Egyptian Journal of Forensic Sciences

ORIGINAL ARTICLE Open Access

Population data and genetic diversity analysis of 17 Y-STR loci in Saudi population

Ahmed Ch. Kassab, Hissah Fahad Mohammed Alaqeel, Safia A. Messaoudi, Saranya R. Babu, Sajjad Ali Shahid and Abdul Rauf Chaudhary*

Abstract

Background: The Y chromosome polymorphism has been widely studied for human migrations, population genetics, forensic applications, and paternity analysis. However, studies regarding genetic lineage and population genetic structure of the Y chromosome in different regions of Saudi Arabia are limited.

Aim: This study aimed to analyze the distribution of Y chromosome haplotypes in a sample of 125 native Saudi males from different geographic regions of Saudi Arabia and compare to previously published Y chromosome haplotype data from Saudi Arabia and some neighboring Arab populations.

Materials and methods: Buccal swabs were collected from 125 healthy unrelated native Saudi males from different geographic regions of Saudi Arabia. Genomic DNA was extracted by Chelex®100; 17 Y-STR loci were amplified using the AmpFℓSTR Yfiler PCR amplification kit and detected on the 3130 Genetic Analyzer™. Allele frequency and gene diversity were calculated with online tool STRAF. The Saudi population data were compared with the neighboring populations using pairwise genetic distances and associated probability values were calculated using the Y Chromosome Haplotype Reference Database Website (YHRD) software.

Results and conclusion: One hundred six YSTR haplotypes and 102 YSTR alleles (excluding 4 null alleles) were identified having a discrimination capacity (DC) of 85.8%. The highest haplotype diversity (HD) and gene diversity (GD) were observed at the loci DYS 458 (0.817) and DYS385b (0.807), respectively. According to our results, the Iraqi and Qena (Egypt) populations appeared to have closer relatedness to the Saudi population as compared with Yemen. The UAE and Kuwait populations showed the same degree of relatedness to the Saudi population followed by Bahrain. On the contrary, the Adnanit and Qahtanit populations of Jordan demonstrated low genetic distance from the Saudi population. In short, studying a population sample of pure Saudi ethnicity enabled us to identify a unique set of haplotypes which may help in establishing genetic relatedness between Saudi and the neighboring Arab populations. The present paper, therefore, highlights the importance of ensuring ethnic originality of the study sample while conducting population genetics studies.

Keywords: Y chromosome STRs, Saudi Arabia, Forensic genetics, Arab population

* Correspondence: AKassab@nauss.edu.sa
Department of Forensic Sciences, College of Criminal Justice, Naif Arab University for Security Sciences, P.O Box: 6830, Riyadh 11452, Kingdom of Saudi Arabia

Background

The Kingdom of Saudi Arabia (KSA) is the largest Arab country constituting the bulk of 80% of the Arabian Peninsula. Before the foundation of the modern Saudi Arabia, it consisted of four distinct regions: Hejaz, Najd, Al-Ahsa, and Asir (Al-Rasheed 2013). Tribes in the KSA are actually the descendants of the peninsula's original ethnic stock; therefore, a certain degree of ethnic heterogeneity is evident among both the sedentary as well as the nomadic populations of modern KSA.

Genetic variations in the KSA were contributed earlier by nomadic or Bedouin tribes and clans (Gordon 2005) living in small groups of Persians, Turks, black Africans, and other ethnicities originating from sub-Saharan Africa along the Red Sea coast (Bowen 2014). In addition, the annual pilgrimage (Hajj) to Mecca has long brought hundreds of thousands of migrants representing various ethnic groups from Arab (Jordan, Iraq, Yemen), Asian, and Far Eastern countries to the KSA who overstayed and settled in and around Makkah, Jeddah, and Medina, (Fig. 1) (Ochsenwald and Philby 2016). Nevertheless, majority of the native Saudi population subgroups in the northern, central (excluding Riyadh), western (excluding Jeddah and Makkah), southern, and eastern regions remained genetically distinct because of their adherence to the consanguineous marriage practice (El-Hazmi et al. 1995).

The Y chromosome polymorphism has been studied widely for human migrations, forensic applications, and paternity analysis (Jobling and Tyler-Smith 2000; Quintana-Murci et al. 2001). The Y-STR markers are inherited without recombination down the paternal line with a little mutation and gene conversion (Rozen et al. 2003; Trombetta et al. 2010). These markers not only provide information on the male lineage relationship (Lowery et al. 2013) but also help in studying the local population structure and its demographic history

Fig. 1 Map of Saudi Arabia showing neighboring Arab populations

(Roewer et al. 2005). Y-STR typing has become an important tool in forensic investigations because of its discrimination power and marked genetic variations which produced highly informative Y chromosome STR haplotypes. Due to the greater sensitivity of non-recombining Y chromosomal markers to founder effects and genetic drift, Y-STRs are very powerful in detecting genetic differences between populations (Heraclides et al. 2017; Iacovacci et al. 2017; Li et al. 2016).

Studies regarding Y chromosome genetic lineage and population genetic structure in Saudi Arabia are limited (Abu-Amero et al. 2009; Alshamali et al. 2009; Khurbani et al. 2018, 2019). In the present paper, we present analysis of Y chromosome haplotypes in 125 native Saudi males from different geographic regions of Saudi Arabia, using the AmpFℓSTR® YFiler® Amplification kit (Life Technologies, USA). We also compared our Y chromosome STR haplotypes to previously published Y chromosome haplotype data from Saudi Arabia and seven neighboring Arab populations (Fig. 1). It is hoped that findings of this study will add to the existing state of knowledge about the population genetics and distribution of Y-STR haplotypes in Saudi Arabia.

Materials and methodology

Sample collection

Approval of the Institutional Ethical Committee to conduct this study was obtained well in advance. Buccal swabs were collected from 125 parentally unrelated, fully informed and consented, as per Helsinki Declaration, native (until three generations), and healthy Saudi males from all the regions of Saudi Arabia (Fig. 2), including Riyadh, Al Qassim in central; Tabuk, Al Jawf, Al-Hudud Al Shimaliyah, Hail in northern; Madinah, Makkah in western; Asir, Jizan, Najran in southern; and Dammam, Al-Khobar, Jubail in Eastern provinces. Their 3-generation ethnicity was established by looking at their respective national identification (ID) cards. Information regarding their birth places were provided by the donor. All buccal swab donors were adults and came from different walks of life including teachers, businessmen, policemen, and university students. They were recruited from universities, schools, police stations, and shopping centers. None of the donors underwent bone marrow transplant, radiotherapy, frequent blood transfusion, and chemotherapy in the near past. Most of them were married and none of the participants had any known Y chromosome abnormality.

