Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution xxx (2007) xxx–xxx
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic–Neolithic encounter

Cristian Capelli [a,b,*], Francesca Brisighelli [a], Francesca Scarnicci [a], Barbara Arredi [a,1], Alessandra Caglia' [a,2], Giuseppe Vetrugno [a], Sergio Tofanelli [c], Valerio Onofri [d], Adriano Tagliabracci [d], Giorgio Paoli [c], Vincenzo L. Pascali [a]

[a] *Institute of Legal Medicine, Universita' Cattolica del Sacro Cuore, Rome, Italy*
[b] *Department of Zoology, University of Oxford, South Park Road, Oxford OX1 3PS, UK*
[c] *Department of Biology, Anthropology Unit, University of Pisa, Pisa, Italy*
[d] *Institute of Legal Medicine, Universita' Politecnica delle Marche, Policlinico Torrette, Ancona, Italy*

Received 25 July 2006; revised 30 October 2006; accepted 21 November 2006

Abstract

The Italian peninsula, given its geographical location in the middle of the Mediterranean basin, was involved in the process of the peopling of Europe since the very beginning, with first settlements dating to the Upper Paleolithic. Later on, the Neolithic revolution left clear evidence in the archeological record, with findings going back to 7000 B.C. We have investigated the demographic consequences of the agriculture revolution in this area by genotyping Y chromosome markers for almost 700 individuals from 12 different regions. Data analysis showed a non-random distribution of the observed genetic variation, with more than 70% of the Y chromosome diversity distributed along a North–South axis. While the Greek colonisation during classical time appears to have left no significant contribution, the results support a male demic diffusion model, even if population replacement was not complete and the degree of Neolithic admixture with Mesolithic inhabitants was different in different areas of Italy.

Keywords: Y chromosome; STRs; SNPs; Italian peninsula; Neolithic revolution

1. Introduction

The hominid presence in the Italian peninsula has been complex and extended in time. *Homo sapiens* probably made his first appearance in this area around 30–40 K years before present (ybp) (Cunliffe, 2001). Around 11,000 ybp in the Fertile Crescent new resources became available to humans in the means of domesticated crops and animals. The new technology was now able to support large communities and provided the resources for a demographic expansion (Cunliffe, 2001). Technology spread quite fast across the European peninsula, reaching the western fringes just 4000 years later (Ammerman and Cavalli-Sforza, 1984). The related demographic impact is still a matter of debate, but a consensus seems to have been reached on substantial Neolithic contribution in the Mediterranean area (Semino et al., 2000; Chikhi et al., 2001; Simoni et al., 2000). In Italy, Apulia, Calabria and Eastern Sicily were involved since the very beginning in this process as testified by the first Neolithic archaeological remains dating to around 9000 ybp. Farming technology appeared in Central Italy, on both sides

* Corresponding author. Fax: +44 (0) 1865 310447.
E-mail address: cristian.capelli@zoo.ox.ac.uk (C. Capelli).
[1] Present address: BMR Genomics, Padova, Italy.
[2] Present address: DAC Servizio Polizia Scientifica Sezione Genetica Forense, Via Tuscolana 1548-00173, Roma.

1055-7903/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2006.11.030

of the Apennines, and in the North East, in the Po and Adige Valleys, only 1000 years later. In the remaining areas, i.e. North and Central West Italy, farming technology arrived later, around 6.5K ybp and was characterised by a marked continuity with earlier Mesolithic groups. Indigenous communities in fact tended to select specific aspects of the new technology and integrate them with their existing ways of life (Cunliffe, 2001). This led to the presence of two well-defined farming groups in the peninsula: a North Italian–Tyrrhenian one and a South Italian-Adriatic one (Cunliffe, 2001). Distribution of genetic variation in Italy has only been investigated by a few studies. Barbujani and Sokal (1991) investigated the presence of genetic barriers within the peninsula by looking at classic polymorphisms. Interestingly, the zones of sharp changes in gene frequency were reflected at both linguistic and geographical level. MtDNA analysis revealed the presence of clines within the peninsula (Barbujani et al., 1995), as previously shown also by classical markers (Cavalli-Sforza et al., 1994). More recently, a Y chromosome investigation identified a single North–South major cline across the peninsula, (Di Giacomo et al., 2003), but pointed to local drift and founder effect as the main explanations for the observed distribution of genetic diversity.

Despite the archeological relevance and the fact that its position in the middle of the Mediterranean but connected to central Europe provides a preferential location for testing hypothesis related to the peopling of the continent itself, little efforts have been expended testing demographic scenarios shaping the currently observed genetic variation in Italian peninsula. Recently, Barbujani and Goldstein (2004) and Currat and Excoffier (2005), have both proposed two major models regarding the peopling of Europe: the demic diffusion model (DD) and the cultural diffusion model (CD), the major difference among the two being the demographic impact that Near Eastern farmers had on the European peninsula. The two models have different expectations in respect of the pattern of genetic variation and so concordance with observed results can be tested.

We investigated Italian Y chromosome variation by sampling a total of 699 chromosomes in 12 different areas along the peninsula. We genotyped these chromosomes using 10 microsatellites and 17 unique event polymorphisms, defining haplotypes (hpts) and haplogroups (hgs), respectively. In comparison to the most recent research in this area (Di Giacomo et al., 2003), we included a much larger set of markers, a more focused sampling scheme and larger average population size, a strategy that made possible a more comprehensive evaluation of the data. Results point to a distribution of genetic variation along a North–South axis and support the demic diffusion model. We discuss how this scenario could be explained by the admixture of two different groups: the Mesolithic original inhabitants of the peninsula and the incoming Neolithic farmers. Implications at continental levels are also discussed.

2. Materials and methods

2.1. Samples

Samples were selected according to father's place of origin. In order to have larger samples, we clustered samples smaller than 30 individuals with the closest sampling point within a 30-km radius. Collection was performed using buccal swabs and blood draws. DNA was extracted by a modified salting out protocol (Miller et al., 1988). Sample locations are described in Fig. 1. Sample sizes are in Table 1. TLB, CMA and SAP Y-STR haplotypes have been previously described (TL, MA and SA, respectively in Capelli et al., 2006a). For simplicity, as all samples except VLB were from the Italian peninsula, when referring to Italy we meant the geographic area of the peninsula, unless otherwise stated. European UEP and STR available data were included for comparison (Semino et al., 2000; Flores et al., 2004; Goncalves et al., 2005; Cinnioglu et al., 2004; Bosch et al., 2000, 2001; Capelli et al., 2006b; Roewer et al., 2005).

2.2. Genotypings

Microsatellite variation was investigated by the analysis of the following 10 microsatellites: DYS 388, 393, 392, 19, 390, 391, 389 I and II and 385—which is a double allele locus. PCR amplification was performed in two different multiplexes as previously described (Capelli et al., 2006a). Genotyped UEPs were as follows: M9, M17, M26, M35, M78, M81, M89, M173, M170, M172, M201, 92R7, 12f2, SRY10831, YAP, RPS4Y, tat. We first genotyped all the samples using a multiplex containing M173, M17, M172 and M170, following the protocol described in Capelli et al. (2006b). Then, we genotyped one after the other M35, M201, 12f2, M89, M9 (Capelli et al., 2006b). The M201 marker was genotyped as described in Flores et al. (2003). Samples ancestral at these markers, were additionally tested for YAP, RPS4Y and SRY10831 (Thomas et al., 1999; Capelli et al., 2001). Samples derived at M35 were additionally tested for M78 and M81. The two markers were co-amplified using following primers M78F 5′-tggctgtatgggtttctttga-3′ (label with Hex dye), M78R 5′-cttattttgaaatatttggaagagc-3′, M81F 5′-gcactatcatactcagctacacatctc-3′ and M81R 5′-ctataatattcagtacaatagtgtcgcac-3′ (label with Ned dye). The PCR products were then digested using Bsr B1and HpyCH 4 IV restriction enzymes and scored using a 373 AB automated sequencer. Samples derived at M9 but not derived at M173 were additionally tested for tat and 92R7 (Rosser et al., 2000; Capelli et al., 2006b). A subset of the samples was SNP genotyped as described in Onofri et al. (2006). Phylogenetic relationships between markers and nomenclature follow the Y Chromosome Consortium (YCC, 2002). In our investigation a number of Y chromosome STR duplications were found (see Supplementary material). In particular, we note that, differing from the previously described DYS19 duplicated chro-

