
Title: Population genetics data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit

Author: Hussain M Alsafiah

PII: S1872-4973(17)30202-8
DOI: https://doi.org/10.1016/j.fsigen.2017.09.014
Reference: FSIGEN 1788

To appear in: *Forensic Science International: Genetics*

Authors: William H Goodwin, Sibte Hadi

PII: S1872-4973(17)30202-8
DOI: https://doi.org/10.1016/j.fsigen.2017.09.014
Reference: FSIGEN 1788

To appear in: *Forensic Science International: Genetics*

Author: Mohammed A. Alshaikhi

PII: S1872-4973(17)30202-8
DOI: https://doi.org/10.1016/j.fsigen.2017.09.014
Reference: FSIGEN 1788

To appear in: *Forensic Science International: Genetics*

Author: Pet-Paul Wepeba

PII: S1872-4973(17)30202-8
DOI: https://doi.org/10.1016/j.fsigen.2017.09.014
Reference: FSIGEN 1788

To appear in: *Forensic Science International: Genetics*

Received date: 14-7-2017

Genetics https://doi.org/10.1016/j.fsigen.2017.09.014

# Population genetics data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit.

Hussain M Alsafiah[1,2], William H Goodwin [2], Sibte Hadi [2], Mohammed A Alshaikhi [3], Pet-Paul Wepeba [2].

[1] *Forensic Genetics Laboratory, General Administration of Criminal Evidences, Public Security, Ministry of Interior, Kingdom of Saudi Arabia.*

[2] *School of Forensic and Applied Sciences, University of Central Lancashire, Preston, United Kingdom*

[3] *Forensic Medicine Department, Security Forces Hospitals Programme/Dammam, Ministry of Interior, Kingdom of Saudi Arabia.*

**Highlights**:

- GlobalFiler® PCR amplification kit was used to generate a reference dataset for the population of Saudi Arabia.
- SE33, D12S391, and D1S1656 in this kit are more informative for the population of Saudi Arabia than any locus in the AmpFlSTR® Identifiler® PCR amplification kit.

**Keywords:** STRs; Population genetics data; Arabian Peninsula; Saudi Arabia; GlobalFiler® PCR amplification kit.

**Dear Editor,**

Saudi Arabia, in the southwest region of Asia, occupies the majority of the Arabian Peninsula. It shares borders with eight Arab countries: Bahrain, Qatar, and the United Arab Emirates (UAE) to the east; Oman and Yemen to the south; Jordan and Iraq to the north and Kuwait to the northeast. Saudi Arabia is divided into 13 administrative provinces, and as of the 2016 census, its population is 31,742,308 (20,064,970 are Saudi and 11,677,338 are non-Saudi). Half of the population resides in two administrative provinces of Riyadh and Mecca [1].

Saudi Arabia is an Arab country where African and Asian surrounding populations have influenced the genetic structure of its population [2,3]. The majority of Saudi Arabian Y-chromosome composition was estimated to be of Levantine origins (69%); with significant contributions from the east via Iran (17%), and Africa (14%) [3].

The intermediate location between Africa and Asia, and the coastal borders of the Red Sea and the Gulf Sea, have facilitated migrations between Africa and Asia, and trading between neighbouring areas. In addition, the Arabian Peninsula is connected to the Levant by long landlocked area that has contained important routes for trading caravans and migration. The movement of people has been enhanced over the last few centuries through the presence of the holy cities of Mecca and Al-Madinah, which have received millions of Muslims performing the Haj for more than 1,400 years, some of whom have remained for many generations.

The first forensic genetics laboratory in the Criminal Evidences Administration was established in 1991, in the capital city of Riyadh. Due to the contribution of the first Saudi Forensic Genetics Laboratory to fighting terrorism and solving crime, another 12 forensic genetics laboratories have been established. Currently, ten of the laboratories are accredited to ISO17025:2005. AmpFlSTR® Identifiler®Plus is the standard STR kit in Saudi Arabia that provides a typical match probability of $2.2278 \text{ X } 10^{-18}$. D19S433 is the most informative locus, and TPOX is the least informative locus.

