Expanded Croatian 12 X-STR loci database with an overview of aberrant profiles

Gordan Mršić [a,b], Petar Ozretić [c], Josip Crnjac [d], Siniša Merkaš [a,b], Viktorija Sukser [a,b], Ivana Račić [a,b], Sara Rožić [a,b], Lucija Barbarić [a,b], Maja Popović [e], Marina Korolija[a,b]

[a] Forensic Science Centre "Ivan Vučetić", Ilica 335, 10000 Zagreb, Croatia
[b] Forensic Science Office, University of Zagreb, Ulica Ivana Lučića 5-6, 10000 Zagreb, Croatia
[c] Ruđer Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia
[d] University Department for Forensic Sciences, University of Split, Ruđera Boškovića 31, 21000 Split, Croatia
[e] Faculty of Veterinary Medicine, University of Zagreb, Heinzelova ulica 55, 10000 Zagreb, Croatia

Corresponding author:
Marina Korolija, Forensic Science Centre "Ivan Vučetić", Ilica 335, 10000 Zagreb, Croatia
e-mail: mkorolija@mup.hr
tel: +385 1 3787 181
fax: +385 1 3788 051

Expanded Croatian 12 X-STR loci database with an overview of aberrant profiles

ABSTRACT

In order to implement X-chromosome short tandem repeat (X-STR) typing into routine forensic practice, reference database of a given population should be established. Therefore we extended already published data with additional 397 blood samples from unrelated Croatian citizens, and analyzed the total of 995 samples (549 male and 446 female) typed by Investigator® Argus X-12 Kit. To test genetic homogeneity of consecutively processed five historic-cultural regions covering the entire national territory, we calculated pairwise Fst genetic distances between regions based on allele and full haplotype frequencies. Since the comparison did not yield any statistically significant difference, we integrated STR profile information from all regions and used the whole data set to calculate forensic parameters. The most informative marker is DXS10135 (polymorphism information content (PIC)=0.929) and the most informative linkage group (LG) is LG1 (PIC=0.996). Cumulative power of discrimination across 12 analyzed loci amounted to 0.9999999999999996 in females and 0.9999999995 in males. We confirmed linkage disequilibrium for seven marker pairs belonging to LG2, LG3 and LG4. We also compared Croatia with 13 European populations based on haplotype frequencies and detected no statistically significant Fst values after Bonferroni correction in any LG. Multi-dimensional scaling plot revealed tight grouping of four Croatian regions amongst populations of southern, central and northern Europe, with the exception of northern Croatia. In this study we gave the first extensive overview of aberrant profiles encountered during Investigator® Argus X-12 typing. We found ten profiles consistent with single locus duplication followed by tetranucleotide tract length polymorphism. Locus DXS10079 is by far the most frequently affected one, presumably mutated in eight samples. We also found four profiles consistent with X-chromosome aneuploidy (three profiles with XXX pattern and one profile with XXY pattern). In conclusion, we established integral forensic Croatian X-chromosome database, proved forensic pertinence of Investigator® Argus X-12 Kit for the entire Croatian population and identified locus DXS10079 as mutational hotspot.

Keywords: Investigator® Argus X-12; forensic X-STR markers; Croatian X-STR database; aberrant X-STR profiles; X chromosome duplication; locus DXS10079

INTRODUCTION

Usefulness of X-chromosome short tandem repeat (X-STR) typing has so far been very well described and established for both identification and kinship testing in the specific forensic contexts [1]. Implementation of X-STR analysis into routine forensic casework primarily requires general population database for match probability calculations. As the purpose of this study is to build comprehensive forensic database based on 12 X-STR loci for the Croatian population, we expanded data set of the already published 598 profiles [2-4] with additional 397 profiles and consolidated Investigator® Argus X-12 Kit typing results. The ensuing comprehensive database encompasses 995 individuals, originating from five historic-cultural regions of the northern [2], central [3], southern [4], eastern and western part of the country **(Figure 1).**

The Republic of Croatia is a European country bordering with Slovenia and Hungary in the north, Serbia and Bosnia and Herzegovina in the east, Montenegro in the south and Italy in the south-west. It is situated on the Adriatic Sea and positioned between Central and Southeast Europe **(Figure 1).** Regarding its ethnic composition, Croatia is a homogenous country. The population of 4.3 million mostly consists of Croats (90.4%), while minorities include Serbian (4.4%) and 21 other ethnicities (less than 1% each) [5]. Croatia in general went through a dynamic demographic history and particular geographic regions experienced quite different demographic perturbations over the centuries [6]. In that respect, we assumed that genetic substructuring at the X-STR loci might have occurred within the Croatian population. Although genetic diversity on the X-chromosome is

presumably lower in comparison to autosomes due to lower mutation rate and reduced effective population size, the latter feature ascertains faster genetic drift for the X-chromosome than for autosomes. As a consequence, X-chromosome exhibits more pronounced differentiation value for intra- and inter-population comparison [7]. In this regard, it is important to evaluate potential regional substructuring within the population under forensic investigation, in order to establish separate forensic databases, if needed. We therefore aimed to characterize the genetic diversity of 12 X-STRs in the Croatian population.

With the advance of high-throughput genomics techniques, indications emerged that X-chromosome duplication events might have been underestimated previously [8]. It is therefore possible that yet undetected minor duplications of X-chromosome fragments, which do not produce obvious clinical features, appear with increased frequency in the population compared to other chromosomes. A fraction of duplication events might be detected indirectly during forensic X-STR typing, as already shown in forensic typing of autosomes [9] and Y chromosome [10, 11]. Since duplications of X-STR markers might complicate forensic interpretation due to generation of additional electropherogram peaks (male biallelic and female triallelic profiles), it seems important to report such events encountered during creation of population X-STR databases. Not only forensic community would benefit from that kind of information, but also genetic anthropology and molecular medicine.

Within the scope of this study, we constructed comprehensive database to achieve final goal of establishing X-STR typing as a routine forensic practice in Croatian Forensic Science Centre "Ivan Vučetić". We computed forensic parameters to test the forensic pertinence of Investigator® Argus X-12 Kit in Croatian population and tested potential population substructuring to assess the universality of the comprehensive database. We also put Croatia into wider, inter-population context by comparing our data with previously published data for 13 European populations. Finally, in order to help forensic experts in profile evaluation and subsequent interpretation, we gave a detailed characterization of X-chromosome instability at the population level, manifested as X-STR aberrant profiles.