DNA extraction

Genomic DNA was extracted from buccal swabs using Chelex® 100 as described by Walsh et al. (1991) and quantified in the 7500 Real-Time PCR System using Quantifiler® Duo DNA Quantification Kit (Applied

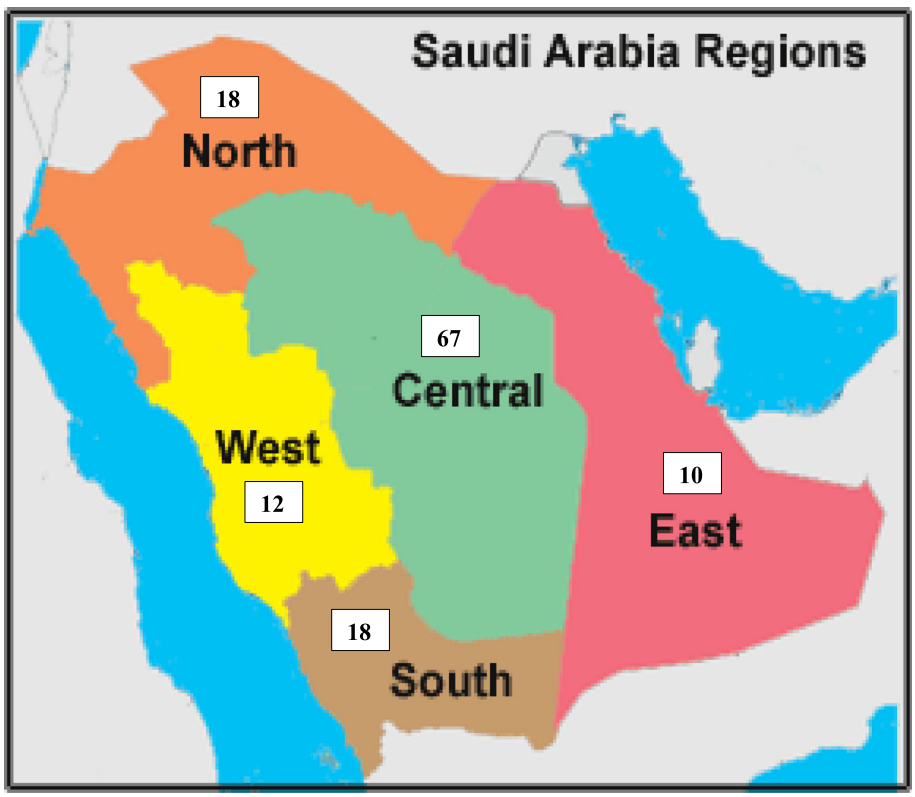

Fig. 2 Map of Saudi Arabia showing 5 geographic regions with corresponding sample sizes collected from those regions

Biosystems, USA) to regulate the input quantity of DNA for PCR amplification of respective Y-STR loci.

PCR and capillary electrophoresis

Extracted DNA were amplified for 17 Y chromosomal STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, DYS439, DYS437, DYS448, DYS458, DYS456, DYS635, and Y-GATA-H4) by the multiplex assay using AmpFℓSTR® YFiler® Amplification kit (Life Technologies, USA) in HID Veriti® 96-Well Thermal Cycler. Amplified Y-STR fragments were size separated by capillary electrophoresis (CE) using the 3130 Genetic Analyzer® (Life Technologies, USA) following the manufacture's protocol. GeneScan 500 LIZ was used as an internal size standard. Fragment size of amplified fragments was determined by *GeneMapper ID-X* Version 1.4 (Applied Biosystems, USA). Allele designation was based on comparison with the allelic ladder provided in the AmpFℓSTR® YFiler® Amplification kit (Life Technologies, USA). Amplified fragment analysis and YSTR typing were carried out according to the quality assurance standards recommended by the Scientific Working Group on DNA Analysis Methods (SWGDAM, 2014).

Statistical analysis

An online tool, "STR Analysis for Forensics (STRAF)", developed by Gouy and Zieger (2017), was used to calculate the Y-STR allele frequency, gene diversity (GD), haplotype diversity (HD), and discrimination capacity (DC). Using the online Y Chromosome Haplotype Reference Database (YHRD) tool (Willuweit and Roewer 2015; https://yhrd.org/), population pairwise genetic distance (R_{ST}) and associated probability values $p < 0.05$ were calculated using AMOVA (Analysis of molecular variance) tool (Roewer et al. 1996) and visualized in a multi-dimensional scaling (MDS) plot for the following neighboring Arab populations: Jordan (Qahtanit), 114; Jordan (Adnanit), 50; Iraq, 124; Kuwait, 285; UAE, 191; Bahrain, 156; Yemen, 128; and Egypt (Qena), 52 (Table 1). The DYS389I was subtracted from DYS389II as recommended by YHRD to calculate AMOVA and MDS.

Results and discussion

YSTR profiling has been considered as a vital tool for forensic investigation of cases like sexual assault (Maiquilla et al. 2011), missing persons (Coble et al. 2009), and kinship (Barra et al. 2015). Other applications include population genetics, anthropology, and epidemiology studies investigating the risk of prostate cancer (Paracchini et al. 2003; Hameed et al. 2015). Because of its crucial geographical location in the Arabian Peninsula and in the Gulf of Oman, several authors have studied Y chromosome diversity in native Saudi population employing Y-STR technology (Cadenas et al. 2008; Abu-Amero et al. 2009; Alshamali et al. 2009; Khurbani et al. 2018, 2019). The current report presents the population data for 17 Y-STR loci among 125 adult, native, Saudi male volunteers recruited from different geographic regions of Saudi Arabia (Fig. 2).