To date, four studies have described the genetic diversity of forensic STRs in the Saudi population. The first was a study of 207 samples with eight STR loci [4]; another two studies investigated 13 STR loci in Saudi individuals residing in Dubai (94 samples) [5] and 15 STR loci in individuals residing in Kuwait (250 samples) [6]. The most recent study was carried out in 2015, testing 190 individuals from the Riyadh province using the AmpFlSTR® Identifiler® PCR amplification kit [7].

However, Saudi individuals residing in Dubai, Kuwait or even in the Riyadh province are not necessarily representative of the entire population of Saudi Arabia. In addition, the high percentage of consanguinity (56%) is a major factor in shaping the genetic structure of the Saudi population [8,9]. This increases the importance of studying the genetic diversity of the Saudi population to evaluate to what extent new STR markers can be utilised for crime scene and kinship cases.

Therefore, this study aims to examine additional STR loci by using the GlobalFiler® PCR amplification kit (ThermoFisher, USA), and to enlarge the sample size (previous studies only used at most 250 samples). This study received ethical approval from the Security Forces Hospitals Programme in Saudi Arabia and from the Ethics Committee of the University of Central Lancashire, UK.

Blood samples spotted on FTA cards (Whatman, UK) were collected from 500 unrelated representatives of the population of Saudi Arabia. Samples were collected from six cities in different provinces of Saudi Arabia: Mecca, Al-Madinah, Riyadh, Dammam, Tabuk and Abha. DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, Germany) from five (2 mm diameter) punches per sample. DNA concentrations were estimated using a Qubit® Fluorometer 3.0 with the Qubit® dsDNA HS Kit (ThermoFisher, USA).

Twenty-one autosomal STR loci, a Y-STR locus (DYS391), a Y-indel locus (Yq11.221) and amelogenin were amplified using an internally validated 12.5 µL reaction (3.75 µL Master Mix & 1.25 µL Primer Set) and 7.5 µL (0.5 ng) of extracted DNA using Veriti thermal cycler (ThermoFisher, USA) following the 29-cycle protocol. The 21 autosomal STRs were D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S391 and D2S1338.

PCR amplicons and an internal Size Standard dye of 600 LIZ™ v2 were separated using an ABI 3500 DNA Genetic Analyser (ThermoFisher, USA) following the manufacturer guidelines, except that POP-6™ together with 50 cm capillaries were utilised. The run time was increased in the Data Collection Software v3 (ThermoFisher, USA) from 1550 to 3500 s, which was validated internally.

Following the published nomenclatures and the guidelines of the International Society for Forensic Genetics (ISFG) [10], alleles from the 21 STRs were identified using GeneMapper™ *ID-X* Software v1.2 and an allelic ladder (ThermoFisher, USA).

PowerStat v 1.2 (Promega, USA) was employed for calculating statistical parameters: power of discrimination, power of exclusion, matching probability, polymorphism information content, observed homozygosity and observed heterozygosity (Supplementary Table 1). Arlequin v3.5.2.1 Software was used to calculate the expected heterozygosity, to conduct a Hardy-Weinberg equilibrium exact test (Supplementary Table 2) and to perform a population differentiation exact test (Supplementary Table 3) [11].

The 21 STRs combined have a $1.42091 \times 10^{-26}$ combined match probability, a 0.99999999999999999999999986 combined power of discrimination (PoD) and a 0.999997405 combined power of exclusion. Most of the 21 STRs (18/21) show ≥ 0.9 PoD; SE33 is the most informative locus with 0.993 PoD, and TPOX is the least informative locus with 0.84 PoD. Allele ranges varied from 6 alleles in

TH01 to 44 alleles in SE33; some alleles show very high frequencies in the Saudi population; for example, allele 8 in the TPOX and allele 15 in the D22S1045 scored the highest frequencies of 0.520 and 0.463 respectively. The observed heterozygosity ranged from 0.660 in the TPOX to 0.914 in the SE33 (Supplementary Table 1).

