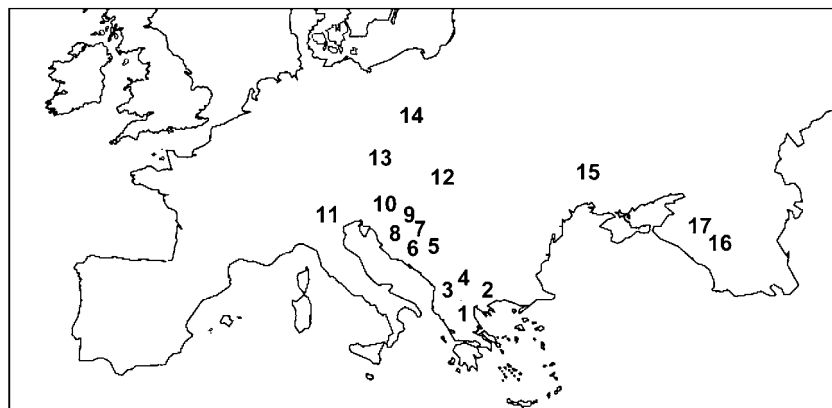
### Samples

The sample consists of 1206 unrelated male individuals from 17 population samples (Figure 1). Two-hundred and thirty-five of these, namely 64 Albanians from Former Yugoslavia Republic of Macedonia, 29 Croats from Osijek, 75 Slovenians and 67 northeast Italians (from the province of Trento), are reported here for the first time. The remaining include samples reported earlier,<sup>23,27,33</sup> and consist of 104 Caucasians (38 Balkarians and 66 Georgians), 149 Greeks (92 from Athens and 57 from Macedonia), 55 Albanians (collected in Tirana), 89 Croats, 99 Polish, 75 Czechs, 92 Ukrainians, 53 Hungarians and 255 Bosnia-Herzegovinians (84 Bosniacs, 90 Croats and 81 Serbs). Blood samples were collected from healthy unrelated adults after obtaining informed consent. DNA was extracted from whole blood according to the standard phenol/chloroform procedure, followed by ethanol precipitation.

In addition, P37.2\* samples identified from a screening of the Sorenson Molecular Genealogy Foundation collection (over 14 000 Y chromosomes from more than 100 countries) were also included.

### Examined bi-allelic markers and microsatellites

Eighty Y-specific unique event polymorphisms (Figure 2) were examined in hierarchical order. Two new mutations



**Figure 1** Geographic location of the studied samples: 1, Greeks; 2, Macedonian Greeks; 3, Albanians; 4, Albanians from Former Yugoslavia Republic of Macedonia (FYROM); 5, Bosniacs; 6, Bosnia-Croats; 7, Bosnia-Serbs; 8, Croats; 9, Croats from Osijek; 10, Slovenians; 11, northeast Italians; 12, Hungarians; 13, Czechs; 14, Poles; 15, Ukrainians; 16, Georgians; and 17, Balkarians.

were discovered: M507 (AC006376.2: g.148697 T→G) and M521 (AC012068.5: g.5219 C→T). The first is associated with I-M253 and was discovered while typing M227, the second is associated with E-M78\* and was ascertained while typing V12.

Bi-allelic markers were analysed by PCR/AFLPs (YAP<sup>34</sup> and 12f2<sup>13</sup>), by PCR/RFLPs (P37,<sup>35</sup> M9, M269,<sup>36</sup> V12, V13, V22, V27, V32 and V65<sup>28,30</sup>) or by PCR/DHPLC (M241, M253, M267, M285, M287, M304, M343, M365, M367–M369,<sup>29</sup> M410,<sup>37</sup> M423, M429, M438<sup>31</sup> and M406,<sup>38</sup> and those reported in the YCC<sup>39</sup>).

Nomenclature used for Hg labelling is in agreement with YCC<sup>39</sup> conventions and recent updating.<sup>31,32,38</sup> Subsets of samples belonging to E-V13, J-M241 and I-M423 Hgs were also analysed at 12 STR loci: DYS19, YCAIIa/b, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS439 and DYS460 by using multiplex reactions according to STRBase information (<http://www.cstl.nist.gov/biotech/strbase/y20prim.htm>) and ABI PRISM<sup>®</sup> 3100 DNA Sequencer, internal size standard and GeneScan fragment analysis software. DYS445<sup>38</sup> was analysed separately in J-M410 samples.

### Statistical analysis

Hg diversity (H) was computed using the standard method of Nei.<sup>40</sup> Principal components (PC) analysis was performed on Hg frequencies using Excel software implemented by Xlstat. The relative amount of accumulated diversity, as a function of geography, was evaluated through the mean microsatellite variance estimated for each population with a sample size of at least five individuals. Hg frequency and variance maps were generated by using Surfer Golden software following the Kriging procedure.<sup>41</sup> Within specific Hgs, Median-Joining (MJ) networks<sup>42</sup> were constructed using Network 4.5.0.0 program (Fluxus Engineering, <http://www.fluxus-technology.com>). Networks were calculated by the MJ method, where  $\epsilon = 0$  and microsatellite loci were weighted proportionally to the inverse of the repeat variance observed in each Hg and after having processed the data with the reduced-median method. The age of microsatellite variation within Hgs was evaluated employing the methodology described by Zhivotovsky *et al*<sup>43</sup> as modified according to Sengupta *et al*.<sup>37</sup> A microsatellite evolutionary effective mutation rate of  $6.9 \times 10^{-4}$  was chosen as it is suitable for use in situations where the elapsed time frame is  $\geq 1000$  years or  $\sim 40$  generations, clearly appropriate given the prehistoric time depths being explored in this study.<sup>44</sup>

It is worth mentioning that ambiguities related to past episodes of population history (eg, size fluctuations, bottlenecks, etc.) create inherent uncertainties in the calibration of the YSTR molecular clock; thus the estimated ages of microsatellite variation should be considered with caution.

