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:

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)
Allele and haplotype diversity of 12 X-STRs in Sardinia

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

\textsuperscript{a} Dipartimento di Scienze della Sanità Pubblica e Pediatriche, Università di Torino, Italy
\textsuperscript{b} Dipartimento di Scienze Mediche, Università di Torino, Italy
\textsuperscript{c} Italian Institute for Genomic Medicine, Torino, Italy
\textsuperscript{d} Dipartimento di Scienze Biomediche, Università di Cagliari, Italy
\textsuperscript{e} Dipartimento di Scienze della Vita e dell'Ambiente, Università di Cagliari, Italy

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

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

Abstract

The analysis of clusters of tightly linked X-chromosome short tandem repeat (STR) markers can assist the interpretation of complex kinship cases. However, when linkage disequilibrium (LD) is present in the population of origin of tested individuals, haplotype rather than allele frequencies should be used in likelihood calculations. The diversity of twelve X-STRs arranged in four linkage groups (I: DXS10148-DXS10135-DXS8378; II: DXS7132-DXS10079-DXS10074; III: DXS10103-HPRTB-DXS10101; IV: DXS10146-DXS10134-DXS7423) was tested in a Sardinian population sample (n=516) including three open populations from the Northern, Central and Southern part of the island, and three isolates (Benetutti, Desulo, Carloforte). Evidence of LD was detected in Sardinia within each linkage group. Significant differences in haplotype and allele frequency distribution of X-STR markers was seen between isolates and open populations, which on the contrary appeared highly homogeneous.
The percentage of Sardinian haplotypes previously unobserved in a similar dataset compiled for the Italian population was: 76.3% (linkage group I), 61.3% (linkage group II), 54.1% (linkage group III), 58.9% (linkage group IV). Significant heterogeneity in haplotype distribution between Sardinians and mainland Italians was observed at linkage group IV.

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

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

1. Population

Sardinia is the second largest island in the Mediterranean Sea, with a population of over 1.6 million according to the latest Italian census [1]. A combination of geographic, historical and environmental factors contributed to the substantial differentiation of modern Sardinians from European continental populations, testified by distinctive archaeological, linguistic, and cultural records [2]. Sardinia is also a well-known outlier within the European [3] and Italian [4] genetic landscape. Genetic isolation of Sardinia must therefore be carefully considered in the context of forensic investigations, especially when uniparental markers are employed [5,6]. Among haploid markers, X-chromosome short tandem repeat (STR) loci play a relevant role in forensics, assisting the interpretation of complex kinship cases in addition or alternatively to autosomal STRs [7]. The use of clusters of tightly linked X-STRs forming highly informative haplotypes is particularly profitable in such cases [8], but the possibility for linkage disequilibrium (LD) must then be taken into account in the following biostatistical evaluations [9,10]. Previous studies focusing on the X
chromosome indicated the presence of high levels of LD in Sardinia, explained by the combined
effect of genetic drift, peculiar demographic history, and slow population growth [11,12]. However,
allele/haplotype frequency data of forensically relevant X-STR markers in Sardinians are extremely
limited, at present [13,14]. To fill this gap, samples from both open and isolated Sardinian
populations were tested with a commercial kit (Investigator Argus X-12 Kit, Qiagen) including four
clusters of closely linked X-STR triplets.

2. Sample

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

3. Extraction, PCR amplification, genotyping and statistical analyses

Genomic DNA was extracted using the QIAamp DNA Mini kit (Qiagen). The Investigator Argus
X-12 Kit was used, according to the manufacturer’s instructions, to amplify 12 X-STR loci arranged
in four linkage groups: I (DXS10148-DXS10135-DXS8378); II (DXS7132-DXS10079-
DXS10074); III (DXS10103-HPRTB-DXS10101); IV (DXS10146-DXS10134-DXS7423).
Detection and separation of PCR products were carried out using the ABI Prism 3500 Genetic Analyzer and GeneMapper software (Thermo Fisher Scientific).