The quality of the study sample greatly affects the outcome of the population genetics studies. For example, Shringarpure and Xing (2014) reported that the accuracy of population stratification and recovery of individual ancestry are greatly affected by the sampling bias in the data collection process. Other studies have shown that sample selection bias can affect population structure analysis of genotype data, genetic ancestry of individuals, and evolutionary history of a certain population (Rosenberg et al. 2002; Patterson et al. 2006). Most of the studies carried out in the Saudi population (Cadenas et al. 2008; Abu-Amero et al. 2009; Alshamali et al. 2009; Khurbani et al. 2018, 2019) are based upon the sample collected either from Saudi blood banks, hospitals, forensic casework samples, or from native Saudis living abroad who are mostly self-declared and are not subjected to any type of further verification therefore, lacking the reliable ethnic or demographic originality that may affect, to some extent, the outcome of population genetic parameters.

The present study is the first study from Saudi Arabia in which samples were collected through a well-designed questionnaire served by a trained field worker assuring the acquisition of accurate ethnic data up to three

Table 1 Neighboring Arab population's reported Y-STR haplotypes and corresponding YHRD accession number

Countries	No. of reported Y-STR haplotypes	YHRD accession no.
UAE	191	YA003765
Kuwait	285	YA003763
Jordan [Qahtanit]	114	YA003523
Jordan [Adnanit]	50	YA003522
Yemen	128	YA003764-1
Iraq	124	YA003858
Bahrain	156	YA004278
Egypt (Qena)	52	YA003403

generations to confirm the actual geographic descent. Moreover, the geographic location of each participant was not recorded on the basis of his current place of residence (as done in most of the previous studies) but rather on the basis of the birthplace of the volunteer's great grandfather. Therefore, slight differences in certain population genetic parameters are expected in the present study.

Distribution of Y-STR haplotypes in a sample of 125 native, unrelated Saudi individuals were analyzed, and 102 different Y-STR alleles and 106 Y-haplotypes were observed. Ninety-one (85.8%) of the 106 haplotypes were unique, while the remaining 15 (14.2%) were shared; 12/125 (9.6%) haplotypes were repeated twice and 3/125 (2.4%) haplotypes were shared by two individuals. The most frequent haplotype was H23 (14,10,30,23,13,11,12, 13/18,10,11,14, 20,14,19,21,11) which was shared by four (3.2%) individuals (Table 2). Although the Arabian Peninsula is the region where numerous migrations between Africa and Asia took place since ancient times, our results showed an average degree of haplotype diversity among the Saudi Arabian population most probably due to consanguinity practice and moderate sample size.

Table 3 shows the distribution of YSTR alleles, their corresponding allele frequency, gene diversity (GD), haplotype diversity (HD), and F_{ST} or genetic distance. The maximum number of YSTR alleles ($n = 11$) was seen at the locus DYS 385b followed by DYS 385a ($n = 10$) and DYS 635 ($n = 8$) indicating their high degree of polymorphism. The least polymorphic YSTR loci were DYS 3891,391,437 and YGATA-H4 with each locus having 4 alleles. The maximum HD (0.817) was observed at the locus DYS 458 followed by the locus DS385b (0.787) and DYS 392 (0.684). The locus DYS 437 showed the least HD (0.155). The discrimination capacity (DC) calculated for 17 YSTR loci in the Saudi male population was 85.85%. In a recent report, Khurbani et al. (2018) studied a sample of 597 Saudi individuals from 5 geographic regions of Saudi Arabia using 27-YSTR Yfiler® plus and reported a DC of 95.3%. However, when they studied the same sample using 17-YSTR Yfiler® kit, their population DC declined to 74.7% which is considerably lower than what we have reported in the present study (85.85%) using the same 17-YSTR Yfiler® kit. This may be due to ethnic authenticity of our studied sample compared with the study of Khurbani et al. (2018) which had 15% of their Saudi volunteers recruited from the UK.

In the present study, YSTR locus DYS385b showed the highest gene diversity (GD) (0.807) followed by DYS458 (0.800) and DYS385a (0.686). The loci with the least GD were DYS437 (0.222) preceded by DYS392 (0.299) and DYS389I (0.355) (Table 3). The diversity of the Y chromosome is affected by factors such as the effective male population size, genetic drift, male behavior, marriage systems, and male patterns of migration (Jobling and Tyler-Smith 2003). The range of polymorphism and associated mutational properties makes Y chromosome the best candidate to answer many forensic, anthropological, population genetics, and evolutionary questions (de Knijff 2000; Jobling and Tyler-Smith 2003). Previous studies suggest that Saudi Arabia has a strategic position between Asian and African populations (Luis et al. 2004). The genetic structure of Saudi Arabia has been modulated by gene flow from Asian and African surroundings (Abu-Amero et al. 2009).

A total of four null alleles appeared in our study, one each in the haplotype H30, H31, and H64 at the locus DYS 458 and one in H75 at the locus DYS 456 (Table 2). A previous study by Chandler has shown that the YSTR locus DYS 458 has the highest mutation rate of 0.00814 followed by the locus DYS 456 showing a mutation rate of 0.00735 (Chandler 2006). As well as in a worldwide collaborative study, 137 null alleles were identified at 17 of the 23 Y-STR loci. The occurrence of null alleles has been associated with the mutation rate of the locus in question.

It was also observed that DYS385a/b to be the most informative marker having 21 complete alleles. In addition, it also showed micro variant allele 17.1 at DYS 385a indicating one base pair deletion within or far from the repeat regions (Butler 2011). Such partial repeat variant occurring at a low frequency may be useful in understanding the Y chromosome diversity and recent migrations.