Please cite this article in press as: Capelli, C. et al., Y chromosome genetic variation in the Italian peninsula is clinal ..., Mol. Phylogenet. Evol. (2007), doi:10.1016/j.ympev.2006.11.030

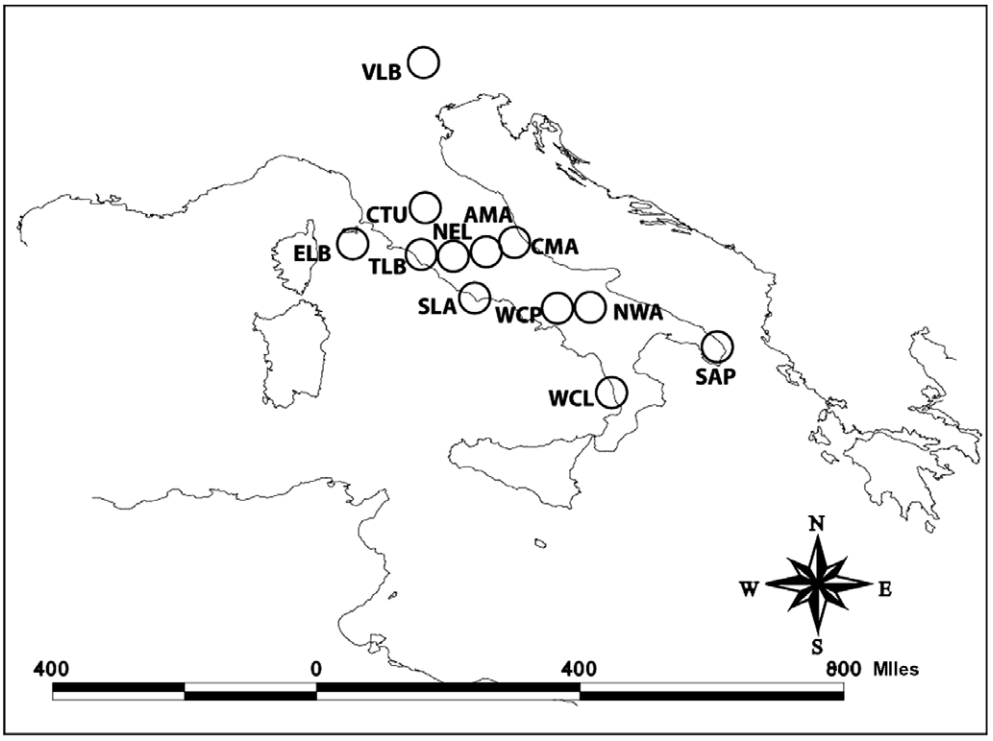

Fig. 1. Geographical location of the investigated Italian samples. Open circles showed the sample location areas. Sample codes are as follows: AMA, Apennine Marche; CMA, Central Marche; CTU, Central Tuscany; ELB, Elba Island (Tuscany); NEL, North–East Latium; NWA, North–West Apulia; SAP, South Apulia; SLA, South Latium; TLB, Tuscany–Latium border; VLB, Val Badia (Alto Adige); WCL, West Calabria; WCP, West Campania.

mosomes, the ones here described were part of hg G instead of C3b (Zerjal et al., 2003; Nasidze et al., 2005). This will be presented in a more comprehensive way in a future publication. Haplotypes containing duplicated loci were not included in analysis based on STR variation.

2.3. Data analysis

Haplogroup frequencies were estimated by chromosome counting. European SNP and STRs data included Anatolia (Cinnioglu et al., 2004), Iberia (Goncalves et al., 2005; Flores et al., 2004; Bosch et al., 2000, 2001), Cyprus, Sicily, Sardinia, Malta (Capelli et al., 2006b), Albania (Semino et al., 2000; Pericic et al., 2004; Robino et al., 2002), Greece (Semino et al., 2000; Parreira et al., 2002). Principal component analysis was performed by POPSTR (Henry Harpending, personal communication). AMOVA analysis was performed by the Arlequin package (Schneider et al., 2000). Spatial correlation analysis was performed using hg frequencies using the AIDA software (Bertorelle and Barbujani, 1995). Geographic pattern of genetic variation were also investigated using Barrier 2.2 (Manni et al., 2004). Average squared distance (ASD) values were estimated by MICROSAT (Minch, 1996). The ADMIX software by Bertorelle and Excoffier (1998) was applied to estimate the Anatolian vs. West European contribution to the Italian samples (see Section 3) and performed on haplogroup frequencies.

3. Results

3.1. Italian Y chromosome haplogroup composition

The 699 chromosomes were SNP genotyped and clustered in 13 different haplogroups (Table 1) following the Y chromosome genealogy (YCC, 2002). Four haplogroups were not observed even if markers defining those were tested: A, C, N3 and Px(R). Haplogroups R1*(xR1a1), J2, G and E3b1 comprised more than 80% of the total chromosomes. The frequencies in the entire sample set were 40%, 20%, 11% and 10%, respectively. The remaining haplogroups had frequencies below 7% in the total sample, and never above 9% in a single population (except I*xI1b2 in the ELB sample at 16%). In all samples except two, WCL and SLA, R1*(xR1a1) was the most common haplogroup. Y-STRs were used to estimate intra-haplogroup diversity. Locus DYS385 has a duplicated allele pattern that can not be resolved assigning each allele to the corresponding locus. We thus decided to exclude DYS385 from STR variance estimation. Similarly, to avoid double estimation of locus variation, repeat number at locus DYS389 II was calculated by subtracting the number of repeats at DSY389 I. Intra-haplogroup STR diversity was estimated as the ASD (Slatkin, 1995; Goldstein et al., 1995a). The values for the four most common haplogroups are reported in Table 2.

Please cite this article in press as: Capelli, C. et al., Y chromosome genetic variation in the Italian peninsula is clinal ..., Mol. Phylogenet. Evol. (2007), doi:10.1016/j.ympev.2006.11.030

