

ORIGINAL ARTICLE

Evidence of high genetic variation among linguistically diverse populations on a micro-geographic scale: a case study of the Italian Alps

Valentina Coia¹, Ilaria Boschi², Federica Trombetta², Fabio Cavulli¹, Francesco Montinaro^{3,4}, Giovanni Destro-Bisol^{3,4}, Stefano Grimaldi¹ and Annaluisa Pedrotti¹

Although essential for the fine-scale reconstruction of genetic structure, only a few micro-geographic studies have been carried out in European populations. This study analyzes mitochondrial variation (651 bp of the hypervariable region plus 17 single-nucleotide polymorphisms) in 393 samples from nine populations from Trentino (Eastern Italian Alps), a small area characterized by a complex geography and high linguistic diversity. A high level of genetic variation, comparable to geographically dispersed European groups, was observed. We found a difference in the intensity of peopling processes between two longitudinal areas, as populations from the west-central part of the region show stronger signatures of expansion, whereas those from the eastern area are closer to the expectations of a stationary demographic state. This may be explained by geomorphological factors and is also supported by archeological data. Finally, our results reveal a striking difference in the way in which the two linguistically isolated populations are genetically related to the neighboring groups. The Ladin speakers were found to be genetically close to the Italian-speaking populations and differentiated from the other Dolomitic Ladins, whereas the German-speaking Cimbri behave as an outlier, showing signatures of founder effects and low growth rate.

Journal of Human Genetics (2012) 57, 254–260; doi:10.1038/jhg.2012.14; published online 15 March 2012

Keywords: Cimbri; demography; Ladin; mitochondrial DNA; Trentino

INTRODUCTION

Human mitochondrial DNA (mtDNA) variation has been intensely scrutinized in European populations since the late 1980s.^{1–3} The considerable data accumulated so far has allowed researchers to reconstruct the phylogeography of principal lineages and make inferences concerning key events in the prehistory of European populations.⁴ However, mtDNA variation has been little studied at micro-geographic level through dense sampling^{5–7} despite the fact that exploring how genetic diversity is distributed in geographically close populations, we may complement studies which aim to identify patterns of variation on a larger scale (macro-regional, continental or global), identifying local areas of genetic discontinuity and populations with peculiar gene pools.

In this context, Trentino and its populations may be regarded as a valuable case study. In fact, its geographical complexity and linguistic diversity provide us with the opportunity to assess the role of physical and cultural factors on human genetic variation in a micro-geographic framework. It is a relatively small area (6200 km²) located in North-Eastern Italy (Trentino-Alto Adige region) and a significant portion of its territory overlaps the eastern alpine system (Figure 1). Its complex

territory, with 60% of the land located at an altitude > 1000 m above sea level, may have influenced human mobility. However, the presence of numerous valleys, the Adige river and the Garda lake has always provided communication routes, favoring population movements and interactions. Such ambivalence is also reflected in the contrast between the central-western part of Trentino (for example, Adige, Non and Sarca valleys) where flat valley bottoms are more suitable for human occupation compared with the eastern zone (for example, Primiero, Fassa valleys and Lagorai group), which is composed mostly of mountains with steep slopes and thickly wooded areas.^{8,9,10} Trentino is also characterized by the presence of groups with a noticeable linguistic diversity. It hosts the Italian-speaking communities of the romance language (~500 000 people), which speak five different dialects: *Semi-ladino*; *Trentino occidentale*; *Trentino orientale*, *Trentino centrale* and *Trentino dell'Avisio*,^{11,12} and the romance-speaking group of Ladins (belonging to the Rhaetian sub-branch) from the Fassa Valley (about 9000 people),¹³ which is part of the Dolomitic Ladins (*Ladin*) settled in other valleys of the Italian oriental Alps (Gardena and Badia in south Tyrol and Fodom and Ampezzo in Veneto).¹⁴ Finally, it hosts the German-speaking group of Cimbri who are

¹Department of Philosophy, History and Cultural Heritage, University of Trento, Trento, Italy; ²Institute of Legal Medicine, Catholic University, Rome, Italy; ³Department of Environmental Biology, Sapienza University of Rome, Rome, Italy and ⁴Italian Institute of Anthropology, Rome, Italy

Correspondence: Dr V Coia, Department of Philosophy, History and Cultural Heritage, Laboratory 'B. Bagolini', Prehistory, Medieval Archaeology and Historical Geography, University of Trento, Piazza Venezia, 41, 38100 Trento, Italy.

E-mail: valentina.coia@uniroma1.it

Received 16 May 2011; revised 21 December 2011; accepted 10 January 2012; published online 15 March 2012

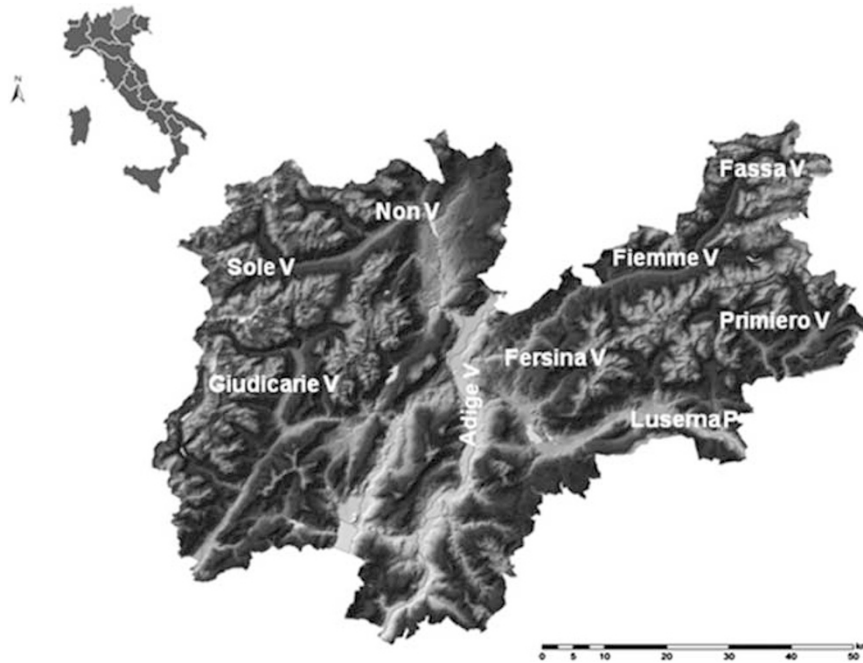


Figure 1 Geographic location of the populations from Trentino analyzed in this study. P, plateau, V, valley, Linguistic groups: Italian-speaking [*Semi Ladino* (Non and Sole valleys), *Trentino occidentale* (Giudicarie valley), *Trentino orientale* (Primiero and Fersina valleys), *Trentino centrale* (Adige valley) and *Trentino dell'Avisio* (Fiemme valley) dialects]; Ladin-speaking (Fassa valley); German-speaking (Luserna plateau).