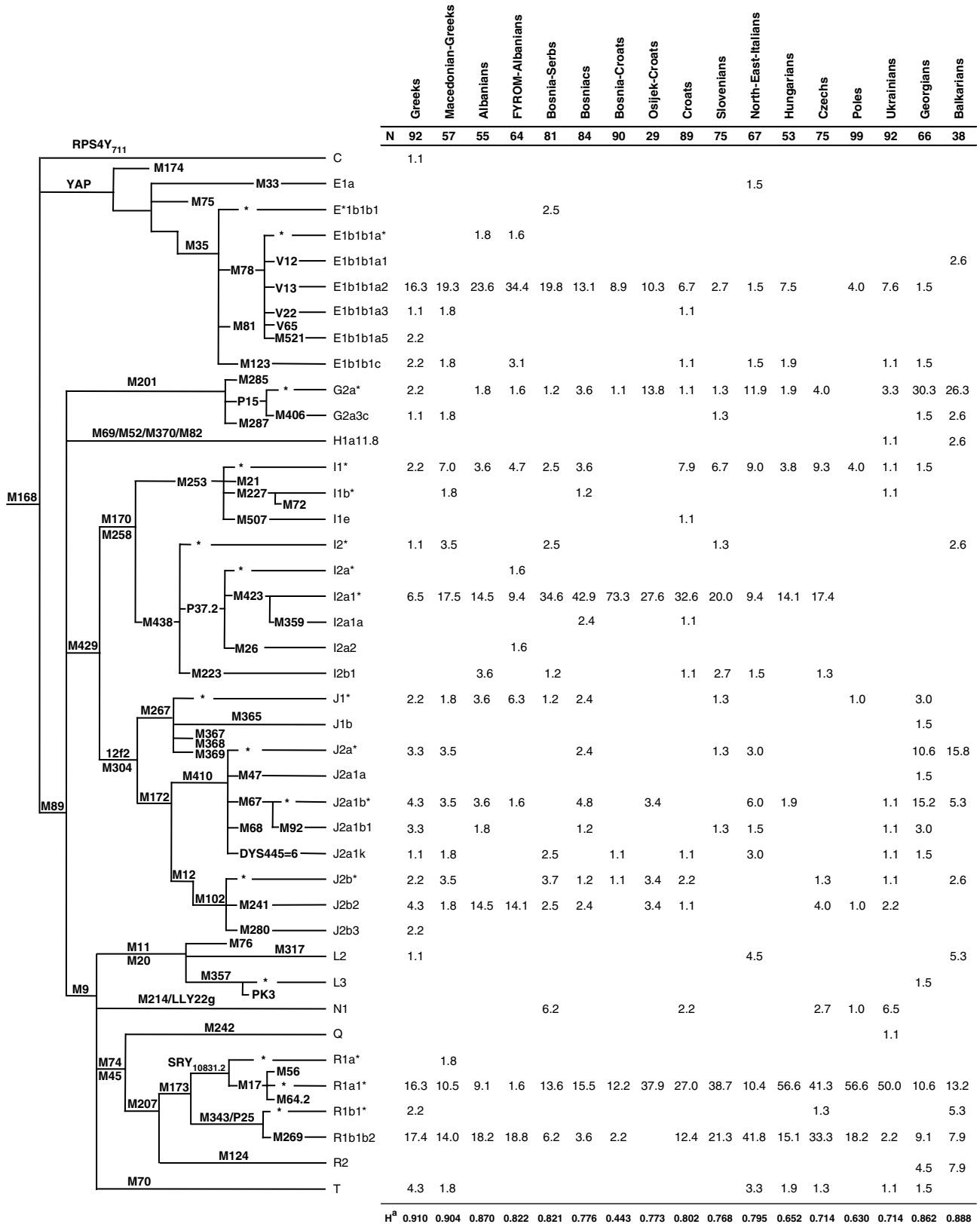
### Results

Figure 2 illustrates the phylogenetic relationships of Y-chromosome Hgs and their distribution in the examined southeast European populations. The main Hgs observed in Europe<sup>45</sup> (E, I, J, R1a and R1b) contribute differently to the gene pool of the various East European areas, Hg I and Hg R being the most represented always and Hg E and Hg J being mainly frequent in the southern Balkan populations.

Hg I is restricted to western Eurasia<sup>23</sup> and is particularly frequent in the Balkans where it characterizes 36.3% of the total Y chromosomes. Two of its branches, I-M223 and I-M253, are scarcely represented in southeast Europe, the first being only sporadically observed and the second showing frequencies around 5%, with higher values (around 9%) in Macedonian Greeks and Croats. Differently, the recently described M423 SNP,<sup>31</sup> which characterizes the previously paraphyletic P37 clade, accounts for the majority (77.2%) of the East European Hg I chromosomes. Its diffusion seems not to have affected the neighbouring North Italian populations, where low incidences (0–2%) are observed. The I-M423 sub-clade is characterized by a frequency distribution with high Central Balkan values ( $> 70\%$  in Bosnia-Herzegovina) and decreasing frequencies moving from the southern Dinaric Alps to northern Croatia. Although I-M423 comprises virtually all the I-P37.2 Balkan-related chromosomes reported earlier,<sup>31</sup> we have also detected one I-P37.2\* Albanian subject and, from a screening of previously identified P37.2 chromosomes (Rootsi *et al*<sup>23</sup> and SMGF collection), 30 further P37.2\* subjects, two from Moldavia<sup>23</sup> and 28 of either documented or presumed western European ancestry, were identified.

Hg R1 is common throughout western Eurasia<sup>22</sup> and accounts for more than 30% of the Balkan Y-chromosome pool. With the exception of one R1a\*-SRY<sub>10831.2</sub> and five R1b1-M343\* individuals, all the remaining R1 lineages belong to R1a1-M17\* and R1b1b2-M269. These two sub-clades, which show in Europe opposite-frequency gradients with maximum incidences in eastern and western regions, respectively, still display high values in northern Balkans and sensibly decrease southward. R-M269 chromosomes are common in the Balkans and Anatolia and, according to the observed internal divergence of their 49a,f branches,<sup>29</sup> most likely predated the origin of agriculture. However, the current lack of informative Hg sub-division within these populations (Figure 2) does not allow, at this time, to evaluate the role of R-M269 chromosomes during the transition to agriculture.

