AL-KUNOOZE SCIENTIFIC JOURNAL
ISSN: 2706-6231 E ,2706-6223 P
Vol.11 No.4 (2025)

Direct sequence of D13S317 loci for analysis and identification of STR structure in Iraqis

Maysoon Khudheyer Derraj Almyahi
Department of Medical Laboratory Technology, College of Health and Medical Techniques, Southern Technical University.

***Corresponding Author Email Address**: myalmyahi@stu.edu.iq
ORCID ID:

Abstract

The analysis of DNA sequences plays a critical role in genetic research, but challenges persist in attaining high precision with mixed sequences. This study aims to evaluate the efficiency of the Mixed Sequence Analysis application in analyzing DNA sequences with heterozygous base calling as well as manual analysis. DNA was extracted from 31 unrelated Iraqi individuals, and D13S317 is one specific STR locus that is highly variable among individuals and was amplified. The mixed sequence was then analyzed using the Mixed Sequence Analysis tool, complemented by manual chromatogram interpretation. The most frequent allele detected was TATC 12 (52.27%), followed by TATC 9 (20.45%). The least frequent allele was TATC10, which was 4.55%; also, different alleles were detected with high variability. The analysis revealed that while the application was beneficial, there were challenges in distinctive variations at the base level due to noise in the chromatogram. Despite its limitations, this research highlights the importance of developing accurate sequencing analysis techniques, particularly in situations where other methods are unavailable.

Key words: Genetic Variation, STR, D13S317 Locus, DNA sequencing

Introduction

Short Tandem Repeats (STRs) are repeating sequences of 2-6 base pairs of DNA that are broadly spread across the human genome. These loci are extremely polymorphic, since they exhibit a high degree of difference in their number of repeats among individuals, making them valuable tools for genetic analysis. STRs are widely used in paternity testing, genetic mapping, forensic science, and anthropological studies due to their consistency, ease of analysis, and presence in non-coding regions of the genome, which guarantees that they do not impact phenotypic traits [1]. This large diversity can be utilized to help disambiguate population structure or to infer historical human migration [2].

D13S317 is one specific STR locus, located on chromosome 13, that is part of the set of loci commonly analyzed in forensic DNA profiling. D13S317 is a highly polymorphic marker that involves a repeating sequence of 4 base pairs (TATC) [3]. The number of repeats at this locus can differ significantly between individuals, making it an exceptional marker for identifying genetic alterations [4;5]. The inclusion of D13S317 in forensic panels helps in providing a more consistent and precise genetic profile, mainly when analyzing complex mixtures of DNA samples.

The D13S317 locus has also been studied in the context of population genetics, where its variation is used to evaluate genetic diversity among different populations. Furthermore, the D13S317 marker is one of the many loci examined in the widely used CODIS (Combined DNA Index System) database, which assists law enforcement agencies in comparing and matching DNA samples from crime scenes with identified individuals [1].

Mixed DNA samples, containing genetic material from multiple donors, pose significant challenges in genetic analysis, particularly when using Sanger sequencing.

In such mixtures, heterozygous loci often produce overlapping peaks in chromatograms, which complicates accurate allele interpretation and individual DNA profile resolution. This complication is especially relevant in forensic science, clinical diagnostics, and microbiological studies, where accurate DNA identification is critical.

A free web-based program named **Mixed Sequence Reader (MSR) (http://msr.cs.nthu.edu.tw/)**, designed to analyze DNA sequences from heterozygous base-calling fluorescence chromatograms, typically in .abi file format [6]. MSR compares these sequences with reference data to identify variations such as indels, SNPs, and STRs. It can physically discover indels and STRs and also predict microsatellite patterns. This study explores whether the Mixed Sequence Reader (MSR) software can precisely resolve mixed STR alleles in the D13S317 locus of Iraqi individuals.

Methods

Sample Collection and DNA Isolation

A total of 31 blood samples were collected from unrelated Iraqi individuals. Genomic DNA was isolated from the whole blood using a commercial DNA extraction kit (Genaid, Taiwan) according to the manufacturer's instructions.

PCR Amplification of the D13S317 Flanking Region

The D13S317 flanking region was amplified using the following primers [7]: Forward (F): GGGTTGCTGGACATGGTATC; Reverse (R): TCCGAGGTTCTCTTCCTGTG

PCR was achieved in a thermal cycler (Applied Biosystem, US) in a total volume of 25 µL containing: 12.5 µL from 10X PCR GoTaq Green master mix (Promega, USA), 1 µL (10 pmol)of each primer, and 2 µL genomic DNA; the volume was completed with D.D. water. The PCR amplification was achieved under the following conditions:

initial denaturation at 95°C for 5 minutes, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 58°C for 30 seconds, and extension at 72°C for 1 minute. A final extension step was carried out at 72°C for 7 minutes.

