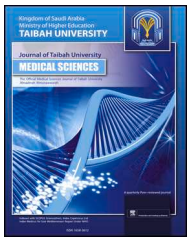
Journal of Taibah University Medical Sciences (2026) 21(3), 452–459

Taibah University

Journal of Taibah University Medical Sciences

www.sciencedirect.com



Original Article

Genetic variation of 15 autosomal STR loci in the southern region population of KSA: A forensic DNA analysis study

Husein Alhatim, PhD [a], AbdulRahman Muthanna, PhD [b,c,*], Muhammad N. Hakim Abdullah, PhD [b] and Mohd N. Mohd Desa, PhD [b]

[a] *Department of Forensic Sciences, College of Criminal Justice, Naif Arab University for Security Sciences, Riyadh, KSA*
[b] *Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia Serdang, Malaysia*
[c] *Office of Postgraduate Studies, UCSI University Kuala Lumpur, Malaysia*

Received 9 December 2025; revised 21 February 2026; accepted 11 April 2026

الملخص

أهداف البحث: تُعدّ التكرارات الترادفية القصيرة واسمات وراثية عالية التعدد الشكلي، ولها دور أساسي في الاستدلال الجنائي واختبارات إثبات النسب. ويُعدّ فهم التنوع الوراثي لهذه المواضع في جماعات سكانية محددة أمرًا بالغ الأهمية لضمان دقة التطبيقات الجنائية. هدفت هذه الدراسة إلى تقييم التباين الوراثي ومعايير الكفاءة الجنائية لخمسة عشر موضعًا جسميًا من مواضع التكرارات الترادفية القصيرة لدى سكان المنطقة الجنوبية من المملكة العربية السعودية.

طرق البحث: جُمعت مسحات شدقية من 150 فردًا سعوديًا أصيلًا غير ذوي قرابة من المنطقة الجنوبية. واستُخلص الحمض النووي باستخدام مادة Chelex 100، ثم جرى تقدير كميته باستخدام عدة Quantifiler® Duo DNA Kit. كما أُجري تضخيم مواضع التكرارات الترادفية القصيرة باستخدام عدة AmpFlSTR® Identifiler® Plus Kit، وأُجري الرحلان الكهربائي الشعري باستخدام جهاز Genetic Analyzer 3130. وتم إجراء التحليلات الإحصائية باستخدام برنامجي GeneAlex 6.5 وPowerstats v12.

النتائج: لوحظ ما مجموعه 140 أليلًا مختلفًا عبر المواضع الخمسة عشر، مع 6.57073827 × 10^24 من التراكيب الجينية المحتملة. وبلغت القدرة التمييزية المجمعة، واحتمال الاستبعاد، واحتمال التطابق المجمع، واحتمال التطابق العشوائي 0.999788998 و0.999989779 و2.83258e-16 و3.53034e+15 على التوالي. وأظهرت جميع المواضع قيمًا مرتفعة لمحتوى المعلومات متعدد الأشكال، باستثناء الموضع D7S820.

الاستنتاجات: تُظهر الدرجة المرتفعة من المعلوماتية التي أبدتها المواضع الخمسة عشر للتكرارات الترادفية القصيرة لدى سكان جنوب المملكة العربية السعودية فائدتها في تطبيقات الاستدلال الجنائي واختبارات إثبات النسب.

الكلمات المفتاحية: التكرارات الترادفية القصيرة؛ الوراثة الجنائية؛ المملكة العربية السعودية؛ الوراثة السكانية، البصمة الوراثية

Abstract

Objective: Short tandem repeats (STRs) are highly polymorphic genetic markers with essential roles in forensic identification and paternity testing. Understanding the genetic diversity of STR loci in specific populations is crucial for accurate forensic applications. The aims of this study were to evaluate the genetic variation and forensic efficiency parameters for 15 autosomal STR loci in the southern region population of KSA.

Methods: Buccal swabs were collected from 150 unrelated native Saudi individuals from the southern region. DNA extraction was performed using Chelex 100, followed by quantification using a Quantifiler® Duo DNA Kit. STR amplification was conducted using an AmpFlSTR® Identifiler® Plus Kit, and capillary electrophoresis was performed with a Genetic Analyzer 3130. Statistical analyses were conducted using GeneAlex 6.5 and Powerstats v12 software.

Results: In total, 140 different alleles were observed across the 15 STR loci, with $6.57073827 \times 10^{24}$ possible genotype combinations. The combined power of discrimination, probability of exclusion, combined match probability, and random match probability were 0.999788998, 0.999989779, 2.83258e-16, and 3.53034e+15, respectively.

* Corresponding address: Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Malaysia.
E-mail: mansoor@upm.edu.my (A. Muthanna)

Peer review under responsibility of Taibah University.

ELSEVIER Production and hosting by Elsevier

 https://doi.org/10.1016/j.jtumed.2026.04.002

All loci had high polymorphic information content values, except for D7S820.

Conclusion: The high informativity observed for the 15 STR loci in the southern Saudi population demonstrates their usefulness for forensic and paternity testing applications.

