

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Allele and haplotype diversity of 12 X-STRs in Sardinia

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1660961> since 2020-02-28T17:01:19Z

Published version:

DOI:10.1016/j.fsigen.2017.12.002

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

1 **Allele and haplotype diversity of 12 X-STRs in Sardinia**

2

3 **Robino C^{a*}, Lacerenza D^a, Aneli S^{b,c}, Di Gaetano C^{b,c}, Matullo G^{b,c}, Robledo R^d, Calò C^e**

4

5 ^a Dipartimento di Scienze della Sanità Pubblica e Pediatriche, Università di Torino, Italy

6 ^b Dipartimento di Scienze Mediche, Università di Torino, Italy

7 ^c Italian Institute for Genomic Medicine, Torino, Italy

8 ^d Dipartimento di Scienze Biomediche, Università di Cagliari, Italy

9 ^e Dipartimento di Scienze della Vita e dell'Ambiente, Università di Cagliari, Italy

10 * **Corresponding author:** Dipartimento di Scienze della Sanità Pubblica e Pediatriche, Università
11 di Torino, Corso Galileo Galilei 22, 10126 Torino, Italy

12 e-mail: carlo.robino@unito.it (C. Robino)

13

14 **Abstract**

15 The analysis of clusters of tightly linked X-chromosome short tandem repeat (STR) markers can
16 assist the interpretation of complex kinship cases. However, when linkage disequilibrium (LD) is
17 present in the population of origin of tested individuals, haplotype rather than allele frequencies
18 should be used in likelihood calculations. The diversity of twelve X-STRs arranged in four linkage
19 groups (I: DXS10148-DXS10135-DXS8378; II: DXS7132-DXS10079-DXS10074; III: DXS10103-
20 HPRTB-DXS10101; IV: DXS10146-DXS10134-DXS7423) was tested in a Sardinian population
21 sample (n=516) including three open populations from the Northern, Central and Southern part of
22 the island, and three isolates (Benetutti, Desulo, Carloforte). Evidence of LD was detected in
23 Sardinia within each linkage group. Significant differences in haplotype and allele frequency
24 distribution of X-STR markers was seen between isolates and open populations, which on the
25 contrary appeared highly homogeneous.

26 The percentage of Sardinian haplotypes previously unobserved in a similar dataset compiled for the
27 Italian population was: 76.3% (linkage group I), 61.3% (linkage group II), 54.1% (linkage group
28 III), 58.9% (linkage group IV). Significant heterogeneity in haplotype distribution between
29 Sardinians and mainland Italians was observed at linkage group IV.

30 The study confirms the presence of high levels and complex patterns of LD along the X
31 chromosome in Sardinia, and provides population-specific haplotype data for biostatistical
32 evaluation in kinship testing.

33

34 **Keywords**

35 Sardinia, X chromosome, X-STR, linkage disequilibrium, kinship testing

36

37 **1. Population**

38 Sardinia is the second largest island in the Mediterranean Sea, with a population of over 1.6 million
39 according to the latest Italian census [1]. A combination of geographic, historical and environmental
40 factors contributed to the substantial differentiation of modern Sardinians from European
41 continental populations, testified by distinctive archaeological, linguistic, and cultural records [2].
42 Sardinia is also a well-known outlier within the European [3] and Italian [4] genetic landscape.
43 Genetic isolation of Sardinia must therefore be carefully considered in the context of forensic
44 investigations, especially when uniparental markers are employed [5,6]. Among haploid markers,
45 X-chromosome short tandem repeat (STR) loci play a relevant role in forensics, assisting the
46 interpretation of complex kinship cases in addition or alternatively to autosomal STRs [7]. The use
47 of clusters of tightly linked X-STRs forming highly informative haplotypes is particularly profitable
48 in such cases [8], but the possibility for linkage disequilibrium (LD) must then be taken into
49 account in the following biostatistical evaluations [9,10]. Previous studies focusing on the X

50 chromosome indicated the presence of high levels of LD in Sardinia, explained by the combined
51 effect of genetic drift, peculiar demographic history, and slow population growth [11,12]. However,
52 allele/haplotype frequency data of forensically relevant X-STR markers in Sardinians are extremely
53 limited, at present [13,14]. To fill this gap, samples from both open and isolated Sardinian
54 populations were tested with a commercial kit (Investigator Argus X-12 Kit, Qiagen) including four
55 clusters of closely linked X-STR triplets.