The haplotypes seen in our studied regions of Saudi Arabia were compared with the published data haplotypes of seven neighboring Arab populations using the YHRD database. As observed in the present study, the R_{ST} values of the Egyptian (Qena) and Iraqi populations are closer to the Saudi population and are at equigenetic distance (R_{ST} 0.0018) with Saudi Arabia. Yemenites are slightly distant (R_{ST} 0.0022) from Saudi Arabia as reported by Abu-Amero et al. (2009) and Alshamali et al. (2009), but still closer than Abu Dhabi (R_{ST} 0.0028) and Kuwaiti populations (R_{ST} 0.0028) which is parallel to Triki-Fendri et al. (2010). Bahrain, although being the geographically nearest country to the kingdom of Saudi Arabia, yet genetically the most distant country (R_{ST} 0.0155) from the Saudi population (Table 4, Fig. 3). The most distant population from Saudi Arabia are the Arab Qahtanits in Jordan showing an R_{ST} value of 0.0146, the highest in the present study, followed by the Adnanit Jordanians (R_{ST} 0.0106). Al-Zahery et al. (2011) described a common ancestral origin of the Marsh Iraqi Arabs from South Arabian Peninsula. Moreover, Jordan has 70% of its population of Palestinian origin (González et al. 2008; Flores et al. 2005). This therefore was reason for its high genetic distance.

Table 2 Y chromosome haplotype analysis of 17-YSTR markers in male Saudi population ($n = 125$)

No. of haplotypes	Haplotype	Sample no.	Sample location	DYS 19	DYS 391	DYS 389II	DYS 390	DYS 389I	DYS 392	DYS 393	DYS 385	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	YGATA H4
1	H1	1	Central	15	10	31	21	13	11	13	17,19	11	11	14	21	15	15	21	12
1	H2	2	Central	14	11	30	22	13	11	12	13,18	10	12	14	20	14	18	22	11
1	H3	3	Central	14	11	30	24	13	11	12	13,19	10	11	14	20	15	19	21	11
2	H4	4	Central	15	11	29	25	13	13	13	15,19	10	12	14	18	16	16	21	12
		9	Central	15	11	29	25	13	13	13	15,19	10	12	14	18	16	16	21	12
1	H5	5	Central	14	11	30	23	12	11	12	13,19	10	12	14	20	14	19	21	11
1	H6	6	Central	14	10	31	24	13	11	13	16,17	10	12	14	20	16	18	21	11
1	H7	7	Central	14	9	29	22	13	11	12	13,16	9	11	15	21	15	16	23	11
1	H8	8	Central	16	10	31	23	13	11	13	10,11	10	12	14	20	13	17	22	11
3	H9	10	Central	14	10	30	23	13	11	12	13,18	10	11	14	20	15	18	21	11
		67	Central	14	10	30	23	13	11	12	13,18	10	11	14	20	15	18	21	11
		70	Eastern	14	10	30	23	13	11	12	13,18	10	11	14	20	15	18	21	11
2	H10	11	Central	14	11	29	23	13	11	12	13,19	10	12	14	20	14	18	23	11
		13	Central	14	11	29	23	13	11	12	13,19	10	12	14	20	14	18	23	11
1	H11	12	Central	14	10	30	23	13	13	12	13,18	10	11	14	20	15	18	21	11
1	H12	14	Central	13	10	29	24	13	11	13	16,17	10	12	15	20	14	15	22	12
1	H13	15	Central	14	11	31	23	13	11	12	13,19	9	13	14	18	14	18	21	12
1	H14	16	Central	15	10	29	22	12	11	13	15,15	10	11	16	20	16	16	21	11
1	H15	17	Central	15	10	28	24	14	11	14	10,11	11	11	14	20	13	19	0	11
2	H16	18	Central	14	10	30	23	14	11	13	16,17	10	12	14	21	15	16	21	11
		45	Central	14	10	30	23	14	11	13	16,17	10	12	14	21	15	16	21	11
1	H17	19	Central	13	10	30	23	14	13	15	16,17	9	9	14	19	15	16	21	11
1	H18	20	Central	14	12	31	23	13	11	12	13,19	10	11	14	20	14	18	21	12
1	H19	21	Central	14	10	30	23	14	11	13	16,17	10	12	14	20	15	16	21	11
1	H20	22	Central	14	11	30	23	13	11	12	13,19	10	11	14	20	15	19	22	12
2	H21	23	Central	14	11	30	23	13	11	12	13,19	10	11	14	20	14	19	22	12
		42	Central	14	11	30	23	13	11	12	13,19	10	11	14	20	14	19	22	12
1	H22	24	Central	14	11	28	23	12	11	12	13,18	10	11	13	21	14	18	21	11
4	H23	25	Central	14	10	30	23	13	11	12	13,18	10	11	14	20	14	19	21	11
		90	Northern	14	10	30	23	13	11	12	13,18	10	11	14	20	14	19	21	11
		105	Southern	14	10	30	23	13	11	12	13,18	10	11	14	20	14	19	21	11
		119	Southern	14	10	30	23	13	11	12	13,18	10	11	14	20	14	19	21	11
1	H24	26	Central	15	9	31	21	14	11	13	12,13	10	13	16	22	18	17	20	12
1	H25	27	Central	15	11	28	23	12	13	13	15,15	13	13	14	18	15	18	23	13
1	H26	28	Central	14	10	31	24	13	11	13	17,17.1	10	11	14	20	16	18	21	11
1	H27	29	Central	14	11	30	23	13	11	12	13,19	10	11	14	20	14	20	21	11
1	H28	30	Central	14	10	30	23	13	11	12	13,17	10	11	14	20	15	18	20	11
2	H29	31	Central	15	10	28	23	13	11	13	13,16	9	12	15	22	15	15	21	12
		123	Eastern	15	10	28	23	13	11	13	13,16	9	12	15	22	15	15	21	12
1	H30	32	Central	11	10	31	24	13	12	13	16,19	12	11	14	21	15		22	12
1	H31	33	Central	14	10	30	23	13	11	13	14,15	10	12	14	20	14		21	13
3	H32	34	Central	14	11	30	23	13	11	12	13,18	10	11	14	20	14	18	21	11
		82	Northern	14	11	30	23	13	11	12	13,18	10	11	14	20	14	18	21	11

Table 2 Y chromosome haplotype analysis of 17-YSTR markers in male Saudi population ($n = 125$) *(Continued)*

