J. Biol. Res. 2012 n.1 - Vol. LXXXV - Rubbettino - Soveria Mannelli

A Y Variant Which Traces the Genetic Heritage of *Ligures* Tribes

S. Bertoncini[1], G. Ferri[2], G. Busby[3], L. Taglioli[1], M. Alù[2], C. Capelli[3], G. Paoli[1], S. Tofanelli[1]

[1] Dipartimento di Biologia, Università di Pisa, Italia
[2] Dipartimento ad Attività Integrata di Laboratori Anatomia Patologica e Medicina Legale, Università di Modena e Reggio Emilia, Italia
[3] Department of Zoology, University of Oxford, UK. E-mail: stofanelli@biologia.unipi.it

KEY WORDS: *Ligures*, Y chromosome, S28/U152, Urnfield culture.

Introduction

About the current debate concerning the expansion of the major European Y chromosome lineage, R-M269, Busby *et al.* (2011) showed that the estimated time of origin and related demographic reconstructions are strongly influenced by the choice of the mutation rate, the panel of STR markers and the initial population size. For this reason, it is currently an hard task to discriminate prehistoric from historical contributions.

Nonetheless, there is substantial agreement that the northwest–southeast frequency cline of R-M269 chromosomes doesn't hold when considering the frequency map of some R-M269 sub-lineages, whose spatial distribution is somewhat radial and centered in different areas. When the R-M269 variation is apportioned into subclades, the distribution pattern of each Y lineage is more easily associated" to either a demographic events or to the archeological/historical/linguistic evidence.

The distribution map of the R-S28/U152 sub-lineage covers Central Europe with frequency peaks around the Western Alps and Northern Thyrrenum (Fig. 1). Discordant molecular dates for the origin and expansion of R-S28/U152 mutation have been estimated from academic and genealogical data ranging from 11,400 and 3,700 years BP (in Faux 2011). As well, a number of populations have been claimed to be the vector which spread R-S28/U152 chromosomes over Europe: Celts, Flemings, Vikings, Angle, Cimbri, Illyrians among others.

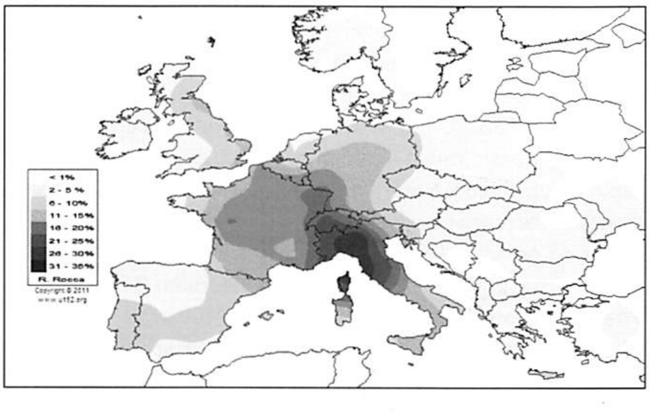

Fig. 1. *Maps of distribution of R-S28/U152 lineage in Europe (from U152.org).*

With the aim to further define the demographic and cultural framework of R-S28/U152 diffusion we investigated the STR variability in R-S28/U152 Y chromosomes from native donors residing in the territories once inhabited by *Ligures Apuani* tribes (Garfagnana and Lunigiana valleys, North Italy) and where *Ligures* would have been massively deported under the Roman rule according to classic historians (Sannio - South Italy). A better knowledge of R-S28/U152 variability will provide scholars with a key marker to interpret movements of populations and ideas out and across the Alpine area.

Materials and Methods

We collected buccal swabs in four isolated villages of the Apuan Alps (North Tuscany):

Vagli (N = 23) and Arni (N = 11) from the Garfagnana valley side; Pruno (N = 9) and Terrinca (N = 13) from the Thyrrenian side (Fig. 2). Samples from each village were selected among male unrelated donors (according to the three generations rule) bearing one of the founding surnames recorded on parishes registers since the XVI