DNA sequencing

Sequencing of the PCR-amplified D13S317 region was done using the Sanger method at an expert genomic facility in South Korea. The purified PCR products were exposed to cycle sequencing using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), and the sequencing was achieved on an ABI 3500 Genetic Analyzer.

Sequences interpretation

The resulting chromatogram files, provided in .abi format, were analyzed using the MSR software to identify genomic variations like indels and microsatellite repeats. MSR aligns sequences with reference data to identify variation positions, ensuring accurate identification even for ambiguous chromatogram traces. The Mixed Sequence Reader (MSR) software was used with default parameters for sequence alignment and analysis. The GRCh37/hg19 human genome assembly in FASTA format and uploaded as the reference file within the MSR platform was used as the reference sequence for the D13S317 locus. The chromatogram files (.abi) created from Sanger sequencing were aligned against this reference to recognize the number and structure of TATC repeats.

The MSR software was configured with the following cutoff settings:

- Minimum STR repeat length: 4 bp
- Mismatch tolerance: 2 bases
- Signal threshold for base calling: 20% above background
- IUPAC ambiguity codes were used to distinguish overlapping heterozygous peaks

Manual curation of each sample's sequence alignment was achieved in parallel to confirm

repeat number and structure.

Results

The amplified D13S317 flanking region from 31 blood samples is shown in Figure 1, and the PCR products were analyzed using the Mixed Sequencer Reader (MSR) software, accompanied by manual analysis. The analysis exposed mixed short tandem repeat (STR) sequences (Figure 2a), which were analyzed to determine the repeat number and frequency (Figure 2b).

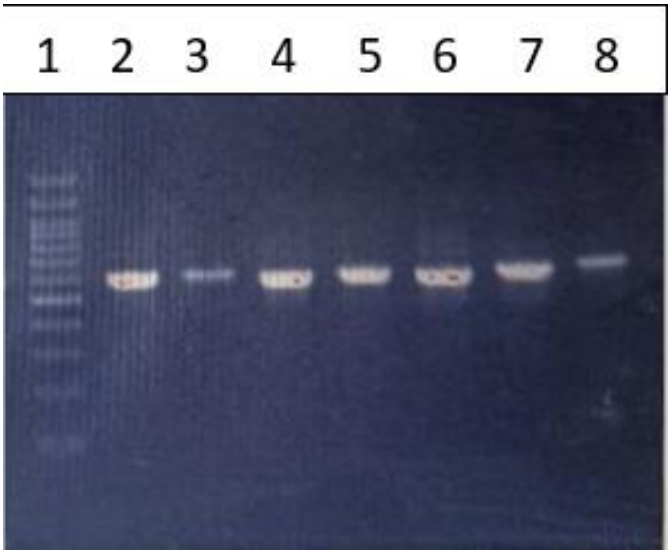

Figure 1: The electrophoresis image shows separate bands for samples 2-7, with a 603 bp PCR product for the D13S317 region. Lane 1 is 100 bp ladder for size reference. Lane 8 shows a shorter allele, representing a different PCR product size compared to the others.

In this study, the mixed chromatograms from the direct sequencing of the D13S317 region were analyzed, indicating two different alleles with a heterozygous genotype. Using manual aids and MSR software, the sequence data were compared to the GRCh37 reference, identifying variations in the number of TATC repeats and confirming the allele lengths, therefore providing insights into the genetic variation at this microsatellite locus.

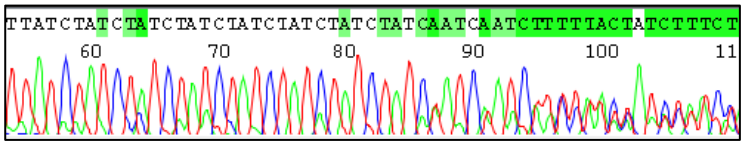

Figure 2: a. The direct sequencing of the D13S317 region displayed a mixed chromatogram, representing a heterozygous genotype with two different alleles (TATC8 and TATC12) present in the sample.

Seq: TTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCWATCWATCWWYYWWYYAWYYW

Seq1: TTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCAATCAATCATCTATCTATCTTTCT

Seq2: TTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCAATCAATC

Figure 2: b. Seq1 has 8 TATC repeats, while Seq2 has 12 TATC repeats, resulting in a difference in length between the two sequences. IUPAC symbols represent overlapped traces.