thought to derive from the Bavarian populations that colonized a vast territory, including some Eastern Italian Valleys (*Veneto*, 1053 AD) and Lavarone, Folgaria and the Luserna plateau in Trentino¹⁵ (1216 AD). Their language (*Zimbar*) belongs to the Southern Bavarian-Austrian dialects, which today is only spoken in Luserna (300 inhabitants).¹⁶

In this study, we investigate the variation of mtDNA in nine populations from Trentino, which are distributed throughout the territory and cover most of its linguistic diversity. Our objectives are to evaluate the extent of genetic variation across the Trentino region and understand whether and how geographic factors and linguistic diversity could have played a role in determining the genetic structure and the demographic profile of the populations under study. The results obtained highlight the importance of historical, cultural and environmental knowledge when studying the genetic structure of human populations, even when focusing on small geographical areas.

MATERIALS AND METHODS

Sampling and laboratory analyses

The data set comprises 393 individuals from nine locations: Adige valley (central region), Giudicarie, Non and Sole valleys (western region), Fassa, Fersina, Fiemme and Primiero valleys and the Luserna plateau (eastern region; see Figure 1). Sampling sites are located at an average distance of 71 km. In the framework of the BIOSTRE project (Biodiversity and history of populations from Trentino: <http://laboratoriobagolini.it/en/search/projects/biostre/>), blood samples and buccal swabs were collected in apparently healthy and unrelated individuals selected according to the place of birth of the sampled individual and of their parents and grandparents (at least, three generations). The procedure and informed consent were reviewed and approved by the 'Comitato Etico per la Sperimentazione con l'Essere Umano' of the University of Trento (<http://www.unitn.it/en/ateneo/2640/university-human-research-ethics-committee>).

The DNA was extracted using the 'Nucleic Acid Isolation System' by QuickGene-810 instrument following the standard protocols for blood and swabs samples (FUJIFILM; http://products.fujifilm.eu/products/life_science/nucleic_acids/index.html).

Following the protocol of Quintans *et al.*,¹⁷ we analyzed the variation at 17 single-nucleotide polymorphisms of the mtDNA coding region (3010, 3915, 3992, 4216, 4336, 4529, 4580, 4769, 4793, 6776, 7028, 10398, 10400, 10873, 12308, 12705 and 14766) in order to classify samples in 21 principal informative haplogroups (H*, H1, H2, H3, H4, H5, H6, H7, HV0, V, HV*, J/T, J1, J2, U*, K, R*, M*, N*, I, and L3*). A region of 651 nucleotides was sequenced, (control region from 16033 to 114), encompassing the hypervariable region 1 and part of region 2 [HVR-1 (from 16033 to 16569) and HVR-2 (from 57 to 114)] according to the Cambridge Reference Sequence rCRS, GenBank accession NC_012920. A fragment of about 1.4 kb was amplified using the primers L15564 and H429, whereas L15990 and H155 were used as internal primers for sequencing (Cagri, unpublished; Supplementary Table S1). To classify the sequences belonging to the K haplogroup in the two sub-haplogroups K1 and K2, three additional positions of the HVR-2 (146, 152 and 263) were investigated by using the additional primer H429 (Supplementary Table S1). Fragment sizes were detected by the ABI PRISM 310 genetic Analyzer (Applied Biosystem, Life Technologies Corporation, Carlsbad, CA, USA).

The quality of the mtDNA data sequences was checked using the method suggested by Bandelt *et al.*¹⁸ (see Supplementary Table S2). The mtDNA sequences reported herein have been submitted to GenBank (accession numbers JQ623511-JQ623902).

Statistical analyses

To evaluate the overall level of interpopulation variation, we calculated the Fst parameter among the nine populations analyzed using the Arlequin software (version 3.5.1.2).¹⁹

The genetic relationships among populations were explored by different methods in order to detect possible genetic patterns related to geographic or linguistic diversity. First, we estimated the populations pairwise genetic distances (Fst, Reynolds' distance)²⁰ using the Arlequin software¹⁹ and the obtained matrix was visualized in a Multi-Dimensional Scaling plot²¹ obtained by using the SPSS software (release 16.0.1 for windows, SPSS Inc.). Second, we performed a Correspondence analysis, which is based on the haplogroup frequencies (see Supplementary Table S2), using the Past program (version 1.88).²² Finally, we partitioned genetic variance at different hierarchical levels of population subdivision according to geographic and language groups (see the

legend to Figure 1) by means of a molecular analysis of variance²³ using the Arlequin software.¹⁹

To quantify the correlation between geographical distances and genetic relatedness among populations, the Spearman's rho correlation value (r^2) between geographic and genetic distance matrices was obtained using the Mantel program.²⁴ Geographical distances (bidimensional matrix of the cumulative costs) were calculated as the shortest and easiest walking routes (cost surface) between one location and another chosen taking into account the morphology of the landscape (based on the DTM 40 m resolution) and the presence of natural barriers (slope inclination as a friction value) using the Grass GIS software with the algorithms *r.walk* and *r.drain*.^{23,25} This approach was preferred to traditional approaches based on the measures of the great circle or air geographic distances among populations, as it is more appropriate for studies carried out on a large geographic scale in mountainous environment.^{8,9}