Keywords: DNA profiling; Forensic genetics; Population genetics; KSA; STR

Introduction

Forensic DNA analysis has revolutionized criminal investigations and paternity testing worldwide.[1,2] Short tandem repeats (STRs) are the gold standard for human identification due to their high polymorphism, robust amplification characteristics, and minimal degradation in compromised samples.[3] Developing population-specific allele frequency databases is essential for the accurate statistical interpretation of DNA evidence and ensuring reliable forensic applications.[4]

The KSA has diverse regional populations, and thus comprehensive genetic studies are required to establish robust forensic DNA databases.[5] The southern region of KSA is a significant demographic area with unique genetic characteristics that may differ from other regions within the kingdom. Understanding these regional variations is crucial for ensuring accurate forensic interpretations and establishing appropriate statistical parameters for legal proceedings.[6]

STR loci are highly variable DNA sequences consisting of 2–6 base pair repeat units.[7] The AmpFlSTR® Identifiler® Plus Kit amplifies 15 autosomal STR loci (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, VWA, TPOX, D18S51, D5S818, and FGA) plus the amelogenin gender determination marker.[8] These loci have been extensively validated and are widely used in forensic laboratories globally, providing high discriminatory power for individual identification and paternity testing.[9]

Population genetic studies are crucial for understanding allele frequency distributions, which directly impact forensic statistical calculations.[10] The power of discrimination (PD), probability of exclusion (PE), and match probability (MP) are key parameters that determine the forensic utility of STR loci in specific populations. In addition, Hardy–Weinberg equilibrium (HWE) testing ensures the validity of population assumptions that underlie forensic statistics, which is particularly important in populations with potential consanguineous marriages.[11]

Previous studies in KSA focused on the central and western regions, where they demonstrated significant genetic diversity and the forensic utility of STR markers.[12,13] However, comprehensive data are limited from the southern region. Regional genetic studies from neighboring Arab countries detected considerable population-specific variations in allele frequencies, emphasizing the importance of establishing region-specific databases.[14,15]

International forensic DNA databases, including CODIS (USA) and the ENFSI STRidER platform, are built on population-specific allele frequency data collected at the regional level.[16,17] Studies from diverse populations in the Levant, North Africa, and the Gulf have demonstrated that relying on national average allele frequencies can introduce measurable errors into likelihood ratios in the presence of significant regional substructure.[18,19] Southern KSA is a particularly distinct context, characterized by documented tribal endogamy and geographic isolation, thereby increasing the likelihood of the population's substructure diverging from the national average.[20,21] Empirical confirmation of this divergence was obtained by analyzing 21 autosomal STRs across five Saudi regions, which showed that the genetic distances were greatest between the northern and southern regions, with heterozygote deficiency observed at nearly all loci, and attributed to high rates of consanguineous marriage.[22] Similarly, extensive geographic and social structure in Saudi paternal lineages has been shown to directly affect the utility and interpretation of forensic DNA databases.[23] Due to the expanding role of DNA profiling in Saudi judicial proceedings, relying on a national frequency database without a southern region-specific data set risks the production of inaccurate likelihood ratio statistics in forensic casework. Therefore, the present study aimed to address this deficiency by providing a comprehensive population-specific data set for 15 autosomal STR loci in the southern Saudi population, contributing to the establishment of a national forensic DNA database, enhancing the accuracy of forensic interpretations in the region, and supporting the development of standardized protocols for DNA analysis in Saudi Arabian forensic laboratories.

Materials and Methods

Sample collection

A cross-sectional descriptive study was conducted among native Saudi males and females from the southern region of KSA at the Armed Forces Hospital, Southern Region, KSA. The study included participants who self-identified as native Saudi individuals with paternal and maternal ancestry from the southern region for at least two generations, but with no known familial relationship to any other participant, and no history of blood transfusion or bone marrow transplantation, which could potentially compromise the integrity of DNA profiles. Individuals who did not meet these criteria or who declined to provide informed written consent were excluded from the study. Buccal swab samples were collected from the inner cheek of each participant using sterile cotton swabs. Samples were air dried in a sterile environment for 1–2 h and stored at −20 °C until DNA was extracted.

DNA extraction

DNA was extracted using the Chelex 100 method. A solution of 5% Chelex 100 was prepared in the laboratory. The cotton swab head was cut off and placed in a 1.5 mL

microcentrifuge tube with 500 μL of deionized water. After vortexing for 10 s, the sample was incubated at room temperature for 15 min with periodic gentle mixing.

The samples were centrifuged at 1300 rpm for 5 min at room temperature. The supernatant was removed, leaving a volume of 20–30 μL. This washing step was repeated once more. Subsequently, 100 μL of Chelex 100 solution was added and samples were incubated at 56 °C for 30 min, followed by 100 °C for 8 min. After final centrifugation, the extracted DNA was stored at 4 °C.

DNA quantification

DNA was quantified using a Quantifiler® Duo DNA Quantification Kit and Real-Time PCR System 7500. Standard dilution series were prepared and 23 μL of reaction mixture (containing 10.5 μL primer mix and 12.5 μL reaction mix) was combined with 2 μL of DNA sample or standard in 96-well optical reaction plates.

