bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genetic diversity and haplotypic structure of a Saudi population sample using Investigator Argus X-12 amplification kit.

Safia A. Messaoudi[1*], Saranya R. Babu[1], Abrar B. Alsaleh[1], Mohamed Albajjah[1], Noora AlSnan[2]

[1]Forensic Biology Department, Naïf Arab University for Security Sciences, Riyadh, 1142, KSA

[2]Forensic Science Laboratory, Directorate of Forensic Science, General Directorate of Criminal Investigation and Forensic Science, Ministry of Interior, Kingdom of Bahrain

[*]Corresponding Author:

Dr Safia A. Messaoudi

Forensic Biology Department, Naif Arab University for Security Sciences, P.O Box: 6830 Riyadh 11452

Riyadh – Kingdom of Saudi Arabia

Email: smassoudi@nauss.edu.sa

Co-Authors emails:

Saranya Babu sramesh@nauss.edu.sa

Abrar Alsaleh aalsaleh@nauss.edu.sa

Mohamed Albajja malbajjah@nauss.edu.sa

Noora Al Snan alsnann@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Background:

X-chromosome short tandem repeat (X-STR) markers have shown a great capability in forensic identity investigations and paternity testing involving kinship analysis.

Material and methods:

In the current study, the distribution of 12 X-STR loci (DXS7132, DXS7423, DXS8378, DXS10074, DXS10079, DXS10101, DXS10103, DXS10134, DXS10135, DXS10146, DXS10148, and HPRTB) located in four linkage groups (LG)s was evaluated using Investigator® Argus X-12 Amplification Kit in 200 unrelated healthy individuals (105 males, and 95 females) from the central region of the Saudi Arabia in order to create a DNA database.

Results:

No significant difference was recorded in the allele frequencies of males and females. Our results indicated that DXS10146 locus was the most informative with 21 alleles while DXS8378 locus was the least with 5 alleles. Hardy–Weinberg equilibrium (HWE) was applied and confirmed for all loci in the female samples, except for DXS10074, DXS10101, DXS10135 and DXS10148 ($p > 0.05/12 = 0.0042$). Forensic parameters showed that all X-STRs loci either as individual markers or as linkage groups provide genetic information with high discrimination that is appropriate for forensic purposes with Paternity Informed Consent (PIC), Power of exclusion (PE), and Paternity index (PI) varied from 0.61211 to 0.917979, 0.38722 to 0.842949, and 0.038416 to 0.16367, respectively. A significant Linkage disequilibrium (LD) with *p*-value after Bonferroni

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

correction $p \leq 0.05/66 = 0.0008$ was observed for 17 pairs of loci in male samples and 4 pairs of loci in female. In the male group, LG3 showed relatively high values of Haplotype diversity (HD). This indicated that these LGs were quite informative in the studied Saudi group and would have high application value in forensic sciences. The pairwise genetic distance fixation index (Fst) results showed that the Saudi population is genetically close to the Egyptian and Emirati populations and distant to the Turkish population. Genetic distances were represented in a non-metric MDS plot representing that Saudis cluster with Middle East populations and are clearly separated from European and East Asian populations.

Conclusion:

The current study revealed that Investigator® Argus 12 X-STR kit would support forensic application, kinship testing involving female offspring, and human identification in Saudi populations.

Keywords: Forensic genetics, DNA typing, X-STRs, haplotypes, Saudi population.

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The use of X-chromosome short tandem repeat (X-STR) markers has been greatly increased in the forensic setting over the last decade (1). They have shown great forensic ability and are, therefore, recognized as efficient tools in forensic investigations of kinship cases (2, 3). A variety of X-STR systems have been developed and added to the list of conventional markers for better results in complex kinship analysis and cases involving mother-son and father/daughter relationship (4-6).

This forensic and human identity testing potential of X-STRs is due to the inheritance pattern of the X-chromosome compared to other genetic markers (2). Any forensically questioned relationship, where at least one female is involved, may benefit from the use of X-STRs which can be applied to cases of missing persons, criminal incest, immigration, deficiency paternity or other similar situation (7).