E-M35 is the only branch of Hg E observed in this survey. It is prevalently represented by E-M78 chromosomes, almost completely ( $> 90\%$ ) belonging to the recently described<sup>30</sup> E-V13 sub-clade. Only four E-M78\*, which do not belong to any already described sub-clade, have been observed in the southern Balkans. Two of them (from

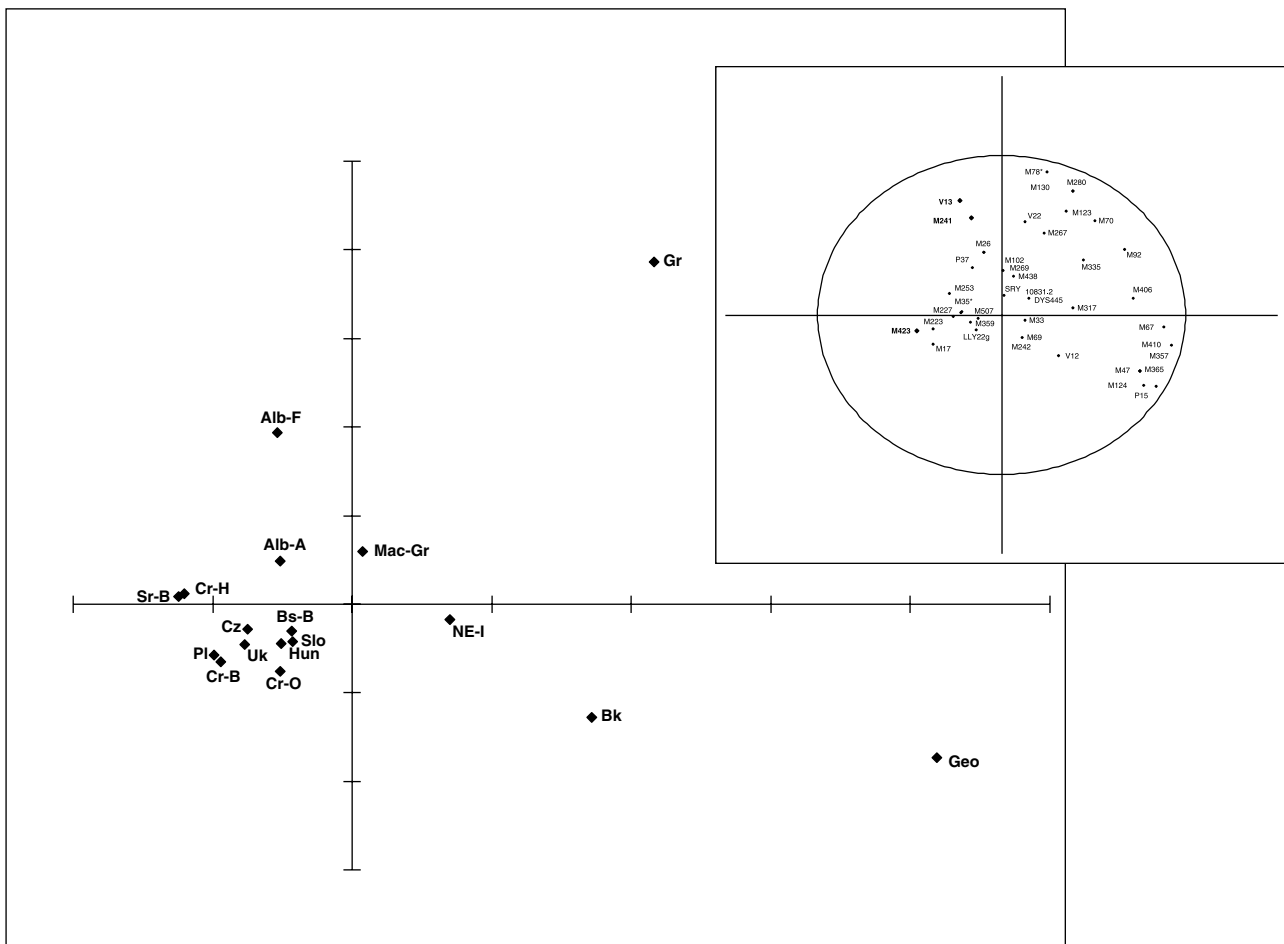


Greece) turned out to be characterized by the mutation M521 and therefore represent a new M78 lineage.

The majority of the Balkan Hg J Y chromosomes belong to the J-M172 sub-Hg and range from 2% to 20%. Both its main branches, J-M410 and J-M12/M102\*, were observed; although the first is scattered in different sub-clades (J-M67, J-M92 and J-DYS445-6) with distinct local patterns, the second is most represented by J-M241.

The PC analysis, from the perspective of population Hg frequencies (Figure 3), reveals a tight cluster of populations not comprising southern Balkan and Caucasian

groups. Common to this cluster are lower frequencies of Hgs, G-M201 and J-M410, and higher frequencies of Hgs, I-M423, E-V13 and J-M241. Whereas the first two are primarily Middle Eastern Hgs and have been shown to be associated with the early Neolithic colonization of Crete,<sup>38,46</sup> Italy,<sup>47,48</sup> and southern Caucasus, I-M423, E-V13 and J-M241, in spite of parallel Balkan patterns of distribution, have clearly different origins.<sup>30,31,38</sup> Their comparison can therefore provide insights into the complex interaction between the European Mesolithic foragers and the Middle Eastern Neolithic farmers