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

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

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

4. Quality control

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

5. Results

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

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

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

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

The Sardinian haplotype dataset was compared with that compiled by Bini et al., consisting of 200 Italians (including 12 Sardinians) typed for the Investigator Argus X-12 loci [14]. The percentage of Sardinians haplotypes which were not previously observed in Italians [14] was: 76.3% (linkage
group I), 61.3% (linkage group II), 54.1% (linkage group III), 58.9% (linkage group IV).

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

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

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

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

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

References


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).

Table S1 see attachment…
<table>
<thead>
<tr>
<th>Linkage group I</th>
<th>SAN</th>
<th>BEN</th>
<th>SAC</th>
<th>DES</th>
<th>SAS</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAN</td>
<td>0.006</td>
<td>0.000</td>
<td>0.012</td>
<td>0.000</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>BEN</td>
<td>0.018</td>
<td>0.006</td>
<td>0.019</td>
<td>0.004</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>SAC</td>
<td>0.315</td>
<td>0.018</td>
<td>0.016</td>
<td>0.002</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>DES</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.015</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>SAS</td>
<td>0.450</td>
<td>&lt;0.001</td>
<td>0.063</td>
<td>&lt;0.001</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>0.135</td>
<td>0.018</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linkage group II</th>
<th>SAN</th>
<th>BEN</th>
<th>SAC</th>
<th>DES</th>
<th>SAS</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAN</td>
<td>0.000</td>
<td>-0.001</td>
<td>0.002</td>
<td>-0.002</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>BEN</td>
<td>0.496</td>
<td>0.003</td>
<td>0.005</td>
<td>0.002</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>SAC</td>
<td>0.676</td>
<td>0.117</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>DES</td>
<td>0.261</td>
<td>0.054</td>
<td>0.396</td>
<td>0.000</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>SAS</td>
<td>0.901</td>
<td>0.117</td>
<td>0.333</td>
<td>0.432</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>0.360</td>
<td>0.072</td>
<td>0.315</td>
<td>0.072</td>
<td>0.117</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linkage group III</th>
<th>SAN</th>
<th>BEN</th>
<th>SAC</th>
<th>DES</th>
<th>SAS</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAN</td>
<td>0.008</td>
<td>0.007</td>
<td>0.016</td>
<td>0.008</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>BEN</td>
<td>0.036</td>
<td>0.006</td>
<td>0.013</td>
<td>0.008</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>SAC</td>
<td>0.036</td>
<td>0.018</td>
<td>0.008</td>
<td>0.003</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>DES</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.045</td>
<td>0.010</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>SAS</td>
<td>0.018</td>
<td>&lt;0.001</td>
<td>0.135</td>
<td>0.009</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>0.009</td>
<td>0.036</td>
<td>0.027</td>
<td>0.036</td>
<td>0.351</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linkage group IV</th>
<th>SAN</th>
<th>BEN</th>
<th>SAC</th>
<th>DES</th>
<th>SAS</th>
<th>CAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAN</td>
<td>0.004</td>
<td>0.000</td>
<td>0.012</td>
<td>-0.001</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td>BEN</td>
<td>0.135</td>
<td>0.007</td>
<td>0.016</td>
<td>0.007</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>SAC</td>
<td>0.514</td>
<td>&lt;0.001</td>
<td>0.007</td>
<td>0.002</td>
<td>0.006</td>
<td></td>
</tr>
<tr>
<td>DES</td>
<td>&lt;0.001</td>
<td>0.009</td>
<td>&lt;0.001</td>
<td>0.013</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>SAS</td>
<td>0.721</td>
<td>&lt;0.001</td>
<td>0.144</td>
<td>&lt;0.001</td>
<td>0.012</td>
<td></td>
</tr>
<tr>
<td>CAR</td>
<td>0.018</td>
<td>&lt;0.001</td>
<td>0.045</td>
<td>0.009</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

**Table S2** Pairwise test of LD in the Sardinian population sample. The cells containing comparisons between markers within linkage groups are shaded in grey. Significant p-values after Bonferroni correction for multiple testing are shown in bold.
Table S3 Pairwise genetic distances for haplotype data: $F_{ST}$ and $p$-values are reported above and below the diagonal, respectively. Significant results are shown in bold.

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