Table 1
Italian samples studied

Location	Sample ID	Latitude	R1a1		R1* (xR1a1)		K* (xN3,P)		J2		J* (xJ2)		I1b2		I* (xI1b2)		G		F (*xG,I,J,K)		E3b2		E3b1		E3b* (xE3b1,E3b2)		DE* (xE3b)		n
Val Badia (Alto Adige)	VLB	46.47			0.68	23	0.09	3	0.09	3					0.06	2	0.03	1					0.06	2					34
Central Tuscany	CTU	43.27	0.05	2	0.46	19	0.02	1	0.17	7	0.07	3			0.05	2	0.07	3					0.07	3	0.02	1			41
Central Marche	CMA	42.51	0.02	1	0.37	22			0.36	21	0.05	3			0.08	5	0.07	4					0.05	3					59
Elba Island (Tuscany)	ELB	42.48	0.01	1	0.53	50			0.08	8			0.04	4	0.16	15	0.11	10					0.06	6	0.01	1			95
Apennine Marche	AMA	42.31	0.07	2	0.33	9	0.04	1	0.22	6					0.04	1	0.15	4					0.07	2	0.07	2			27
Tuscany–Latium border	TLB	42.25	0.05	4	0.41	32	0.04	3	0.19	15					0.05	4	0.15	12					0.08	6	0.04	3			79
North–East Latium	NEL	42.21	0.02	1	0.38	21	0.02	1	0.15	8					0.05	3	0.13	7			0.02	1	0.22	12	0.02	1			55
South Latium	SLA	41.29	0.04	2	0.37	19	0.08	4	0.25	13			0.02	1	0.06	3	0.12	6					0.06	3					51
North–West Apulia	NWA	41.09	0.07	3	0.52	24	0.04	2	0.17	8	0.02	1					0.13	6			0.04	2							46
West Campania	WCP	41.07	0.02	2	0.29	24	0.08	7	0.17	14	0.05	4	0.02	2	0.08	7	0.10	8			0.02	2	0.12	10	0.04	3	0.01	1	84
South Apulia	SAP	39.50	0.01	1	0.27	19	0.06	4	0.24	17	0.01	1			0.03	2	0.11	8	0.04	3			0.20	14	0.03	2			71
West Calabria	WCL	39.21	0.02	1	0.32	18			0.35	20	0.04	2			0.02	1	0.11	6					0.16	9					57
			0.03	20	0.40	280	0.04	26	0.20	140	0.02	14	0.01	7	0.06	45	0.11	75	0.00	3	0.01	5	0.10	70	0.02	13	0.00	1	699

The table reports in order: the sample location, the sample code, the latitude of the sampling area and the haplogroup frequencies in percentage (left column) and number of individuals (right column). In the last raw are reported the sum of the hg frequencies and the related number of chromosomes.

3.2. Population structure

Population relationships among Italian samples were investigated by principal component analysis (data not shown). Thirty percent of the total variation was displayed by principal component 1 (PC1). Haplogroups R1*(xR1a1), J2, and E3b1 had the highest loading factors (+0.21, −0.14, −0.14, respectively) along this axis. We plotted PC1 sample values vs. sample latitudes obtaining a correlation between the two ($R^2 = 0.55$). In order to check for the presence of a geographic pattern of haplogroup distribution, we plotted the frequencies of the three hgs underlying PC1 in each population vs. the sample latitude. The regression lines are shown in Fig. 2A. R1*(xR1a1) frequencies tend to increase linearly moving north, while E3b1 and J2 decrease along the same direction. We additionally investigated the distribution of genetic variation as expressed by ASD within each of the three haplogroups (Table 2) and plotted these values vs. the sample latitudes, as shown in Fig. 2B. The regression line for R1*(xR1a1) showed very little correlation between ASD and latitude, while both E3b1 and J2 showed a linear decrease for ASD with increasing latitude. We additionally tested how Italian samples are related to each other within a European context. Principal component analysis was performed including European data (Semino et al., 2000; Flores et al., 2004; Goncalves et al., 2005; Cinnioglu et al., 2004; Bosch et al., 2000, 2001; Capelli et al., 2006b) (Fig. 3). European populations are distributed along axis one following a W–SE direction, a result already shown by others (Semino et al., 2000; Rosser et al., 2000). This pattern has been interpreted as the result of the admixture of Neolithic Near Eastern farmers with the European Mesolithic groups following the demographic expansion that was likely associated with the development of agricultural technology (Semino et al., 2000; Rosser et al., 2000). When included, the Italian samples do not cluster all together. Northern and Southern samples in fact tend to show along axis one negative and positive values respectively, suggesting a closer affinity to Western European populations for the former and to South–East and South–Central Europe for the latter, with few exceptions (see below) (Fig. 3). In order to test this interpretation, we performed the analysis of the distribution of genetic variation (AMOVA) using available Y-STR European data (Roewer et al., 2005; Cinnioglu et al., 2004; Parreira et al., 2002; Robino et al., 2002; Pericic et al., 2004), including the Italian samples. When South Italian samples were clustered with SouthEast and South-Central European samples and the Northern groups with West Europe, the percentage of between group variation was 6.04% and the within group 1.70% (Φ_{sc} 0.018, $P \ll 0.01$; Φ_{ct} 0.06, $P \ll 0.01$), while the exchange of the Italian groups decreased the between groups variation and increased the within group value (3.77% and 2.67% respectively, Φ_{sc} 0.027, $P \ll 0.01$; Φ_{ct} 0.037, $P \ll 0.01$).

The evidence for differential clustering tendency for North and South samples was confirmed when AMOVA

Table 2
ASD (average squared distance) values and associated Standard Error (SE) for the four most frequent hgs in each Italian sample

Code	Latitude	R1*(xR1a1)	SE	J2	SE	G	SE	E3b1	SE
VLB	46.47	0.467	0.148	0.278	0.073	0.000	0.000	0.125	0.075
CTU	43.27	0.486	0.140	0.612	0.135	0.611	0.198	0.111	0.068
CMA	42.51	0.298	0.055	0.629	0.263	0.167	0.075	0.167	0.075
ELB	42.48	0.544	0.057	0.758	0.157	0.615	0.222	0.257	0.059
AMA	42.31	0.623	0.166	0.583	0.149	0.406	0.160	0.125	0.078
TLB	42.25	0.701	0.121	0.656	0.232	0.707	0.135	0.347	0.090
NEL	42.21	0.529	0.140	0.410	0.093	0.750	0.334	0.297	0.176
SLA	41.29	0.414	0.073	0.837	0.280	0.486	0.153	0.167	0.073
NWA	41.09	0.297	0.058	0.840	0.450	0.120	0.072	—	—
WCP	41.07	0.527	0.106	0.892	0.233	0.667	0.226	0.210	0.107
SAP	39.50	0.533	0.126	0.979	0.185	0.383	0.091	0.457	0.140
WCL	39.21	0.500	0.089	0.754	0.201	0.472	0.123	1.062	0.319