56

57 **2. Sample**

58 Samples from 516 Sardinians (318 males and 198 females) were collected by means of
59 venipuncture or buccal swab. Donors belonged to three open populations from Northern (n = 50),
60 Central (n = 118) and Southern (n = 197) Sardinia, and three isolated populations: the mountain
61 villages of Benetutti (n = 44) and Desulo (n = 34), and the linguistic enclave of Carloforte (n = 73)
62 inhabited by the descendants of Genoese settlers still speaking a distinctive archaic form of the
63 Ligurian dialect [15-17] (Supplementary material, Fig. S1). All individuals were unrelated,
64 apparently healthy, born and resident in the selected villages, or areas, for at least three generations.
65 The study was reviewed and approved by the University of Cagliari Ethical Committee and all
66 voluntary participants read and signed an informed consent form.

67

68 **3. Extraction, PCR amplification, genotyping and statistical analyses**

69 Genomic DNA was extracted using the QIAamp DNA Mini kit (Qiagen). The Investigator Argus
70 X-12 Kit was used, according to the manufacturer's instructions, to amplify 12 X-STR loci arranged
71 in four linkage groups: I (DXS10148-DXS10135-DXS8378); II (DXS7132-DXS10079-
72 DXS10074); III (DXS10103-HPRTB-DXS10101); IV (DXS10146-DXS10134-DXS7423).

73 Detection and separation of PCR products were carried out using the ABI Prism 3500 Genetic
74 Analyzer and GeneMapper software (Thermo Fisher Scientific).

75 Statistical parameters of forensic interest were calculated using the on-line functions provided by
76 the ChrX-STR.org 2.0 database (<http://www.chrx-str.org>) [18].

77 Test of Hardy–Weinberg equilibrium (HWE) for genotypic data (female subsample), haplotype
78 diversity, pairwise test of LD, analysis of molecular variance (AMOVA), and pairwise genetic
79 distances (FST) for haplotype data (male subsample) were performed with Arlequin software
80 version 3.5 [19]. To account for multiple testing, Bonferroni correction was applied to adjust
81 threshold p-value (α level).

82 Matrixes of Slatkin's linearized pairwise genetic distances were calculated from allele frequencies
83 with Arlequin software, averaged, and represented by multidimensional scaling (MDS) analysis
84 using the *isomds* function as implemented in MASS package, and *vegan* package of R v.3.3.0 [20].

85

86 **4. Quality control**

87 XX28 DNA included in the Investigator Argus X-12 Kit was used as control DNA for allele
88 assignment. This manuscript follows the guidelines for the publication of population data indicated
89 by the journal [21].

90

91 **5. Results**

92 Allele and haplotype frequencies in the Sardinian sample are displayed in Supplementary material,
93 Table S1, together with statistical parameters of forensic interest calculated for each X-STR marker.

94

95 **6. Other remarks**

96 Based on the observed and expected distribution of genotypes in the Sardinian female subsample,
97 all the tested X-STR loci were found to be in HWE ($\alpha=0.004$).

98 Pairwise test of LD delivered statistically significant results ($\alpha =0.0008$) exclusively for pairs of
99 markers located within linkage groups (Supplementary material, Table S2). All the four linkage
100 groups were interested by the presence of LD in the Sardinian population sample. At subpopulation
101 level, LD between markers DXS10103-DXS10101 of linkage group III was observed in Northern
102 Sardinia ($p = 0.0002$), Central Sardinia ($p < 0.0001$), Desulo ($p = 0.0007$) and Carloforte ($p <$
103 0.0001). Moreover LD was found between DXS10148-DXS10135 of linkage group I in Desulo ($p <$
104 0.0001) and DXS10146-DXS10134 of linkage group IV in Carloforte ($p < 0.0001$).

105 When considering the distribution of haplotypes, the three isolates of Benetutti, Desulo and
106 Carloforte generally showed lower haplotype diversity values compared to open populations
107 (Supplementary material, Table S1). Pairwise comparisons between subpopulations showed
108 multiple significant F_{ST} values ($\alpha=0.003$) at linkage group I, III and IV (Supplementary material,
109 Table S3). Notably, significant comparisons always involved isolates, whereas no evidence of
110 differentiation was seen between the open populations from Northern, Central and Southern
111 Sardinia. AMOVA, performed after grouping the three open populations in a single group,
112 evidenced heterogeneity for linkage group IV (2.31% of the observed variation among population
113 groups; $F_{ST} = 0.020$; $p < 0.05$). Remarkably, linkage group IV is located within Xq28, close to the
114 Glucose-6-phosphate-dehydrogenase (G6PD) gene. Mutation in G6PD was positively selected by
115 malaria, that affected lowland and costal areas of Sardinia, but not the elevated interior regions of
116 the island [22].