STR amplification

DNA was amplified using an AmpFlSTR® Identifiler® Plus PCR Amplification Kit. The optimal DNA concentration for amplification was 0.5–1.5 ng/μL in a volume of 10 μL. Samples with concentrations exceeding 1.5 ng/μL were diluted with TE buffer. PCR amplification was conducted using a GeneAmp PCR System Veriti 9700 thermal cycler with the following program: initial denaturation at 95 °C for 11 min, followed by 28 cycles at 94 °C for 20 s, 59 °C for 3 min, and 60 °C for 10 min, with a final hold at 4 °C.

Capillary electrophoresis

STR products were prepared for capillary electrophoresis by combining 8.3 μL Hi-Di™ Formamide, 0.4 μL GeneScan™-500 LIZ™ internal size standard, and 1 μL amplified DNA. Samples were denatured at 95 °C for 3 min and cooled to 4 °C for 3 min, before injecting into the Genetic Analyzer 3130 system. The instrument was configured with POP-7™ Polymer and a 36 cm capillary array. Electrophoresis was performed using GeneMapper® ID-X v1.4 with an allele calling threshold of 50 relative fluorescence units. Each analytical run included a positive control (9947A human DNA standard) and negative control (reagent blank) to monitor the amplification quality and detect potential contamination (Supplementary 1). All samples with off-ladder alleles, spectral pull-up, or other electrophoretic artifacts were flagged and re-analyzed. Only profiles that satisfied the quality criteria for all 15 loci were retained for statistical analysis.

Statistical analysis

Statistical analyses were performed using GeneAlex v6.5 for allele frequency calculations, observed heterozygosity (H_o) and expected heterozygosity (H_e), HWE testing, and Fisher's exact tests. Powerstats v12 was used to calculate forensic parameters, including the polymorphic information content (PIC), PD, PE, MP, and typical paternity index (TPI). Given that HWE was tested across 15 loci simultaneously, a Bonferroni-corrected significance threshold of $\alpha = 0.05/15 = 0.0033$ was applied to reduce the risk of type I errors from multiple testing. This corrected threshold is referenced throughout the following Results section when interpreting the patterns of HWE deviations.

Results

Allele frequency distribution

Genetic analysis of 150 individuals from the southern Saudi population identified 140 distinct alleles among the 15 STR loci examined, representing 69% of the 203 possible alleles documented in the literature. Allelic diversity varied considerably among loci, where the number of observed alleles ranged from six at TH01 to 14 at D21S11. The most polymorphic locus was D21S11 with 14 different alleles, followed by D2S1338 with 13 alleles. By contrast, TH01 exhibited the lowest polymorphism with only six observed alleles. The denominator of 203 alleles was derived from the total number of alleles catalogued across the 15 loci in the AmpFlSTR® Identifiler® Plus allelic ladder, as documented in the kit user's manual[8] and consistent with the values summarized by Butler[1] (Supplementary 2).

Complete STR profiles were successfully generated for all 150 samples across all 15 loci (Supplementary 3). Table 1 presents the comprehensive allele frequency data and basic statistical parameters for all 15 STR loci analyzed in this study (Supplementary 4). The allele frequencies ranged from 0.002 to 0.600, where TPOX had the highest frequency for its most common allele (allele 8 at 0.600), whereas D21S11 had the lowest frequency for its rarest allele (allele 35.2 at 0.002). The allele size ranges varied from 6 to 10 repeats at TH01 to 24.2–35.2 repeats at D21S11. Most loci had relatively balanced allele distributions, although some specific alleles clearly predominated, such as TPOX with allele 8 comprising 60% of all observations.

HWE analysis

HWE testing revealed population structure variations across the loci examined. Four loci (D21S11, D3S1358, TPOX, and D5S818) deviated significantly from HWE ($p < 0.001$), while two additional loci (D13S317 and D2S1338) exhibited moderate but statistically significant deviations ($p < 0.05$). HWE was maintained for the remaining nine loci ($p > 0.05$), indicating random mating patterns for these genetic markers. In particular, after applying Bonferroni correction for multiple testing (adjusted threshold $\alpha = 0.0033$), only four loci with $p < 0.001$ retained statistically significant. D13S317 ($p = 0.029$) and D2S1338 ($p = 0.047$) no longer satisfied the corrected threshold, and their apparent deviations may have indicated type I errors.

Table 2 presents the detailed HWE test results for all 15 STR loci. The chi-square values ranged from 2.11 (TH01) to 25.67 (D21S11), where the latter was the most pronounced deviation from equilibrium expectations. The observed de-

Table 1: Allele frequencies and basic statistical parameters for 15 STR loci.