Eight allelic ladder variants were detected at the SE33: allele 7.3 (10 samples), allele 13.3 (two samples), allele 17.2 (one sample), allele 22 (8 samples), allele 23 (3 samples), allele 28 (one sample), allele 33 (two samples) and allele 34 (5 samples). In addition, three variants at the D1S1656—allele 7 (one sample), allele 8 (one sample), and allele 18 (3 samples)—were also detected; all these have already been observed and are reported in STRBase [12].

Non-reported variants were also detected in SE33: allele 14.3 (one sample), allele 20.3 (one sample), allele 33.1 (one sample) and allele 38 (one sample).

One sample showed three alleles (9, 12, OL) at the D7S820, and showed homozygous allele (16) at the SE33 (Supplementary Figure 1A). This suggests that the OL allele belongs to the D7S820 forming a triplet allele phenomenon, or is an unusual short allele belonging to the SE33 forming a heterozygote genotype (OL,16). Therefore, the D7S820 was genotyped again by the PowerPlex® 21 System (Promega, USA) following the manufacturer guidelines, and gave only two alleles (9, 12) (Supplementary Figure 1B). This demonstrates that the OL allele belongs to the SE33 locus and because of the adjacent locations of the D7S820 and the SE33 in the GlobalFiler® PCR amplification kit, the OL allele appeared within the allelic window of D7S820.

Based on the sizes of the OL allele (296.85 bp) and allele 4.2 (306.55 bp) in the allelic ladder (Supplementary Figure 2), the OL allele was designated as allele 2 which has not been reported in STRBase [12]. Therefore, the genotype of this sample at the SE33 was designated (2, 16) rather than (16, 16).

Although the D18S51, D2S441, D22S1045, D7S820 and the SE33 have shown deviation from the Hardy–Weinberg equilibrium (HWE) ($P$-value < 0.05), no significant deviation was detected after applying Bonferroni correction ($P$-value < 0.002) (Supplementary Table 2).

After the Bonferroni correction the population differentiation exact test showed that the data of the Saudi populations previously reported in [4,6] is consistent with the data reported in this study, i.e. no significant pairwise differences were observed.

However, the data of the Saudi population in Dubai [5], showed significant difference in the TH01 locus ($P$-value = 0), which was due in part to the notable

differences in alleles frequencies at this locus. For example, allele 7 had 0.179 frequency in the current study while it had 0.08 frequency in [5], which is over 2-fold higher. This inconsistency may be due to the small number of Saudi participants (94 samples) in this study leading to an exaggerated sampling effect.

There were also significant differences with the data from the Riyadh province [7], at three loci (vWA, CSF1PO and TH01). Despite the relatively small sample size (190 samples), alleles 5.3, 7.3, and 8.3 at TH01 locus were observed which have not been observed in the current study or in previous studies of the Saudi population. In addition, this study found that 9 out of 15 loci had significant deviation from HWE (P-value < 0.05), which was attributed to the prevalence of consanguinity in Saudi Arabia. The general percentage of consanguinity in the Riyadh province is 60%, which is higher than the average rate of Saudi Arabia (56%), and is even higher (74.3%) in rural areas [9].

As expected, the differentiation between the data obtained in this study and the data from the Yemeni, Omani [5], Kuwaiti, Egyptian, Iraqi, Iranian, Indian [6], Qatari [13] and UAE populations [14] varies, with a general trend of more significant differences being detected as the populations become more geographically separated. For example, there was no significant difference observed between the Saudi and the Kuwaiti populations whereas there were significant differences between the Indian and the Saudi populations in 13 out of the 15 loci compared (Supplementary Table 3).