117 The Sardinian haplotype dataset was compared with that compiled by Bini et al., consisting of 200
118 Italians (including 12 Sardinians) typed for the Investigator Argus X-12 loci [14]. The percentage of
119 Sardinians haplotypes which were not previously observed in Italians [14] was: 76.3% (linkage

120 group I), 61.3% (linkage group II), 54.1% (linkage group III), 58.9% (linkage group IV).
121 Conversely, the percentage of Italian haplotypes listed in [14] not found in the Sardinian sample
122 was: 67.9% (linkage group I), 49.6% (linkage group II), 46.5% (linkage group III), 56.3% (linkage
123 group IV). In every linkage group, the most frequent haplotype found in the Sardinian sample was
124 observed at least once in Italians, and vice versa.

125 AMOVA evidenced significant variation among groups (whole Sardinian sample vs Italy) at
126 linkage group IV (1.47%; $F_{ST} = 0.008$; $p < 0.05$). Among Sardinian subpopulations, the isolate of
127 Benetutti mostly contributed to this result, as reflected by the highly significant genetic distance
128 from the Italian population ($F_{ST} = 0.029$; $p < 0.001$).

129 MDS analysis was used to summarize genetic differences between Sardinians and other relevant
130 populations from Europe, Northern Africa, and the Middle East [14,23-29] (Supplementary
131 material, Figure S2). A loose cluster including Sardinia, Central and Western Mediterranean
132 populations, and Algeria was observed, confirming the overall genetic homogeneity previously
133 described for X-chromosome biallelic markers in that geographical area, with the notable exception
134 of Morocco [30]. Also confirmed was the outlier position of Albania in the Mediterranean context,
135 previously seen for X-STRs included in the Investigator Argus X-12 kit [29].

136 In its recent guidelines on the use of X-STRs in kinship analysis [31], the DNA Commission of the
137 International Society for Forensic Genetics recommends that haplotype frequencies should be used
138 for likelihood calculations when LD exists. The obtained results confirm that complex patterns of
139 LD along the X chromosome are present in Sardinia, which also involve forensically relevant X-
140 STR markers. Accordingly, the present study provides haplotype database of suitable size [10] for
141 the computation of likelihoods in kinship tests carried out on individuals with Sardinian ancestry.

142

143 **Acknowledgments**

144 This work was supported by Dipartimento di Scienze Mediche, Università di Torino, funding
145 “Progetti di Ricerca finanziati dall’Università degli Studi di Torino (ex 60%)” – Anno 2015” to
146 C.D.G.

147

148 **References**

149 [1] www.istat.it/it/files/2013/01/Sardegna_completo.pdf

150 [2] G. Paulis, Lingue e popoli della Sardegna preistorica e protostorica, *Antropol. Contemp.* 18
151 (1995): 37–40.

152 [3] M. Sikora, M.L. Carpenter, A. Moreno-Estrada, B.M. Henn, P.A. Underhill, et al., Population
153 genomic analysis of ancient and modern genomes yields new insights into the genetic ancestry of
154 the Tyrolean Iceman and the genetic structure of Europe, *PLoS Genet.* 10 (2014) e1004353.

155 [4] C. Di Gaetano, F. Voglino, S. Guarrera, G. Fiorito, F. Rosa, et al., An overview of the genetic
156 structure within the Italian population from genome-wide data. *PLoS One* 7 (2012) e43759.

157 [5] D. Lacerenza, S. Aneli S, C. Di Gaetano, R. Critelli, A. Piazza, et al., Investigation of extended
158 Y chromosome STR haplotypes in Sardinia, *Forensic Sci. Int. Genet.* 27 (2017) 172-174.

159 [6] C. Rapone, E. D'Atanasio, A. Agostino, M. Mariano, M.T. Papaluca, et al. Forensic genetic
160 value of a 27 Y-STR loci multiplex (Yfiler® Plus kit) in an Italian population sample, *Forensic*
161 *Sci. Int. Genet.* 21 (2016) e1-5.