No. of haplotypes	Haplotype	Sample no.	Sample location	DYS 19	DYS 391	DYS 389II	DYS 390	DYS 389I	DYS 392	DYS 393	DYS 385	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	YGATA H4
		111	Southern	14	11	30	23	13	11	12	13,18	10	11	14	20	14	18	21	11
1	H33	35	Central	15	10	31	21	13	12	14	14,17	11	12	13	20	15	17	22	11
1	H34	36	Central	12	10	29	22	13	15	13	15,16	11	12	14	19	15	19	22	11
1	H35	37	Central	14	10	29	23	13	13	13	14,16	9	11	14	19	17	18	21	10
1	H36	38	Central	14	10	28	24	13	11	14	13,16	11	10	16	19	14	18	25	11
1	H37	39	Central	16	10	29	22	12	11	14	12,15	10	12	16	22	15	18	21	11
1	H38	40	Central	16	9	30	21	13	11	13	12,13	11	12	16	21	15	16	0	12
2	H39	41	Central	13	10	29	23	13	11	12	9,21	10	11	14	20	14	18	21	11
		59	Central	13	10	29	23	13	11	12	9,21	10	11	14	20	14	18	21	11
1	H40	43	Central	14	10	29	24	12	11	14	15,16	10	12	14	19	15	17	20	12
1	H41	44	Central	13	10	30	23	13	11	14	16,17	11	12	14	20	15	16	23	11
2	H42	46	Central	14	10	29	23	13	11	13	12,19	10	12	14	20	14	20	21	13
		118	Southern	14	10	29	23	13	11	13	12,19	10	12	14	20	14	20	21	13
1	H43	47	Central	15	10	29	25	13	12	13	10,11	11	11	14	20	14	16	22	11
1	H44	48	Central	15	10	30	21	13	11	15	17,18	11	11	14	20	16	16	21	12
1	H45	49	Central	14	11	27	24	12	13	12	11,16	12	13	15	19	15	15	23	13
1	H46	50	Central	13	10	32	24	13	11	13	16,17	10	12	14	20	15	18	26	11
1	H47	51	Central	16	11	31	25	14	11	13	11,14	11	11	14	20	16	15	23	11
1	H48	52	Central	13	10	30	23	13	11	14	16,19	10	11	14	20	15	15	24	11
1	H49	53	Central	14	11	31	23	14	11	12	13,19	10	12	14	20	15	19	21	11
1	H50	54	Central	14	10	30	23	14	11	13	17,18	10	12	14	21	15	16	21	11
1	H51	55	Central	15	10	30	24	12	11	13	16,17	12	12	15	19	15	19	22	12
2	H52	56	Central	14	11	29	23	13	11	12	13,17	10	11	14	20	14	18	21	11
		83	Northern	14	11	29	23	13	11	12	13,17	10	11	14	20	14	18	21	11
1	H53	57	Central	14	11	30	24	13	11	12	16,17	10	11	14	20	15	19	21	11
1	H54	58	Central	14	10	29	23	13	11	12	13,18	10	11	14	20	14	18	21	11
1	H55	60	Central	14	10	30	23	13	11	12	12,20	10	12	14	20	14	18	21	11
1	H56	61	Central	14	10	30	24	13	11	13	16,17	10	11	14	20	17	17	21	11
1	H57	62	Central	11	10	30	23	13	12	13	16,19	11	11	14	21	15	15	22	11
1	H58	63	Central	14	11	30	23	13	13	12	13,19	10	11	14	20	14	19	21	11
1	H59	64	Central	14	11	29	23	13	11	13	16,18	11	12	14	19	15	17	21	11
2	H60	65	Central	14	10	30	23	13	11	12	13,17.1	10	11	14	20	15	18	21	11
		66	Central	14	10	30	23	13	11	12	13,17.1	10	11	14	20	15	18	21	11
1	H61	68	Central	14	10	29	25	13	13	14	15,17	9	11	14	19	16	17	22	12
1	H62	69	Eastern	14	9	28	23	12	14	11	13,15	10	12	14	19	16	16	24	11
1	H63	71	Eastern	14	10	31	23	13	11	12	13,17	10	11	14	20	15	17	20	11
1	H64	72	Eastern	11	10	30	24	13	12	13	16,19	11	11	14	21	15		21	11
1	H65	73	Northern	14	11	30	23	13	13	13	15,16	9	11	14	18	15	19	22	11
1	H66	74	Northern	14	11	29	24	13	11	12	13,18	10	11	14	20	14	18	21	11
1	H67	75	Northern	14	10	30	23	13	11	12	13,17	10	12	14	20	14	17	21	11
1	H68	76	Northern	14	11	31	23	13	11	12	13,18	10	11	14	20	14	17	22	11
1	H69	77	Northern	24	11	29	23	13	11	12	13,18	10	11	14	20	14	19	21	11
1	H70	78	Northern	16	10	29	25	13	13	14	13,14	12	12	14	19	15	17	23	13

Table 2 Y chromosome haplotype analysis of 17-YSTR markers in male Saudi population ($n = 125$) *(Continued)*