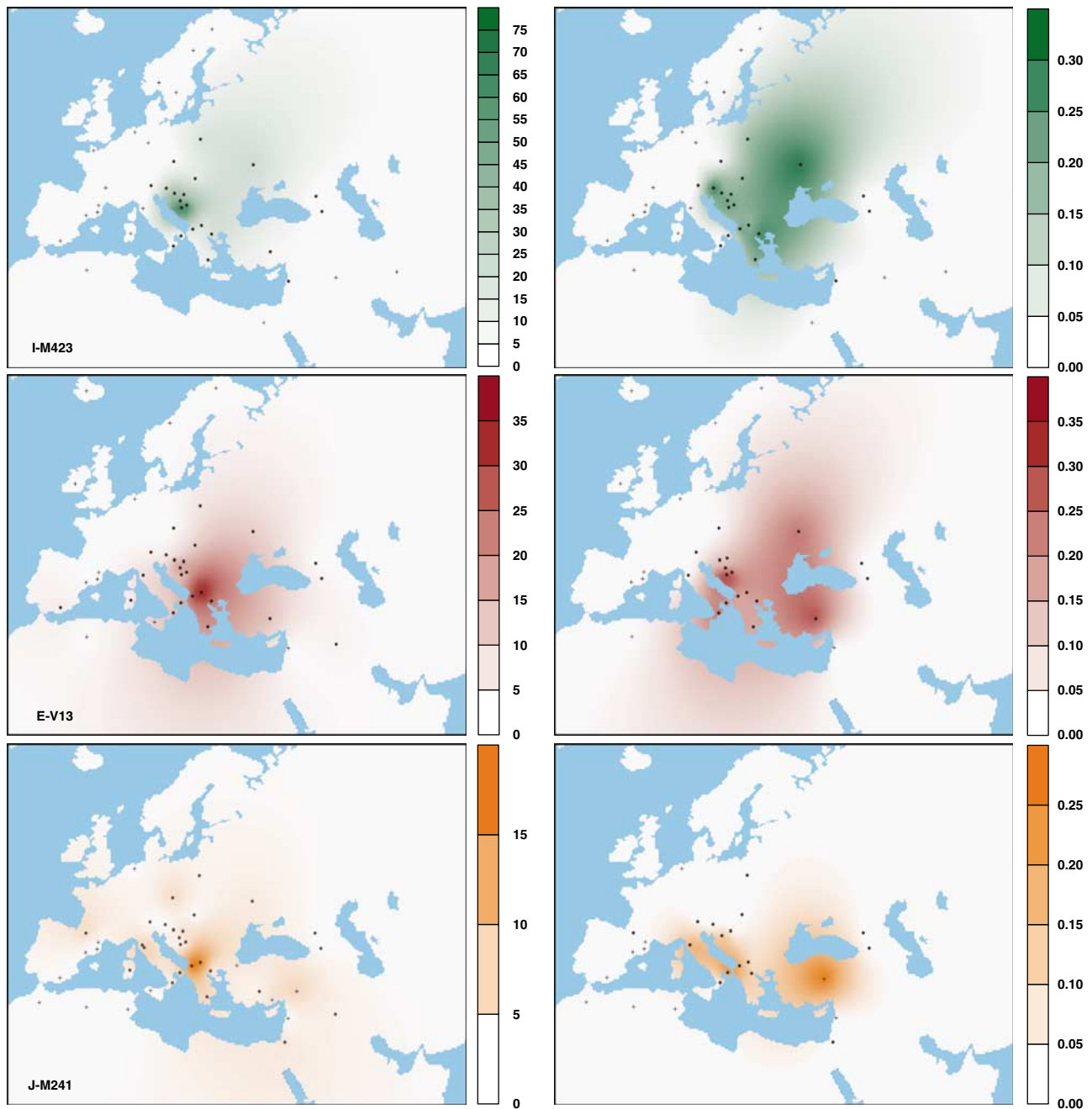
**Figure 3** PC analysis performed using haplogroup frequencies in the populations of this study. Gr, Greeks; Mac-Gr, Macedonian Greeks; Alb-A, Albanians; Alb-F, Albanians from FYROM; Sr-B, Bosnia-Serbs; Bs-B, Bosniacs; Cr-B, Bosnia-Croats; Cr-O, Croats from Osijek; Cr-H, Croats of Croatia; Slo, Slovenians; NE-I, northeast Italians; Hun, Hungarians; Cz, Czechs; Pl, Poles; Uk, Ukrainians; Geo, Georgians; Bk, Balkarians. Thirty-four percent of the total variance is represented. Insert illustrates the contribution of each haplogroup.

**Figure 2** Phylogeny of Y-chromosome haplogroups and their frequencies (%) in the examined populations. Nomenclature and haplogroup labelling according to the Y Chromosome Consortium (<http://ycc.biosci.arizona.edu/>) updated according to Karafet *et al.*<sup>32</sup> \*Paragroups: Y chromosomes not defined by any phylogenetic downstream-reported and -examined mutation. <sup>a</sup>Intrapopulation haplogroup diversity. The terminal markers of haplogroups E-V12 and E-V13 (V32 and V27, respectively) were typed but did not show any variation.

during the transition to farming society in the Balkans. These Hgs, although characterized by different distribution patterns of frequency and variance (Figure 4), display networks of microsatellite haplotypic variation (Supplementary Figures S1–S3), all consistent with a Balkan expansion.

## Discussion

Various episodes of population movement have affected southeast Europe, and the role of the Balkans as a long-standing gateway to Europe from the Near East is illustrated by the phylogenetic unification of Hgs I and J by the basal M429 mutation.<sup>31</sup> This evidence of common



**Figure 4** Frequency (left) and variance (right) distributions of the main Y-chromosome haplogroups, I-M423, E-V13 and J-M241, observed in this survey. Frequency data are reported in Figure 2, variance data are relative to the examined microsatellite reported in the Supplementary Table S2. We acknowledge that interpolated spatial frequency surfaces should be viewed with caution because of sample size.<sup>41</sup> ● Data from this study. Frequency and variance values were assigned to sample-collection places (dots). Population samples (geographically close) with less than five observations were pooled and the corresponding variance assigned to a middle position of the pooled sample locations. + Data from the literature.<sup>13,23,27,28,36,45,49–54</sup>

ancestry suggests that ancestral IJ-M429\* Y chromosomes probably entered Europe through the Balkan route sometime before the Last Glacial Maximum. They subsequently evolved into Hg J in the Middle East and Hg I in Europe in a typical disjunctive phylogeographic pattern. Such a geographic corridor is likely to have experienced additional subsequent gene flows, including the migration of agricultural colonists from the Middle East. Pottery is a useful proxy for the spread of farming both spatially and temporally. The first appearance of pottery in the Adriatic region was in Corfu at 6500 BC and reached the northern most Adriatic ~1000 years later.<sup>21</sup> Its dispersal provides a comparative template for spatial and temporal patterns of Y chromosome Hg diversity observed in this area.