Since the X-chromosome recombines only in females (4), therefore, X-STRs show stronger linkage disequilibrium (LD) than autosomal STRs.

Szibor *et al*. (4) described four original linkage groups contained the following markers: DXS8378 and DXS9902 in linkage group 1; DXS7132, DXS6789, DXS101, DXS7424, and GATA172D05 in linkage group 2; HPRTB in linkage group 3; and DXS7423 in linkage group 4.

LD between X-STR loci has proven to be population-specific and is further dependent on factors such as random drift, founder effect, mutation rates, selection, and population admixture (8, 9).

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Among haploid markers and in addition to STRs, stable haplotypes of closely associated X-chromosome markers have proven to be a powerful, highly informative and particularly profitable tool in the analysis of complex kinship cases (2, 10).

Saudi Arabia is the largest country in the Arabian Peninsula with a population exceeding 32 million where people are non-uniformly distributed (11). Nearly half of the population resides in two major administrative provinces, Riyadh and Mecca (12). Indigenous Arab people who seem to constitute ~63% of the population are historically organized into geographically-differentiated patrilineal descent groups, or tribes (13), with a known tradition of consanguinity (14). In addition, African and Asian surrounding populations have also influenced the genetic structure of Arab population (15, 16).

The majority of Saudi Arabian Y-chromosome composition has been reported to be of Levantine origins (69%); with significant contributions from the East via Iran (17%), and Africa (14%) (16).

This geographical and social organization might be expected to have an effect on patterns of genetic diversity. However, allele/haplotype frequency data of forensically relevant X-STR markers in Saudi Arabia are extremely limited.

The present study is the first to characterize X-STRs-based genetic diversity of Saudi Arabian population from the Central region using Investigator™ Argus X-12 amplification kit. Twelve gonosomal STR markers were studied to estimate the allelic frequencies and evaluate different forensic and genetic population parameters in a sample of 105 unrelated 105 Saudi male and 95 female volunteers. It is hoped that present investigations will contribute significantly to the existing state of knowledge regarding the X-STR genetic diversity in Saudi population.

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Materials and methods

The present study was undertaken to determine allele frequencies and statistical parameters of forensic interest of 12 X-STR loci (DXS7132, DXS7423, DXS8378, DXS10074, DXS10079, DXS10101, DXS10103, DXS10134, DXS10135, DXS10146, DXS10148, and HPRTB) in a randomly selected cohort of unrelated Saudi individuals from the central region of the kingdom of Saudi Arabia for the purpose of constructing closely linked STR markers groups on the X-chromosome.

Samples collection

Buccal swabs were obtained from 200 natives (up to three generations), unrelated healthy adult, fully informed and consented male (n=105) and female (n=95) volunteers living in the central region of Saudi Arabia. Demographic and past medical history was collected through a standard questionnaire. Individuals having documented genetic disorders, bone marrow transplant, and recent whole blood transfusion were excluded. The study was approved by the Ethical Committee for Scientific Research at the Naif Arab University of Security Sciences, Riyadh, Saudi Arabia.

DNA extraction and quantification

Genomic DNA was isolated from buccal swabs using the Chelex extraction method (17). Extracted DNA samples were quantified following the Quantifiler Duo Human DNA Quantification Kit (Thermo Fisher Scientific, Inc., Waltham, MA, USA) as recommended by the manufacturer in 7500 Real-time PCR System (Thermo Fisher Scientific Company, Carlsbad, USA) to normalize the input DNA quantity in multiplex PCR reaction.

DNA amplification and fragment detection

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The 12 X-STR loci were simultaneously amplified by the multiplex PCR using the Investigator Argus X-12 Amplification Kit (QIAGEN). Multiplex PCR was performed on an Applied Biosystems Veriti Thermal Cycler ((Thermo Fisher Scientific, Inc., Waltham, MA, USA)) in a 10.6 µl reaction volume containing 10 µL of Hi-Di formamide, with appropriate 0.1 µL size standard and 0.5 µL of the amplified PCR product. The samples were denatured for 3 min at 95□ and snap cooled on ice before loading onto the instrument. Amplified fragments were separated by capillary electrophoresis using 3130 Genetic Analyzer ((Thermo Fisher Scientific, Inc., Waltham, MA, USA), following the manufacturer's instructions and size separated with internal lane size standard BTO-5 following manufacturer's protocols (Qiagen N.V., Venlo, Netherlands).