MATERIALS AND METHODS

The total of 203 samples (103 male and 100 female) from eastern Croatian region and the total of 194 samples (146 male and 48 female) from western Croatian region were analyzed. Sampling was performed in an attempt to account for all subpopulation variations by choosing unrelated participants from the entire region covering Požega-Slavonia, Brod-Posavina, Osijek-Baranja and Vukovar-Srijem Counties of eastern Croatia, and Primorje-Gorski Kotar, Lika-Senj and Istria Counties of western Croatia. All samples were collected during routine forensic work by the staff of Forensic Science Centre "Ivan Vučetić", and their use in the study was approved by the Ethics Committee of the Institute for Medical Research and Occupational Health, Zagreb, Croatia.

Genomic DNA was extracted from Flinders Technology Associates (FTA) cards (Whatman, Maidstone, Kent, UK) using Chelex-100 method [12]. Concentration of the extracted DNA was determined by Investigator® Quantiplex Kit (Qiagen GmbH, Hilden, Germany). Normalization of samples was carried out to approximately 1 ng/µL. Amplification was performed by Investigator® Argus X-12 Kit (Qiagen GmbH, Hilden, Germany) in multiplex PCR, containing primers for amelogenin (sex determination) and 12 X-STR markers belonging to four different linkage groups (LGs): LG1 (DXS10148, DXS10135, DXS8378), LG2 (DXS7132, DXS10079, DXS10074), LG3 (DXS10103, HPRTB, DXS10101), and LG4 (DXS10146, DXS10134, DXS7423). Positive and negative controls were also amplified in each PCR batch. Amplification products were analyzed on 3500 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Data obtained from capillary electrophoresis were analyzed using GeneMapper ID-X software (version 1.4, Applied Biosystems). Peak threshold values of 100 RFU and 200 RFU were applied for heterozygous and homozygous alleles, respectively. All samples containing variant alleles, female samples with triallelic patterns **(Supplementary material 3 Figures S1-S4)** and male samples with biallelic pattern **(Supplementary material 3 Figures S5-S10)** were confirmed by re-extraction followed by amplification and capillary electrophoresis. To exclude the possibility of contamination,

four profiles consistent with aneuploidy of X-chromosome **(Supplementary material 3 Figures S11-S14)** were analyzed by AmpFlSTR® *SEfiler Plus*™ PCR Amplification Kit **(Supplementary material 4 Figures S1-S4)**. All procedures and protocols were carried out following manufacturers' instructions.

Statistical analysis

Allele frequencies for all samples and haplotype frequencies for male samples were determined by counting the number of times each allele/haplotype was observed in all given samples. For biallelic male and triallelic female samples, allele(s) with highest frequencies were selected for further calculations. Testing for a departure from Hardy-Weinberg Equilibrium (HWE), including observed heterozygosity (Ho) and expected heterozygosity (He), was performed only for female samples. Presence of pairwise linkage disequilibrium (LD) between loci was tested by likelihood-ratio test using the Expectation-Maximization algorithm for female, and by the exact test using a Markov chain for male samples. Genetic heterogeneity within population was estimated as gene i.e. haplotype diversity (H) for male haplotype data. For testing of genetic homogeneity among the five Croatian regions, analysis of molecular variance (AMOVA) and pairwise genetic distances (Fst) were calculated based on both allele frequencies (male and female samples) and full haplotypes of male samples. Pairwise genetic distances were also calculated for inter-population comparison of haplotype frequencies between pooled Croatian population samples and 13 neighboring European populations: Slovenia, Italy, Hungary, Albania, Czechia, Germany, Greece, West Mediterranean, Lithuania, Belarus, Portugal, Denmark and Sweden [13-24]. All aforementioned computations were performed using Arlequin software v3.5.2.2 [25], while significance level for all statistical tests was set to 0.05 and corrected for multiple comparisons using Bonferroni adjustment.

Forensic parameters encompassing polymorphism information content (PIC), power of exclusion (PE), power of discrimination (PD) for males and for females, mean exclusion chance (MEC) for deficiency cases (Krüger's formula), MEC for normal trios consisting of a mother, a daughter and a putative father (Kishida's formula), and MEC for duos consisting of a daughter and a putative father (Desmarais' formula), were computed based on allele frequencies data using on-line tool available at ChrX-STR.org web page [26]. PIC for LGs was calculated using the R script available as supporting information from Zidkova et al. [16].

To additionally examine the relationship with neighboring European populations, pairwise genetic distances with sample size correction (Fst^{*}) [27] were calculated based on allele frequencies using POPTREE2 software [28]. Inter-population comparison was presented by multi-dimensional scaling (MDS) plot constructed with IBM SPSS Statistics for Windows, Version 23.0 (IBM Corp., Armonk, NY, USA).

RESULTS AND DISCUSSION

Intra-population comparison

In addition to the already published data for northern [2], central [3] and southern Croatia [4], allele and haplotype frequencies for eastern Croatia (100 females and 103 males) and western Croatia (48 females and 146 males) are given in **Supplementary material 1 (Tables S1-S4)**. In both additional regions, HWE was confirmed for all loci. All off-ladder alleles found in the sample pool of western Croatia (DXS7423 9; DXS10074 16.3, 17.2; DXS10079 19.2; DXS10101 27.3; DXS10135 15.1, 20.3; DXS10146 36.2, 48.2; DXS10148 14, 17, 22, 27.2, 32.1) and eastern Croatia (DXS10079 13, 19.2, 24; DXS10135 24.1; DXS10146 38.2, 47.2; DXS10148 22) have been reported by others. The most informative marker in both eastern and western Croatian population is DXS10135 (PIC=0.93 and 0.92, respectively), while the least informative one is DXS8378 (PIC=0.62 and 0.63, respectively).

In order to test genetic homogeneity, pairwise Fst genetic distances between five Croatian regions were calculated based on both allele frequencies (male and female samples) and full haplotypes of male samples. No significant Fst values ($p<0.005$) after Bonferroni correction were detected in either case (**Supplementary material 1 Table S5).** Locus-by-locus AMOVA also confirmed these results (data

not shown). Therefore, all subsequent calculations were performed using a complete Croatian dataset of 995 samples.