This study has shown that the 21 STR loci of the GlobalFiler™ PCR Amplification Kit, as expected, provided a much higher discrimination power in the population of Saudi Arabia compared to the currently used STR kit. Three of the additional STR loci (SE33, D12S391, and D1S1656) in this kit are more informative than any locus in the AmpFlSTR® Identifiler® PCR amplification kit. Therefore, adapting this kit in the forensic genetics laboratories of Saudi Arabia as a standard STR kit would be very beneficial. Data of this study can be requested from HMHAlsafiah@uclan.ac.uk. or WHGoodwin@uclan.ac.uk.

**Conflicts of Interest**

The authors declare that there are no conflicts of interest.

**Acknowledgements**

We would like to thank the Security Forces Hospitals Programme and staff for allowing us to utilise its services for sampling. This study has been funded by Saudi Arabian Cultural Bureau in London.

**References**

[1] General Authority for Statistics in Saudi Arabia, Demographic Research Bulletin,https://www.stats.gov.sa/sites/default/files/en-demographic-research-2016_4.pdf, 2016 (accessed 15.05.17).

[2] K. Abu-Amero, A. Gonzalez, J. Larruga, T. Bosley, V. Cabrera, Eurasian and African mitochondrial DNA influences in the Saudi Arabian population, BMC Evol.Biol. 7 (2007) 32-46.

[3] K. Abu-Amero, A. Hellani, A. Gonzalez, J. Larruga, V. Cabrera, P. Underhill, Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions, BMC Genet. 10 (2009) 59-67.

[4] S. Sinha, M. Amjad, C. Rogers, J.E. Hamby, U.A. Tahir, K. Balamurugan, et al., Typing of eight short tandem repeat (STR) loci in a Saudi Arabian population, Forensic Sci.Int. 104 (1999) 143-146.

[5] F. Alshamali, A.Q. Alkhayat, B. Budowle, N.D. Watson, STR population diversity in nine ethnic populations living in Dubai, Forensic Sci.Int. 152 (2005) 267-279.

[6] M. Al-Enizi, J. Ge, S. Ismael, H. Al-Enezi, A. Al-Awadhi, W. Al-Duaij, et al., Population genetic analyses of 15 STR loci from seven forensically-relevant populations residing in the state of Kuwait, Forensic.Sci.Int.Genet. 7 (2013) e106-e107.

[7] A. Osman , H. Alsafar, G. Tay, J. Theyab,M. Mubasher, N. Eltayeb-El Sheikh, H. AlHarthi, M. Crawford, G. El Ghazali, Autosomal Short Tandem Repeat (STR) Variation Based on 15 Loci in a Population from the Central Region (Riyadh Province) of Saudi Arabia, J Forensic Res. 6 (2015) 267-271.

[8] A. Warsy, M. Al-Jaser, A. Albdass, S. Al-Daihan, M. Alanazi, Is consanguinity prevalence decreasing in Saudis?: A study in two generations, Afr.Health.Sci. 14 (2014) 314-321.

[9] M. El-Mouzan, A. Al-Salloum, A. Al-Herbish, M. Qurachi, A. Al-Omar, Regional variations in the prevalence of consanguinity in Saudi Arabia, Saudi Med.J. 28 (2007) 1881-1884.

[10] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci.Int. 165 (2007) 238-243.

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

[12] C. Ruitberg, D. Reeder, J. Butler, STRBase: a short tandem repeat DNA database for the human identity testing community, Nucleic Acids Res. 29 (2001) 320-322.

[13] A. Perez-Miranda, M. Alfonso-Sanchez, J. Pena, R. Herrera, Qatari DNA variation at a crossroad of human migrations, Hum.Hered. 61 (2006) 67-79.

[14] R. Jones, W. Tayyare, G. Tay, H. Alsafar, W. Goodwin, Population data for 21 autosomal short tandem repeat markers in the Arabic population of the United Arab Emirates, Forensic.Sci.Int.Genet. 28 (2017) e41-e42.