Alleles were assigned according to the International Society of Forensic Genetics (ISFG) guidelines for forensic STR by comparing to the reference allelic ladder included in the kit using GeneMapper ID-X software v.1.2 (Thermo Fisher Scientific, Inc., Waltham, MA, USA). A peak detection threshold of 50 RFUs was used for allele designation.

Forensic and population genetic parameters

Allele frequencies and haplotype frequencies were estimated for each locus using StatsX v2.0 software (18). Fisher's exact tests to evaluate the Hardy–Weinberg equilibrium (HWE) by locus and linkage disequilibrium (LD) between pair of loci were estimated with STRAF - a convenient online tool for STR data evaluation in forensic genetics (19). Mean exclusion chance in Duos (MECD), PD for females (PDf), and PD in males (PDm) were estimated using ChrX-STR.org 2.0 website (20).

Interpopulation pairwise genetic distances based on Fst between central region Saudi population and the rest of populations extracted from the literature were calculated using

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

POPTREE2 software (21) and represented by a nonmetric multidimensional scaling (NM-MDS) analysis using IBM SPSS Statistics v21.0 Software. Phylogenetic tree was constructed from allele frequency data by using the neighbour-joining method (22) *via* MEGA X: Molecular Evolutionary Genetics Analysis (23).

Results and Discussion:

Intrapopulation Genetic diversity

In the present study, allele frequencies, observed heterozygosity (Ho), expected heterozygosity (He) and *p*-value for Hardy-Weinberg equilibrium (HWE) were determined for the twelve X-chromosomal Short Tandem Repeat (STR) markers DXS8378, DXS10146, DXS7423, DXS10134, HPRTB, DXS10101, DXS10103, DXS10079, DXS10074, DXS7132, DXS10148 and DXS10135 within DNA samples from the population of central Saudi Arabia.

Out of 200 samples, 146 unique alleles were observed with the allele frequencies ranging from 0.0034 to 0.4542. Allele frequencies for each of the 12 X-STR markers in the population groups (male, female) are shown in Table 1. There were no significant differences in the distribution of allele frequencies between male and female samples ($p > 0.05$) using the exact test, which proved that these X-STR loci had no gender bias of allele distribution in the studied group. Therefore, we pooled male and female samples together for calculating forensic parameters.

The most polymorphic locus was DXS10146 with 21 alleles while DXS8378 was the least polymorphic locus with 5 alleles. The highest Gene Diversity (GD) was observed for locus DXS10135 with 0.9937 while the smallest for DXS7423 with a GD value of 0.661355 (Table 1)

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

HWE test was performed on female samples. All loci were in HWE, in the female Saudi population, with the exception of DXS10074, DXS10101, DXS10135 and DXS10148 even after Bonferroni correction ($p > 0.05/12 = 0.0042$) (Table 1). This deviation could be due to the consanguinity high rate and the smaller sample size.

Forensic efficiency

Based on allele frequencies of the pooled data, we further determined the statistical parameters of forensic interest. The most informative locus was DXS10135 with 20 alleles while DXS8378 was lowest with 5 alleles. For the present studied population, the PIC, PE, and PI varied from 0.61211 to 0.917979, 0.38722 to 0.842949, and 0.038416 to 0.16367, respectively (Table 1).

The combined values of each forensic parameter, ${}_{c}PD_{M}$, ${}_{c}PD_{F}$, ${}_{c}$MECK Rüger, cMEC Kishida, and cMEC Desmarais as well as ${}_{c}$MEC Desmarais Duo, were 0.999999987962839, 0.999999999999998, 0.999999701236302, 0.999999972467329, 0.999999987434766 and 0.999999813797358, respectively (Supplementary Table).