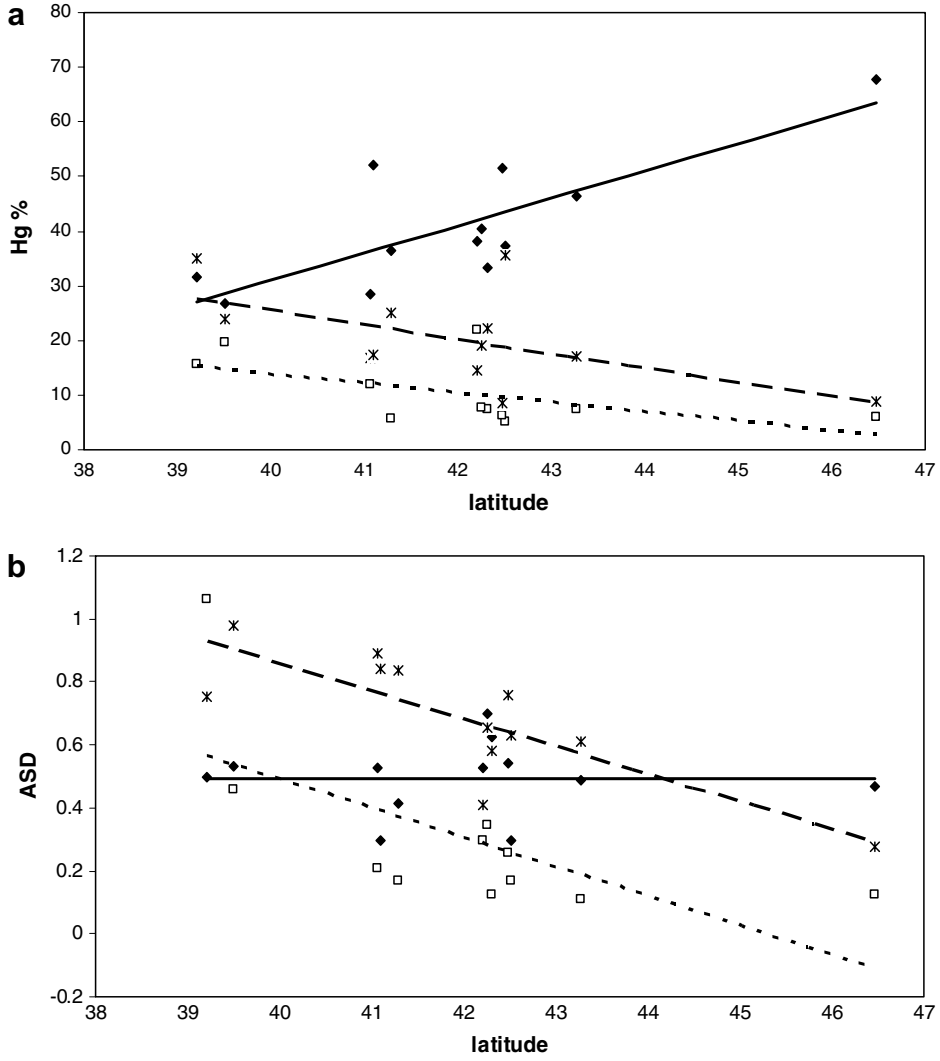

Fig. 2. Regression lines of the hg frequencies/average squared distance (ASD) versus latitude. (a) The regression lines of the latitude values of the selected samples (*X* axis) are plotted versus hg frequencies (*Y* axis); thick line, hg R1(*xR1a1), dashed line, hg J2; dotted line hg E3b1. E3b1 hg was absent from the NWA sample: this data point was not considered; (b) the latitude values (*X* axis) are plotted vs. ASD (*Y* axis) within selected hg; thick line, hg R1*(xR1a1); dashed line, hg J2; dotted line, hg E3b1. The E3b1 hg was absent from the NWA sample: this data point was not considered.

was conducted on the Italian samples only. Samples were divided in two groups, broadly defined as North and South, on the basis of their latitude (above and below 41.50° latitude, Table 3). This grouping displayed a negative value for the between groups variation (−0.19%, Φ_{ct} −0.001, $P = 0.46$), with a within group variation of

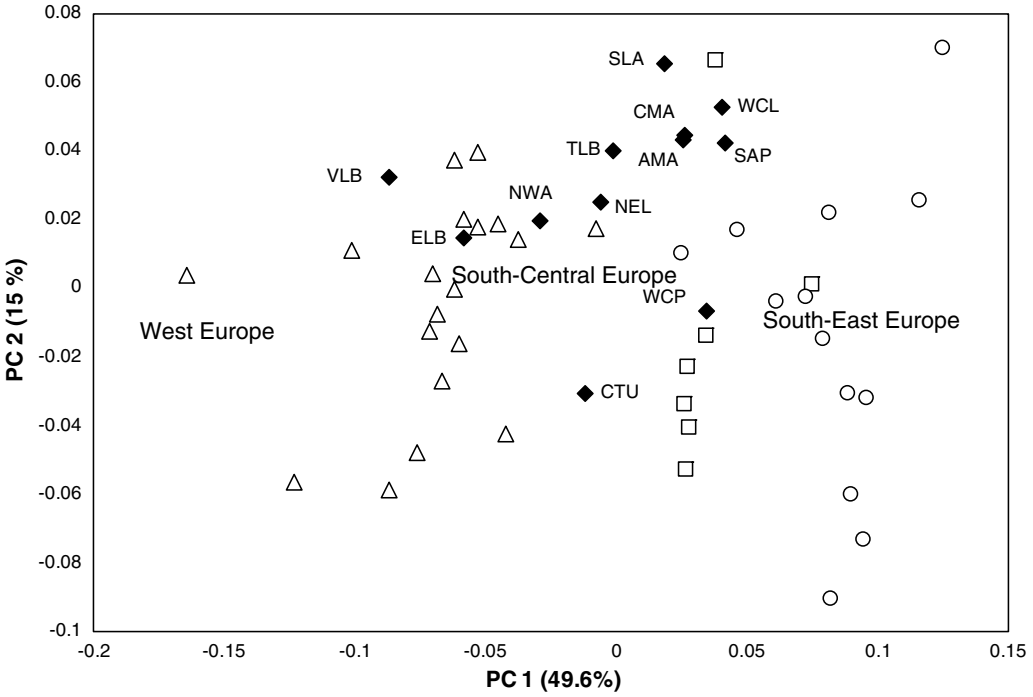

Fig. 3. Plot of the principal component analysis conducted on the haplogroup frequencies of the Italian and European samples; Italian samples investigated in this study are indicated by black diamonds, codes as in Table 1. Open circles, South–East European samples; open squares, South–Central European samples; open triangle, West European samples.

Table 3
Grouping schemes followed for the AMOVA analyses (see text)

	Latitude		PC	
	North	South	North	South
VLB	x		x	
CTU	x		x	
CMA	x			x
ELB	x		x	
AMA	x			x
TLB	x		x	
NEL	x		x	
SLA		x		x
NWA		x	x	
WCP		x		x
SAP		x		x
WCL		x		x

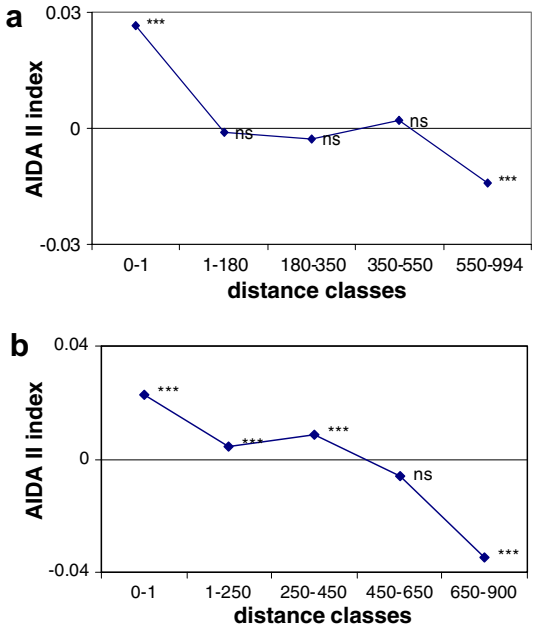

Fig. 4. Spatial autocorrelation analysis as performed by AIDA software: (a) complete Italian dataset; (b) Italian dataset excluding the NWA and CMA samples. $^{***}P < 0.005$; ns, not significant.

2.49% (Φ_{sc} 0.023, $P \ll 0.01$). We then tested a modified grouping (Table 3, PC column) as suggested by PC plot (Fig. 3, positive vs. negative values along axis 1) obtaining a significant percentage of genetic variation between groups of 2.68% (Φ_{ct} 0.026, $P < 0.01$) and within groups 0.88% (Φ_{sc} 0.015, $P \ll 0.01$). The main differences between the two groupings tested (latitude vs. principal component based) are the positioning of the central Italian samples AMA and CMA, and of the southern sample NWA. Starting from the first grouping tested (by latitude), we checked how moving samples across the groups influenced the genetic variation distribution. We moved the samples singularly from one group to the other and estimated the amount of genetic variation between and within groups. The movement of the NWA and CMA samples changed the between group variation to positive