162 [7] C. Gomes, M. Magalhães, C. Alves, A. Amorim, N. Pinto, et al., Comparative evaluation of
163 alternative batteries of genetic markers to complement autosomal STRs in kinship investigations:
164 autosomal indels vs. X-chromosome STRs, *Int. J. Legal Med.* 126 (2012) 917–921.

165 [8] R. Szibor, X-chromosomal markers: past, present and future, *Forensic Sci. Int. Genet.* 1 (2007)
166 93–99.

- 167 [9] A.O. Tillmar, T. Egeland, B. Lindblom, G. Holmlund, P. Mostad, Using X-chromosomal
168 markers in relationship testing: calculation of likelihood ratios taking both linkage and linkage
169 disequilibrium into account, *Forensic Sci. Int. Genet.* 5 (2011) 506-11.
- 170 [10] D. Kling, B. Dell'Amico, A.O. Tillmar, FamLinkX - implementation of a general model for
171 likelihood computations for X-chromosomal marker data, *Forensic Sci. Int. Genet.* 17 (2015) 1-7.
- 172 [11] P. Zavattari, E. Deidda, M. Whalen, R. Lampis, A. Mulargia, et al, Major factors influencing
173 linkage disequilibrium by analysis of different chromosome regions in distinct populations:
174 demography, chromosome recombination frequency and selection, *Hum. Mol. Genet.* 9 (2000)
175 2947-2957.
- 176 [12] A. Angius, D. Bebbere, E. Petretto, M. Falchi, P. Forabosco P, et al., Not all isolates are equal:
177 linkage disequilibrium analysis on Xq13.3 reveals different patterns in Sardinian sub-populations,
178 *Hum. Genet.* 111 (2002) 9-15.
- 179 [13] N. Cerri, A. Verzeletti, F. Gasparini, A. Poglio, F. De Ferrari. Population data for 8 X-
180 chromosome STR loci in a population sample from Northern Italy and from the Sardinia island,
181 *Forensic Sci. Int. Genet. Suppl. Ser. 1* (2008) 173-175.
- 182 [14] C. Bini, L.N. Riccardi, S. Ceccardi, F. Carano, S. Sarno, et al., Expanding X-chromosomal
183 forensic haplotype frequencies database: Italian population data of four linkage groups, *Forensic*
184 *Sci. Int. Genet.* 15 (2015) 127–130.
- 185 [15] M. Capocasa, P. Anagnostou, V. Bachis, C. Battaglia, S. Bertoncini, et al., Linguistic,
186 geographic and genetic isolation: a collaborative study of Italian populations, *J. Anthropol. Sci.* 92
187 (2014) 1–32.
- 188 [16] S. Sartoris, O. Varetto, N. Migone, N. Cappello, A. Piazza, et al., Mitochondrial DNA
189 polymorphism in four Sardinian villages, *Ann. Hum. Genet.* 4 (1988) 327–340.

- 190 [17] R. Robledo, L. Corrias, V. Bachis, N. Puddu, A. Mameli, et al., Analysis of a genetic isolate:
191 the case of Carloforte (Italy), *Hum. Biol.* 84 (2012) 735–754.
- 192 [18] R. Szibor, S. Hering, J. Edelmann, A new Web site compiling forensic chromosome X
193 research is now online, *Int. J. Legal Med.* 120 (2006) 252–254.
- 194 [19] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform
195 population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564-567.
- 196 [20] R_Core_Team (2015). “R: A Language and Environment for Statistical Computing”.
- 197 [21] A. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, et al., New guidelines for the
198 publication of genetic population data, *Forensic Sci. Int. Genet.* 7 (2013) 217-220.
- 199 [22] A. Falchi, I. Piras, Genetic effects of malaria in Sardinia (Italy), in: C. Calò, G. Vona (Eds.),
200 Human genetic isolates, Research Signpost, Trivandrum, 2006, pp. 151-176.
- 201 [23] L. Poulsen, C. Tomas, K. Drobnič, V. Ivanova, H.S. Mogensen, et al., NGMSElect and
202 Investigator Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia,
203 and Turkey, *Forensic Sci. Int. Genet.* 22 (2016) 110–112.
- 204 [24] A. Bekada, S. Benhamamouch, A. Boudjema, M. Fodil, S. Menegon, et al., Analysis of 21 X-
205 chromosomal STRs in an Algerian population sample, *Int. J. Legal Med.* 124 (2010) 287–294.
- 206 [25] J.F. Ferragut, K. Bentayebi, J.A. Castro, C. Ramon, A. Picornell, Genetic analysis of 12 X-
207 chromosome STRs in Western Mediterranean populations, *Int. J. Legal Med.* 129 (2015) 253-255.
- 208 [26] J. Edelmann, S. Lutz-Bonengel, J. Naue, S. Hering, X-chromosomal haplotype frequencies of
209 four linkage groups using the Investigator Argus X-12 Kit, *Forensic Sci. Int. Genet.* 6 (2012) e24–
210 e34.