The present results indicate that all X-STRs loci either as individual markers or as linkage groups provide genetic information with high discrimination that is appropriate for forensic purposes. The high values of combined PDm and combined PDf, as well as high values of combined MEC in deficiency cases, normal trios, and duo indicated that these 12 X-STR loci of the Investigator® Argus X-12 kit would be suitable package as a complement of autosomal STRs for forensic identity and paternity testing involving kinship analysis when the offspring is female or investigating father-daughter relationships in the Saudi population.

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Linkage disequilibrium analysis

Tests of linkage disequilibrium (LD) were performed for female and male samples for all pairs of loci in the group and the results were presented in Table 2 and 3. Significant LD with p-value ≤ 0.05 was observed for 19 pairs of loci in female and 29 pairs of loci in male samples. Moreover, and in order to improve the level of significance, Bonferroni correction was applied and significant results with $p \leq 0.05/66$ (0.0008) were observed in 4 pairs of loci in female; DXS10103-DXS10148, DXS10146-DXS7132, HPRTB-DXS10079 and HPRTB-DXS10134 and 17 pairs of loci in male; DXS7132 showed the highest number of LD with the following 6 markers: DXS8378, DXS7423, DXS10103, DXS10079, DXS10135 and DXS10148, while DXS10134 showed LD with only one marker DXS10074. DXS7423 and HPRTB loci were in LD with DXS10146, DXS10148, DXS10079, DXS10103 and DXS10148, DXS10101, DXS8378, respectively. DXS8378 was in LD with DXS10148 and DXS10101 as well as with DXS7132 and HPRTB.

The 12 X-STR markers are clustered into four linkage groups based on their physical localizations (Diegoli, 2015; Szibor et al., 2003): LG1 (Xp22) contains DXS8378-DXS10135- DXS10148, LG2 (Xq11) contains DXS7132-DXS10074-DXS10079, LG3 (Xq26) contains DXS10101-DXS10103-HPRTB and LG4 (Xq28) contains DXS7423-DXS10134-DXS10146.

In male group, a significant association was identified between the pairwise LD and the markers located in the same LG (Table 3). The tight linkage between the X-STR trios in the four linkage groups are as follows: LG1 (DXS10148 and DXS8378), LG2 (DXS7132 and DXS10079), LG3 (DXS10101 and HPRTB) and LG4 (DXS7423 and DXS10146).

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Whereas, no pairwise LD was found among the four LGs for the female group (Table 2). The lack of association between some of the markers inside linkage groups has advantages as it contributes to the high PDs value of this multiplex in the present population (Gomes et al., 2017).

Linkage expectations were only based on physical distances between loci, but LD may also result from random genetic drift, founder effects, mutations, selection and population admixture or stratification (24, 25).

Previous studies have demonstrated that loci inside each of the four X-STR linkage groups show reduced, but non-zero recombination rates (26). In addition, recombination between linkage groups is less than 50% (6, 27, 28). In fact, Diegoli et al. (2016)(6) argues against a simple model of independent X chromosomal linkage groups for these reasons.

Haplotypic structure

Haplotype diversity (HD) of four LGs composed of the closely linked X-STRs were tested in the male group (n=105 samples) and listed in Table 4. Each cluster of three X-STR markers is considered as one haplotype. The numbers of observed haplotypes in each of the 4 linkage groups- LG1, LG2, LG3, LG4 were 70, 69, 81 and 69, respectively, while HD values were 0.99063, 0.99042, 0.99353, and 0.990787, respectively.

The LG3 showed relatively high values of HD compared to other LGs. This indicated that these linkage groups were quite informative in the studied Saudi group and would have high application value in forensic sciences.

The three most common haplotype for LG1 was 12-29-26.1 displaying 4.76 % of haplotype frequency, in LG2 two sets of haplotypes 12-8-20 and 16-15-20 was observed

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

each with a frequency of 3.8%. LG3 and LG4 presented haplotypes 29.2-19-13 and 15-34-45.2 with 3.8% and 4.76% of haplotype frequency respectively.