No. of haplotypes	Haplotype	Sample no.	Sample location	DYS 19	DYS 391	DYS 389II	DYS 390	DYS 389I	DYS 392	DYS 393	DYS 385	DYS 438	DYS 439	DYS 437	DYS 448	DYS 456	DYS 458	DYS 635	YGATA H4
1	H71	79	Northern	14	10	30	24	13	11	13	16,17	10	12	14	20	17	17	21	11
1	H72	80	Northern	15	10	28	22	12	11	13	15,16	10	11	16	21	16	16	21	11
1	H73	81	Northern	13	10	30	23	13	11	13	16,16	10	14	14	20	15	17	21	11
1	H74	84	Northern	14	10	31	23	13	11	12	13,18	10	11	14	20	14	18	22	11
1	H75	85	Northern	14	11	30	23	13	11	16	13,17	10	11	14	20		18	21	11
2	H76	86	Northern	14	10	30	23	13	11	12	13,17	10	11	14	20	14	18	22	11
		87	Northern	14	10	30	23	13	11	12	13,17	10	11	14	20	14	18	22	11
1	H77	88	Northern	14	10	30	23	13	11	12	13,18	10	12	14	20	14	17	21	11
1	H78	89	Northern	14	11	32	23	13	11	12	13,19	10	11	14	20	14	18	21	11
2	H79	91	Western	14	11	30	23	13	11	12	13,19	10	11	14	20	14	18	21	11
		116	Southern	14	11	30	23	13	11	12	13,19	10	11	14	20	14	18	21	11
1	H80	92	Western	16	10	27	22	12	10	14	13,14	10	11	15	18	13	20	21	11
1	H81	93	Western	15	11	30	23	13	11	12	13,17	10	11	14	20	15	18	21	10
1	H82	94	Western	14	11	30	23	13	11	12	13,19	10	11	14	20	14	19	21	12
1	H83	95	Western	14	10	30	23	13	11	14	13,18	10	11	14	20	14	20	21	11
1	H84	96	Western	13	10	30	24	12	11	13	16,17	10	12	14	20	15	17	24	11
1	H85	97	Western	15	10	30	22	13	11	12	15,16	10	12	14	19	15	17	21	12
1	H86	98	Western	14	11	30	23	13	11	12	13,18	10	11	14	20	14	17	21	11
1	H87	99	Western	14	10	33	24	15	11	13	16,17	10	12	14	20	16	17	21	11
1	H88	100	Western	15	10	30	23	13	11	12	13,18	10	11	14	20	14	19	21	11
1	H89	101	Western	14	10	31	24	13	11	13	16,17	10	11	14	20	16	17	21	11
1	H90	102	Western	14	11	29	22	13	11	12	13,19	10	11	14	19	13	19	21	11
1	H91	103	Southern	13	10	32	25	13	11	13	16,18	10	13	14	21	15	17	23	11
1	H92	104	Southern	14	10	30	24	13	11	12	18,19	10	13	14	20	15	17	21	11
1	H93	106	Southern	14	11	30	22	13	11	12	13,19	10	11	14	19	13	19	21	11
1	H94	107	Southern	14	11	31	23	13	10	12	13,17	10	11	14	19	14	17	22	11
1	H95	108	Southern	14	11	29	22	13	11	12	13,19	10	11	14	20	13	17	21	11
1	H96	109	Southern	14	11	30	23	13	11	12	13,17	10	11	14	19	14	19	21	11
1	H97	110	Southern	14	9	31	23	14	11	12	13,18	10	11	14	20	14	19	22	10
1	H98	112	Southern	13	10	29	23	12	11	12	14,18	10	12	14	19	17	19	21	12
1	H99	113	Southern	15	10	28	25	12	11	12	18,19	9	12	14	18	14	16	21	12
1	H100	114	Southern	15	11	30	22	14	11	12	13,19	10	11	14	20	13	19	21	11
1	H101	115	Southern	14	10	30	23	13	11	12	13,18	10	11	14	20	14	20	21	10
1	H102	117	Southern	15	11	31	21	13	11	12	15,16	11	12	14	21	15	16	21	12
1	H103	120	Eastern	14	11	31	22	13	13	13	16,17	10	11	14	20	14	19	21	11
1	H104	121	Eastern	16	10	31	21	13	11	13	17,18	11	11	14	21	15	15	21	12
1	H105	122	Eastern	13	10	31	23	13	11	13	17,18	10	12	14	19	14	17	22	11
2	H106	124	Eastern	15	10	28	23	13	11	13	13,16	9	12	15	22	15	15	21	12
		125	Eastern	15	10	28	23	13	11	13	13,16	9	12	15	22	15	15	21	12

Table 3 Frequency Distribution of 17Y-STR haplotypes among native male Saudi Population (n=125)

Alleles	DYS019	DYS385a	DYS385b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS635	YGATA H4
0															0.024	0.008	
9		0.016					0.040				0.072	0.008					
10		0.024					0.584	0.016			0.768	0.008					0.032
11	0.024	0.016	0.024				0.368	0.832	0.008		0.120	0.584					0.720
12	0.008	0.056		0.112			0.008	0.040	0.544		0.032	0.344					0.200
13	0.096	0.520	0.016	0.792				0.096	0.336	0.016	0.008	0.048		0.056			0.048
14	0.648	0.032	0.024	0.088				0.008	0.088	0.880		0.008		0.424			
15	0.160	0.088	0.040	0.008				0.008	0.016	0.056				0.376	0.072		
16	0.056	0.184	0.104						0.008	0.048				0.096	0.136		
17		0.048	0.232											0.032	0.192		
17.1			0.024														
18		0.016	0.248										0.064	0.008	0.304		
19			0.264										0.144		0.224		
20			0.008										0.648		0.048	0.040	
21			0.016			0.056							0.112			0.664	
22						0.104							0.032			0.168	
23						0.616										0.072	
24	0.008					0.160										0.024	
25						0.064										0.008	
26																0.008	
27					0.016												
28					0.072												
29					0.216												
30					0.496												
31					0.168												
32					0.024												
33					0.008												
No. of alleles	**7**	**10**	**11**	**4**	**7**	**5**	**4**	**6**	**6**	**4**	**5**	**6**	**5**	**7**	**9**	**8**	**4**
GD	0.546	0.686	0.807	0.355	0.678	0.582	0.526	0.299	0.588	0.222	0.393	0.542	0.546	0.671	0.800	0.527	0.442
F_{ST}	− 0.021	0.024	− 0.008	− 0.002	− 0.010	− 0.008	− 0.016	0.019	0.070	− 0.005	0.026	0.012	− 0.005	0.070	0.058	0.012	0.043
HD	0.552	0.597	0.787	0.312	0.684	0.560	0.539	0.337	0.583	0.155	0.331	0.511	00.597	0 0.676	0.0.817	00.499	0.408

GD gene diversity, F_{ST} genetic distance, *HD* haplotype diversity

Table 4 Matrix of the pairwise F_{ST} genetic distances between native Saudi population and eight neighboring Arab populations (below diagonal) obtained for 10,100 permutations (s.e. ≤ 0.0038)