Locus	No. of Alleles	Allele Range	Most Common Allele (Frequency)	Least Common Allele (Frequency)	H_o	H_e
D8S1179	10	10–16	13 (0.280)	10 (0.007)	0.820	0.837
D21S11	14	24.2–35.2	30 (0.267)	35.2 (0.002)	0.833	0.824
D7S820	7	7–13	10 (0.340)	13 (0.013)	0.740	0.769
CSF1PO	9	8–15	12 (0.293)	15 (0.007)	0.667	0.715
D3S1358	7	12–19	15 (0.320)	19 (0.013)	0.747	0.752
TH01	6	6–10	9.3 (0.367)	6 (0.020)	0.807	0.763
D13S317	7	8–15	11 (0.307)	15 (0.020)	0.667	0.768
D16S539	8	8–15	11 (0.273)	15 (0.020)	0.773	0.764
D2S1338	13	15–28	19 (0.173)	28 (0.007)	0.793	0.878
D19S433	12	9–17.2	14 (0.200)	17.2 (0.007)	0.847	0.859
VWA	7	14–20	17 (0.273)	20 (0.027)	0.827	0.778
TPOX	7	6–12	8 (0.600)	6 (0.007)	0.520	0.591
D18S51	12	9–22	14 (0.160)	22 (0.007)	0.833	0.850
D5S818	9	7–16	11 (0.293)	16 (0.007)	0.753	0.764
FGA	12	17–26	22 (0.153)	17 (0.020)	0.847	0.862

H_o: observed heterozygosity, H_e: expected heterozygosity.

viations may reflect population substructure, consanguineous mating patterns, or technical factors affecting genotype calling. In particular, loci with significant deviations (D21S11, D3S1358, TPOX, and D5S818) will require careful consideration in forensic applications and population genetic analyses.

Heterozygosity patterns

Heterozygosity analysis detected substantial variations in genetic diversity among loci. The H_o values ranged from 0.520 at TPOX to 0.847 at both D19S433 and FGA, while the H_e values varied from 0.591 at TPOX to 0.878 at D2S1338. Most loci had H_o values slightly below the H_e values, consistent with HWE predictions and typical population patterns. However, four loci (D21S11, TH01, D16S539, and VWA) had H_o values that marginally exceeded the H_e values, potentially suggesting technical artifacts or population-specific factors influenced these measurements.

Table 2: Hardy–Weinberg equilibrium (HWE) test results.

Locus	Chi-square	*p*-value	HWE Status
D8S1179	8.45	0.133	Equilibrium
D21S11	25.67	<0.001[a]	Deviation
D7S820	4.23	0.238	Equilibrium
CSF1PO	7.89	0.162	Equilibrium
D3S1358	18.34	<0.001[a]	Deviation
TH01	2.11	0.348	Equilibrium
D13S317	12.45	0.029[a]	Deviation
D16S539	6.78	0.148	Equilibrium
D2S1338	14.23	0.047[a]	Deviation
D19S433	9.67	0.085	Equilibrium
VWA	5.43	0.143	Equilibrium
TPOX	22.89	<0.001[a]	Deviation
D18S51	8.91	0.112	Equilibrium
D5S818	19.78	<0.001[a]	Deviation
FGA	7.34	0.119	Equilibrium

[a] Significant deviation from Hardy–Weinberg equilibrium.

Forensic utility assessment

The forensic statistical parameters demonstrated the high discriminatory power of the 15 STR loci for identification and paternity testing applications in the southern Saudi population. Table 3 shows the comprehensive forensic statistical parameters for all of the loci examined in this study. The PD values ranged from 0.768 (D7S820) to 0.969 (D2S1338), where 12 of the 15 loci had PD values above 0.900, indicating their excellent capacity for discrimination.

Table 3: Forensic statistical parameters for 15 STR loci.

Locus	PD	PE	PIC	MP	TPI
D8S1179	0.949	0.719	0.826	0.051	2.78
D21S11	0.964	0.533	0.815	0.036	2.13
D7S820	0.768	0.542	0.538	0.232	2.18
CSF1PO	0.838	0.623	0.683	0.162	2.49
D3S1358	0.894	0.507	0.721	0.106	2.03
TH01	0.884	0.669	0.736	0.116	2.68
D13S317	0.917	0.682	0.738	0.083	2.73
D16S539	0.909	0.716	0.732	0.091	2.76
D2S1338	0.969	0.756	0.868	0.031	3.21
D19S433	0.921	0.756	0.846	0.079	3.21
VWA	0.897	0.731	0.748	0.103	2.86
TPOX	0.794	0.317	0.559	0.206	1.46
D18S51	0.949	0.703	0.835	0.051	2.81
D5S818	0.908	0.539	0.736	0.092	2.17
FGA	0.959	0.756	0.850	0.041	3.21

PD = power of discrimination, PE = probability of exclusion, PIC = polymorphic information content, MP = match probability, TPI = typical paternity index.

Table 4: Combined forensic statistical parameters.

Parameter	Value
Combined power of discrimination	0.999788998
Combined probability of exclusion	0.999989779
Combined match probability	2.83258e-16
Random match probability	3.53034e+15
Total number of possible genotypes	$6.57073827 \times 10^{24}$

Table 5: Comparison of key parameters with other Saudi regions.