Interpopulation diversity

To our knowledge, this is the first report from Saudi Arabia on these 12 markers, so no Saudi data is available for comparison. In the present study, Saudi allele frequencies were compared with eight worldwide populations from the literature to elucidate the genetic relationships between Saudis, Emiratis (29), Egyptians (30), Turkish (31), Algerians (32), Jewish (33), Filipinas (34), Bangladeshis (35) and Indians (36) using the population pairwise genetic distance fixation index (Fst). Fst was used to perform the comparison, with differences being statistically significant for *p*-values < 0.001 after Bonferroni correction (Table 5).

The results of population Fst for 12 X-STRs was substantial, ranging from 0.004 to 0.052 indicating differences in allele frequencies between the studied populations that varied geographically, linguistically and socially.

The comparison with published data showed that Saudi population in this study had lower pairwise Fst values with those populations that are geographically close, Egypt and UAE and higher genetic distance with the Turkish population from different continental origin (Turkey, 2018) (Table 5). Therefore, North African and Central Asian influences seem to be much higher in Saudi Arabia than in East Asia

Sample bias-corrected Fst distances were obtained and were represented in multidimensional scaling (MDS) plots (Figure 1). In the plot, Saudis cluster with Middle East populations; Egyptians, UAE and Jewish while Algerian population is found be

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

close to the Middle Eastern population and they are clearly separated from European Turkish; and East Asian populations.

As shown, Saudi, Egyptian, UAE, Jewish and Algerian populations are positioned in the upper right cluster, whereas Indian, Bangladeshi and Filipina populations were clustered together. Turkish population was found in separate clusters apart.

NJ (neighbour-joining) tree (Figure 2) has also been generated using Nei's Fst distances of X-STR loci estimated among the 9 populations. This tree graphically represents genetic distance among these 9 populations based on the genotyped 12 X-STR loci. The NJ tree showed that the Saudi population is in close relationship with the Middle Eastern populations (Egyptian, Emirati, and Jewish).

The comparison showed a genetic diversity between Saudi population, European and East Asian populations which may be due to geographic and continental origin distinction. Therefore, it is important to develop population data for forensic analysis.

Conclusion

The current study was reported to be the first literature exploring the allele frequency and forensic parameters of the 12 X-chromosome STR markers in the Saudi population. The results revealed that X-STRs analysis showed high $_{c}$PD and $_{c}$MECs providing extreme advantages in kinship testing involving female offspring, forensic casework, and human identification. Further, to accomplish the detailed understanding of the genetic structure, intra and inter-population relationships, larger sample sizes must be included from a wider geographic area.