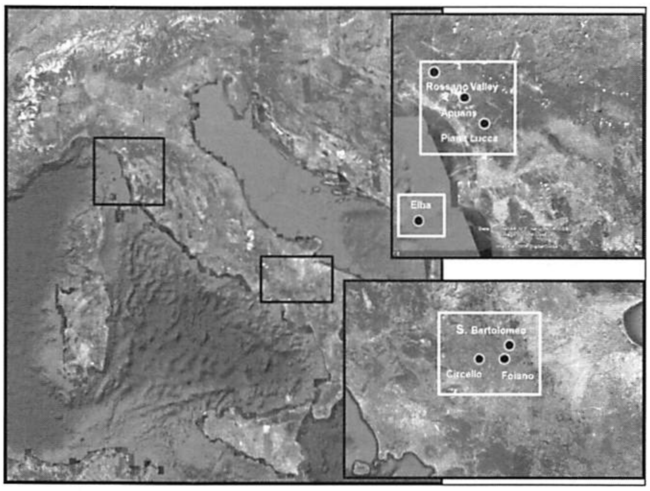

Fig. 2. *Map of sample locations.*

century. We also collected samples from isolated hamlets in the Lunigiana valley (Rossano, N = 14), from Lucca surroundings (N = 51), from Elba island (N = 95) and from isolated villages in the Sannio (Circello N = 34, Foiano N = 13, San Bartolomeo N = 19) (Fig. 2).

Genomic DNA was extracted and genotyped at 17-locus Y-STR using the AmpFl STR ® Yfiler ® kit (Applied Biosystems, USA) and at 43 Y-SNPs using the Snapshot Multiplex Kit (Applied Biosystems, USA). In particular, we genotyped by two mini-sequencing reactions 10 SNPs downstream to M269 (L51/S167, L11/S127, U106/S21, S116/P312, M167/SRY2627, S28/U152, M126, M160, L2/S139, L21/S145). In order to minimize the statistical bias we selected 10 STR loci upon criteria of information content and evolutionary stability: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439.

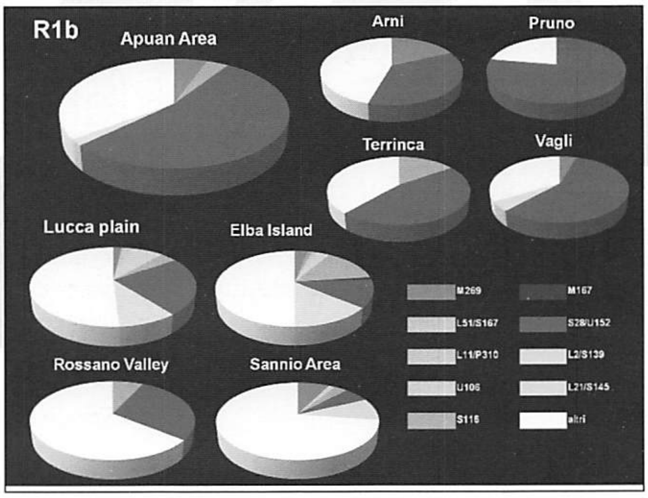

Fig. 3. *R-M269 downstream lineages in the sampled populations.*

Statistical analysis

Genetic distances were calculated using Arlequin 3.5.1.2. (Excoffier *et al.*, 2007). The relationship between populations was assessed by pairwise Fst distances based on haplotype frequencies, graphically represented through a multidimensional scaling (MDS) analysis performed with STATISTICA 7.0 (StatSoft, Inc., Tulsa, OK, USA).

The spatial autocorrelation analysis, used to describe the relationships between data observed at different geographic locations, was computed by SAAP (Wartenberg, 1989) and AIDA software (Bertorelle and Barbujani 1995). TMRCAs were calculated by the Network (http://www.fluxus-engineering.com/) and the BATWING software (http://www.maths.abdn.ac.uk/). The divergence between the S28/U152 high/low frequency groups under a "split and growth" model was simulated by ASHEs (http://ashes.codeplex.com/) according to Tofanelli *et al.* (2009).