Hg J is most common (~50%) in the Middle East and Anatolia,<sup>27,29,47</sup> with a spread zone spanning from northwest Africa to India.<sup>12,55</sup> It has been related to different Middle Eastern migrations.<sup>12,56</sup> In addition to Hg J-M410, Hg G-P15 chromosomes, which are also common in Anatolia,<sup>29</sup> have been implicated in the colonization and subsequent expansion of early farmers in Crete, the Aegean and Italy.<sup>38,46–48</sup> Earlier studies have concluded that the J-M410 sub-clades, J-DYS445-6 and J-M67, are linked to the spread of farming in the Mediterranean Basin,<sup>38,47</sup> with a likely origin in Anatolia.<sup>29</sup> Interestingly, J-DYS445-6 and J-M92 (a sub-lineage of M67), both have expansion times between 7000 and 8000 years ago (Table 1), consistent with the dating of the arrival of the first farmers to the Balkans. The first detection of milk residue in ceramic pottery occurs in sites from northwest Anatolia 7000–8500 years ago,<sup>58</sup> an age that approximates the Hg-expansion times.

Regarding Hg J-M12/M102, which is discernable from India to Europe, the M12/M102\* chromosomes display a very high YSTR diversity, whereas on the other hand, the J-M241 sub-lineage has low diversity (Table 1) in the Balkans, indicating different demographic histories. Although Hg J-M241 shows high variance in India,<sup>37</sup> its place of origin is still uncertain. As J-M241 has older expansion times in Sicily, Apulia and Turkey (Table 2), it may have arrived in the Balkans from elsewhere.

On the other hand, the expansion times of Hg V13 (Table 3) are consistent with a late Mesolithic time frame. The Greek Mesolithic, although different in its material culture from the Natufian Mesolithic of the Levant, bears some resemblance to the Mesolithic of southern Anatolia.<sup>60</sup> This archaeological congruence between the Mesolithic of the Balkans and southern Anatolia may mirror the similar E-V13 expansion times observed for Konya, Franchthi Cave and Macedonian Greece, all approximately 9000 years ago. Moreover, E-V13 YSTR-related data from Bulgaria and Macedonia,<sup>28</sup> both with a variances of 0.28, suggest an expansion time of approximately 10000 years ago. It is likely that the origin of V13 occurred somewhere within the zone of these sample collections. In addition, it is also worth noting that in the Anatolian region

**Table 1** Ages of microsatellite variation and mean variance of microsatellite loci within haplogroups

Haplogroup (N)	Age of microsatellite variation <sup>a</sup> ± SE <sup>b</sup> (ky)	Mean variance <sup>a</sup>
I-P37*-DYS388-15 (28)	4.0 ± 1.7	0.107
I-M423 (219)	7.8 ± 2.3	0.198
E-V13 (92)	5.9 ± 1.7	0.155
J-M410* (23)	15.8 ± 3.8	0.410
J-M67 (42)	11.0 ± 2.4	0.287
J-M67* (32)	11.2 ± 2.9	0.298
J-M92 (10)	7.0 ± 1.9	0.187
J-DYS445-6 (10)	7.7 ± 2.6	0.164
J-M12 (51)	7.4 ± 1.4	0.191
J-M12/102* (15)	12.3 ± 2.7	0.332
J-M241 (34)	4.8 ± 1.3	0.121

<sup>a</sup>Values computed over the following microsatellite loci: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439 and DYS460 (values of YCAa,b were excluded from the calculation because of the impossibility of assigning each allele to one or the other locus of this ambiguous system). Ages of microsatellite variation were estimated as reported in Sengupta *et al.*<sup>37</sup> Owing to the small sample sizes, the estimates of the expansion time should be considered indicative.

<sup>b</sup>SE was computed over loci.<sup>43,57</sup>

**Table 2** Ages of microsatellite variation and mean variance of microsatellite loci within J-M241 haplogroup in Turkey, the Balkans and Italy

Region (N)	Age of microsatellite variation <sup>a</sup> ± SE <sup>b</sup> (ky)
Turkey <sup>c</sup> (5)	10.1 ± 3.4
Albania <sup>d</sup> (8)	5.4 ± 2.5
FYROM <sup>d</sup> (9)	2.4 ± 1.2
Greece <sup>d,e</sup> (5)	2.9 ± 1.2
Central Italy <sup>d</sup> (5)	5.8 ± 1.4
Apulia <sup>d</sup> (5)	9.6 ± 3.4
Sicily <sup>f</sup> (11)	10.0 ± 5.4

<sup>a</sup>Values computed over the following microsatellite loci: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439 and DYS460 (values of YCAa,b were excluded from the calculation because of the impossibility of assigning each allele to one or the other locus of this ambiguous system). Ages of microsatellite variation were estimated as reported in Sengupta *et al.*<sup>37</sup> Owing to the small sample sizes, the estimates of the expansion time should be considered indicative.

<sup>b</sup>SE was computed over loci.<sup>43,57</sup>

<sup>c</sup>Cinnioglu *et al.*<sup>29</sup> (region 5 (*n* = 2), region 6 (*n* = 1), region 9 (*n* = 2)); values computed with DYS461 in place of DYS460.

<sup>d</sup>Present paper.

<sup>e</sup>Greeks (*n* = 4), Macedonian Greek (*n* = 1).

<sup>f</sup>Di Gaetano *et al.*<sup>59</sup>

of supposed Einkorn wheat origin<sup>61</sup> (region 5 of Cinnioglu *et al.*<sup>29</sup>), only one V13 chromosome out of 43 is found (PA Underhill, unpublished data). Therefore, as no evidence at present supports the association of E-V13 Hg with the attested origin of farming in southeast Anatolia, the possibility of farming adoption by Balkan E-V13-associated people is plausible. The low E-V13 frequency

**Table 3** Ages of microsatellite variation and mean variance of microsatellite loci within E-V13 haplogroup in Turkey and Greece

Region (N)	Age of microsatellite variation <sup>a</sup> ± SE <sup>b</sup> (ky)
Konya-Turkey <sup>c</sup> (10)	9.4 ± 2.9
Macedonian Greece <sup>c</sup> (8)	8.5 ± 4.5
Greece <sup>c</sup> (8)	4.8 ± 1.6
Nea Nikomedeia (North Greece) <sup>d</sup> (6)	8.6 ± 4.0
Seklo and Dimini (Central Greece) <sup>d</sup> (20)	4.3 ± 1.8
Lerna and Franchthi Cave (South Greece) <sup>d</sup> (20)	9.2 ± 4.3

<sup>a</sup>Values computed over the following microsatellite loci: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439 and DYS460 (values of YCAa,b were excluded from the calculation because of the impossibility of assigning each allele to one or the other locus of this ambiguous system). Ages of microsatellite variation were estimated as reported in Sengupta *et al.*<sup>37</sup> Owing to the small sample sizes, the estimates of the expansion time should be considered indicative.