The level of intrapopulation genetic variation was analyzed through the calculation of different indices, including haplotype diversity (HD), the mean number of pairwise differences (MNPDs) and the number of different haplotypes (H) using the Arlequin software.¹⁹ Furthermore, to overcome some limitations of the MNPD, which could overestimate diversity in populations that might have experienced effective size reductions (for example, bottlenecks), we also calculated the weighted mean pairwise difference within each haplogroup (weighted mean intralinear mean pairwise differences, WIMPs) as reported in Hurles *et al.*²⁶

As a final step, we applied different methods to characterize the demographic profile of populations under study. These include the Fu's F_s neutrality test²⁷ and the mismatch distribution (that is, the distribution of the number of pairwise differences between haplotypes),^{28,29} which were obtained using the Arlequin software.¹⁹ Moreover, a Bayesian approach was used in order to infer the effective population size and the growth rate in Trentino populations. A Markov Chain Monte Carlo coalescent-based algorithm implemented in the Beast software (v1.4.8)³⁰ was used and for each population we set an upper limit of 100 000 for population size and a mutation rate of 4.3×10^{-6} mutations per site per generation (25 years) as reported in Soares *et al.*³¹ A Hasegawa Kishino Yano + G model of nucleotide substitution³² was used under the demographic model (exponential growth or constant population), which seemed to be the most appropriate considering the demographic inferences obtained through the methods indicated above (see Table 2). We performed a 200 000 000 Monte Carlo sampling among possible genealogical trees (20 000 000 burn-in) to ensure that convergence had been achieved and that effective sample sizes were adequately high. Mean value, confidence intervals and median values were inferred by Tracer 1.5 software (<http://tree.bio.ed.ac.uk/software/tracer/>), the modal value was inferred with R software (ver. 2.11.1, <http://www.r-project.org/>).

RESULTS

Mitochondrial variation in Trentino populations

A total of 191 different haplotypes were detected in 393 individuals (see Supplementary Table S2). Transitions occurred in 102 out of 110 total polymorphic sites, transversions in five (16188G, 16227C, 16305T, 16317T, 16318T) and both mutation types in three sites (16129G-C, 16147C-A, 16355C-A). Moreover, one deletion (del 16162) and two insertions (+44C and +57C) were detected in one, five and two subjects, respectively. Heteroplasmy was observed in two individuals (position 16263C/T and 16327C/T).

The haplotypes were classified into 20 main haplogroups based on single-nucleotide polymorphism analysis (H*, H1, H2, H3, H4, H5, H6, H7, HV0, HV*, V, J/T, J1, J2, U*, K, R*, M*, N* and I) and were further classified into 92 sub-haplogroups (Supplementary Table S2) according to the updated phylogenetic tree of global human mtDNA³³ (Build 10, <http://www.phylotree.org>) and the HaploGrep software³⁴ (<http://haplogrep.uibk.ac.at/>). Five haplogroups (H, U, T, J and HV) occur in our data set at frequencies >5%. The prevalent mtDNA haplogroups are the same that are found in most European populations.³⁵ The haplogroup H accounts for 41% of the total sequences

Table 1 Intrapopulation diversity parameters in the nine populations analyzed

Populations	N	H (%)	HD	s.e.	MNPD	s.e.	WIMP
V. Adige	56	44 (79)	0.990	0.006	6.472	3.109	4.157
V. Fassa	47	26 (55)	0.934	0.026	5.458	2.675	2.796
V. Fersina	25	15 (60)	0.937	0.031	5.627	2.793	3.849
V. Fiemme	41	23 (56)	0.962	0.015	5.834	2.847	2.026
V. Giudicarie	52	38 (73)	0.983	0.008	6.347	3.058	2.784
P. Luserna	21	11 (52)	0.919	0.034	6.762	3.320	2.550
V. Non	48	33 (69)	0.970	0.014	5.588	2.730	3.127
V. Primiero	40	28 (70)	0.981	0.010	5.260	2.596	2.364
V. Sole	63	36 (57)	0.971	0.009	5.595	2.722	3.577

Abbreviations: H, number of different haplotypes; HD, haplotype diversity and its standard error; MNPD, mean number of pairwise differences and its standard error; N, sample size; WIMP, weighted mean intralinear mean pairwise differences; P, plateau; V, valley.

(from 35 to 41% in Europe), the haplogroup U for 16% (from 18 to 22%), whereas the haplogroup T reaches a frequency of 12% (from 9 to 11%) and J arrives at 8% (from 8 to 9%). However, the H6 haplogroup represents an interesting finding. In fact, it is rather rare in Europe (average of ~1%, recalculated from Brandstätter *et al.*³⁶) and mostly restricted to the near-East and the south Caucasus.³⁷ However, H6 is widespread in the Cimbri and the Fersina valley (~20%), where it reaches a statistically higher frequency than in the other groups from Trentino (from 0 to 4%; $P < 0.05$ in the differentiation test between all pairs of populations). Another observation worth noting regarding the near-eastern haplogroup K, which was found at frequency of 19% in the Cimbri community, a value which is very distant from those observed in the other populations from Trentino (from 0 to 5%; $P < 0.05$ in the differentiation test in four out of eight comparisons) and eastern-central European populations (from ~3% in Romania to 6% in Austria and Bavaria; recalculated from Achilli *et al.*)³⁸

Intrapopulation diversity parameters are reported in Table 1. The populations from the Adige valley and Luserna show values at the two extremes for the number of different haplotypes, H [44 (79%) versus 11 (52%)] and HD (0.990 ± 0.006 versus 0.919 ± 0.034). It is worth noting that one of the highest MNPD values was observed among the Cimbri. Interestingly, the apparent contrast between the two measures of within-population diversity for the Cimbri disappeared when MNPD was replaced by WIMP. In fact, the Cimbri show one of the lowest values of weighted mean pairwise difference within each haplogroup (2.550) among those found in the total database (from 2.364 to 4.157). This supports the notion that WIMP may be a more appropriate measure of intrapopulation genetic diversity than MNPD when dealing with populations that are subject to genetic drift (see Discussion).³⁹

Interpopulation variation

The percentage of variation among Trentino populations was found to be close to 1% ($F_{st} = 0.0115$; P -value = 0.0005). Using the jackknife procedure, this parameter reached the minimum and the maximum values when the population from Luserna and the Adige valley were removed from the data set, respectively ($F_{st} = 0.0089$ and 0.0137).