Population	Saudi Arabia	Bahrain	Egypt (Qena)	Iraq	Jordan (Arab-Adnanit)	Jordan (Arab-Qahtanit)	Kuwait (Kuwait City)	Abu Dhabi (UAE)	Yemen (Sana)
Saudi Arabia	-	-	-	-	-	-	-	-	-
Bahrain	0.0051	-	-	-	-	-	-	-	-
Egypt (Qena)	0.0018	0.0035	-	-	-	-	-	-	-
Iraq	0.0018	0.0035	0.0001	-	-	-	-	-	-
Jordan(Arab-Adnanit)	0.0106	0.0123	0.0091	0.0090	-	-	-	-	-
Jordan (Arab-Qahtanit)	0.0146	0.163	0.133	0.0131	0.0220	-	-	-	-
Kuwait (Kuwait City)	0.0028	0.0044	0.0013	0.0014	0.0101	0.142	-	-	-
Abu Dhabi (UAE)	0.0028	0.0044	0.0011	0.0012	0.0100	0.0141	0.0023	-	-
Yemen (Sana)	0.0022	0.0039	0.0005	0.0006	0.0095	0.0137	0.0018	0.0012	-

Conclusion

By providing the population data on the genetic variations at 17 YSTR loci in a sample of the native Saudi male population ($n = 125$), an attempt has been made to develop an understanding about the genetic relationship between Saudi Arabia and the neighboring Arab population. Our results show that the Saudi population is genetically closer to the Iraqi, Qena (Egypt), and Yemen (Sana) populations than the Kuwaiti, Abu Dhabi (UAE), Bahrain, and Jordan population. According to our findings, the Saudi population lacks patrilineal homogeneity across the entire region, being homogeneous at one place and partly heterogeneous in others (data not presented here). This may be due to the highly conserved social culture, practice of consanguineous marriages in certain regions, and religious or historical migration to Makkah, Medina, and Jeddah. Unfortunately, because of the limited sample size from different geographic regions of Saudi Arabia, an independent forensic and population statistics could not be performed.

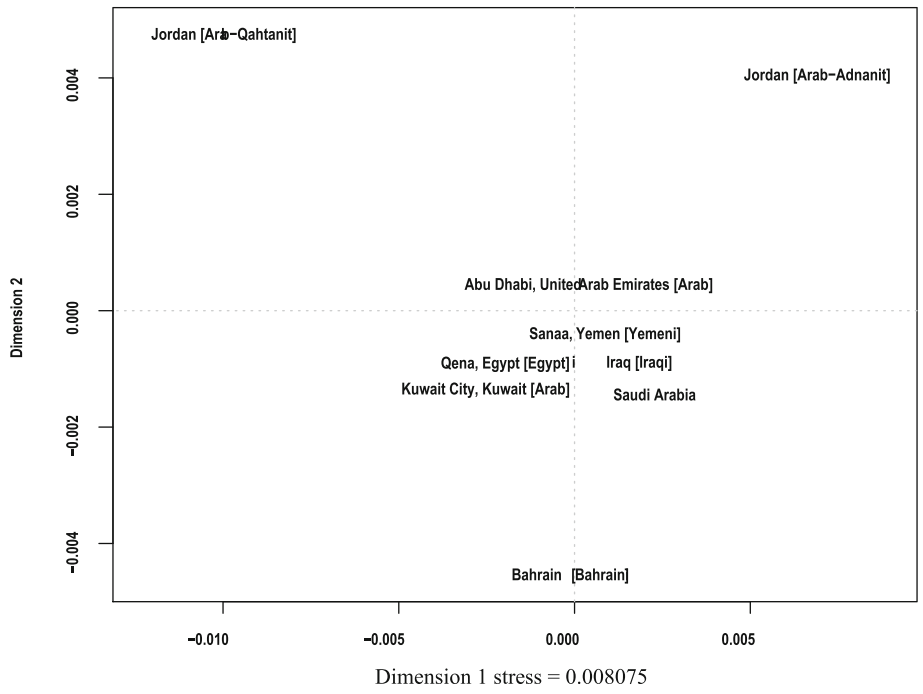

Fig. 3 Multidimensional scaling (MSD) based on pairwise R_{ST} genetic distances of native Saudi populations and seven other neighboring Arab populations (Egypt, Iraq, Jordan, UAE, Yemen, Kuwait, and Bahrain)

Further studies are, therefore, needed to establish precise patrilineal inheritance in the Saudi population and explore its relationship with neighboring Arab countries.

Abbreviations

STR: Short tandem repeats; AMOVA: Analysis of molecular variance; YHRD: Y Haplotype Reference Database; R_{ST}: Population pairwise genetic distance; GD: Gene diversity; HD: Haplotype diversity; DC: Discrimination capacity; MDS: Multidimensional scaling plot

Acknowledgements

Authors would like to acknowledge all the volunteers for donating their samples and full cooperation throughout the study. At the same time, we are grateful to Mr. Mohammed Al-Bujjah and Ms. Abrar Al-Saleh from the Department of Forensic Science, Naif Arab University for Security Sciences for their laboratory assistance in this study.

Adherence to National and International regulations

Not applicable

Authors' contributions

ACK was responsible for the design, supervision, and preparation of manuscript. HFMA did the collection of sample and demographic data. SAM was responsible for the statistical analysis. SRB Performed labwork and preparation of the manuscript. SAS performed labwork. ARC was responsible for the analysis, interpretation, and preparation of the manuscript. The author(s) read and approved the final manuscript.

Funding

This project was funded by King Abdul-Aziz City for Science and technology (KACST)- Kingdom of Saudi Arabia under the grant number AT-35-152.

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author.

Ethics approval

Ethical approval was obtained from the institutional Ethics Committee.

Consent for publication

All the authors have given written consent for publication of this manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 9 February 2020 Accepted: 25 August 2020
Published online: 10 September 2020

References

Abu-Amero KK, Hellani A, Gonzalez AM, Larruga JM, Cabrera VM, Underhill PA (2009) Saudi Arabian Y-chromosome diversity and its relationship with nearby regions. BMC Genet 10:59

Al-Rasheed M (2013) A most masculine state: gender, politics and religion in Saudi Arabia. Cambridge University Press, New York

Alshamali F, Pereira L, Budowle B, Poloni ES, Currat M (2009) Local population structure in Arabian Peninsula revealed by Y-STR diversity. Hum Hered 68(1):45

Al-Zahery N, Pala M, Battaglia V, Grugni V, Hamod MA, Kashani BH, Semino O (2011) In search of the genetic footprints of Sumerians: a survey of Y-chromosome and mtDNA variation in the Marsh Arabs of Iraq. BMC Evol Biol 11:288

Barra GB, Santa Rita TH, Chianca CF, Velasco LF, de Sousa CF, Nery LF, Costa SS (2015) Fetal male lineage determination by analysis of Y-chromosome STR haplotype in maternal plasma. Forensic Sci Inter: Genet 1(15):105–110

Bowen WH. 2014. The history of Saudi Arabia. Greenwood; 2nd edition. Series, "The Greenwood Histories of the Modern Nations". ABC-CLIO, LLC, Santa Barbara, California (USA): ISBN-13: 978-1610698764.