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Gomes C, Magalhães M, Alves C, Amorim A, Pinto N, Gusmão L. Comparative evaluation of alternative batteries of genetic markers to complement autosomal STRs in kinship investigations: autosomal indels vs. X-chromosome STRs. International journal of legal medicine. 2012;126(6):917-21.
2. Szibor R. X-chromosomal markers: past, present and future. Forensic Science International: Genetics. 2007;1(2):93-9.
3. Messoussi M, Prieto-Fernández E, Baeta M, Núñez C, Ammar-El Gaaied AB, de Pancorbo MM, et al. Genetic variation of 17 X-chromosome STR loci in Tunisian population of Nabeul. International journal of legal medicine. 2018:1-4.
4. Szibor R, Hering S, Kuhlisch E, Plate I, Demberger S, Krawczak M, et al. Haplotyping of STR cluster DXS6801–DXS6809–DXS6789 on Xq21 provides a powerful tool for kinship testing. International journal of legal medicine. 2005;119(6):363-9.
5. A. Barbaro PC, A. Barbaro. X-STR typing for an identification casework. International Congress Series. 2006;1288:513-5.
6. Diegoli TM, Rohde H, Borowski S, Krawczak M, Coble MD, Nothnagel M. Genetic mapping of 15 human X chromosomal forensic short tandem repeat (STR) loci by means of multi-core parallelization. Forensic Science International: Genetics. 2016;25:39-44.
7. Diegoli TM, Coble MD. Characterization of X Chromosomal Short Tandem Repeat Markers for Forensic Use. 2015.
8. Lee HY, Park MJ, Chung U, Lee HY, Yang WI, Cho SH, et al. Haplotypes and mutation analysis of 22 Y-chromosomal STRs in Korean father-son pairs. International journal of legal medicine. 2007;121(2):128-35.
9. Pereira R, Gomes I, Amorim A, Gusmao L. Genetic diversity of 10 X chromosome STRs in northern Portugal. International journal of legal medicine. 2007;121(3):192-7.
10. Hering S, Augustin C, Edelmann J, Heidel M, Dressler J, Rodig H, et al. DXS10079, DXS10074 and DXS10075 are STRs located within a 280-kb region of Xq12 and provide stable haplotypes useful for complex kinship cases. International journal of legal medicine. 2006;120(6):337-45.
11. Khubrani YM, Wetton JH, Jobling MA. Analysis of 21 autosomal STRs in Saudi Arabia reveals population structure and the influence of consanguinity. Forensic Sci Int Genet. 2019;39:97-102.
12. Arabia GAfSiS. Demographic Research Bulletin. 2016 [cited 2019 24/4]; Available from: https://www.stats.gov.sa/.
13. Al-Rasheed M. A History of Saudi Arabia. . Cambridge Cambridge University Press.; 2010.
14. Al-Gazali L, Hamamy H, Al-Arrayad S. Genetic disorders in the Arab world. Bmj. 2006;333(7573):831-4.
15. Abu-Amero KK, Gonzalez AM, Larruga JM, Bosley TM, Cabrera VM. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evol Biol. 2007;7:32.
16. Abu-Amero KK, Hellani A, Gonzalez AM, Larruga JM, Cabrera VM, Underhill PA. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. BMC Genet. 2009;10:59.
17. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques. 1991;10(4):506-13.

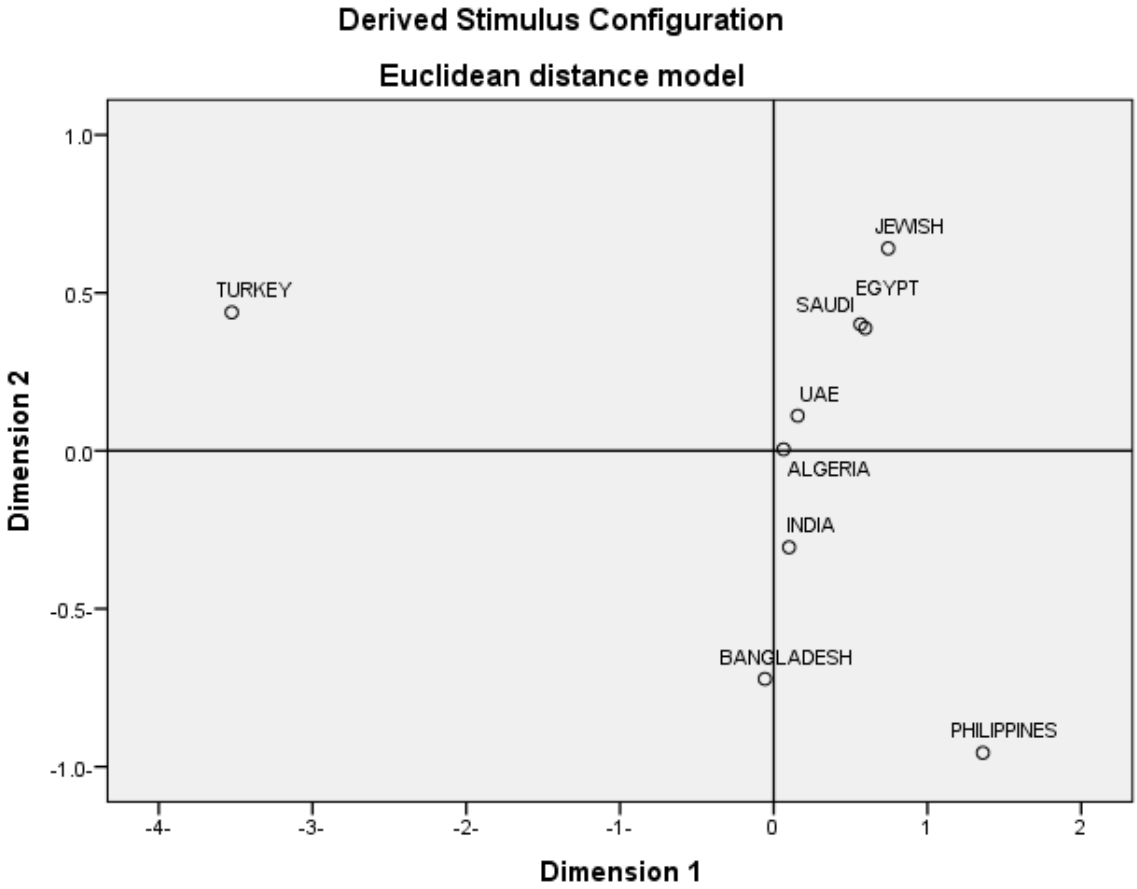
bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