The genetic distances among populations were visualized in a bidimensional Multi-Dimensional Scaling plot (Figure 2). We can distinguish one central group composed of most populations under study (11/36 statistically significant genetic distances, see the legend to Figure 2). There are two populations that are external to this cluster; the Sole valley and, more markedly, the Cimbri. This pattern was also confirmed when the genetic relationships between populations were

analyzed by the Correspondence analysis, which is based on haplo-group frequencies (Supplementary Figure S1).

The populations analyzed were then compared with another 37 European populations (we considered the region from np 16033 to 16383 of the HVR-1; see Supplementary Figure S2 and Supplementary Table S3). Together with the populations from the Pusteria, Isarco and Gardena valleys (Ladins) from south Tyrol⁶ and the other Ladin group from Veneto⁴⁰ (eastern Italy), the Cimbri from Luserna again behave as an outlier (with 40/45 significant genetic distances with an average of $F_{st}=0.036$). The remaining Trentino populations are scattered in the central group showing a certain degree of genetic heterogeneity (14/28; average F_{st} , from 0.013 to 0.018). It is interesting to note that the Ladins from the Fassa valley analyzed in this study are genetically separated from the other Ladin groups studied by Thomas *et al.*⁶ and Vernesi *et al.*,⁴⁰ with all the pairwise genetic distances among these groups of Ladins being statistically significant. Our data set was also compared with a subset of 25 European populations, six of which, in previous studies, had been reported to be isolated (Aromuns from Albany and Macedonia,⁴¹ the Csango from Hungary,⁴² the Sardi from Urzulei⁴³ and the Ladin from the Badia and Gardena valleys in south Tyrol,⁶ Supplementary Figure S3 and Supplementary Table S3). The genetic differentiation of the Cimbri is among the highest we observed. In fact, only three populations [(Csango (CSA) Aromuns from Macedonia (Aro5) and Sardi (Sar1) in Supplementary Figure S3)]

show a higher number of significant genetic distances (24/24; average F_{st} from 0.027 to 0.041) than observed in our group (22/24; average $F_{st}=0.038$).

The analysis of molecular variance showed that the percentage of variation obtained subdividing the data set into geographic or linguistic groups (see the legend to Figure 1) was very low and not statistically significant (data not shown). Accordingly, a substantial lack of correlation between genetic and geographic distance matrices was observed ($r^2=0.0119$, $P=0.344$).

Demographic inferences

The demographic parameter estimates in the Trentino populations are reported in Table 2 and Figure 3. The F_u 's F_s and the mismatch distribution (with associated Harpending's Raggedness index value, HRI) have been shown to be sensitive to the demographic history of the populations.^{27,29} Population demographic expansion generally leads to significant large negative values of F_u 's F_s and to unimodal distribution of nucleotide differences (with lower HRI), whereas stationary populations display reduced negative values of F_u 's F_s and multimodal distribution with larger HRI.^{44,45} Based on these parameters, the populations from Luserna and Fersina valleys are close to the expectations of a stationary state (not significant F_u 's F_s and highest HRI values). On the contrary, we have observed clear signs of demographic expansion, although with different intensities, in the other populations. In fact, Adige, Giudicarie, Non, Sole and Primiero valleys display stronger signals than Fiemme and Fassa valleys. This pattern is mirrored by the estimates of effective population size, which ranged from 3293 (Luserna) to 73 217 (Adige), reaching relatively low values in the populations from Fersina, Fassa and Fiemme valleys (from 3550 to 11 006). Similarly, growth rate estimates, which were not calculated for Cimbri and Fersina because of the lack of signatures of expansion, attained a minimum of 1.8×10^{-3} and 1.9×10^{-3} in Fassa and Fiemme valleys, respectively, and the maximum of 3.5×10^{-3} in Non Valley (Table 2).

The small number of samples analyzed for the population from Luserna (21) and Fersina valley (25) compared with the other groups (from 40 to 63) may give rise to the possibility that these results could have been biased by sample size. To verify this possibility, we extracted 1000 random samples of 21 individuals for the seven populations with the highest number of samples analyzed (see Table 2), and recalculated the F_u 's F_s , as well as the HD and the MNPD parameters. The distribution values obtained for these parameters are provided in the Supplementary Figures S4, S5 and S6. Even after this procedure, the populations with the strongest signal of demographic expansion (Adige, Giudicarie, Non and Sole valleys) show statistically significant

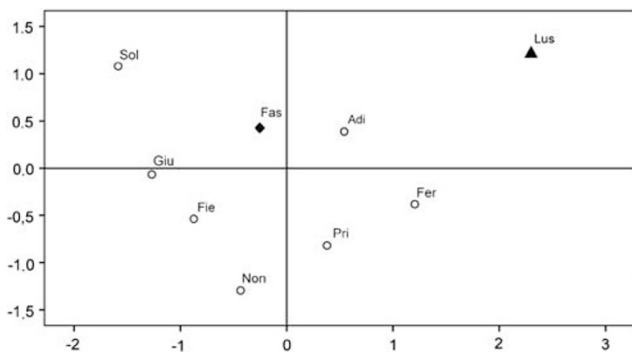


Figure 2 Multi-dimensional scaling plot (MDS) of the F_{st} genetic distances in the nine populations analyzed. Stress value: 0.167. Adi, Adige; Fas, Fassa; Fer, Fersina; Fie, Fiemme; Giu, Giudicarie; Lus, Luserna; Non, Non; Pri, Primiero; Sol, Sole. Symbols: circle (Italian-speaking), triangle (German-speaking), rumble (Ladin-speaking). Statistically significant genetic distances: Giu/Fer, Sol/Adi, Sol/Pri, Sol/Non, Sol/Fer, Sol/Lus, Lus/Pri; Lus/Giu; Lus/Non, Lus/Fie, Lus/Fas.