211 [27] C. Tomas, I. Skitsa, E. Steinmeier, L. Poulsen, A. Ampati, et al., Results for five sets of
212 forensic genetic markers studied in a Greek population sample, *Forensic Sci. Int. Genet.* 16 (2015)
213 132–137.

214 [28] K. Bentayebi, A. Picornell, M. Bouabdeallah, J.A. Castro, R. Aboukhalid, et al., Genetic
215 diversity of 12 X-chromosomal short tandem repeats in the Moroccan population, *Forensic Sci. Int.*
216 *Genet.* 6 (2012) e48-e49.

217 [29] P. Mršić, P. Ozretić, J. Crnjac, S. Merkaš, I. Račić, et al., Analysis of 12 X-STR loci in the
218 population of south Croatia, *Mol. Biol. Rep.* 44 (2017) 183-189.

219 [30] C. Tomas, J.J. Sanchez, A. Barbaro, C. Brandt-Casadevall, A. Hernandez, et al., X-
220 chromosome SNP analyses in 11 human Mediterranean populations show a high overall genetic
221 homogeneity except in North-west Africans (Moroccans), *BMC Evol. Biol.* 8 (2008) 75.

222 [31] A.O. Tillmar, D. Kling, J.M. Butler, W. Parson, M. Prinz M, et al., DNA Commission of the
223 International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship
224 analysis, *Forensic Sci. Int. Genet.* 29 (2017) 269-275.