Our results are in concordance with previous study where the variability of Croatian population had been analyzed based on Y-chromosomal haplogroup distribution, and strong similarity between the five aforementioned Croatian regions was found [6]. Consequently, there is no need for separate databases in forensic interpretation of both X-STR and Y-STR profiles obtained by corresponding kits. However, it is noteworthy that samples from small, isolated populations (e.g. islands and remote villages), where more pronounced genetic difference is expected, have not been accounted for establishing the common X-STR forensic database.

Alleles and haplotypes in forensic context

In the entire Croatian data set of 995 samples (1441 chromosomes), no significant departure from HWE was detected at any locus. The total of 223 alleles across all loci was detected with frequency span from 0.00069 to 0.42679 **(Supplementary material 2 Table S1)**. Marker DXS10135, comprising 35 alleles in total, is the most informative (PIC=0.93), with the allele 25 as the most common (frequency of 0.103). Marker DXS8378, with the total of 7 alleles, is the least informative (PIC=0.93), and the sum of frequencies of three alleles alone (10, 11, 12) equals 0.949. Both DXS10135 and DXS8378 belong to the LG1. As expected, results for allele variability are very similar to the already published data for other European populations [13, 16, 17, 20, 24].

Haplotype frequencies, estimated directly from the pooled population data for each LG, are presented in **Supplementary material 2 Table S2**. The most informative linkage group is LG1 (PIC= 0.9955), followed by LG4, LG2 and LG3, with frequency of the most common haplotypes equaling 0.0128, 0.0182, 0.031 and 0.0364, respectively. All LGs display gene diversity value of over 0.99, with LG1 being the most diverse (0.9974), which is further corroborated by the highest PIC value of 0.9955 for LG1 **(Supplementary material 2 Table S3)**. LG1 is also the most polymorphic in other worldwide populations typed by Investigator® Argus X-12 Kit [13, 14, 16, 17, 24]. Of all possible haplotypes for each linkage group, 320 (4.4%), 189 (11.4%), 201 (8.9%) and 269 (3.9%) were observed in LG1, LG2, LG3 and LG4, respectively. Of the observed haplotypes, 60.9% of them are unique for LG1, 50% for LG2, 46.8% for LG3 and 57.6% for LG4. In the entire sample pool, no shared 12-marker haplotype was found.

To test the suitability of Investigator® Argus X-12 Kit for forensic casework in Croatian population, forensic parameters were calculated and summarized in **Table 1.** All tested parameters demonstrated high combined values, especially PD in females (0.9999999999999996), which proves the applicability of Investigator® Argus X-12 Kit in both forensic identification and kinship analysis.

Linkage disequilibrium analysis

In the male sample pool, linkage disequilibrium was confirmed ($p<0.0008$ after Bonferroni correction) for: LG2 marker pair DXS10079-DXS10074, all three marker pairs within LG3, and LG4 marker pairs DXS10134-DXS10146 and DXS7423-DXS10134 **(Supplementary material 2 Table S4)**. Confirmation of the obtained results was found in female samples for LG3 marker pairs DXS10101-DXS10103 and DXS10101-HPRTB, as well as for LG4 marker pair DXS10134-DXS10146. Moreover, one additional marker pair from LG4 (DXS7423-DXS10146) was found in females **(Supplementary material 2 Table S5)**. In total, seven out of 12 possible marker pairs across all LGs are in linkage disequilibrium. LD is confirmed for all marker pairs of LG3 and LG4. No statistically significant LD occurred between any marker pair belonging to different LGs.

Investigator® Argus X-12 Kit consists of four groups of closely linked STR markers which should be treated as haplotypes that are passed to the offspring unchanged [29]. Non-random association of alleles should be tested at the population level, using the actual haplotype frequencies in male samples. In populations with high genetic variability, large sample size is needed to establish all allele pairs that are in LD. This is well illustrated by typing Croatian population region by region, where the evidence for LD is demonstrated for maximum three marker pairs on a smaller sample size

(approximately 100 male samples). By enlargement of sample size to 549 male profiles, LD is demonstrated for six marker pairs in male samples and one additional in female samples. Lack of LD evidence for any marker pair within LG1 is probably due to the fact that LG1 is the most diverse group, including by far the greatest number of haplotypes. Therefore, in this case, even a larger sample size would be needed to confirm LD for all linked markers.

Inter-population genetic distance

Haplotype frequencies were used to compare Croatian population with 13 European populations also typed by Investigator® Argus X-12 Kit**.** After Bonferroni correction, no significant difference ($p<0.00078$) was found in any LG **(Supplementary material 2 Table S6)**. It was already shown, by us and others who addressed this issue, that European populations are indeed homogenous according to Argus X-12 markers [4, 16, 20]. There is no notable genetic variability, except for the populations that underwent specific population processes (genetic drift, interbreeding etc.), like Ibiza [19], Greenland [22], Sardinia and Southern Italy [24]. For the purpose of inter-population comparison, we chose not to include populations of different continental origins, because we had already shown correlation between genetic and geographical distance for southern Croatian population [4]. Instead, in order to obtain better resolution and determine positioning of different Croatian regions within European context of more geographically closer populations, we created a MDS plot including northern, central, southern, eastern and western Croatian populations along with 13 European populations **(Figure 2)**. Left-to-right positioning of the dots representing populations along the first dimension axis of MDS plot correlates with south-to-north geographic distribution. The first dimension zero axis aligns with south-to-central Europe dividing line. Southern Croatia is positioned within the more scattered group of south European populations, while eastern, western and central Croatia belong to a tighter group consisted of central and north European populations. These results fit well with previous similar inter-population comparisons [4, 20]. Interestingly, Denmark and northern Croatia stand out as outliers at the rightmost end of the plot. Indications of more distant position of northern Croatia are visible from intra-population comparisons, where the lowest Fst p-values (statistically significant prior to Bonferroni correction) were established between northern and eastern Croatia **(Supplementary material 1 Table S5a)**, as well as northern and southern Croatia **(Supplementary material 1 Table S5b)**. Besides its distant position on the plot, it is interesting to notice that northern Croatia lies nearest to Hungary. That might reflect geographic proximity and historical coexistence of Hungarian and north Croatian entities within the Austro-Hungarian Empire.