Table 2 Estimates of some demographic parameters in the nine populations analysed

Populations	N	F_u 's F_s	P	HRI	P	EPS	5%	95%	GR	5%	95%
V. Adige	56	-25.162	0.000	0.008	0.837	73217	38314	99998	2.6E-03	1.6E-03	3.9E-03
V. Fassa	47	-11.836	0.000	0.016	0.445	11006	4849	23958	1.8E-03	6.50E-04	3.4E-03
V. Fersina	25	-3.833	0.060	0.059	0.035	3550	1930	6161			
V. Fiemme	41	-8.576	0.004	0.017	0.447	7339	3087	17025	1.9E-03	5.20E-04	3.9E-03
V. Giudicarie	52	-25.176	0.000	0.014	0.294	27310	10856	69248	2.6E-03	1.30E-03	4.5E-03
P. Luserna	21	-0.446	0.441	0.091	0.004	3293	1724	5825			
V. Non	48	-23.082	0.000	0.013	0.552	29132	10497	80139	3.5E-03	1.68E-03	5.9E-03
V. Primiero	40	-18.680	0.000	0.027	0.058	15893	5216	47301	3.3E-03	1.40E-03	6.1E-03
V. Sole	63	-22.745	0.000	0.009	0.789	17916	8049	37144	2.1E-03	1.00E-03	3.7E-03

Abbreviations: EPS, effective population size (Mode values); F_u 's F_s , F_u 's F_s test and its P -value; GR, growth rate (Mode values); HRI, Harpending's Raggedness index and its P -value; N, samples size; P, plateau; V, valley. Non-significant P -values are marked in bold.

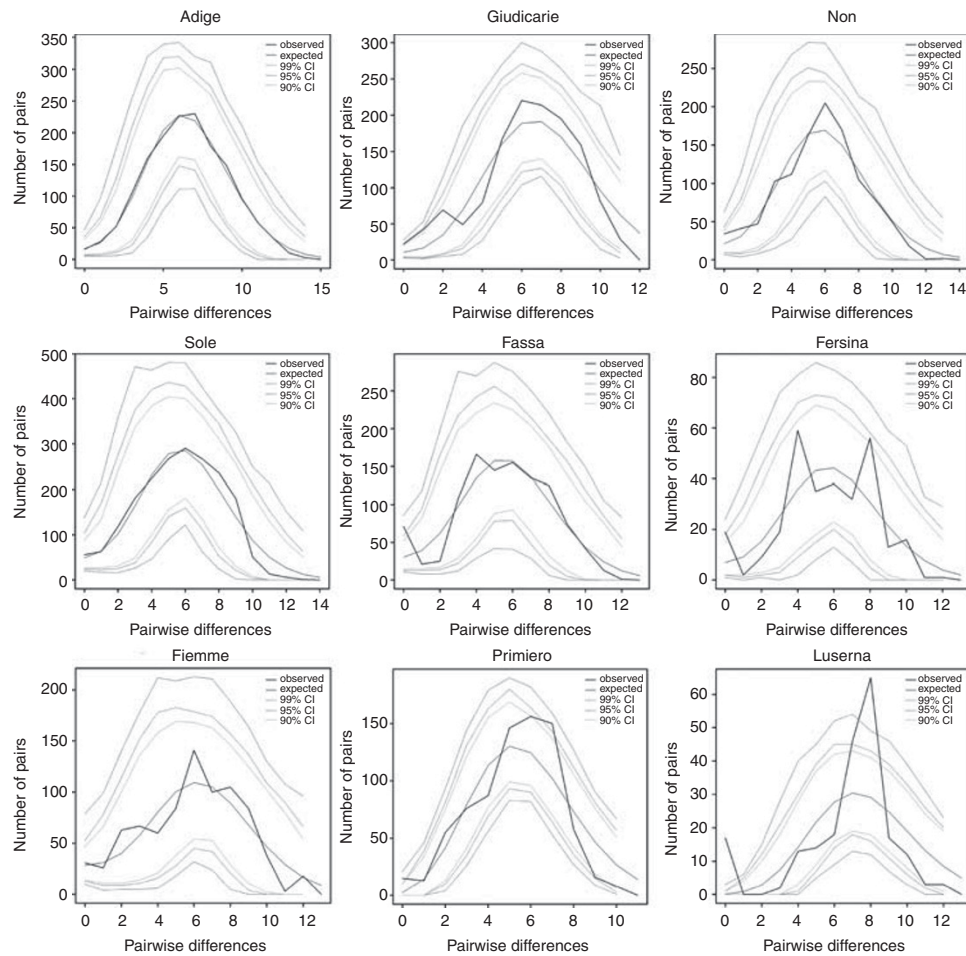


Figure 3 Mismatch distribution in the nine populations analyzed (see also Table 2).

values of F_s (Supplementary Figure S4), which makes us confident that the results obtained for Luserna and the Fersina valley are reliable.

DISCUSSION

The genetic structure of Trentino populations

Understanding the way in which genetic variation is structured at micro-geographic level is an important objective for anthropologists and human geneticists. In fact, the reconstruction of the genetic structure at a finer scale may help in gaining a better understanding of whether geographical and/or cultural factors have influenced the genetic relations among populations in a short geographic range. This is even more true for areas such as Trentino where geographical and linguistic factors could have played an important role in shaping genetic diversity.

On the whole, Trentino populations show a level of genetic differentiation, which is comparable to or even higher than widely distributed European populations. This can be easily inferred from the Multi-Dimensional Scaling of the genetic distances among 46 European populations, where the populations analyzed are dispersed in the plot (see Supplementary Figure S2). From a quantitative point of view, the mitochondrial variation among Trentino populations is higher ($F_{st}=0.0117$, recalculated considering the region spanning from np 16033 to 16383) than that obtained for nine geographically distant European populations (from $F_{st}=0.0041$ to 0.0098 using different data sets).

We could detect no statistically significant correlation between genetic and geographic distances, which indicates that the genetic differentiation among populations does not decrease consistently according to their geographic distances. This is in accordance with previous mtDNA studies conducted on a macro-geographic scale in Europe.^{46,47} Nonetheless, our study supports a role for geography in shaping the genetic structure of Trentino populations, although in a more subtle way than would be expected. In fact, through the estimates of the demographic parameters, it was possible to highlight important differences among the central western and the eastern parts of the region. The populations settled in the former area (the Adige, Giudicarie, Sole and Non valleys) show strong signs of expansion, whereas most inferences obtained for the latter zone fit a stationary demographic state (Luserna and Fersina valley) or indicate a lower degree of expansion (Fassa, Fiemme). The only exception is represented by the Primiero sample, which shows strong sign of expansion and growth rate despite it belongs to the eastern area.