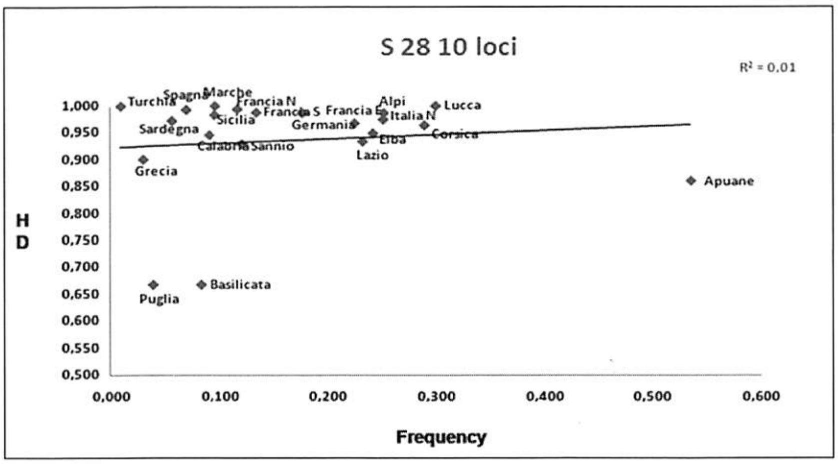

Fig. 4. *Correlation between frequency/haplotype diversity of R-S28 lineage.*

Results and Discussion

The haplogroup profiles show that S28/U152 chromosomes are predominant in the Apuan area, irrespective of the valley side (Fig. 3). Averaged (51,8%) and individual (Pruno 77,8%) frequency values are by far the highest found so far. Only in the mountainous area of Northern Tuscany the S28/U152 variant sums barely to the totality of R-M269 positive Y chromosomes and its L2/S139 subclade is poorly represented. This suggests that surviving Y chromosomes of the *Ligures* deported in the Sannio by the Romans, if any, are diluted in the current male pool.

To better understand the distribution of the S28/U152 lineage, frequency/diversity values have been plotted on a graph (Fig. 4). The lack of significant correlation between the two variables denotes no clear-cut directions of the migrations. The defect of variability showed by "Puglia" and "Basilicata" might be due to the low sample size whereas that of "Apuans" is most likely the consequence of long-term drift or founder effects.

The MDS space (Fig. 5) groups genetically homogeneous but geographically heterogeneous populations characterized by low-frequency of S28/U152 and places a number of satellite populations characterized by high-frequency of S28/U152 gravitating around it. Such a pattern supports a basic radial-like demographic expansion of this haplogroup. In some marginal areas, however, centers of genetic differentiation have arisen. One of these centers seems to coincide with the territories inhabited by the tribes of the *Ligures Apuani* in the I millennium BCE.

The spatial autocorrelation analysis (Moran's I, Fig. 6) gave a bimodal distribution with an intermediate depression. It suggests two plausible demographic scenarios: 1) the formation of two groups by two independent events of spatial expansion; 2) a "surfing" effect during a single event of spatial expansion, namely the increase of a mutation by its serial propagation on the edge of a spatial wave of expansion (Excoffier and Ray, 2008).

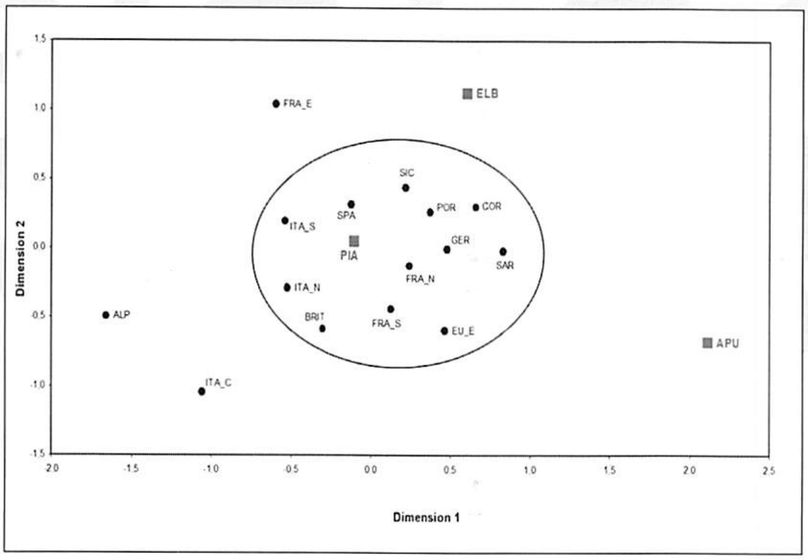

Fig. 5. *MDS based on a Fst distance matrix calculated on 10-locus Y-STR haplotypes R-S28/U152. Stress:0,169.*