Aberrant X-STR profiles

Out of 100 analyzed female profiles from eastern Croatia, two of them exhibited unexpected electropherogram patterns. One profile displayed three triple-peak patterns: at LG2 loci (DXS10079 and DXS10074) and LG3 locus DXS10101 **(Supplementary material 3 Figure S11)**. The other profile displayed eight triple-peak patterns: at LG1 loci (DXS10135 and DXS8378), all LG2 loci (DXS7132, DXS10079 and DXS10074), LG3 locus DXS10101 and LG4 loci (DXS10134 and DXS7423) **(Supplementary material 3 Figure S12)**.

Out of 48 analyzed female profiles from western Croatia, one of them exhibited six triple-peak patterns: at LG1 loci (DXS10148 and DXS10135), LG2 locus DXS7132, LG3 loci (HPRTB and DXS10101) and LG4 locus DXS10146 **(Supplementary material 3 Figure S13)**. Out of 146 analyzed male profiles from western Croatia, two of them showed unexpected electropherogram patterns. One profile displayed eight double-peak patterns: at LG1 loci (DXS10135 and DXS8378), LG2 loci (DXS7132 and DXS10079), LG3 loci (DXS10103 and HPRTB) and LG4 loci (DXS10146 and DXS10134) **(Supplementary material 3 Figure S14)**. The other profile displayed two peaks at locus DXS10079, corresponding to alleles 20 and 21 **(Table 3)**.

Taking into account all 995 Croatian samples genotyped over time by the Investigator® Argus X-12 Kit in our laboratory, as much as 14 profiles (1.4%) displayed additional peaks at either single or multiple loci. It is obvious that such percentage should not be ignored because in forensic context, aberrant

profiles can add complexity to data interpretation, especially in the interpretation of partial and mixed profiles. Characterization and quantification of X-chromosome instability at the population level would therefore be of help in subsequent DNA casework. In Croatian sample pool, out of 14 aberrant profiles, four of them had additional peaks at multiple loci indicating aneuploidy **(Supplementary material 3 Figures S11-S14)**, while ten profiles had additional peaks at a single locus **(Supplementary material 3 Figures S1-S10)**. Single locus events are summarized in **Table 2**, and the phenomena of by far the most represented aberration (biallelic pattern in six male profiles and triallelic pattern in two female profiles at locus DXS10079) was elaborated in our recent study [4]. This exemplify the importance of documenting mutational hotspots within population, considering the fact that the additional peak at DXS10079 might be expected in 0.45% female and even 1.1% male profiles in Croatian population.

Locus DXS10079 is a tetranucleotide STR marker located in pericentromeric region of Xq12, flanked by upstream marker DXS7132 and downstream marker DXS10074, with approximate distance of 2.3Mb between each other. In humans, pericentromeric regions are by six to seven fold enriched with duplicated sequences [30]. Mutation patterns at DXS10079 in Croatian population indicate Xq12 duplication of possibly rather small region between loci DXS7132 and DXS10074, followed by tetranucleotide tract length mutation. The size of duplication is assumed given that flanking markers never display duplication/mutation pattern, and that larger duplications would likely have profound clinical impact on an affected person, as has been documented for approximately 9Mb duplication comprising Xq12 region [9, 31]. Aforementioned authors hypothesize that Xq12-Xq13.3 duplication causes increased dosage of several duplicated genes that contribute to neurobehavioral phenotype [9, 31]. One of those genes, androgen receptor (AR), is located 48 kb downstream of locus DXS10079 and directly upstream of locus DXS10074. It makes AR the most probable coding candidate to be duplicated in biallelic/triallelic DXS10079 samples. Expression of AR was linked to male aggression in both humans and animals [32, 33]. Although it is impossible for us to further investigate potential AR gene duplication in DXS10079 biallelic samples because of ethical considerations, it is most appealing to speculate that the reference casework sample collection might be enriched by individuals with increased androgen receptor dosage, resulting with enhanced testosterone signaling and consequent aggressive behavior. In the context of potential functional disomy, it would also be interesting to establish frequency of duplication events without subsequent addition/loss of tetranucleotide repeats. The frequency would presumably be higher, with more profound effect at the population level. However, since definitive confirmation of such events cannot be carried out by routine forensic DNA typing methods, subsequent analysis such as comparative genomic hybridization should be performed.

Defining DXS10079 as a duplication hotspot might go beyond Croatian population, since the occurrence of mutational events are documented in Greek [18] and Cabo Verde [34] populations on a much smaller sample size compared to Croatia. Moreover, the inauguration study for DXS10079 locus described three families, out of approximately 333 parent-child trios from German population, with apparent small duplication encompassing DXS10079 [35]. In total, three female triallelic profiles and one male biallelic profile were found with one-step or two-step shift in the repeat number [35]. In our study, where no parental samples were available, we simply excluded the possibility of close relatedness and endogamy of individuals from Croatian sample pool bearing DXS10079 mutations [4]. Nevertheless, it is interesting to notice that all cases except one are found in central and southern Croatia **(Table 2),** and two remaining triallelic profiles at loci DXS10134 and DXS10146 are found in southern Croatia **(Table 2).** It indicates some kind of regional localization, but more population and experimental data, including mitochondrial DNA and Y-chromosome haplotyping would be needed to confirm potential non-random distribution. In any case, forensic community should be encouraged to report “duplication” profiles that presently cannot be taken into account for statistical STR calculations. That practice would enhance the development of statistical calculation strategies able to account for the phenomena present in actual populations.

One male and three female X-STR profiles exhibited patterns consistent with aneuploidy **(Table 3)**. In all female samples we detected several triallelic loci, which indicates the presence of an additional X

chromosome (Triple X syndrome - 47, XXX) **(Supplementary material 3 Figures S11-S13)**. In the male sample profile, we found eight biallelic loci which indicates the presence of an additional X-chromosome (Klinefelter syndrome - 47, XXY) **(Supplementary material 3 Figure S14)**. Distribution of aberrant loci across different LGs, which are scattered over the entire chromosome, suggests an extra X, rather than duplication of a large chromosomal fragment **(Table 3)**. This finding immediately raises the question of potential contamination that might have arisen upon re-processing of stored FTA cards. However, it is highly unlikely because all reference samples were initially genotyped when collected during casework process, and no contamination was found. Nevertheless, after repeated X-STR analysis, we performed additional AmpFLSR®SEfiler Plus™ amplification, which resulted in normal, diploid profiles with homo- or heterozygous loci in all autosomal STR markers **(Supplementary material 4 Figures S1-S4)**. In addition, amelogenin peak heights also reflect an increased dosage of X-chromosome in all profiles **(Supplementary material 4 Figures S1-S4).**

Triple X syndrome in human population is not a rare disorder, affecting about 1 in 1000 female births. It often goes undiagnosed due to mild symptoms [36]. Our results of 0.67% female individuals affected by Trisomy X considerably exceed the established incidence rate of 0.1%. Again, as discussed for DXS10079 duplication, sample size is not sufficient neither to exclude high occurrence by chance, nor to speculate that increased percentage of affected individuals that are at risk of developing social-emotional difficulties [37] might be overrepresented in reference sample pool from casework. The finding of a single male profile indicating Klinefelter syndrome among the 549 analyzed is consistent with worldwide estimations ranging from 1 case in 500 to 1 case in 1000 live male births.