18. Lang Y, Guo F, Niu Q. StatsX v2.0: the interactive graphical software for population statistics on X-STR. International journal of legal medicine. 2019;133(1):39-44.
19. Gouy A, Zieger M. STRAF-A convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet. 2017;30:148-51.
20. ChrX-STR.org. ChrX-STR.org 2.0. [cited 2019 14 march]; Available from: http://www.chrx-str.org/.
21. Takezaki N, Nei M, Tamura K. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. Mol Biol Evol. 2010;27(4):747-52.
22. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406-25.
23. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547-9.
24. Pereira V, Gusmão L, Valente C, Pereira R, Carneiro J, Gomes I, et al. Refining the genetic portrait of Portuguese Roma through X-chromosomal markers. American journal of physical anthropology. 2012;148(3):389-94.
25. Chakravarti A. Population genetics—making sense out of sequence. Nature genetics. 1999;21(1s):56.
26. Castañeda M, Mijares V, Riancho JA, Zarrabeitia MT. Haplotypic Blocks of X-linked STRs for Forensic Cases: Study of Recombination and Mutation Rates. Journal of forensic sciences. 2012;57(1):192-5.
27. Nothnagel M, Szibor R, Vollrath O, Augustin C, Edelmann J, Geppert M, et al. Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome. Forensic Science International: Genetics. 2012;6(6):778-84.
28. Pamjav H, Kugler R, Zalán A, Völgyi A, Straky Z, Endrédy P, et al. X chromosomal recombination study in three-generation families in Hungary. Forensic Science International: Genetics. 2012;6(3):e95-e6.
29. Almarri MA, Lootah RA. Allelic and haplotype diversity of 12 X-STRs in the United Arab Emirates. Forensic Science International: Genetics. 2018;33:e4-e6.
30. Elakkary S, Hoffmeister-Ullerich S, Schulze C, Seif E, Sheta A, Hering S, et al. Genetic polymorphisms of twelve X-STRs of the investigator Argus X-12 kit and additional six X-STR centromere region loci in an Egyptian population sample. Forensic Science International: Genetics. 2014;11:26-30.
31. Acar E, Bulbul O, Rayimoglu G, Shahzad M, Argac D, Altuncul H, et al. Optimization and validation studies of the Mentype® Argus X-8 kit for paternity cases. Forensic Science International: Genetics Supplement Series. 2009;2(1):47-8.
32. Bekada A, Benhamamouch S, Boudjema A, Fodil M, Menegon S, Torre C, et al. Analysis of 21 X-chromosomal STRs in an Algerian population sample. International journal of legal medicine. 2010;124(4):287-94.
33. Ferragut J, Bentayebi K, Pereira R, Castro J, Amorim A, Ramon C, et al. Genetic portrait of Jewish populations based on three sets of X-chromosome markers: Indels, Alu insertions and STRs. Forensic Science International: Genetics. 2017;31:e5-e11.
34. Salvador JM, Apaga DLT, Delfin FC, Calacal GC, Dennis SE, De Ungria MCA. Filipino DNA variation at 12 X-chromosome short tandem repeat markers. Forensic Science International: Genetics. 2018;36:e8-e12.
35. Sufian A, Hosen MI, Fatema K, Hossain T, Hasan MM, Mazumder AK, et al. Genetic diversity study on 12 X-STR loci of investigator® Argus X STR kit in Bangladeshi population. International journal of legal medicine. 2017;131(4):963-5.