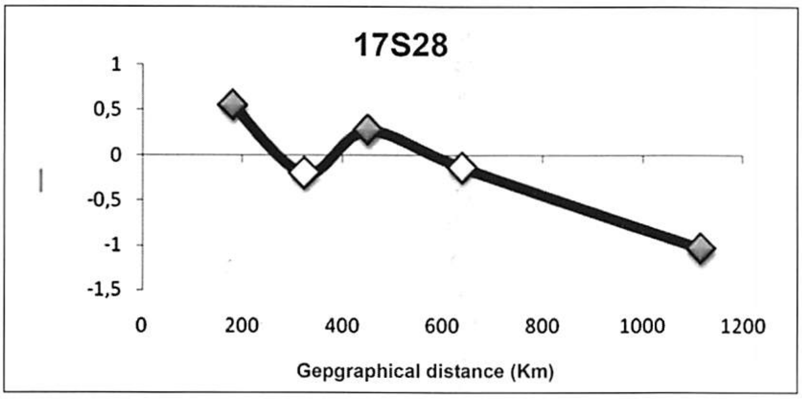

Fig. 6. *Spatial autocorrelation (17- locus Y-STR).*

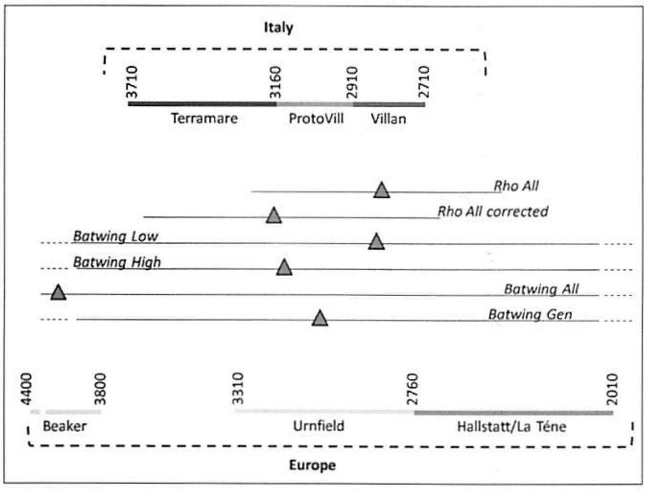

Fig. 7. *Estimated times since the most common ancestor for R-S28/U152 10-locus haplotypes. Rho all: 2,834 ± 396 (3,230-2,438 years ago) (1,220-428 BCE); Rho all correct: 3,174 ± 478 (3,652-2,696 years ago) (1,642-686 BCE); Batwing low: 2,825 ± 6,372 years ago (815 BCE); Batwing high: 3,112 ± 3,062 years ago (1,102 BCE); Batwing all: 4,257 ± 6,192 years ago (2,247 BCE).*

It has been verified by extensive simulations (Klopfstein *et al.*, 2006) that there is a negative relationship between the bearing capacity of an area, the size of the demes and the probability that a neutral mutation moves within the wave of expansion reaching high frequencies.

The fact that the highest frequencies of the S28/U152 mutation are observed in populations that have a long history of adaptation to mountain environments, with low bearing capacity and low effective size (200-500 units) is highly compatible with the latter scenario.

To give insights on the time since the S28/U152 expansion, TMRCAs were calculated both using a locus by locus average mutation rate (2.78×10^{-3} mut/locus/gen) following Ballantyne *et al.* (2010) (Fig. 7). Most of the estimations, ranging from 1,300 to 750 BCE, fall within the Urnfield period, an aboriginal Central European culture of the Middle Bronze Age which introduced to Italy the practice of body incineration.

Lastly, we simulated the divergence between the S28/U152 high/low frequency groups under a "split and growth" model (Fig. 8). More precisely, we analyzed how the DHS genetic distance varies in the Urnfield generation time window depending on the initial effective size N_e of the two groups. The results support the "surfing" theory.

The observed DHS values for the group at higher frequency is compatible with a small initial N_e (N_0 = 500), while those for the group at lower frequencies is consistent with three times as large deme size (N_0 = 1500).