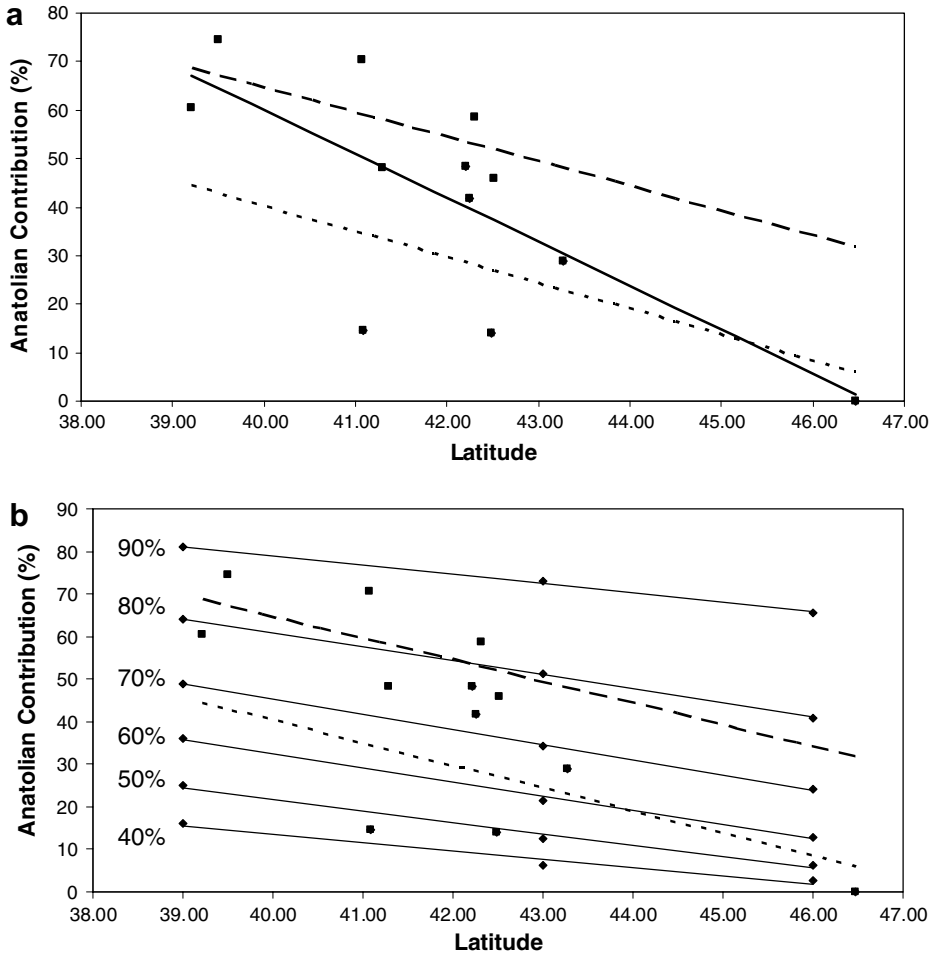

Fig. 5. Admixture analysis: (a) regression lines for the Anatolian genetic contribution to the Italian samples as estimated by the ADMIX software. Dashed line, regression line estimated for southern Italian samples; dotted line, regression line estimated for Northern Italian samples; continuous line, regression line estimated for the entire Italian dataset; (b) regression lines estimated for three admixture event with different newcomer contribution, as described in the Section 3. Admixture proportions are indicated beside each line (see text for description). Dashed and dotted lines are the regression lines estimated for Southern and Northern Italian samples, respectively.

values (0.69% and 1.50%, respectively) while any other shuffling maintained negative values. Moving both at the same time increased the between groups values to 2.89% (Φ_{ct} 0.028, $P \ll 0.01$) and decreased the within groups value to 0.77% (Φ_{sc} 0.007, $P \ll 0.01$). The shuffle of the AMA sample between the two groups did not have any relevant effect. CMA and NWA samples result clearly separated from geographically surrounding samples when investigated by the Barrier 2.2 Software (data not shown).This seems to suggest that for samples east of the Apennines (CMA) the clustering of southern samples is supported for higher latitudes than west of the Apennines. The NWA sample is clearly representing a geographical out-group in respect to the other southern samples (see below). We therefore considered CMA and NWA as part of South and North Italian peninsula respectively in the subsequent analysis. Given the inconclusive indication of the AMOVA analysis for the AMA sample, we decided to follow the PC results and considered this sample as part of the South.

3.3. Spatial correlation analysis

In order to properly test for the presence of clinally distributed genetic variation, we performed spatial correlation analysis as implemented in the AIDA software (Bertorelle and Barbujani, 1995), using hg frequencies. Five classes of distance were selected in order to have similar number of comparisons within each class (Fig. 4a).The results are significant (considering within population diversity) for distances above 500 km. Shorter distance classes were all non significant. We also evaluated the influence of each sample on the observed pattern by excluding the samples one at a time. The exclusion of CMA or the NWA sample resulted in a clinal pattern significant for all distance classes, while the alternative exclusion of the other samples resulted in one or more than one class not being significant (data not shown). The simultaneous exclusion of both CMA and NWA resulted in a very clear clinal pattern (Fig. 4b), confirming the absence of correlation with latitude for these samples (see above).

Please cite this article in press as: Capelli, C. et al., Y chromosome genetic variation in the Italian peninsula is clinal ..., Mol. Phylogenet. Evol. (2007), doi:10.1016/j.ympev.2006.11.030

3.4. Admixture analysis

Given the current proposed models for agricultural technology spread in Europe [CD vs DD; Barbujani and Goldstein, 2004, Currat and Excoffier, 2005], we investigated the genetic impact that Near Eastern populations might have had on Italian populations by admixture analysis performed using the ADMIX program (Bertorelle and Excoffier, 1998). This software estimates the contribution of two source population to the tested population, using allele frequencies. In admixture analysis, a major problem is to correctly identify the source groups. Chikhi et al. (2002) used data of Basques, Lebanese, Syrians and Turks (Semino et al., 2000) to investigate the degree of Near Eastern introgression into a large European dataset by the use of a likelihood approach (Chikhi et al., 2001). Results suggested that all European populations probably experienced Neolithic introgression, with admixture values that must have been larger than 70%. We previously suggested that current Levantine population might have experienced very strong Arab introgression (Capelli et al., 2006b). For this reason, we alternatively selected a large Anatolian data set as representative of newcomer Farmers (Cinnioglu et al., 2004). Linguistic and genetic data support the Basques as descendents from the Palaeolithic inhabitants of Europe (Gamkrelidze and Ivanov, 1990; Cavalli-Sforza et al., 1994; Richards et al., 2000; Semino et al., 2000; Wilson et al., 2001; Chikhi et al., 2002) even if some concerns have been raised (Alonso et al., 2005). However, it has to be noted that drift and bottlenecks might have introduced severe distortions in current Basque populations compared to the ancient ones (Chikhi et al., 2002; Alonso et al., 2005). In order to minimise such occurrences, and following previous results describing limited genetic variation across Iberian samples (Flores et al., 2004), we have decided to pull together data from Iberia (Flores et al., 2004; Bosch et al., 2001) as a representative of the original European inhabitants. The genetic contribution of Anatolian and Iberian source populations were estimated in all the Italian samples by the use of ADMIX. Anatolian contribution estimates were plotted vs. the corresponding sample latitudes. The regression line from these data shows an inverse correlation between Anatolian contribution and latitude (see Fig. 5a). We additionally estimated the regression lines of the samples when clustered according the PC plot results (Table 3, PC column). Interestingly the Northern and Southern samples have two different regression lines, pointing to different degree of genetic impact of the newcomers on the Italian populations. As noted by Chikhi et al. (2002), admixture expectations can be calculated considering a very simple admixture model. This model takes into consideration a number of admixture steps in which the newcomer populations mix with the original inhabitants and create a new community, from which a new "admixture wave" is generated. With P_N the proportion of farmers in the admixed population, the newcomer contribution to each location will decrease in a geometric way from P_N to P_N^n, where n is the number of admixture events (Chikhi et al., 2002). For example, given a P_N of 0.9, and considering a minimum of three admixture events occurring, the proportion of newcomer "genes" will be 81%, 73% and 66%. Following this approach, we plotted the estimated Newcomers contribution vs. three latitude points, evaluating the effect of different P_N admixture proportions (Fig. 5b). If compared to the regression lines estimated from North and South Italian samples, it appears that the different position of the two lines is related to different admixture contributions, with Southern samples showing higher contribution than Northern ones (Fig. 5b, Mann–Whitney test, $P < 0.05$).