These differences may be explained by geomorphological factors and also by archeological data.⁴⁸

Compared to the eastern area, the geographical settings of the central-western region is characterized by wider valleys, which are more suitable for agricultural activities, lower altitudes and more accessible mountain passes. These characteristics, as well as the remarkable presence of good quality flint sources, may explain why

the right side of the Adige valley has been densely populated since ancient times, as indicated by the abundance of archeological records in the Adige, Non and Giudicarie valleys.⁴⁸ Accordingly, considerable archeological evidence dated from Paleolithic to protohistoric times has been recovered, leading scholars to suggest that human groups were constantly present during the Holocene all along the Adige valley.⁴⁸ Moreover, the presence of exogenous raw materials (for example, obsidian, green stone), ornaments or symbolic items (for example, the shapes and decoration styles of the pottery), indicates contacts between communities located in the Adige valley and those settled on the Po plain and/or in the northern Alps.⁴⁸ By contrast, geomorphological features have always represented an obstacle to the massive peopling of the eastern area. In fact, the Valsugana Valley is the only densely populated area in this part of the territory.⁴⁹

Linguistic diversity was found to be a poor predictor of mitochondrial genetic diversity in Trentino, a fact that has been observed on a continental scale,⁵⁰ contrasting with previous studies of autosomal variation,^{51,52} whereas the analysis of the European Y-chromosome gave conflicting results.^{53,54}

This is particularly evident when comparing the relationships between the Ladins from the Fassa valley and the Cimbri from Luserna with other populations from Trentino and Europe. The Ladins show a substantial genetic affinity with the Italian-speaking groups from Trentino. In fact, they are separated from the neighbor population from Fiemme valley by their lowest pairwise *F_{st}* genetic distance (0.006), and share the highest number of haplotypes (9). Unfortunately, groups that are geographically close to Ladins have not been sampled in previous studies,^{6,40} which hampers any comparison with the results obtained in the course of the present investigation. However, the genetic distance between Ladins from Fassa valley and those previously studied is relatively high [(from 0.02 (with Ladins from Badia valley, LadB) to 0.08 (with Ladins from Veneto, LadV), see Supplementary Figure S2)]. Furthermore, the number of haplotypes shared by Ladins from Fassa valleys and those from other alpine areas is relatively low (four with LadB, four with LadG and zero with LadV) and only one haplotype is shared simultaneously among all Ladin groups. This is consistent with the high diversity among Dolomitic Ladin communities reported by Thomas *et al.*⁶ On the whole, the data accumulated so far suggest that the genetic heterogeneity among Ladins could be due to the combined effect of a scarce reciprocal genetic exchange, probably due to geographic isolation, and genetic drift.

The Cimbri seem to be set apart from all the other populations, both from Trentino and Europe, including linguistically related German groups from Bavaria (Supplementary Figure S2). Considering our data set, the Cimbri show the lowest values of the parameters, which describe within-population genetic variation (HD, number of haplotypes and WIMPs). They also show one of the lowest levels of HD among 60 other European populations (see Supplementary Table S3). Finally, they lack three of the most frequent European lineages (H*, H1 and H3) while they bear rare haplogroups at remarkably high frequencies (H6 and K1, see Supplementary Table S2).

Certainly, one should consider that the small number of individuals analyzed for the Cimbri, the lowest of our data set, could have biased the result. However, applying the resampling procedure we showed that our results cannot be simply accounted for by a low sample size. Moreover, we sampled a portion of 7% of the current population of Luserna, the highest value for the data set (from 0.05% for the Adige valley to 2.5% for the Fersina valley, see Supplementary Table S4). Furthermore, a peculiar mitochondrial profile is not to be unexpected for this group, in which genetic drift could have played an important

role in shaping the genetic structure in two different ways. First, the Cimbri from Luserna could have experienced a founder effect, a possibility indicated by historical sources that trace the origin of the community back (XIII century) to a few families that moved from Lavarone, the first Trentino area that was colonized by the Cimbri, to the Luserna plateau.¹⁵ Moreover, the community has always been characterized by a very small number of individuals, which reached a modest peak of 1005 inhabitants in the 1920s.¹⁵ Finally, cultural isolation could have increased the differentiation between the Cimbri and the other populations, a possibility supported by some lexical data, which indicate that the Cimbri language has maintained some words (for example, Libar) that have been taken from the neighboring romance-speaking populations during the first contacts (XIII century).⁵⁵

In conclusion, through this study, we highlighted some important aspects of the genetic structure of populations of Trentino. The genetic data obtained in the course of this study made it possible to infer different demographic patterns in two longitudinal areas, which are seemingly related to geomorphological factors. Furthermore, although no evident relation between linguistic and genetic diversity was observed, the two linguistic isolates, Ladins and Cimbri, proved to be the key populations when drawing inferences of anthropological significance. The different way in which Ladins and Cimbri are genetically related to the populations from Trentino is of particular interest. In fact, this evidence points to the importance of carrying out a systematic study among the linguistic isolates of the Italian Alps to understand how they fit into the surrounding genetic background and identify the signatures of their original genetic ancestry.^{56–58}

ACKNOWLEDGEMENTS

We are greatly indebted to the blood donors, whose participation made this research possible. We are also indebted to the medical and nursing staff of the Azienda Provinciale per i Servizi Sanitari and in particular to Drs Massimo Ripamonti and Alberto Marcolla. We are also grateful to the volunteers of the Associazione Volontari Italiani Sangue del Trentino (AVIS), to Biblioteca Intercomunale di Primiero, Istituto Culturale Ladin, Kulturinstitut Luserna and to all the other people who kindly contributed to the collection of the samples (<http://laboratoriobagolini.it/ricerca/progetti/biostre/ringraziamenti/>). We also thank Professors Patrizia Cordin, Ermenegildo Bidese and Enrico Cavada for their suggestions regarding the linguistic and historical aspects, and Paolo Anagnostou for his technical support and advice concerning the paper. This study was financed by the Autonomous Province of Trento (Provincia Autonoma di Trento; bando post-doc 2006; BIOSTRE project to V.C.). We also received support from the Ministero della Ricerca Scientifica (Progetti di Ricerca di interesse nazionale 2007-2009, prot.n. 2007TYXE3X; 2009-2011, prot.n. 200975T9EW) and the Italian Institute of Anthropology (Isita, project "Atlante Bioculturale Italiano"). V.C. designed the research and collected the samples; V.C., I.B. and F.T. performed the experiments; V.C. analyzed the molecular data together with F.M., who did the computation using the Beast program, and F.C., who calculated the geographic distances; V.C. together with G.D.B., S.G. and A.P. wrote the paper. All authors have read and approved the final manuscript.