CONCLUSION

In an effort to introduce X-STR analysis for identification and kinship testing, an extensive database including 995 profiles typed by Investigator® Argus X-12 Kit has been established for the Croatian population. We conclude that Investigator® Argus X-12 Kit is suitable for forensic casework and the reference database is universally applicable in the entire Croatia.

Some STR markers incorporated in Investigator® Argus X-12 Kit are affected by the occurrence of additional electropherogram peaks. In Croatian population, locus DXS10079 exhibits the greatest instability manifested as biallelic pattern in males and triallelic pattern in females. Hence we conclude that locus DXS10079 is potential mutational hotspot.

In order to avoid misinterpretation of DNA evidence in criminal casework, information on pattern and frequency of X-STR profile aberrations should be available to forensic community.

The DNA Analysis Unit of the Forensic Science Centre “Ivan Vučetić” is accredited according to ISO/IEC 17025, and regularly participates in quality control proficiency testing provided by the German DNA Profiling group (GEDNAP). This article follows the population data publication guidelines set by the journal [38-40].

Acknowledgements

This work was supported by Ministry of the Interior of Croatia.

We are particularly grateful to out late colleague Branka Gršković, who unselfishly devoted her time and expertize to realization of this research.

Conflict of interest: none.

Figure 1. Map of Croatian regions and position of Croatia in European geographical context.

Figure 2. A two-dimensional multidimensional scaling plot drawn from sample bias corrected Fst* genetic distances calculated from the allele frequencies of 12 X-chromosome STRs included in Investigator Argus X-12 kit with the POPTREE2 software. Stress=0.1336/RSQ=0.9287. NCroatia -

northern Croatia; SCroatia - southern Croatia; CCroatia - central Croatia; WCroatia - western Croatia; ECroatia - eastern Croatia.

[1] T.M. Diegoli, Forensic typing of short tandem repeat markers on the X and Y chromosomes, Forensic Sci Int Genet 18 (2015) 140-51. 10.1016/j.fsigen.2015.03.013

[2] J. Crnjac, P. Ozretic, S. Merkas, M. Ratko, M. Lozancic, M. Korolija, M. Popovic, G. Mrsic, Investigator Argus X-12 study on the population of northern Croatia, Genet Mol Biol 40(1) (2017) 80-83.

[3] J. Crnjac, P. Ozretic, S. Merkas, M. Ratko, M. Lozancic, S. Rozic, D. Spoljaric, M. Korolija, M. Popovic, G. Mrsic, Analysis of 12 X-chromosomal markers in the population of central Croatia, Legal Medicine 21 (2016) 77-84.

[4] G. Mrsic, P. Ozretic, J. Crnjac, S. Merkas, I. Racic, S. Rozic, V. Sukser, M. Popovic, M. Korolija, Analysis of 12 X-STR loci in the population of south Croatia, Molecular Biology Reports 44(1) (2017) 183-189.

[5] Croatian Bureau of Statistics, Census of Population, Households and Dwellings 2011, Population by Citizenship, Ethnicity, Religion and Mother Tongue, Zagreb, Croatia, 2013.

[6] G. Mrsic, B. Grskovic, A. Vrdoljak, M. Popovic, I. Valpotic, S. Andelinovic, V. Stenzl, E. Ehler, L. Urban, G. Lackovic, P. Underhill, D. Primorac, Croatian national reference Y-STR haplotype database, Molecular Biology Reports 39(7) (2012) 7727-7741.

[7] S.F. Schaffner, The X chromosome in population genetics, Nat Rev Genet 5(1) (2004) 43-51. 10.1038/nrg1247

[8] J.L. Mueller, H. Skaletsky, L.G. Brown, S. Zaghlul, S. Rock, T. Graves, K. Auger, W.C. Warren, R.K. Wilson, D.C. Page, Independent specialization of the human and mouse X chromosomes for the male germ line, Nat Genet 45(9) (2013) 1083-7. 10.1038/ng.2705

[9] P. Prontera, V. Ottaviani, I. Isidori, G. Stangoni, E. Donti, Xq12-q13.3 duplication: evidence of a recurrent syndrome, Ann Neurol 72(5) (2012) 821-2; author reply 822-3. 10.1002/ana.23754

[10] M. Kayser, L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Kruger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. de Knijff, M. Stoneking, A. Sajantila, Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs, Am J Hum Genet 66(5) (2000) 1580-8. 10.1086/302905

[11] P. Balaresque, E.J. Parkin, L. Roewer, D.R. Carvalho-Silva, R.J. Mitchell, R.A. van Oorschot, J. Henke, M. Stoneking, I. Nasidze, J. Wetton, P. de Knijff, C. Tyler-Smith, M.A. Jobling, Genomic complexity of the Y-STR DYS19: inversions, deletions and founder lineages carrying duplications, Int J Legal Med 123(1) (2009) 15-23. 10.1007/s00414-008-0253-3

[12] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex-100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, Biotechniques 10(4) (1991) 506-513.