Parameter	Southern Region	Central Region[a]	Western Region[b]
Highest PD locus	D2S1338 (0.969)	D19S433 (0.959)	D18S51 (0.965)
Lowest PD locus	D7S820 (0.768)	TPOX (0.806)	TPOX (0.840)
Highest PIC locus	D2S1338 (0.868)	FGA (0.85)	D18S51 (0.85)
Lowest PIC locus	D7S820 (0.538)	TPOX (0.57)	TPOX (0.61)
Average H_e	0.763	0.785	0.781
Average H_o	0.738	0.756	0.751

[a] Data from Osman et al.[12]
[b] Data from Alenizi.[13]

The PE values varied from 0.317 (TPOX) to 0.756 (D2S1338, D19S433, and FGA), reflecting the heterozygosity patterns observed at each locus.

The PIC values ranged from 0.538 (D7S820) to 0.868 (D2S1338), where most loci exhibited high informativeness (PIC $>$ 0.700). The MP values were correspondingly low, ranging from 0.031 (D2S1338) to 0.232 (D7S820). The TPI values ranged from 1.46 (TPOX) to 3.21 (D2S1338, D19S433, and FGA), indicating variable but generally adequate power for paternity determination.

Combined forensic efficiency

Collective analysis of all 15 STR loci demonstrated their exceptional forensic utility for the southern Saudi population. Table 4 presents the combined forensic statistical parameters, highlighting the cumulative power of the entire STR panel. The combined PD reached 0.999788998, indicating that the probability of two unrelated individuals sharing identical genotypes across all 15 loci was approximately 1 in 4.7 million. The combined PE was 0.999989779, demonstrating the excellent capacity for use in paternity exclusion cases.

The combined MP was calculated as 2.83258e-16, corresponding to a random MP of 3.53034e+15 to 1, providing overwhelming statistical support for identity determination. The total number of possible genotypes across all loci was estimated as $6.57073827 \times 10^{24}$, reflecting the enormous genetic diversity captured by this STR panel and its suitability for forensic applications in the population considered.

Regional comparison analysis

Comparative analysis with other Saudi regional populations revealed both similarities and notable differences in STR characteristics. Table 5 compares the key parameters between the southern region examined in this study and data reported previously from central and western Saudi regions. D2S1338 was identified as the most discriminating locus for the southern region (PD = 0.969), whereas D19S433 (PD = 0.959) and D18S51 (PD = 0.965) were their most informative markers for the central and western regions, respectively.

Similar loci were identified as least discriminating in all three regions, where D7S820 was the least discriminating marker for the southern region (PD = 0.768), and TPOX for both the central and western regions (PD = 0.806 and 0.840, respectively). The PIC patterns followed similar trends, where D2S1338 had the highest PIC value in the southern region (0.868), but FGA in the central region (0.85) and D18S51 in the western region (0.85).

The average H_e values were slightly lower in the southern region (0.763) compared with the central (0.785) and western regions (0.781), and the patterns were similar for the H_o values (southern: 0.738; central: 0.756; western: 0.751). These regional differences may reflect population history, migration patterns, and varying degrees of genetic isolation or admixture among regional Saudi populations.

Discussion

Comprehensive profiling of 15 autosomal STR loci in the southern Saudi population for the first time in this study identified 140 distinct alleles and high heterozygosity levels, highlighting the considerable genetic diversity in this geographic area. Some loci deviated from HWE, but such deviations are commonly reported in populations with complex demographic histories marked by consanguinity and subpopulation structure.[24,25] The detection of 140 alleles across the 15 loci demonstrated the substantial genetic variation within the southern Saudi population, consistent with observations from other Middle Eastern populations.[26,27] This high level of polymorphism suggests an extraordinary potential for genotype combinations, which were estimated as 6.57×10^{24}, surpassing estimates in previous Saudi studies conducted in the central region.[28] The enhanced genetic diversity observed in the southern region probably reflects historical migration flows and the characteristic rich tribal diversity of this area.[29]

Among the loci studied, D2S1338 was the most informative marker, with a PIC value of 0.868 and PD value of 0.969. These results align well with previous studies of Egyptian and Jordanian populations,[30,31] corroborating the global trend where D2S1338 consistently ranks as one of the most polymorphic STR loci.[32] By contrast, the PIC values are slightly lower for this locus in northern Saudi populations,[24] indicating regional genetic stratification within KSA.