<sup>b</sup>SE was computed over loci.<sup>43,57</sup>

<sup>c</sup>Present paper.

<sup>d</sup>King *et al.*<sup>38</sup>

and STR variation observed in Crete<sup>38</sup> indicate that if the first Neolithic colonists came from central Anatolia, they did not bring this Hg. The two more recent expansion times for V13 for Greece and Sesklo and Dimini (Table 3), dating to the Bronze Age, possibly reflect a more recent integration of some V13 chromosomes into the populations of the first farmers represented by J-M410 and G-M201 lineages. Both the lack of any plausible Middle Eastern source of E-V13 during either the early Neolithic or Bronze Age and the age of microsatellite variation observed are consistent with E-V13 chromosomes reflecting a Mesolithic heritage as suggested by King *et al.*<sup>38</sup>

As reported earlier,<sup>28</sup> both J-M12 and E-V13 radiation patterns overlap geographically in the Balkans (Figure 4). Although J-M12 chromosomes were not genotyped for M241 by Cruciani *et al.*,<sup>28</sup> the low YSTR diversity observed suggests that these are predominantly M241 derivatives. The difference between E-V13 and J-M241 (Table 1) indicates that both E-V13 frequency and haplotype diversity would have been greater than J-M241 components just before the episode of population growth. This also is the case when the dating is carried out by disregarding the mutational steps connecting the three haplotypes that, including Turkish samples (Supplementary Figure S2), can be considered as founders.<sup>62</sup> Whether or not E-V13 and J-M241 participated in the same demography remains uncertain.

The presence of E-M78\* Y chromosomes in the Balkans (two Albanians), previously described virtually only in northeast Africa, upper Nile,<sup>28,63</sup> gives rise to the question of what the original source of the E-M78 may have been.

Correlations between human-occupation sites and radiocarbon-dated climatic fluctuations in the eastern Sahara and Nile Valley during the Holocene<sup>64</sup> provide a framework for interpreting the main southeast European centric distribution of E-V13. A recent archaeological study reveals that during a desiccation period in North Africa, while the eastern Sahara was depopulated, a refugium existed on the border of present-day Sudan and Egypt, near Lake Nubia, until the onset of a humid phase around 8500 BC (radiocarbon-calibrated date). The rapid arrival of wet conditions during this Early Holocene period provided an impetus for population movement into habitat that was quickly settled afterwards.<sup>64</sup> Hg E-M78\* representatives, although rare overall, still occur in Egypt, which is a hub for the distribution of the various geographically localized M78-related sub-clades.<sup>28</sup> The northward-moving rainfall belts during this period could have also spurred a rapid migration of Mesolithic foragers northwards in Africa, the Levant and ultimately onwards to Asia Minor and Europe, where they each eventually differentiated into their regionally distinctive branches.

Differently from the earlier discussed Hgs, I-M423 represents the southeast European autochthonous clade of I-P37.2. Its distribution reaches Anatolia, where, however, it is only sporadically observed (2.6%, updated from Rootsi *et al.*<sup>23</sup>). Also, virtually, all the I-P37.2\* paragroup members identified in this survey harbouring the peculiar DYS388-15 trinucleotide repeat motif (not observed in any other Hg I clade) likely represent a new rare P37.2 sub-clade. Their distribution (Supplementary Table S1) and the associated YSTR variation age of ~4000 years (Table 1) suggest that they expanded demographically, perhaps from central European regions during the Bronze Age. In this scenario, the only I-P37.2\* chromosome observed in Albania, not characterized by the unusual DYS388-15 repeat motif marker, could either represent the consequence of a reversion event back to the ancestral allele or be a rare representative of the ancestral P37.2 state.

The network of the STR haplotypes identified in 222 Y chromosomes belonging to the I-M423 Hg (Supplementary Figure S1) is characterized by a star-like shape centred on the most frequent and diffused haplotype that is present in all Balkan populations. The marginal positions occupied by the three Turkish chromosomes are in agreement with a recent gene flow. The age of accumulated microsatellite variation associated with Hg I-M423 (Table 1) dates to around 8000 years ago (Early Holocene). Thus, although Hgs G and J mark the successful colonization and subsequent demic expansions of Neolithic pioneers to these regions, consistent with a wave of advance,<sup>19,65</sup> the widespread adoption of farming by Mesolithic hunter-gatherers in the Balkans and Central Europe is recorded in the autochthonous Hg I-M423.

These data indicate the complex interactions between farmers and foragers rather than the large-scale



replacement of hunter-gatherers by pioneering agriculturalists during the spread from the Neolithic to the southeast Europe. The data also indicate that I-M423 and probably also E-V13 representatives would have been well established in the Balkans before the arrival of a nucleus of pioneering agriculturalists.

Thus, unlike Crete, southern and central Italy and the southern Caucasus, the cultural transmission of the Neolithic package played an important role. Either the initial G and J2 Hg agriculturalists who colonized the Balkans at first flourished but later diminished in a similar manner to that proposed regarding the Linearbandkeramik in central Europe<sup>66</sup> or the package was rapidly and robustly adopted by local Mesolithic people in the southern Balkans (plausibly characterized by E-V13), who underwent a demic expansion and a subsequent range expansion to the eastern Adriatic. These former foragers who had recently acquired the Neolithic tradition participated in 'leapfrog' colonizations up the Adriatic, where they eventually transmitted agricultural practices to resident Mesolithic populations represented by I-M423 chromosomes.

Interestingly, the derived Y-chromosome scenario strongly recalls the fourth PC synthetic map of Europe calculated on gene frequencies at 95 nuclear loci,<sup>11</sup> which displays a centre in the southern Balkans and a large surrounding area that terminates with a 'propagule' to the northeast of the central Balkans. On the basis of this observation, our assumptions could provide a possible interpretation of the described expansion centred in the southern Balkans.