[13] L. Poulsen, C. Tomas, K. Drobnic, V. Ivanova, H.S. Mogensen, A. Kondili, P. Miniati, D. Bunokiene, J. Jankauskiene, V. Pereira, N. Morling, NGMSElect and Investigator((R)) Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia, and Turkey, Forensic Sci Int Genet 22 (2016) 110-2. 10.1016/j.fsigen.2016.02.004

[14] L. Poulsen, M.S. Farzad, C. Borsting, C. Tomas, V. Pereira, N. Morling, Population and forensic data for three sets of forensic genetic markers in four ethnic groups from Iran: Persians, Lurs, Kurds and Azeris, Forensic Sci Int Genet 17 (2015) 43-6. 10.1016/j.fsigen.2015.03.010

[15] G. Horvath, A. Zalan, Z. Kis, H. Pamjav, A genetic study of 12 X-STR loci in the Hungarian population, Forensic Science International-Genetics 6(1) (2012) E46-E47. 10.1016/j.fsigen.2011.03.007

[16] A. Zidkova, P. Capek, A. Horinek, P. Coufalova, Investigator (R) Argus X-12 study on the population of Czech Republic: Comparison of linked and unlinked X-STRs for kinship analysis, Electrophoresis 35(14) (2014) 1989-1992. 10.1002/elps.201400046

[17] J. Edelmann, S. Lutz-Bonengel, J. Naue, S. Hering, X-chromosomal haplotype frequencies of four linkage groups using the Investigator Argus X-12 Kit, Forensic Science International-Genetics 6(1) (2012) E24-E34. 10.1016/j.fsigen.2011.01.001

[18] C. Tomas, I. Skitsa, E. Steinmeier, L. Poulsen, A. Ampati, C. Borsting, N. Morling, Results for five sets of forensic genetic markers studied in a Greek population sample, Forensic Sci Int Genet 16 (2015) 132-7. 10.1016/j.fsigen.2015.01.001

[19] J.F. Ferragut, K. Bentayebi, J.A. Castro, C. Ramon, A. Picornell, Genetic analysis of 12 X-chromosome STRs in Western Mediterranean populations, Int J Legal Med 129(2) (2015) 253-5. 10.1007/s00414-014-1071-4

[20] K. Rebala, S.A. Kotova, V.I. Rybakova, T.V. Zabauskaya, A.A. Shyla, A.A. Spivak, I.S. Tsybovsky, Z. Szczerkowska, Variation of X-chromosomal microsatellites in Belarus within the context of their genetic diversity in Europe, Forensic Sci Int Genet 16 (2015) 105-11. 10.1016/j.fsigen.2014.12.011

[21] L. Caine, S. Costa, M.F. Pinheiro, Population data of 12 X-STR loci in a North of Portugal sample, Int J Legal Med 127(1) (2013) 63-4. 10.1007/s00414-012-0672-z

[22] C. Tomas, V. Pereira, N. Morling, Analysis of 12 X-STRs in Greenlanders, Danes and Somalis using Argus X-12, International Journal of Legal Medicine 126(1) (2012) 121-128. 10.1007/s00414-011-0609-y

[23] A.O. Tillmar, Population genetic analysis of 12 X-STRs in Swedish population, Forensic Sci Int Genet 6(2) (2012) e80-1. 10.1016/j.fsigen.2011.07.008

[24] C. Bini, L.N. Riccardi, S. Ceccardi, F. Carano, S. Sarno, D. Luiselli, S. Pelotti, Expanding X-chromosomal forensic haplotype frequencies database: Italian population data of four linkage groups, Forensic Sci Int Genet 15 (2015) 127-30. 10.1016/j.fsigen.2014.11.008

[25] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Molecular Ecology Resources 10(3) (2010) 564-567. 10.1111/j.1755-0998.2010.02847.x

[26] R. Szibor, S. Hering, J. Edelmann, A new Web site compiling forensic chromosome X research is now online, International Journal of Legal Medicine 120(4) (2006) 252-254. 10.1007/s00414-005-0029-y

[27] B.D. Latter, Selection in finite populations with multiple alleles. 3. Genetic divergence with centripetal selection and mutation, Genetics 70(3) (1972) 475-90.

[28] N. Takezaki, M. Nei, K. Tamura, POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface, Mol Biol Evol 27(4) (2010) 747-52. 10.1093/molbev/msp312

[29] R. Szibor, M. Krawczak, S. Hering, J. Edelmann, E. Kuhlisch, D. Krause, Use of X-linked markers for forensic purposes, International Journal of Legal Medicine 117(2) (2003) 67-74. 10.1007/s00414-002-0352-5

[30] X. She, Z. Jiang, R.A. Clark, G. Liu, Z. Cheng, E. Tuzun, D.M. Church, G. Sutton, A.L. Halpern, E.E. Eichler, Shotgun sequence assembly and recent segmental duplications within the human genome, Nature 431(7011) (2004) 927-30. 10.1038/nature03062

[31] N. Kaya, D. Colak, A. Albakheet, M. Al-Owain, N. Abu-Dheim, B. Al-Younes, J. Al-Zahrani, N.M. Mukaddes, A. Dervent, N. Al-Dosari, A. Al-Odaib, I.V. Kayaalp, M. Al-Sayed, Z. Al-Hassnan, M.J. Nester, M. Al-Dosari, H. Al-Dhalaan, A. Chedrawi, H. Gunoz, B. Karakas, N. Sakati, F.S. Alkuraya, G.G. Gascon, P.T. Ozand, A novel X-linked disorder with developmental delay and autistic features, Annals of Neurology 71(4) (2012) 498-508.

[32] M.L. Butovskaya, O.E. Lazebny, V.A. Vasilyev, D.A. Dronova, D.V. Karelin, A.Z. Mabulla, D.V. Shibalev, T.K. Shackelford, B. Fink, A.P. Ryskov, Androgen Receptor Gene Polymorphism, Aggression, and Reproduction in Tanzanian Foragers and Pastoralists, PLoS One 10(8) (2015) e0136208. 10.1371/journal.pone.0136208

[33] J. delBarco-Trillo, L.K. Greene, I.B. Goncalves, M. Fenkes, J.H. Wisse, J.A. Drewe, M.B. Manser, T. Clutton-Brock, C.M. Drea, Beyond aggression: Androgen-receptor blockade modulates social interaction in wild meerkats, Horm Behav 78 (2016) 95-106. 10.1016/j.yhbeh.2015.11.001

[34] H. Afonso Costa, P. Morais, C. Vieira da Silva, S. Matos, R. Marques Santos, R. Espinheira, J. Costa Santos, A. Amorim, X-chromosome STR markers data in a Cabo Verde immigrant population of Lisboa, Mol Biol Rep 41(4) (2014) 2559-69. 10.1007/s11033-014-3114-9