- 1 Brega, A., Scozzari, R., Maccioni, L., Iodice, C., Wallace, D. C., Bianco, I. *et al.* Mitochondrial DNA polymorphisms in Italy. I. Population data from Sardinia and Rome. *Ann. Hum. Genet.* **50**, 327–38 (1986).
- 2 Santachiara Benerecetti, A. S., Scozzari, R., Semino, O., Torrioni, A., Brega, A. & Wallace, D. C. Mitochondrial DNA polymorphisms in Italy. II. Molecular analysis of new and rare morphs from Sardinia and Rome. *Ann. Hum. Genet.* **52**, 39–56 (1988).
- 3 Torrioni, A., Achilli, A., Macaulay, V., Richards, M. & Bandelt, H. J. Harvesting the fruit of the human mtDNA tree. *Trends Genet.* **22**, 339–345 (2006).
- 4 Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bandelt, H. J. *et al.* The archaeogenetics of Europe. *Curr. Biol.* **20**, 174–83 (2010).

- 5 Fraumene, C., Petretto, D., Angius, A. & Pirastu, M. Striking differentiation of sub-populations within a genetically homogeneous isolate (Ogliastra) in Sardinia as revealed by mtDNA analysis. *Hum. Genet.* **114**, 1–10 (2003).
- 6 Thomas, M. G., Barnes, I., Weale, M. E., Jones, A. L., Forster, P., Bradman, N. et al. New genetic evidence supports isolation and drift in the Ladin communities of the South Tyrolean Alps but not an ancient origin in the Middle East. *Eur. J. Hum. Genet.* **16**, 124–34 (2008).
- 7 Pichler, I., Mueller, J. C., Stefanov, S. A., De Grandi, A., Volpato, C. B., Pinggera, G. K. et al. Genetic structure in contemporary south Tyrolean isolated populations revealed by analysis of Y-chromosome, mtDNA, and Alu polymorphisms. *Hum. Biol.* **78**, 441–64 (2006) Erratum in *Hum. Biol.* **7**, 641 (2006).
- 8 Cavulli, F. & Grimaldi, S. *To see or not To see. Archaeological data and surface visibility as seen by an AIS (Archaeological Information System) approach. 2007* In: A. Figueiredo & G. Velho (eds), 'The World is in your eyes', Proceedings of the XXXIII Computer Application and Quantitative Methods in Archaeology Conference (March 2005 –Tomar: Portugal, 413–420 2005).
- 9 Cavulli, F. & Grimaldi, S. in *Mesolithic Horizons. Papers presented at the Seventh International Conference on the Mesolithic in Europe* (eds McCartan, S., Schulting, R., Warren, G., & Woodman, P.). 96–101 (Oxbow Books, Belfast, 2005).
- 10 Cavulli, F., Grimaldi, S., Pedrotti, A. & Angelucci, D. E. in *Hidden Landscapes of Mediterranean Europe. Cultural and Methodological Biases in Pre- and Protohistoric Landscape Studies* (eds van Leusen, M., Pizzolo, G., Sarti, L.). Proc. Int. Meeting; Siena, Italy; 25-27 May 2007. British Archaeological Reports, International Series 2320 (Archeopress, Oxford, 2011).
- 11 Bertoluzza, A. *Atti del II Convegno Sui Dialetti Del Trentino. 18-20 Ottobre 1991* (Centro culturale "Fratelli Bronzetti," Trento, 1992).
- 12 Cordin, P. *L'Archivio lessicale dei dialetti trentini* (Università di Trento, Dipartimento di scienze filologiche e storiche: Trento, 2005).
- 13 PAT. *Provincia autonoma di Trento Trentino in numeri 2009* (Servizio Statistica: Trento, 2009).
- 14 Belardi, W. *Profilo storico-politico della lingua e della letteratura ladina* (Il calamo: Roma, 1994).
- 15 Moz, N A *Luserna. Terra di uomini liberi* (Rovereto: Osiride, 2001).
- 16 PAT. *Trentino in numeri 2008* (Provincia autonoma di Trento: Servizio Statistica, 2008).
- 17 Quintáns, B., Alvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M. V. & Carracedo, A. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci. Int.* **140**, 251–7 (2004).
- 18 Bandelt, H. J., Quintana-Murci, L., Salas, A. & Macaulay, V. The fingerprint of phantom mutations in mitochondrial DNA data. *Am. J. Hum. Genet.* **71**, 1150–60 (2002).
- 19 Excoffier, L., Laval, G. & Schneider, S. Arlequin ver. 3.1: an integrated software package for population genetics data analysis. *Evol. Bioinform. Online* **1**, 47–50 (2005).
- 20 Reynolds, J. Estimation of the coancestry coefficient. Basis for a short-term genetic distance. *Genetics* **105**, 767–799 (1983).
- 21 Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a non metric hypothesis. *Psychometrika* **29**, 1–27 (1964).
- 22 Hammer, Ø., Harper, D. A. T. & Ryan, P. D. PAST: paleontological statistical software package for education and data analysis. *Palaeontologia Electronica* **4**, 1–4 (2001).
- 23 Neteler, M & Mitasova, H. *Open Source GIS. A GRASS GIS Approach* (The International Series in Engineering and Computer Science) vol **773**, 406 (Springer: New York, 2008).
- 24 Relethford, J. H. Mantel. *A Microcomputer Program for Computing the Mantel Probability Between Distance Matrix Elements* (State University of New York, College at Oneonta, Department of Anthropology: Oneonta, NY, 1990).
- 25 Conolly, J & Lake, M. W. *GIS in Archaeology* (Cambridge University Press: Cambridge, 2006).
- 26 Hurles, M. E., Nicholson, J., Bosch, D., Renfrew, C., Sykes, B. C. & Jobling, M. A. Y chromosomal evidence for the origins of oceanic-speaking peoples. *Genetics* **160**, 289–303 (2002).
- 27 Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925 (1997).
- 28 Rogers, A. R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