Butler JM (2011) Advanced topics in forensic DNA typing: methodology, Ist edn. Academic Press. Elsevier Inc

Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ (2008) Y-chromosome diversity characterizes the Gulf of Oman. Eur J Hum Genet 16(3):374–386

Chandler JF (2006) Estimating per-locus mutation rates. J Genet Geneal 2:27–33

Coble MD, Loreille OM, Wadhams MJ, Edson SM, Maynard K, Meyer CE, Niederstätter H, Berger C, Berger B, Falsetti AB, Gill P (2009) Mystery solved: the identification of the two missing Romanov children using DNA analysis. PLoS One 4(3):e4838

de Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Ameri J Hum Genet 67(5):1055–1061

El-Hazmi MAF, Al-Swailem AR, Warsy AS, Al-Swailem AM, Sulaimani R, Al-Meshari AA (1995) Consanguinity among Saudi Arabian population. J Med Genet 32:623–626

Flores C, Maca-Meyer N, Larruga JM, Cabrera VM, Karadsheh N, Gonzalez AM (2005) Isolates in a corridor of migrations: a high-resolution analysis of Y-chromosome variation in Jordan. J Hum Genet 50(9):435–441

González AM, Maca-Meyer N, Flores C, Cabrera VM, Larruga JM, Karadsheh N (2008) Mitochondrial DNA variation in Jordanians and their genetic relationship to other middle East populations. Ann Hum Biol 35(2):212–231

Gordon MS (2005) The rise of Islam. Greenwood Publishing Group, London ISBN: 0-313-321522-7

Gouy A, Zieger M (2017) STRAF—a convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet 30:148–151

Hameed IH, Jebor MA, Kareem MA (2015) Allelic frequencies for the seventeen Y-STR loci observed in Iraqi male patients with prostate cancer. Afr J Biotec 14(15):1252–1260

Heraclides A, Bashiardes E, Fernández-Domínguez E, Bertoncini S, Chimonas M, Christofi V, King J, Budowle B, Manoli P, Cariolou MA (2017) Y-chromosomal analysis of Greek Cypriots reveals a primarily common pre-Ottoman paternal ancestry with Turkish Cypriots. PLoS One 12(6):e0179474

Iacovacci G, D'Atanasio E, Marini O, Coppa A, Sellitto D, Trombetta B, Berti A, Cruciani F (2017) Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries. Forensic Sci Int Genet 27:123–131

Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes: the human Y chromosome, disease and selection. Trends Genet 16(8):356–362

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4(8):598

Khurbani YM, Wetton JH, Jobling MA (2018) Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs. Forensic Sci Int Genet 2(33):98–105

Khurbani YM, Wetton JH, Jobling MA (2019) Analysis of 21 autosomal STRs in Saudi Arabia reveals population structure and the influence of consanguinity. Forensic Sci Int Genet 39:97–102

Li L, Yu G, Li S, Jin L, Yan S (2016) Genetic analysis of 17 Y-STR loci from 1019 individuals of six Han populations in East China. Forensic Sci Int Genet 20:101–102

Lowery RK, Herrera K, Uribe G, Reguiero M, Herrera RJ (2013) Sub-population structure evident in forensic Y-STR profiles from Armenian geographical groups. Legal Med 15(2):85–90

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. Ameri J Hum Genet 74(3):532–544

Maiquilla SM, Salvador JM, Calacal GC, Sagum MS, Dalet MR, Delfin FC, Tabbada KA, Franco SA, Perdigon HB, Madrid BJ, Tan MP (2011) Y-STR DNA analysis of 154 female child sexual assault cases in the Philippines. Int J Legal Med 125(6):817–824

Ochsenwald WL, Philby HJB, Teitelbaum J. 2016. Saudi Arabia: Encyclopedia Britannica; [Accessed on June 06, 2017]. https://www.britannica.com/place/Saudi-Arabia.

Paracchini S, Pearce CL, Kolonel LN, Altshuler D, Henderson BE, Tyler-Smith C (2003) A Y chromosomal influence on prostate cancer risk: the multi-ethnic cohort study. J Med Genet 40(11):815–819

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. PLoS Genet 2(12):e190

Quintana-Murci L, Krausz C, McElreavey K (2001) The human Y chromosome: function, evolution and disease. Foren Sci Inter 118(2):169–181

Roewer L, Croucher PJ, Willuweit S, Lu TT, Kayser M, Lessig R, de Peter K, Jobling MA, Tyler-Smith C, Krawczak M (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116(4):279–291

Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, Knijff P (1996) Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. Hum Mol Genet 5(7):1029–1033

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Sci. 298(5602):2381–2385

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Richard KW, David C, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nat. 423(6942):873–876

Scientific Working Group on DNA Analysis Methods Interpretation Guidelines for Y-Chromosome STR Typing. (2014). http://www.swgdam.org/publications. Accessed May 2015.
Shringarpure S, Xing EP (2014) Effects of sample selection bias on the accuracy of population structure and ancestry inference. G3 Genes Geno Genet 4(5):901–911
Triki-Fendri S, Alfadhli S, Ayadi I, Kharrat N, Ayadi H, Rebai A (2010) Genetic structure of Kuwaiti population revealed by Y-STR diversity. Ann Hum Biol 37(6):827–835
Trombetta B, Cruciani F, Underhill PA, Sellitto D, Scozzari R (2010) Footprints of X-to-Y gene conversion in recent human evolution. Mol Biol Evol 27(3):714–725
Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques. 10(4):506–513
Willuweit S, Roewer L (2015) The new Y Chromosome Haplotype Reference Database. Forensic Sci Int Genet 15:43–48

Publisher's Note