4. Discussion

Two alternative models have been proposed for the dispersion of agricultural technology in the European continent. The CD–cultural diffusion–model suggests a gradual acceptance of the new technology by the original inhabitants with little or no admixture with Near Eastern farmers. The DD–demographic diffusion–model instead points to substantial genetic introgression of the Near Eastern populations, supported in their demographic expansion by the resources offered by the new food producing technology (Cunliffe, 2001). The two models have clear expectations that can be tested for concordance with the observed data. In the absence of an incoming population, as in the case of the CD model, local drift and gene flow would be the main forces shaping genetic variation. Random fluctuation across populations would prevent the establishment of gradients and no evidence of admixture would be expected. Gene frequency clines have been indicated as necessary but not sufficient to support the DD model (Currat and Excoffier, 2005). The distribution of genetic variation following a population expansion is associated with a loss of genetic diversity due to a succession of small founder effects (Barbujani et al., 1995). The DD model would then predict that along the direction of dispersal, frequency and diversity clines would be generated. Given the population introgression, genetic evidence of admixture is also expected to be found. We extensively sampled along the peninsula to specifically address the issue of the distribution of Y chromosome genetic variation in the light of agriculture diffusion models. When compared to other European samples, no outgroups were found among the Italian samples (Fig. 3). The samples in fact are distributed within the genetic variation shown by other European and Mediterranean populations. However, a limited degree of separation among Italian groups emerged along the first principal component, with Northern Italian samples closer to western European populations and Southern samples closer to South East and South Central European groups, with few exceptions (Fig. 3). This different affinity was also highlighted by AMOVA analysis conducted on microsatellite variation. Similar results were also shown in the seminal work on classical polymorphisms by Piazza et al.

Please cite this article in press as: Capelli, C. et al., Y chromosome genetic variation in the Italian peninsula is clinal ..., Mol. Phylogenet. Evol. (2007), doi:10.1016/j.ympev.2006.11.030

(1988). The Italian samples here analysed appear to be placed within the ES–NW European cline that a number of previous studies have already described and that has been considered as compatible with a demographic scenario of admixture between the Near East farmers and the long-term European Mesolithic inhabitants (Menozzi et al., 1978; Semino et al., 2000; Rosser et al., 2000). We have identified clines for three haplogroups, two of which showed also diversity gradients (Fig. 2).

In the light of Near Eastern gene flow, admixture analysis revealed Anatolian introgression in most of the Italian samples. Considering the expectations related to the different models proposed for spread of agricultural technology, these results support the DD model. Despite the presence of Neolithic genes in the current male Italian population, the admixture values as estimated by ADMIX suggested a differential impact of the newcomers across the Italian samples. The estimated degree of introgression is in fact not consistent across all areas, with Southern samples experiencing higher Anatolian contribution than Northern samples (Fig. 5a). A very rough estimation based on Fig. 5b would suggest a 70–90% contribution for the former and 50–70% for the latter. However, given the number of assumptions, these values should not be taken as absolute. The selection as a second source population of a recently available larger sample of Basques (Alonso et al., 2005) did not change the observed pattern (data not shown). It is interesting to note that only one sample (SAP) displayed an Anatolian genetic contribution not significantly different from 100% (data not shown). Simulated samples obtained by re-sampling either Iberians or Anatolians and tested using the same source populations confirmed the sensitivity of this approach as these simulated samples were not significantly different from 0% and 100% Anatolian contribution, respectively (data not shown). These results underline that, beside geographically different Near Eastern contributions, population replacement was not complete across the peninsula. It follows that both Neolithic and Mesolithic genetic components can be found in current Italian male gene pool.

It is interesting to note that hg R1*(xR1a1) does show a frequency cline, opposite to the ones shown by J2 and E3b1, but apparently no diversity gradient is associated (Fig. 2). We are aware that conclusions drawn on single haplogroups are subject to bias and should not be equated to those drawn from entire samples, but we note that populations do have different hg composition that might retain signatures of past demographic events. The Mesolithic populations had low population density and possibly limited gene flow across groups (Mithen, 2004). If we assume that Mesolithic population were characterised by high frequencies of hg R1*(xR1a1) (as the case in current Basque groups, usually considered as representative of the original inhabitants of Europe) (Semino et al., 2000; Wilson et al., 2001; but see also Alonso et al., 2005), genetic variation within this haplogroup would be independent of geographic sampling and instead mainly shaped by local demographic history. It follows that R1*(xR1a1) diversity would not be expected to show clines related to latitude but instead would be randomly distributed across populations. The later newcomers as represented by Neolithic farmers, would have expanded and admixed with these Mesolithic groups, and generated, as expected, frequency and diversity clines along the direction of dispersal as indeed shown by their most representative chromosome types, E3b1 and J2 (Rosser et al., 2000; Semino et al., 2000, 2004; Cinnioglu et al., 2004). Other less common haplogroups might have retained the signature of those events but the current limited sampling size might have prevented their detection. Along the direction of dispersal (Barbujani et al., 1995), an opposite frequency cline, but not a diversity one, for hg R1*(xR1a1) would be generated. The observed higher Near Eastern contribution to East Apennines vs West Apennines samples for northern latitudes is consistent with the archaeological separation existing among early agricultural areas (Cunliffe, 2001).

The current set of data also provides a first frame for testing the hypothesis of genetic continuity from Palaeolithic to Mesolithic in Italy through the last Ice age. This would point to the presence of an Italian Pleistocene refugium, postulated for Iberian, Italian and Balkan peninsulae for a number of species (Hewitt, 2001; Brito, 2005), but not proposed for humans (Semino et al., 2000; Rosser et al., 2000). This inconsistency could be possibly due to the fact that so far no specific haplogroups have been identified at Y chromosome level in Italy (Semino et al., 2000). However this could be due to lack of resolution in the current set of markers and other Y chromosome sublineages not yet characterised might represent a specific marker for the Italian area. Taking into consideration current European Y chromosome hg distribution and data presented here, a possible candidate could be within hg R1*(xR1a1). A comparison of the genetic variation estimated as the variance of the repeat scores averaged across loci of this group in both Iberia and Italy did not show significant difference ($P > 0.05$, data not shown, Brion et al., 2004; Bosch et al., 2000, 2001). Looking at the pattern of haplotypes sharing within hg P [that contains hg R1*(xR1a1)], only 28% of those are shared among the two populations. This value is well below the one estimated when comparing Iberia with areas re-peopled after last glacial maximum, as the British Isles (47%) (Capelli et al., 2003). The opposite pattern is instead observed when comparing haplotypes within haplogroups whose dispersion was probably associated with different and more recent events, as hg J (data not shown). This suggests that the R1*(xR1a1) variation present in Italy appear not to be a subset of the Iberian one. More extensive analysis would give the opportunity to test this hypothesis further.