## Conclusion

This study provides a model that elevates the role of migratory foragers with remote eastern Saharan ancestry who, once established in Asia Minor with their own derived genetic signature, were destined to become the earliest converts to farming and the adherents of its further spread into Europe. Such an interpretation finds support in the 'dispersal model' of Impressed Ware in which the 'Neolithic package' was acquired by native groups and subsequently diffused by interactions between farmers and foragers.<sup>67</sup> Although southeast Europe shows considerable archaeological evidence of the Neolithic transition,<sup>20</sup> our Y-chromosome results provide biological evidence of complexity<sup>21</sup> in the transition to farming in terms of the contrasting influences of pioneering agriculturalists and Mesolithic foragers.

## Acknowledgements

We are grateful to all the donors for providing the blood samples and to all the people and institutions that contributed to their collection. These include the Institute for Transfusion and the Institute for Clinical Biochemistry of Sarajevo, the Cantonal Hospital and the Primary Care Institution of Mostar, the Clinical Hospital Centre and

the Institute for Transfusion of Banja Luka, the Regional Hospitals of Doboij and Bijeljina, from Bosnia-Herzegovina, and Professor Simun Andjelinovic and his team from the University Hospital of Split, from Croatia. We are also grateful to Franco Lo Curto for having provided us with the Italian samples. We thank the three anonymous reviewers for useful comments and suggestions. This research was supported by the Italian Ministry of the Foreign Affairs (to OS), Compagnia di San Paolo (to AT and OS), the Italian Ministry of the University: Progetti Ricerca Interesse Nazionale 2007 (to AT and OS) and Estonian Science Foundation Grant no. 7445 (to SR).

## References

- Otte M: The history of European populations as seen by archaeology; in Renfrew C, Boyle K (eds): *Archaeogenetics: DNA and the population prehistory of Europe*. Cambridge, UK: McDonald Institute Monographs, 2000, pp 41–44.
- Pinhasi R, Semal P: The position of the Nazlet Khater specimen among prehistoric and modern African and Levantine populations. *J Hum Evol* 2000; **39**: 269–288.
- Barbujani G, Dupanloup I: DNA variation in Europe: estimating the demographic impact of Neolithic dispersals; in Bellwood P, Renfrew C (eds): *Examining the Farming/Language Dispersal Hypothesis*. Cambridge, UK: McDonald Institute Monographs, 2002, pp 421–431.
- Chikhi L: Admixture and the demic diffusion model in Europe; in Bellwood P, Renfrew C (eds): *Examining the Farming/Language Dispersal Hypothesis*. Cambridge, UK: McDonald Institute Monographs, 2002, pp 435–447.
- Richards M: The Neolithic invasion of Europe. *Annu Rev Anthropol* 2003; **32**: 135–162.
- Dupanloup I, Bertorelle G, Chikhi L, Barbujani G: Estimating the impact of prehistoric admixture on the genome of Europeans. *Mol Biol Evol* 2004; **21**: 1361–1372.
- Ammerman AJ, Cavalli-Sforza LL: *The Neolithic Transition and the Genetics of Population in Europe*. Princeton, NJ: Princeton University Press, 1984.
- Whittle A: *Europe in the Neolithic. The creation of New Worlds*. Cambridge, UK: Cambridge University Press, 1996.
- Colledge S, Conolly J, Shennan S: Archaeobotanical evidence for the spread of farming in the Eastern Mediterranean. *Curr Anthropol* 2004; **45**: S35–S58.
- Zvelebil M: Who were we 6,000 years ago? In search of prehistoric identities; in M Jones M (ed): *Traces of Ancestry: Studies in Honour of Colin Renfrew*. Cambridge, UK: McDonald Institute Monographs, 2004, pp 41–60.
- Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press, 1994.
- Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS: A view of the Neolithic demic diffusion in Europe through two Y-chromosome specific markers. *Am J Hum Genet* 1996; **59**: 964–968.
- Rosser ZH, Zerjal T, Hurles ME et al: Y chromosomal diversity in Europe is clinal and influenced primarily by geography rather than by language. *Am J Hum Genet* 2000; **67**: 1526–1543.
- Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G: Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet* 2000; **66**: 262–278.
- Chikhi L, Nichols RA, Barbujani G, Beaumont MA: Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci USA* 2002; **99**: 11008–11013.
- Barbujani G, Sokal RR, Oden NL: Indo-European origins: a computer-simulation test of five hypotheses. *Am J Phys Anthropol* 1995; **96**: 109–132.
- Fix AG: Gene frequency clines produced by kin-structured founder effects. *Hum Biol* 1997; **69**: 663–673.
- Curat M, Excoffier L: The effect of the Neolithic expansion on European molecular diversity. *Proc Biol Sci* 2005; **272**: 679–688.