- 29 Harpending, H. C. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**, 591–600 (1994).
- 30 Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
- 31 Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A. et al. Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–59 (2009).
- 32 Hasegawa, M., Kishino, H. & Yano, T. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–74 (1985).
- 33 Van Oven, M. & Kayser, M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum. Mutat.* **30**, 386–394 (2009).
- 34 Kloss-Brandstatter, A., Pacher, D., Schoenherr, S., Weissensteiner, H., Binna, R., Specht, G. et al. HaploGrep a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* **32**, 25–32 (2011).
- 35 MITOMAP. A Human Mitochondrial Genome Database, <http://www.mitomap.org> (2011).
- 36 Brandstätter, A., Zimmermann, B., Wagner, J., Göbel, T., Röck, A. W., Salas, A. et al. Timing and deciphering mitochondrial DNA macro-haplogroup R0 variability in Central Europe and Middle East. *BMC Evol. Biol.* **8**, 191 (2008).
- 37 Roostalu, U., Kutuev, I., Loogväli, E. L., Metspalu, D., Tambets, K., Reidla, M. et al. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in West Eurasia: the Near Eastern and Caucasian perspective. *Mol. Biol. Evol.* **24**, 436–448 (2007).
- 38 Achilli, A., Olivieri, A., Pala, M., Metspalu, D., Fornarino, S., Battaglia, V. et al. Mitochondrial DNA variation of modern Tuscans supports the near eastern origin of Etruscans. *Am. J. Hum. Genet.* **80**, 759–768 (2007).
- 39 Alfonso-Sánchez, M. A., Cardoso, S., Martínez-Bouzas, C., Peña, J. A., Herrera, R. J., Castro, A. et al. Mitochondrial DNA haplogroup diversity in Basques: a reassessment based on HVI and HVII polymorphisms. *Am. J. Hum. Biol.* **20**, 154–164 (2008).
- 40 Vernesi, C., Fuselli, S., Castri, L., Bertorelle, G. & Barbujani, G. Mitochondrial diversity in linguistic isolates of the Alps: a reappraisal. *Hum. Biol.* **74**, 725–730 (2002).
- 41 Bosch, D., Calafell, F., González-Neira, A., Flaiz, C., Mateu, D., Scheil, H G. et al. Paternal and maternal lineages in the Balkans show a homogeneous landscape over linguistic barriers, except for the isolated Aromuns. *Ann. Hum. Genet.* **70**, 459–487 (2006).
- 42 Brandstätter, A., Egyed, B., Zimmermann, B., Duftner, N., Padar, Z. & Parson, W. Migration rates and genetic structure of two Hungarian ethnic groups in Transylvania, Romania. *Ann. Hum. Genet.* **71**, 791–803 (2007).
- 43 Fraumene, C., Belle, E. M., Castri, L., Sanna, S., Mancosu, G., Cosso, M. et al. High resolution analysis and phylogenetic network construction using complete mtDNA sequences in sardinian genetic isolates. *Mol. Biol. Evol.* **23**, 2101–2111 (2006).
- 44 Schneider, S. & Excoffier, L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**, 1079–1089 (1999).
- 45 Excoffier, L. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Mol. Ecol.* **13**, 853–864 (2004).
- 46 Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. & Barbujani, G. Geographic patterns of mtDNA diversity in Europe. *Am. J. Hum. Genet.* **66**, 262–278 (2000).
- 47 Barbujani, G., Stenico, M., Excoffier, L. & Nigro, L. Mitochondrial DNA sequence variation across linguistic and geographic boundaries in Italy. *Hum. Biol.* **68**, 201–215 (1996).
- 48 Lanzinger, M., Marzatico, F. & Pedrotti, A. *Storia del Trentino. Vol I. La preistoria e la protostoria* (Il Mulino: Bologna, 2000).
- 49 Cavada. *Storia del Trentino. Vol II. L'età romana. Il territorio: Popolamento, abitati, necropoli* (Il Mulino: Bologna, 2000).
- 50 Sajantila, A., Lahermo, P. & Anttinen, T. Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.* **5**, 42–52 (1995).
- 51 Barbujani, G. & Sokal, R. R. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc. Natl Acad. Sci. USA* **87**, 1816–1819 (1990).
- 52 Cavalli-Sforza, L. L., Minch, E. & Mountain, J. L. Coevolution of genes and languages revisited. *Proc. Natl Acad. Sci. USA* **89**, 5620–5624 (1992).
- 53 Poloni, E. S., Semino, O., Passarino, G., Santachiara-Benecetti, A. S., Dupanloup, I., Langaney, A. et al. Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am. J. Hum. Genet.* **61**, 1015–1035 (1997).
- 54 Rosser, Z. H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A. et al. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* **67**, 1526–1543 (2000).
- 55 Gamillscheg, E. *Studien zur Vorgeschichte einer romanischen Tempuslehre* (A. Hölder Wien, 1913).
- 56 Destro-Bisol, G., Jobling, M. A., Rocha, J., Novembre, J., Richards, M. B., Mulligan, C. et al. Molecular anthropology in the genomic era. *J. Anthropol. Sci.* **88**, 93–112 (2010).
- 57 Destro Bisol, G., Anagnostou, P., Batini, C., Battaglia, C., Bertoincini, S., Boattini, A. et al. Italian isolates today: geographic and linguistic factors shaping human biodiversity. *J. Anthropol. Sci.* **86**, 179–188 (2008).
- 58 Boattini, A., Griso, C. & Pettener, D. Linguistic versus genetic isolation. The strange case of the Walser from Upper Lys Valley (Italian Western Alps). *J. Anthropol. Sci.* **89**, 161–175 (2011).

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