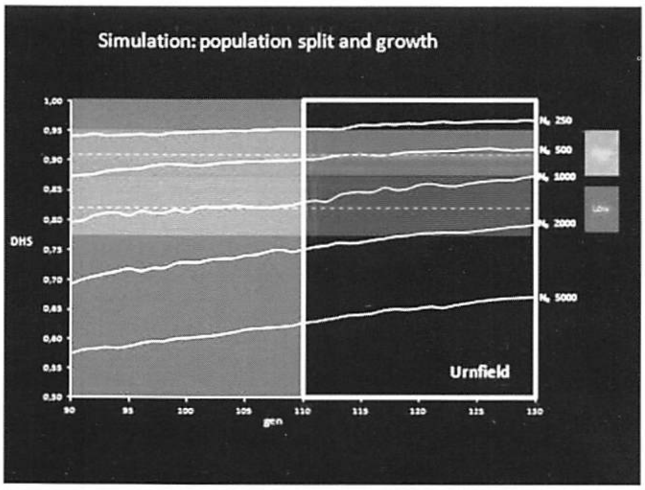

Fig. 8. *Simulated and observed DHS values ±SD. Fixed simulation parameters were used (I = 0.9 M = 1.02 $H_0 = N_0/2$ with N_0 variable).*

Conclusions

The R-S28 mutation represents a key component of the current European genetic landscape, in particular of the pre-Roman Italian landscape. The main contribute

to the current distribution in cisalpine regions can be more confidently attributed to population movements occurred along with the diffusion of the Urnfield culture (3,300 - 2,750 BP) rather than to Paleolithic, Neolithic and Chalcolithic migrations, as previously suggested. Here, a surfing effect may have increased the frequency of this mutation in a limited area (Western Alps and Upper Tyrrhenum), due to the fragmentation of the populations in tribes with a mixed economy (proto-Villanova), who refunded the communities previously settled in the same area (Terramare) and colonized new valleys.
According to this new perspective the Proto-Celtic people of the Canegrate (13th century BCE) and Proto-Golasecca (12th-10th centuries BCE) cultures can be considered the best candidates for of the diffusion.

Acknoledgements

This research was funded by PRIN2007 and PRIN2009 grants to G.P and S.T.

References

Ballantyne K.N., Goedbloed M., Fang R., Schaap O., Lao O., Wollstein A., Choi Y., Van Duijn K., Vermeulen M., Brauer S., Decorte R., Poetsch M., Von Wurmb-Schwark N., De Knijff P., Labuda D., Vézina H., Knoblauch H., Lessig R., Roewer L., Ploski R., Dobosz T., Henke L., Henke J., Furtado M.R., Kayser M. 2010. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.*, 87: 341-353.

Bertorelle G., Barbujani G. 1995. Analysis of DNA diversity by spatial autocorrelation. *Genetics*, 140: 811-819.

Busby G.B., Brisighelli F., Sánchez-Diz P., Ramos-Luis E., Martinez-Cadenas C., Thomas M.G., Bradley D.G., Gusmão L., Winney B., Bodmer W., Vennemann M., Coia V., Scarnicci F., Tofanelli S., Vona G., Ploski R., Vecchiotti C., Zemunik T., Rudan I., Karachanak S., Toncheva D., Anagnostou P., Ferri G., Rapone C., Hervig T., Moen T., Wilson J.F., Capelli C. 2011. The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. *Proc. Biol. Sci.* In press.

Excoffier L., Laval G., Schneider S. 2007. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evol. Bioinform. Online*, 1: 47-50.

Excoffier L., Ray N. 2008. Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol. Evol.*, 23: 347-351.

Faux K.D. 2011. *Y-chromosome marker S28/U152.* Resource page. http://www.davidkfaux.org/R1b1c10_Resources.pdf

Klopfstein S., Currat M., Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. *Mol. Biol. Evol.*, 23: 482-490.

Tofanelli S., Bertoncini S., Castrì L., Luiselli D., Calafell F., Donati G., Paoli G. 2009. On the origins and admixture of Malagasy: new evidence from high-resolution analyses of paternal and maternal lineages. *Mol. Biol. Evol.*, 26: 2109-2124.

Wartenberg, D.E., 1989. *SAAP: A Spatial Autocorrelation Analysis.* Program. Version 4.3. Exeter Software, Setauket, New York.