Conversely, loci such as TPOX and D7S820 exhibited reduced forensic utility. The lower informativeness of the TPOX locus replicates findings obtained in Riyadh and Jeddah cohorts,[25] where allelic variation was limited. In particular, the high frequency of allele 8 at TPOX (0.600) exceeds European population values,[32] but aligns closely with frequencies reported in Gulf populations such as Kuwait[33] and Bahrain.[34] Similarly, the reduced utility of

forensic parameters for D7S820 is consistent with patterns observed across the Arabian Peninsula, reflecting shared ancestral origins and restricted historical gene flow with continental populations.[29]

HWE deviations in six loci necessitate consideration of underlying demographic factors that characterize Saudi populations. Previous studies in central and eastern KSA also highlighted departures from HWE, which were primarily attributed to high consanguinity rates, and as high as 57.7% nationally. From a forensic practice perspective, the observed HWE deviations and evidence of population substructure have a direct implication for the statistical interpretation of DNA evidence because likelihood ratio calculations based on simple allele frequency products may underestimate match probabilities in a structured population. Therefore, forensic practitioners who apply these frequency data are recommended to incorporate the Balding–Nichols substructure correction factor (θ, theta) in likelihood ratio calculations. A θ value of 0.01–0.03 is commonly recommended for structured populations,[35] and would be appropriate for forensic casework involving individuals from the southern Saudi region, pending formal estimation of θ for this population.[36,37] Consanguineous unions are known to inflate homozygosity and distort expectations under random mating models.[38]

Population substructure within the southern region further contributed to these observations. The southern provinces contain multiple tribal groups with historical settlement patterns, potentially causing genetic subdivision, and this effect has also been reported in UAE and Qatar populations where tribal endogamy influences population genetic structure and HWE patterns.[39,40]

The combined discrimination power of 0.999788998 and low MP (2.83×10^{-16}) affirm the robustness of the 15-loci system for forensic applications in southern KSA. These values exceed those reported in previous studies of other Saudi regions utilizing fewer markers,[24,25] reflecting both the expanded marker panel and the pronounced genetic diversity of the southern population.

International comparisons indicate that the combined discrimination power achieved in this study is among the highest reported for Arab populations, surpassing values obtained for Lebanese and Iraqi populations using similar STR sets.[41,42] The elevated forensic efficacy highlights the appropriateness of these markers for human identification in southern KSA.

Regional genetic stratification was apparent within KSA, where the southern population had higher PIC values at D2S1338 (0.868) relative to central and western populations.[24] The slightly lower heterozygosity observed in the south compared with coastal populations probably reflects divergent demographic histories, including varying historical admixture with African populations facilitated by ancient trading routes.[43]

These regional differences have significant forensic relevance due to variations in allele frequencies that impact likelihood ratio calculations. Previous studies have stressed the need for population-specific databases in the Arabian Peninsula to ensure statistical accuracy in forensic casework.[25,27]

The high combined PE of 0.999989779 supports the utility of this STR panel for paternity testing contexts within the southern Saudi population. Loci such as D2S1338, D19S433, and FGA were found to have exclusion powers (~0.756) exceeding those reported in earlier Saudi studies and comparable to highly polymorphic global populations.[24,32] This performance is critically important given the frequent paternity testing demands in Saudi legal and familial dispute resolutions.

These data are foundational for establishing a comprehensive national forensic DNA database for KSA. The regional genetic differences observed in the present study highlight the importance of representative sampling across all geographic zones to achieve reliable population frequency estimates. Previous recommendations for the Arabian Peninsula advocate for multi-regional data integration to account for internal genetic stratification.[25,27]

The considerable informativeness of these autosomal STR loci supports their continued use in forensic laboratories that service the southern region, as well as indicating a need to expand marker sets with rapidly mutating STRs and insertion–deletion polymorphisms to enhance the discrimination power in complex cases.[44,45]

The main limitations of this study include the relatively small sample size of 150 individuals, which might not have been fully representative of the genetic diversity of the southern Saudi population, particularly given its complex tribal structure and demographic history. The significant HWE deviations observed for four loci (after Bonferroni correction) suggest potential population stratification, which could affect the reliability of forensic statistical calculations. HWE analysis was conducted across 15 loci simultaneously without initial multiple-testing correction, increasing the risk of type I errors. The Bonferroni-corrected interpretation presented in this study addressed this limitation post hoc, but future studies should incorporate this correction prospectively. Details of the demographic characteristics of participants were limited because tribal affiliation, consanguinity levels, and specific geographic origins within the southern region were not recorded systematically, making it difficult to explain observed genetic patterns such as HWE deviations and allele frequencies. In addition, the present study focused exclusively on 15 autosomal STR markers, missing broader genetic variations that could be captured through expanded marker panels, including Y-chromosomal and mitochondrial DNA. Reliance on a single DNA extraction method further limits the generalizability to different forensic laboratory conditions.

Conclusion

This study effectively characterized the genetic variation of 15 autosomal STR loci in the southern Saudi Arabian population, revealing high genetic diversity with 140 alleles and strong forensic statistical parameters, including a combined PD of 0.99979 and PE of 0.99999, which confirmed the robustness of the markers for forensic DNA analysis and paternity testing. Most loci had high PIC values, although deviations from HWE in six loci indicated population-specific genetic traits. The observed regional genetic differences highlight the need for population-specific forensic databases in KSA. By providing essential baseline data, this study contributes to the development of a comprehensive national forensic database, supporting DNA-based identification systems linked to national identity programs, and facilitating applications in paternity testing and disaster

victim identification. Future studies should expand the sample sizes and incorporate Y-chromosomal and mitochondrial DNA markers to allow more comprehensive genetic characterization of Saudi populations, thereby enhancing the accuracy and reliability of forensic and genetic services in the region.