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

36. Shrivastava P, Jain T, Gupta U, Trivedi VB. Genetic polymorphism study on 12 X STR loci of investigator Argus X STR kit in Bhil tribal population of Madhya Pradesh, India. Legal Medicine. 2015;17(3):214-7.

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



For matrix, Stress=0.07098 and Squared Correction (RSQ)=0.98850

Figure 1: Multidimensional scaling (MDS) plots of the Central Saudi population and other 8 populations was built using IBM SPSS Statistics v21.0 software based on the Nei's genetic distances.

bioRxiv preprint doi: https://doi.org/10.1101/760819; this version posted September 8, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

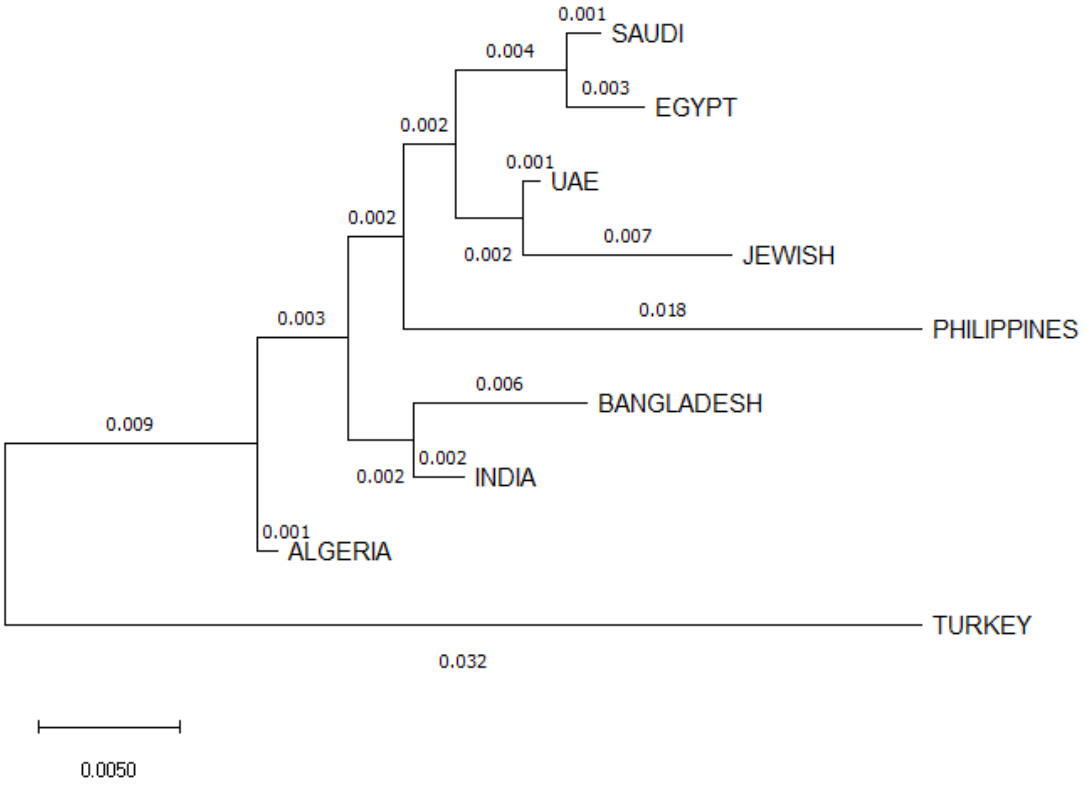


Figure 2: Neighbouring joining tree based on Nei's Fst Distances for the 12 X-STR loci estimated among the 9 populations. The nods are bootstrap values derived from 1000 replications.