Previous studies using autosomal data described the presence of a major North–South Cline within the peninsula. Cavalli-Sforza et al. (1994) showed that 27% of the total genetic variation as shown by classic polymorphisms was summarized along this axis. When compared to European

Please cite this article in press as: Capelli, C. et al., Y chromosome genetic variation in the Italian peninsula is clinal ..., Mol. Phylogenet. Evol. (2007), doi:10.1016/j.ympev.2006.11.030

samples, populations from South Italy clustered with Mediterannean groups, while the others grouped with West and Central European populations (Piazza et al., 1988). Authors suggested the Greek colonization in the South as the major demographic event shaping observed diversity (Piazza et al., 1988) on the basis of compatible historical scenarios. However, this hypothesis was never thoroughly tested, especially in the light of alternative European scenarios proposed by the same authors supporting a Neolithic demic dispersion model (Menozzi et al., 1978; Ammerman and Cavalli-Sforza, 1984; Cavalli-Sforza et al., 1994) and taking in consideration that Greece was the only Mediterranean sample outside Italy included in the PC analysis (Piazza et al., 1988). Assuming the Greek diaspora model, South Italian samples should be genetically closer to Greece than to Anatolia, while the Neolithic model would not show significant differences. We note that D_c and $\Delta\mu$ genetic distances are linearly related to time (Cavalli-Sforza and Edwards, 1967; Goldstein et al., 1995b). We calculated these genetic distances for WCL, WCP and SAP samples vs. Anatolia and Greece (Cinnioglu et al., 2004; Parreira et al., 2002). The three Italian populations were not only closer to Anatolia than to Greece, but all values for Anatolia were smaller than those for Greece (data not shown). This is confirmed also using more specific regional Greek samples (data not shown, Robino et al., 2004). Assuming proper identification of the source populations, these results suggest that in terms of demographic influence on the paternal Italian gene pool, the role of Neolithic farmers was greater than Greek historical colonisers of South Italy.

Similarly, given the sporadic and rare distribution of the E3b2 chromosomes, it is possible to conclude that North African gene flow, if any, left no significant evidence in the current Italian Y chromosome pool (Bosch et al., 2000, 2001; Capelli et al., 2006b).

We finally note that in a recent study, Di Giacomo et al. (2003) genotyped Y chromosome markers for 524 Italians sampled in 17 locations. They found that, excluding R1*(xR1a1), no other clines could be identified and concluded that most of the observed variation was due to drift and founder effects. Local drift has to be expected, due to local demographic histories. This seems the case of the NWA sample, as shown by its reduced genetic variation (Table 2). The NWA sample, despite its localisation in the South, tends clearly to cluster with Northern populations, as shown by various analyses. This is driven by the combination of high hg R1*(xR1a1) frequency and absence/low frequency of hgs E3b1 and J2. In our analyses, some Y chromosome lineages, despite local drift and gene flow, still show the signature of dispersal events in the past. Inspecting the data of Di Giacomo et al., it emerges that despite a lower number of samples points (12 vs. 17) our study is characterised by a larger average sample size, almost twice as much (58 vs. 31). Additionally, while only one of our samples (AMA) is below their average sample size, none of theirs is above or close to our average, with the largest population size in Di Giacomo et al. (2003) being 48. When we included their samples in our analysis, we confirmed the clinal distribution for R1*(xR1a1), J2 and DE hgs (in their study only the YAP marker was genotyped in the DE lineage so this is the closest approximation for E3b1)(data not shown). Besides recognising that drift definitively had a role in shaping current Y chromosome genetic variation, however we concluded that in Italy more than 70% of the observed diversity is distributed along gradients and that Anatolian Farmers did have a different demographic impact on different Italian areas for paternal lienages.

Acknowledgments

We thank all the people those donated their DNA and made the present study feasible. We additionally thank all the people that were involved in the sampling: Sara Partemi, Daniela Vantaggiato, Romolo Donnini, Fidelia Cascini, Angela Reveruzzi, Domenico Alfieri, Leonardo Grimaldi, Semeraro, Roberto Festa, Laura Baldassarri. We additionally would like to thank the AVIS Blood collection directors of Norma, Giuseppe Santucci and Sezze, Ubaldo Brandolini. C.C. thanks Jim Wilson and Martin Richards for useful comments to an early version of the manuscript. This study was funded by the Italian Ministry of University (PRIN-MIUR 2002, N.2002063871).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2006.11.030.

References

Alonso, S., Flores, C., Cabrera, V., Alonso, A., Martin, P., Albarran, C., Izagirre, N., de la Rua, C., Garcia, O., 2005. The place of the Basques in the European Y-chromosome diversity landscape. Eur. J. Hum. Genet. 13, 1293–1302.

Ammerman, A.J., Cavalli-Sforza, L.L., 1984. The Neolothic transition and the genetics of populations in Europe. Princeton University Press, Princeton.

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

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

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

Barbujani, G., Sokal, R.R., Oden, N.L., 1995. Indo-European origins: a computer-simulation test of five hypotheses. Am. J. Phys. Anthropol. 96, 109–132.

Bertorelle, G., Barbujani, G., 1995. Analysis of DNA diversity by spatial autocorrelation. Genetics 140 (2), 811–819.

Bertorelle, G., Excoffier, L., 1998. Inferring admixture proportions from molecular data. Mol. Biol. Evol. 15, 1298–1311.

Bosch, E., Calafell, F., Perez-Lezaun, A., Comas, D., Izaabel, H., Akhayat, O., Sefiani, A., Hariti, G., Dugoujon, J.M., Bertranpetit,

J., 2000. Y chromosome STR haplotypes in four populations from northwest Africa. Int. J. Legal. Med. 114, 36–40.

Bosch, E., Calafell, F., Comas, D., Oefner, P.J., Underhill, P.A., Bertranpetit, J., 2001. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am. J. Hum. Genet. 68, 1019–1029.

Brion, M., Quintans, B., Zarrabeitia, M., Gonzalez-Neira, A., Salas, A., Lareu, V., Tyler-Smith, C., Carracedo, A., 2004. Micro-geographical differentiation in Northern Iberia revealed by Y-chromosomal DNA analysis. Gene 329, 17–25.

Brito, P.H., 2005. The influence of Pleistocene glacial refugia on tawny owl genetic diversity and phylogeography in western Europe. Mol. Ecol. 14, 3077–3094.

Capelli, C., Wilson, J.F., Richards, M., Stumpf, M.P., Gratrix, F., Oppenheimer, S., Underhill, P., Pascali, V.L., Ko, T.M., Goldstein, D.B., 2001. A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. Am. J. Hum. Genet. 68, 432–443.

Capelli, C., Redhead, N., Abernethy, J.K., Gratrix, F., Wilson, J.F., Moen, T., Hervig, T., Richards, M., Stumpf, M.P., Underhill, P.A., Bradshaw, P., Shaha, A., Thomas, M.G., Bradman, N., Goldstein, D.B., 2003. A Y chromosome census of the British Isles. Curr. Biol. 13, 979–984.

Capelli, C., Arredi, B., Baldassari, L., Boschi, I., Brisighelli, F., Caglia, A., Dobosz, M., Scarnicci, F., Vetrugno, G., Pascali, V.L., 2006a. A 9-loci Y chromosome haplotype in three Italian populations. Forensic Sci. Int. 159, 64–70.

Capelli, C., Redhead, N., Romano, V., Cali', F., Lefranc, G., Delague, V., Megarbane, A., Felice, A.E., Pascali, V.L., Neophytou, P.I., Poulli, Z., Novelletto, A., Malaspina, P., Terrenato, L., Berebbi, A., Fellous, M., Thomas, M.G., Goldstein, D.B., 2006b. Population structure in the Mediterranean basin: a Y chromosome perspective. Ann. Hum. Genet. 70, 207–225.

Cavalli-Sforza, L.L., Edwards, A.W., 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 19 (Supp1. 3), 233.

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. The History and Geography of Human Genes. Princeton University Press, Princeton, pp. 277–280.

Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G.L., Lillie, A.S., Roseman, C.C., Lin, A.A., Prince, K., Oefner, P.J., Shen, P., Semino, O., Cavalli-Sforza, L.L., Underhill, P.A., 2004. Excavating Y-chromosome haplotype strata in Anatolia. Hum. Genet. 114, 127–148.

Chikhi, L., Bruford, M.W., Beaumont, M.A., 2001. Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. Genetics 158, 1347–1362.