Source of funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors have no conflict of interest to declare.

Ethical approval

This study was reviewed and approved by the Research Ethics Committee of the Armed Forces Hospital, Southern Region, KSA (Registration Number: H-06-KM-001). Informed consent was obtained from all participants prior to their inclusion in the study.

Authors contributions

Conceptualization, HA and A.M.; methodology, HA and A.M.; software, HA and A.M..; formal analysis, HA and A.M.; investigation, HA, A.M., MNHA and MNMD; data curation, HA, A.M., MNHA and MNMD; writing—original draft preparation, HA and A.M.; writing—review and editing, MNHA and MNMD; visualization, HA, A.M., MNHA and MNMD; supervision, A.M., and MNHA. All authors have read and agreed to the published version of the manuscript.

Acknowledgment

The authors gratefully acknowledge Armed Forces Hospital, Southern Region, Saudi Arabia for donating the samples and issuing the final ethical approval and consent form.

Declaration of generative AI

During the preparation of this work, the authors used Claude Sonnet 4 in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jtumed.2026.04.002.

References

1. Butler JM. *Forensic DNA typing: biology, technology, and genetics of STR markers*. 2nd ed. New York: Academic Press; 2005.
2. Goodwin W, Linacre A, Hadi S. *An introduction to forensic genetics*. 2nd ed. Chichester: Wiley-Blackwell; 2011.
3. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. **Am J Hum Genet 1991**; 49(4): 746–756.
4. National Research Council. *The evaluation of forensic DNA evidence*. Washington DC: National Academy Press; 1996.
5. Kassab A Ch, Alaqeel HF, Messaoudi SA, Babu SR, Shahid SA, Chaudhary AR. Population data and genetic diversity analysis of 17 Y-STR loci in Saudi population. **Egypt J Forensic Sci 2020**; 10(1): 30.
6. Alsafiah HM, Goodwin WH, Hadi S, Alshaikhi MA, Wepeba PP. Population genetic data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler® PCR amplification kit. **Forensic Sci Int Genet 2017**; 31: e59–e61.
7. Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. **Am J Hum Genet 1989**; 44(3): 388–396.
8. Biosystems Applied. *AmpFISTR identifiler PCR amplification kit user's manual*. Foster City, CA: Applied Biosystems; 2006.
9. Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFℓSTR® identifiler® PCR amplification kit. **J Forensic Sci 2004**; 49(6): 1265–1277.
10. Evett IW, Weir BS. *Interpreting DNA evidence: statistical genetics for forensic scientists*. Sunderland: Sinauer Associates; 1998.
11. Hardy GH. Mendelian proportions in a mixed population. **Science 1908**; 28(706): 49–50.
12. Osman AE, Alsafar H, Tay GK, Theyab J, Mubasher M, Sheikh NE, et al. 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 2015**; 6: 267.
13. Alenizi LBTD. *Evaluation of physical genetic trait sites of the population in the western region of the Kingdom of Saudi Arabia (Master's thesis, Naif Arab University for Security Sciences)*; 2016 https://repository.nauss.edu.sa/123456789/64312.
14. Abuidrees AS, Ishaq MJ, En Pu C, Alhamad NA, Alnafea HA, Almehaizea AM. A globally used 15 short tandem repeats (STR) loci in forensic human identification, with their allele frequencies and statistical values in the population of Bahrain. **Arab Gulf J Sci Res 2014, 177**; 32.
15. Al-Obaidli A, Sabbagh A, Pravin S, Krishnamoorthy R. Present day inbreeding does not forbid the forensic utility of commonly explored STR loci: a case study of native qataris. **Forensic Sci Int Genet 2009**; 4(1): e11–e13.
16. Hares DR. Selection and implementation of expanded CODIS core loci in the United States. **Forensic Sci Int Genet 2015**; 17: 33–34.
17. Bodner M, Bastisch I, Butler JM, et al. Recommendations of the DNA commission of the ISFG on quality control of autosomal STR allele frequency databasing (STRidER). **Forensic Sci Int Genet 2016**; 24: 97–102.
18. Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. **J Forensic Sci 2001**; 46: 453–489.
19. S. Mestiri, S. Boussetta, A.J. Pakstis, S. El Kamel, A. Ben Ammar El Gaaied, K.K. Kidd, L. Cherni, New Insight into the human genetic diversity in North African populations by genotyping of SNPs in DRD3, CSMD1 and NRG1 genes, *Mol Genet Genomic Med*, **10**, 2022, e1871.
20. K. Mineta, K. Goto, T. Gojobori, F.S. Alkuraya, Population structure of indigenous inhabitants of Arabia, *PLoS Genet*, **17** (1), 2021, e1009210.
21. El-Hazmi MA, Al-Swailem AR, Warsy AS, et al. Consanguinity among the Saudi Arabian population. **J Med Genet 1995**; 32(8): 623–626.