[35] S. Hering, C. Augustin, J. Edelmann, M. Heidel, J. Dressler, H. Rodig, E. Kuhlisch, R. Szibor, DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases, Int J Legal Med 120(6) (2006) 337-45. 10.1007/s00414-005-0061-y

[36] M. Otter, C.T. Schrander-Stumpel, L.M. Curfs, Triple X syndrome: a review of the literature, Eur J Hum Genet 18(3) (2010) 265-71. 10.1038/ejhg.2009.109

[37] K. Wigby, C. D'Epagnier, S. Howell, A. Reicks, R. Wilson, L. Cordeiro, N. Tartaglia, Expanding the phenotype of Triple X syndrome: A comparison of prenatal versus postnatal diagnosis, Am J Med Genet A 170(11) (2016) 2870-2881. 10.1002/ajmg.a.37688

[38] A. Carracedo, J.M. Butler, L. Gusmao, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, New guidelines for the publication of genetic population data, Forensic Sci Int Genet 7(2) (2013) 217-20. 10.1016/j.fsigen.2013.01.001

[39] A. Carracedo, J.M. Butler, L. Gusmao, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, Update of the guidelines for the publication of genetic population data, Forensic Sci Int Genet 10 (2014) A1-2. 10.1016/j.fsigen.2014.01.004

[40] L. Gusmao, J.M. Butler, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, A. Carracedo, Revised guidelines for the publication of genetic population data, Forensic Sci Int Genet 30 (2017) 160-163. 10.1016/j.fsigen.2017.06.007

Table 1 Forensic parameters.

Linkage Group	Locus	No. of alleles	PIC	PE	PDfemale	PDmale	MEC Krüger	MEC Kishida	MEC Desmarais Duo
LG1	DXS10148	30	0,872603	0,762303	0,975360	0,883723	0,765966	0,872570	0,783675
	DXS10135	35	0,929327	0,863863	0,991558	0,933323	0,864861	0,929306	0,871936
	DXS8378	7	0,632163	0,417471	0,844925	0,692972	0,423151	0,632139	0,486305
LG2	DXS7132	8	0,694091	0,488352	0,887762	0,737336	0,504896	0,694079	0,554401
	DXS10079	13	0,796676	0,636639	0,944287	0,819983	0,646788	0,796664	0,678962
	DXS10074	16	0,812496	0,662387	0,951353	0,833381	0,669797	0,812496	0,699606
LG3	DXS10103	9	0,706566	0,491750	0,899270	0,739366	0,529616	0,706543	0,568508
	HPRTB	10	0,721152	0,525245	0,904084	0,758974	0,540739	0,721140	0,585899
	DXS10101	25	0,895703	0,802640	0,982865	0,903531	0,805530	0,895670	0,818506
LG4	DXS10146	30	0,892907	0,797356	0,982148	0,900948	0,801360	0,892875	0,814508
	DXS10134	33	0,845420	0,714650	0,965826	0,859992	0,723767	0,845386	0,745402
	DXS7423	7	0,655976	0,438535	0,863294	0,706594	0,458952	0,655976	0,512244
Combined			0,99999999812	0,99999817069	0,9999999999999996	0,99999999954	0,99999868605	0,99999999812	0,99999954317

Table 2 Single locus affected cases (10 samples, 1% from the total 995 of analyzed).

		Sample	Locus	Genotype			Croatian Region	Supplementary material 3 Figure	Reference
Triallelic loci	females	JZ26	DXS10134	33	34	38	South	S1	[4]
		JZ101	DXS10146	27	30	31	South	S2	[4]
		SRZ38	DXS10079	14	20	21	Central	S3	[3]
		SRZ91	DXS10079	18	20	23	Central	S4	[3]
Biallelic loci	males	J8	DXS10079		20	21	South	S5	[4]
		J11			20	21	South	S6	[4]
		J57			20	21	South	S7	[4]
		Z111			20	21	West	S8	This study
		J80			20	22	South	S9	[4]
		SR241			21	22	Central	S10	[3]

Table 3 Multiple loci affected cases (four samples, 0.4% from the total of 995 analyzed). Affected loci are marked with "+".

	Sample	LG1			LG 2			LG 3			LG 4			Supplementary material 3 Figure	Croatian Region
		DXS10148	DXS10135	DXS8378	DXS7132	DXS10079	DXS10074	DXS10103	HPRTB	DXS10101	DXS10146	DXS10134	DXS7423		
Triallelic loci in female samples	IZ12					+	+			+				S11	East
	IZ21		+	+	+	+	+			+		+	+	S12	East
	ZZ46	+	+		+				+	+	+			S13	West
Biallelic loci in male sample	Z55		+	+	+	+		+	+		+	+		S14	West