225

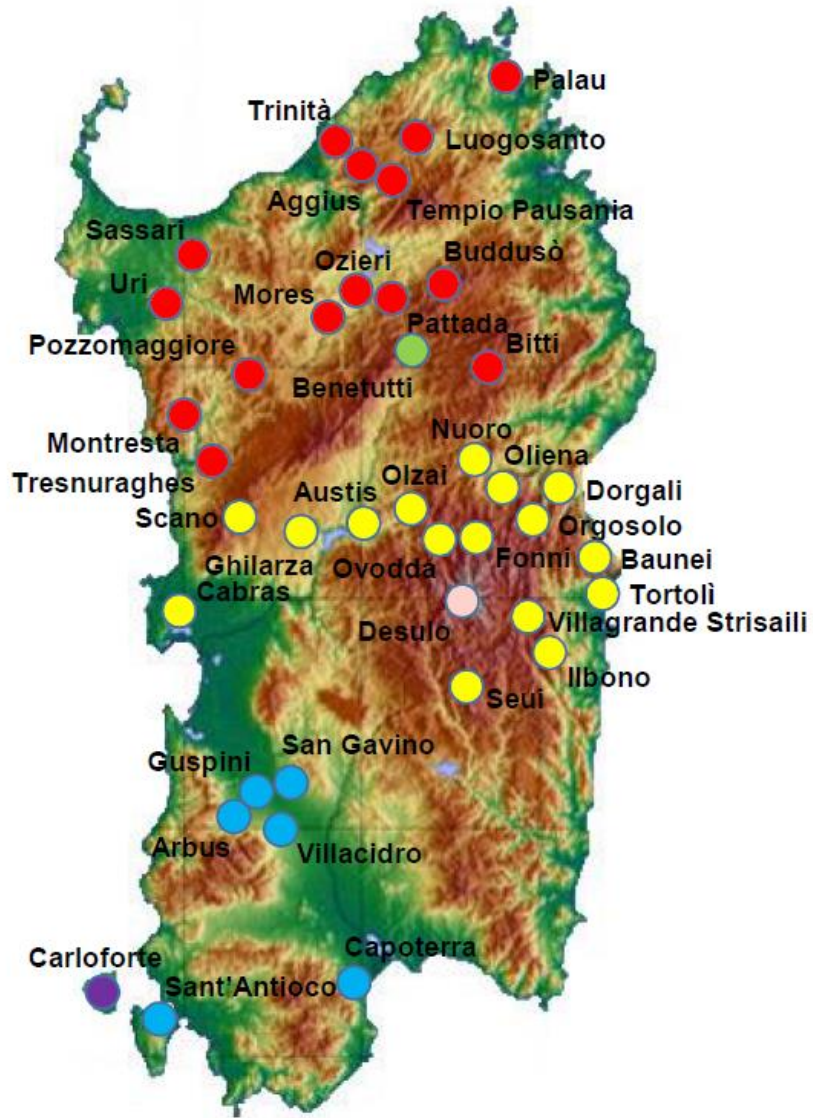


Figure S1 – Geographical location of the tested populations and sampling sites is indicated by colored dots: Northern Sardinia (red); Central Sardinia (yellow); Southern Sardinia (blue); Benetutti (green); Desulo (pink); Carloforte (purple).

226

227 Table S1 [see attachment...](#)

228

	Linkage group I			Linkage group II			Linkage group III		
	DSX10148	DXS10135	DXS8378	DXS7132	DXS10079	DXS10074	DXS10103	HPRTB	DXS10101
DSX10148		<0,0001	0,0861	0,7053	0,2975	0,5169	0,2841	0,3691	0,4999
DXS10135			0,0457	0,3750	0,7361	0,4739	0,6351	0,5651	0,0949
DXS8378				0,9975	0,3759	0,9753	0,0353	0,3544	0,6968
DXS7132					0,4059	<0,0001	0,4145	0,6751	0,7536
DXS10079						<0,0001	0,6017	0,0274	0,0727
DXS10074							0,1755	0,8084	0,7987
DXS10103								<0,0001	<0,0001
HPRTB									<0,0001
DXS10101									
DXS10146									
DXS10134									
DXS7423									

229

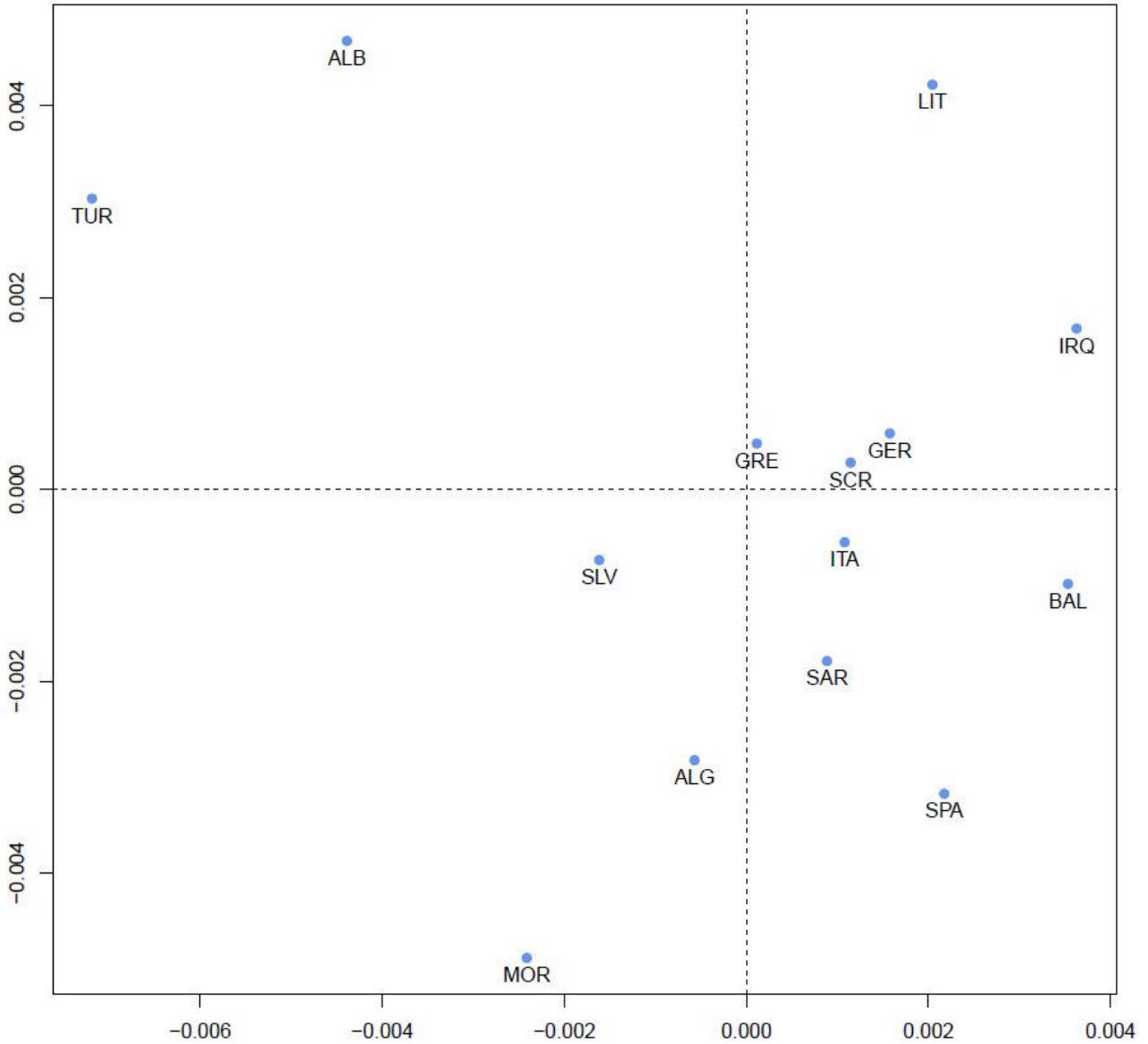
230 **Table S2** Pairwise test of LD in the Sardinian population sample. The cells containing comparisons
231 between markers within linkage groups are shaded in grey. significant p-values after Bonferroni
232 correction for multiple testing are shown in bold.