22. Khubrani YM, Hallast P, Jobling MA, Wetton JH. Massively parallel sequencing of autosomal STRs and identity-informative SNPs highlights consanguinity in Saudi Arabia. **Forensic Sci Int Genet 2019**; 43:102164.
23. Khubrani YM, Wetton JH, Jobling MA. Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs. **Forensic Sci Int Genet 2018**; 33: 98–105.
24. Abdin L, Shimada I, Brinkmann B, Hohoff C. Analysis of 15 short tandem repeats reveals significant differences between the Arabian populations from Morocco and Syria. **Leg Med 2003**; 5: S150–S155.
25. Khubrani YM. *Genetic diversity and population structure of Saudi Arabia*. Doctoral dissertation, University of Leicester; 2020.
26. Hadi A, Hadi S, Almohammed EK, Lazim H. Deciphering the genetic diversity in the Arabian Peninsula and Africa: insights from Y-STR data. **Forensic Sci Med Pathol 2025**: 1–9.
27. Alhmoudi OA, Jones RJ, Tay GK, Alsafar H, Hadi S. Population genetics data for 21 autosomal STR loci for United Arab Emirates (UAE) population using next generation multiplex STR kit. **Forensic Sci Int Genet 2015**; 19: 190–191.
28. 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.
29. Messaoudi S, Alamri MA, Babu SR, Alsaleh AB, Albujja MH, AlSnan N, et al. Evaluation of forensic genetic parameters of 24 STR loci and Y indel in a Southern Region saudi population sample using GlobalFiler™ PCR amplification kit. **Arab J Forensic Sci Forensic Med 2021**; 3(2): 245–259.
30. Omran GA, Rutty GN, Jobling MA. Genetic variation of 15 autosomal STR loci in Upper (Southern) Egyptians. **Forensic Sci Int Genet 2009**; 3(2): e39–e44.
31. Al-Eitan LN, Tubaishat RR. Evaluation of forensic genetic efficiency parameters of 22 autosomal STR markers (PowerPlex® Fusion system) in a population sample of Arab descent from Jordan. **Aust J Forensic Sci 2018**; 50(1): 97–109.
32. Butler JM. *Advanced topics in forensic DNA typing: interpretation*. Academic Press; 2014.
33. Haidar M, Abbas FA, Alsaleh H, Haddrill PR. Population genetics and forensic utility of 23 autosomal PowerPlex Fusion 6C STR loci in the Kuwaiti population. **Sci Rep 2021**; 11(1): 1865.
34. Al-Snan NR, Messaoudi S, Babu S R, Bakhiet M. Population genetic data of the 21 autosomal STRs included in GlobalFiler kit of a population sample from the Kingdom of Bahrain. **PLoS One 2019**; 14(8):e0220620.
35. Buckleton JS, Triggs CM, Walsh SJ, editors. *Forensic DNA evidence interpretation*. 2nd ed. Boca Raton: CRC Press; 2016.
36. Al-Gazali LI, Hamamy H, Al-Arrayad S. Genetic disorders in the Arab world. **BMJ 1997**; 7141: 611–614.
37. Hamamy H. Consanguineous marriages: preconception consultation in primary health care settings. **J Community Genet 2012**; 3(3): 185–192.
38. Tadmouri GO, Nair P, Obeid T, Al Ali MT, Al Khaja N, Hamamy HA. Consanguinity and reproductive health among Arabs. **Reprod Health 2009**; 6(1): 17.
39. Naji M, Damji R, Adan A, Qamar SA, Alghafri R. Population genetics data of 23 autosomal STR loci for three populations in United Arab Emirates. **Forensic Sci Int: Genet Suppl Ser 2019**; 7(1): 187–188.
40. Hunter-Zinck H, Musharoff S, Salit J, Al-Ali KA, Chouchane L, Gohar A, et al. Population genetic structure of the people of Qatar. **Am J Hum Genet 2010**; 87(1): 17–25.
41. Riman S, Ghemrawi M, Borsuk LA, Mahfouz R, Walsh S, Vallone PM. Sequence-based allelic variations and frequencies for 22 autosomal STR loci in the Lebanese population. **Forensic Sci Int Genet 2023**; 65:102872.
42. Farhan MM, Hadi S, Iyengar A, Goodwin WH. Population genetic data for 20 autosomal STR loci in an Iraqi Arab population: application to the identification of human remains. **Forensic Sci Int Genet 2016**; 25: e10–e11.
43. Pfennig A, Petersen LN, Kachambwa P, Lachance J. Evolutionary genetics and admixture in African populations. **Genome Biol Evol 2023**; 15(4): evad054.
44. Vergani D. Forensic identity markers: an overview of forensic autosomal STRs and SNPs. **Forensic DNA Analysis 2021**: 53–76.
45. Murthy V, Jia LF, Samuel VP, Kademane K. Forensic identification by using insertion-deletion polymorphisms. **Int J Hum Genet 2015**; 15(2): 55–59.

How to cite this article: Alhatim H, Muthanna A, Hakim Abdullah MN, Mohd Desa MN. Genetic variation of 15 autosomal STR loci in the southern region population of KSA: A forensic DNA analysis study. J Taibah Univ Med Sc 2026;21(3):452–459.