Chikhi, L., Nichols, R.A., Barbujani, G., Beaumont, M.A., 2002. Y genetic data support the Neolithic demic diffusion model. Proc. Natl. Acad. Sci. USA 99, 11008–11013.

Cunliffe, B., 2001. The Oxford Illustrated History of PreHistoric Europe. Oxford University Press, Oxford.

Currat, M., Excoffier, L., 2005. The effect of the Neolithic expansion on European molecular diversity. Proc. Biol. Sci. 272, 679–688.

Di Giacomo, F., Luca, F., Anagnou, N., Ciavarella, G., Corbo, R.M., Cresta, M., Cucci, F., Di Stasi, L., Agostiano, V., Giparaki, M., Loutradis, A., Mammi', C., Michalodimitrakis, E.N., Papola, F., Pedicini, G., Plata, E., Terrenato, L., Tofanelli, S., Malaspina, P., Novelletto, A., 2003. Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. Mol. Phylogenet. Evol. 28, 387–395.

Flores, C., Maca-Meyer, N., Perez, J.A., Gonzalez, A.M., Larruga, J.M., Cabrera, V.M., 2003. A predominant European ancestry of paternal lineages from Canary Islanders. Ann. Hum. Genet. 67, 138–152.

Flores, C., Maca-Meyer, N., Gonzalez A, M., Oefner, P.J., Shen, P., Perez, J.A., Rojas, A., Larruga, J.M., Underhill, P.A., 2004. Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. Eur. J. Hum. Genet. 10, 855–863.

Gamkrelidze, T., Ivanov, V., 1990. The early history of Indo-European. Sci. Am. 262, 110–116.

Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L.L., Feldman, M.W., 1995a. An evaluation of genetic distances for use with microsatellite loci. Genetics 139, 463–471.

Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L.L., Feldman, M.W., 1995b. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc. Natl. Acad. Sci. USA 92, 6723–6727.

Goncalves, R., Freitas, A., Branco, M., Rosa, A., Fernandes, A.T., Zhivotovsky, L.A., Underhill, P.A., Kivisild, T., Brehm, A., 2005. Y-chromosome lineages from Portugal, Madeira and Acores record elements of Sephardim and Berber ancestry. Ann. Hum. Genet. 69, 443–454.

Hewitt, G.M., 2001. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. Mol. Ecol. 10, 537–549.

Manni, F., Guerard, E., Heyer, E., 2004. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. Hum. Biol. 76, 173–190.

Menozzi, P., Piazza, A., Cavalli-Sforza, L., 1978. Synthetic maps of human gene frequencies in Europeans. Science 201, 786–792.

Miller, S.A., Dykes, D.D., Polesky, H.F., 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16, 1215.

Minch, E., 1996. MICROSAT. Version 1.4. Stanford University Medical Centre, Stanford, USA.

Mithen, S., 2004. After the ice: a global human history 20,000–5000 BC. Phoenix Press, London, UK.

Nasidze, I., Quinque, D., Dupanloup, I., Cordaux, R., Kokshunova, L., Stoneking, M., 2005. Genetic evidence for the Mongolian ancestry of Kalmyks. Am. J. Phys. Anthropol. 128, 846–854.

Onofri, V., Alessandrini, F., Turchi, C., Pesaresi, M., Buscemi, L., Tagliabracci, A., 2006. Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic Sci. Int. 157, 23–35.

Parreira, K.S., Lareu, M.V., Sanchez-Diz, P., Skitsa, I., Carracedo, A., 2002. DNA typing of short tandem repeat loci on Y-chromosome of Greek population. Forensic Sci. Int. 126, 261–264.

Pericic, M., Lauc, L.B., Klaric, I.M., Janicijevic, B., Behluli, I., Rudan, P., 2004. Y chromosome haplotypes in Albanian population from Kosovo. Forensic Sci. Int. 146, 61–64.

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

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Novelletto, A., Oppenheim, A., Norby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozari, R., Torroni, A., Bandelt, H.J., 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am. J. Hum. Genet. 67, 1251–1276.

Robino, C., Gino, S., Ricci, U., Grignani, P., Previdere, C., Torre, C., 2002. Y-chromosomal STR haplotypes in an Albanian population sample. Forensic Sci. Int. 129, 128–130.

Robino, C., Varacalli, S., Gino, S., Chatzikyriakidou, A., Kouvatsi, A., Triantaphyllidis, C., Di Gaetano, C., Crobu, F., Matullo, G., Piazza, A., Torre, C., 2004. Y-chromosomal STR haplotypes in a population sample from continental Greece, and the islands of Crete and Chios. Forensic Sci. Int. 145, 61–64.

Roewer, L., Croucher, P.J., Willuweit, S., Lu, T.T., Kayser, M., Lessig, R., de Knijff, P., Jobling, M.A., Tyler-Smith, C., Krawczak, M., 2005. Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum. Genet. 116, 279–291.

Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D.G.,

Please cite this article in press as: Capelli, C. et al., Y chromosome genetic variation in the Italian peninsula is clinal ..., Mol. Phylogenet. Evol. (2007), doi:10.1016/j.ympev.2006.11.030

Brede, Cooper, G., Corte-Real, H.B., de Knijff, P., Decorte, R., Dubrova, Y.E., Evgrafov, O., Gilissen, A., Glisic, S., Golge, M., Hill, E.W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S.A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L.A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T.A., Mikelsaar, A.V., Mitchell, R.J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P.C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M.J., Previdere, C., Roewer, L., Rootsi, S., Rubinsztein, D.C., Saillard, J., Santos, F.R., Stefanescu, G., Sykes, B.C., Tolun, A., Villems, R., Tyler-Smith, C., Jobling, M.A., . Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. 67, 1526–1543.

Schneider, S., Roessli, D., Excoffier, L., 2000. ARLEQUIN ver. 2.000: A Software for Population Genetic Data Analysis. Genetics and Biometery Laboratory, University of Geneva, Geneva.

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

Semino, O., Magri, C., Benuzzi, G., Lin, A.A., Al-Zahery, N., Battaglia, V., Maccioni, L., Triantaphyllidis, C., Shen, P., Oefner, P.J., Zhivotovsky, L.A., King, R., Torroni, A., Cavalli-Sforza, L.L., Underhill, P.A., Santachiara-Benerecetti, A.S., 2004. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am. J. Hum. Genet. 74, 1023–1034.

Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J., Barbujani, G., 2000. Geographic patterns of mtDNA diversity in Europe. Am. J. Hum. Genet. 66, 262–278.

Slatkin, M., A measure of population subdivision based on microsatellite allele, frequencies, 1995. Genetics 139, 457–462.

Thomas, M.G., Bradman, N., Flinn, H.M., 1999. High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. Hum. Genet. 105, 577–581.

Y Chromosome Consortium., 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res. 12, 339–348.

Wilson, J.F., Weiss, D.A., Richards, M., Thomas, M.G., Bradman, N., Goldstein, D.B., 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc. Natl. Acad. Sci. USA 98, 5078–5083.

Zerjal, T., Xue, Y., Bertorelle, G., Wells, R.S., Bao, W., Zhu, S., Qamar, R., Ayub, Q., Mohyuddin, A., Fu, S., Li, P., Yuldasheva, N., Ruzibakiev, R., Xu, J., Shu, Q., Du, R., Yang, H., Hurles, M.E., Robinson, E., Gerelsaikhan, T., Dashnyam, B., Mehdi, S.Q., Tyler-Smith, C., 2003. The genetic legacy of the Mongols. Am. J. Hum. Genet. 72, 717–721.

Please cite this article in press as: Capelli, C. et al., Y chromosome genetic variation in the Italian peninsula is clinal ..., Mol. Phylogenet. Evol. (2007), doi:10.1016/j.ympev.2006.11.030