233

Linkage group I	SAN	BEN	SAC	DES	SAS	CAR
SAN		0,006	0,000	0,012	0,000	0,003
BEN	0,018		0,006	0,019	0,004	0,008
SAC	0,315	0,018		0,016	0,002	0,006
DES	<0,001	<0,001	<0,001		0,015	0,021
SAS	0,450	<0,001	0,063	<0,001		0,006
CAR	0,135	0,018	<0,001	<0,001	<0,001	
Linkage group II	SAN	BEN	SAC	DES	SAS	CAR
SAN		0,000	-0,001	0,002	-0,002	0,001
BEN	0,496		0,003	0,005	0,002	0,006
SAC	0,676	0,117		0,000	0,001	0,001
DES	0,261	0,054	0,396		0,000	0,005
SAS	0,901	0,117	0,333	0,432		0,002
CAR	0,360	0,072	0,315	0,072	0,117	
Linkage group III	SAN	BEN	SAC	DES	SAS	CAR
SAN		0,008	0,007	0,016	0,008	0,009
BEN	0,036		0,006	0,013	0,008	0,006
SAC	0,036	0,018		0,008	0,003	0,006
DES	<0,001	<0,001	0,045		0,010	0,009
SAS	0,018	<0,001	0,135	0,009		0,001
CAR	0,009	0,036	0,027	0,036	0,351	
Linkage group IV	SAN	BEN	SAC	DES	SAS	CAR
SAN		0,004	0,000	0,012	-0,001	0,008
BEN	0,135		0,007	0,016	0,007	0,012
SAC	0,514	<0,001		0,007	0,002	0,006
DES	<0,001	0,009	<0,001		0,013	0,014
SAS	0,721	<0,001	0,144	<0,001		0,012
CAR	0,018	<0,001	0,045	0,009	<0,001	

234

235 **Table S3** Pairwise genetic distances for haplotype data: F_{ST} and p-values are reported above and
236 below the diagonal, respectively. Significant results are shown in bold.
237



238 **Figure S2** – MDS plot (stress: 0.141; RSQ: 0.98) based on averaged pairwise F_{ST} distances derived
239 from allelic frequencies of X-STR observed in: Sardinia (SAR) (present study), Albania (ALB)
240 [23], Algeria (ALG) [24], Balearic islands (BAL) [25], Germany (GER) [26], Greece (GRE) [27],
241 Iraq (IRQ) [23], Italy (ITA) [14], Lithuania (LIT) [23], Morocco (MOR) [28], Slovenia (SLV) [23],
242 [23], Southern Croatia (SCR) [29], Valencia-Spain (SPA) [25], and Turkey [23].
243
244