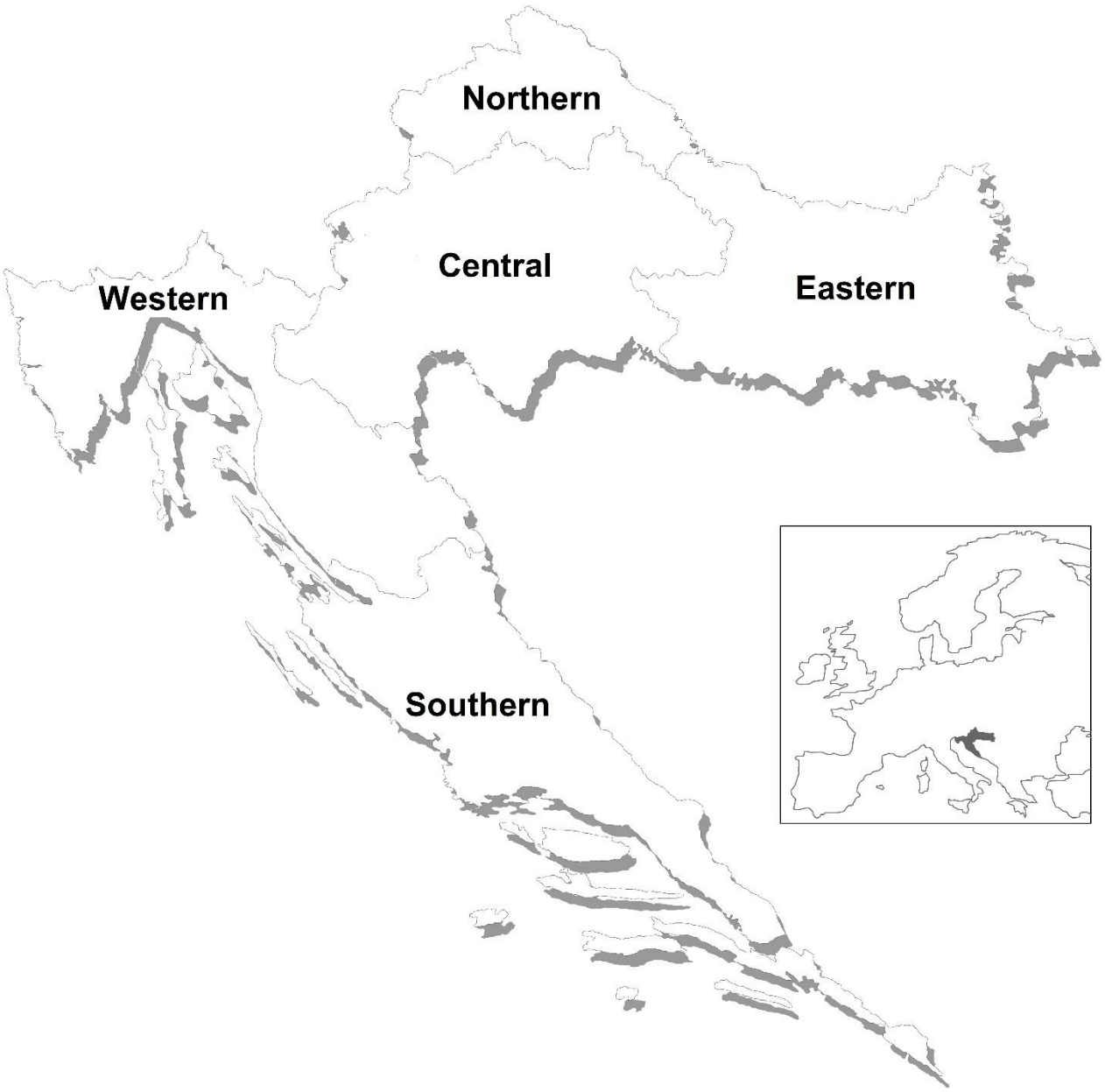

Figure 1. Map of Croatian regions and position of Croatia in European geographical context.

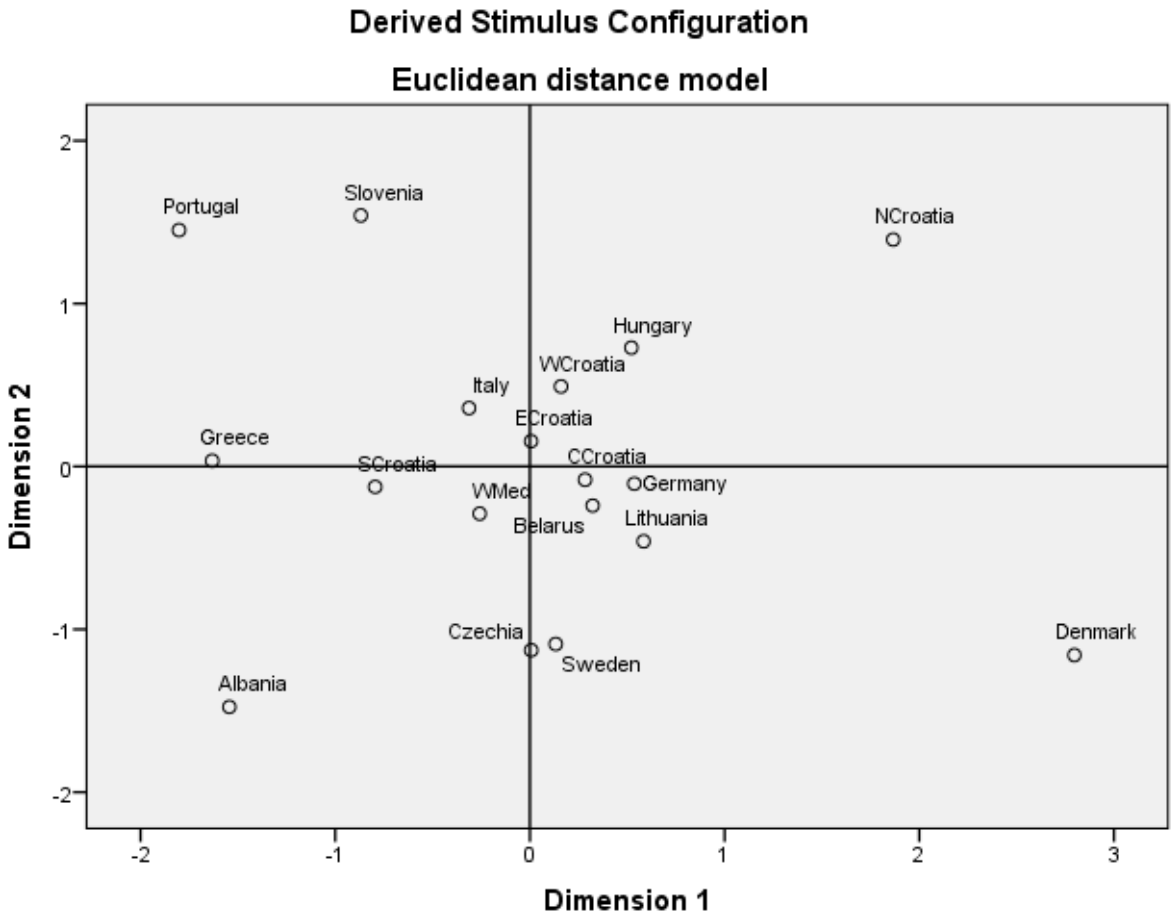

Figure 2. A two-dimensional multidimensional scaling plot drawn from sample bias corrected Fst* genetic distances calculated from the allele frequencies of 12 X-chromosome STRs included in Investigator Argus X-12 kit with the POPTREE2 software. Stress=0.1336/RSQ=0.9287. NCroatia - northern Croatia; SCroatia - southern Croatia; CCroatia - central Croatia; WCroatia - western Croatia; ECroatia - eastern Croatia.