- 19 Klopstein S, Currat M, Excoffier L: The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* 2006; **23**: 482–490.
- 20 Lahr MM, Foley RA, Pinhasi R: Expected regional patterns of Mesolithic-Neolithic human population admixture in Europe based on archaeological evidence; in Renfrew C, Boyle K (eds): *Archaeogenetics: DNA and the Population Prehistory of Europe*. Cambridge, UK: McDonald Institute Monographs, 2000, pp 81–88.
- 21 Forenbaier S, Miracle PT: The spread of farming in the Eastern Adriatic. *Antiquity* 2005; **79**: 514–528.
- 22 Jobling MA, Tyler-Smith C: The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 2003; **8**: 598–612.
- 23 Rootsi S, Magri C, Kivisild T *et al*: Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 2004; **75**: 128–137.
- 24 Perić M, Lauc LB, Klarić IM *et al*: High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. *Mol Biol Evol* 2005; **22**: 1964–1975.
- 25 Hammer MF, Karafet T, Rasanayagam A *et al*: Out of Africa and back again: nested clastic analysis of human Y chromosome variation. *Mol Biol Evol* 1998; **4**: 427–441.
- 26 Scozzari R, Cruciani F, Pangrazio A *et al*: Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum Immunol* 2001; **29**: 871–884.
- 27 Semino O, Magri C, Benuzzi G *et al*: Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and late migratory events in the Mediterranean area. *Am J Hum Genet* 2004; **74**: 1023–1034.
- 28 Cruciani F, La Fratta R, Trombetta B *et al*: Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosomal haplogroups E-M78 and J-M12. *Mol Biol Evol* 2007; **24**: 1300–1311.
- 29 Cinnioğlu C, King R, Kivisild T *et al*: Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* 2004; **114**: 127–148.
- 30 Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R: Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. *Hum Mutat* 2006; **27**: 831–832.
- 31 Underhill PA, Myres NM, Rootsi S *et al*: New phylogenetic relationships for Y-chromosome haplogroup I: reappraising its phylogeography and prehistory; in Mellars P, Boyle K, Bar-Yosef O, Stringer C (eds): *Rethinking the Human Revolution*. Cambridge, UK: McDonald Institute Monographs, 2007, pp 33–42.
- 32 Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF: New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 2008; **18**: 830–838.
- 33 Marjanovic D, Fornarino S, Montagna S *et al*: The peopling of modern Bosnia-Herzegovina: Y-chromosome haplogroups in the three main ethnic groups. *Ann Hum Genet* 2005; **69**: 757–763.
- 34 Hammer MF, Horai S: Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet* 1995; **56**: 951–962.
- 35 Karafet T, Xu L, Du R *et al*: Paternal population history of East Asia: sources, patterns and microevolutionary processes. *Am J Hum Genet* 2001; **69**: 615–628.
- 36 Cruciani F, Santolamazza P, Shen P *et al*: A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am J Hum Genet* 2002; **72**: 1197–1214.
- 37 Sengupta S, Zhivotovsky LA, King R *et al*: Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central Asian pastoralists. *Am J Hum Genet* 2006; **78**: 202–221.
- 38 King RJ, Ozcan S, Carter T *et al*: Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic. *Ann Hum Genet* 2008; **72**: 205–214.
- 39 The Y Chromosome Consortium: A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 2002; **12**: 339–348.
- 40 Nei M: *Molecular Evolutionary Genetics*. New York: Columbia University Press, 1987.
- 41 Relethford JH: Geostatistics and spatial analysis in biological anthropology. *Am J Phys Anthropol* 2008; **136**: 1–10.
- 42 Bandelt HJ, Forster P, Rohl A: Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999; **16**: 37–48.
- 43 Zhivotovsky LA, Underhill PA, Cinnioğlu C *et al*: The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *Am J Hum Genet* 2004; **74**: 50–61.
- 44 Zhivotovsky LA, Underhill PA, Feldman MW: Difference between evolutionarily effective and germ line mutation rate due to stochastically varying haplogroup size. *Mol Biol Evol* 2006; **23**: 2268–2270.
- 45 Semino O, Passarino G, Oefner PJ *et al*: The genetic legacy of Paleolithic *Homo s. sapiens* in extant Europeans: a Y chromosome perspective. *Science* 2000; **290**: 1155–1159.
- 46 Martinez L, Underhill PA, Zhivotovsky LA *et al*: Paleolithic Y-haplogroup heritage predominates in a Cretan highland plateau. *Eur J Hum Genet* 2007; **15**: 485–493.
- 47 Di Giacomo F, Luca F, Anagnou N *et al*: Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Mol Phylogenet Evol* 2003; **28**: 387–395.
- 48 Di Giacomo F, Luca F, Popa LO *et al*: Y chromosomal haplogroup J as a signature of the post-Neolithic colonization of Europe. *Hum Genet* 2004; **115**: 357–371.
- 49 Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Børresen-Dale AL, Underhill PA: Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *Eur J Hum Genet* 2002; **10**: 521–529.
- 50 Al-Zahery N, Semino O, Benuzzi G *et al*: Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol Phylogenet Evol* 2003; **28**: 458–472.
- 51 Capelli C, Redhead N, Abernethy JK *et al*: A Y chromosome census of the British Isles. *Curr Biol* 2003; **13**: 979–984.
- 52 Zei G, Lisa A, Fiorani O *et al*: From surnames to the history of Y chromosomes: the Sardinian population as a paradigm. *Eur J Hum Genet* 2003; **11**: 802–807.
- 53 Arredi B, Poloni ES, Paracchini S *et al*: A predominantly Neolithic origin for Y-chromosomal DNA variation in North Africa. *Am J Hum Genet* 2004; **75**: 338–345.
- 54 Tambets K, Rootsi S, Kivisild T *et al*: The western and eastern roots of the Saami—the story of genetic ‘outliers’ told by mitochondrial DNA and Y chromosomes. *Am J Hum Genet* 2004; **74**: 661–682.
- 55 Quintana-Murci L, Krausz C, Zerjal T *et al*: Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. *Am J Hum Genet* 2001; **68**: 537–542.
- 56 Mitchell RJ, Hammer MF: Human evolution and the Y chromosome. *Curr Opin Genet Dev* 1996; **6**: 737–742.
- 57 Zhivotovsky LA, Underhill PA: On the evolutionary mutation rate at Y-chromosome STRs: comments on paper by Di Giacomo *et al* (2004). *Hum Genet* 2005; **116**: 529–532.
- 58 Evershed RP, Payne S, Sherratt AG *et al*: Earliest date for milk use in the Near East and southeastern Europe linked to cattle herding. *Nature* 2008; **455**: 528–531.
- 59 Di Gaetano C, Cerutti N, Crobu F *et al*: Differential Greek and northern African migrations to Sicily are supported by genetic evidence from the Y chromosome. *Eur J Hum Genet* 2009; **17**: 91–99.

- 60 Perles C: The early Neolithic in Greece: the first farming communities in Europe. *Cambridge, UK: Cambridge University Press* 2001.
- 61 Heun M, Schäfer-Pregl R, Klawan D *et al*: Site of Einkorn Wheat Domestication Identified by DNA Fingerprinting. *Science* 1997; **278**: 1312–1314.
- 62 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.
- 63 Hassan HY, Underhill PA, Cavalli-Sforza LL, Ibrahim ME: Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography and history. *Am J Phys Anthropol* 2008; **137**: 316–323.
- 64 Kuper R, Kröpelin S: Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution. *Science* 2006; **313**: 803–807.
- 65 Edmonds CA, Lillie AS, Cavalli-Sforza LL: Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci USA* 2004; **101**: 975–979.
- 66 Haak W, Forster P, Bramanti B *et al*: Ancient DNA from the first European farmers in 7500-year-old Neolithic sites. *Science* 2005; **310**: 1016–1018.
- 67 Sampietro ML, Lao O, Caramelli D *et al*: Palaeogenetic evidence supports a dual model of Neolithic spreading into Europe. *Proc Biol Sci* 2007; **274**: 2161–2167.

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