The frequencies of the D13S317 alleles in the studied Iraqi population are presented in Table 1. The 12-repeat allele was the most frequent, appearing 23 times (52.27%), followed by the 9-repeat allele (9 times, 20.45%). The less common alleles 8-repeat, 11-repeat, and 10-repeat, were observed with frequencies of 4, 3, and 2, respectively. The repeats arrangement is also demonstrated.

Table 1: The alleles, which were read and analyzed with maximal accuracy using the mixed sequence reader software, are detailed in the table below, attended by the particular allele frequencies or counts. Alleles exhibiting inadequate clarity were excluded from the analysis.

Repeat Number	Frequency	Structure
TATC 8	4 (9.09%)	(TATC)8 (AATC)2 (ATCT)3
TATC 9	9 (20.45%)	(TATC)9 (AATC) (ATCT)3 (TATC)9 (AATC)2 (ATCT)3
TATC 10	2 (4.55%)	(TATC)10 (AATC) (ATCT)3
TATC 11	3 (6.82%)	(TATC)11 (AATC)2 (ATCT)3
TATC 12	23 (52.27%)	(TATC)12 (AATC) (ATCT)2 (TATC)12 (AATC)2 (ATCT)3 (TATC)12 (AATC) (ATCT)4
TATC 13	3 (6.82%)	(TATC)13 (AATC)2 (ATCT)3
Total alleles no.	44	

Discussion

The D13S317 locus is a valuable marker in forensic genetics, used for both individual identification and population genetic studies. In the current study, the allele distribution noticed in the Iraqi population was consistent with common alleles seen in other populations, but also established some unique characteristics. The D13S317 locus showed a drop-out phenotype in critical cases due to variation in primer sites, so this indicates its instability [8].

The most frequent allele in this study was the 12-repeat allele, which was observed 23 times. This is equivalent to previous Iraqi study results [9] and in other populations, where the 12-repeat allele is frequently seen, suggesting a common allele in global genetic diversity [10], unlike Chinese population alleles, among them, the 8-repeat allele was the most frequent [11]. Our results demonstrate different TATC motifs arrangements especially the 12-repeat allele from other populations' alleles [12].

The combination of MSR software and manual analysis was helpful in ensuring the precision of the results. The method used in this study reflects the importance of developing strong analytical skills to make the best use of available resources [13]. This approach can serve as a model for other populations with similar resource limitations. It was found that the structure of Iraqi D13S317 STR alleles matched to the repeat context earlier published by [14] with minor differences. Our study follows the guiding principle for high-quality genetic data reporting [15]. By providing that clear allele frequencies and methods for the Iraqi D13S317 locus, we ensure reliable and reproducible outcomes that support forensic and population genetics applications.

Conclusion

This study offers different and population-specific insights into the sequence structure of the D13S317 locus among Iraqis, contributing to the growing body of forensic and population genetic data. The observed inconsistency in motif arrangements, particularly within alleles of identical length, underscores the limitations of length-based STR typing and highlights the need for sequence-based methods in certain forensic contexts. The successful incorporation of MSR software with manual interpretation reveals a practical and cost-effective strategy for sequence analysis in resource-limited settings. These results are not only valuable

for improving the accuracy of forensic DNA profiling and kinship examination in the Iraqi population but also contribute to larger comparative genomic studies across various populations.

Conflict of Interest: The Author declares there is no conflict of interest.

References:

1- Butler JM. Forensic DNA typing: Biology, technology, and genetics of STR markers. 2nd ed. San Diego: Elsevier Academic Press; 2005.

2- Hosseini SF, Bijanzadeh M, Modheji E. Pattern analysis of short tandem repeats allele frequencies among the population of Khuzestan Province, South of Iran. Avicenna J Med Biotechnol. 2018;10(4):257.

3- Youn BJ, Lee K, Kim CH. A case of single-step mutations at two short tandem repeat loci (D13S317 and DXS10148) among three generations of a Korean family. Biomed Sci Lett. 2022;28(4):327–33.

4- Hedman M. Long D13S317 variant allele: A cautionary case report. Forensic Sci Int Genet. 2015;14:38–41.

5- Singh Negi D, Alam M, Bhavani SA, Nagaraju J. Multistep microsatellite mutation in the maternally transmitted locus D13S317: a case of maternal allele mismatch in the child. International Journal of Legal Medicine. 2006 Sep;120:286-92.

6- Chang CT, Tsai CN, Tang CY, Chen CH, Lian JH, Hu CY, Tsai CL, Chao A, Lai CH, Wang TH, Lee YS. Mixed sequence reader: a program for analyzing DNA sequences with heterozygous base calling. The Scientific World Journal. 2012;2012(1):365104.

7- Al-Hashimi AH, Al-Khazraji SM. Forensic STR identification of human teeth samples exposed to various acidic and alkaline chemical conditions in the Iraqi population. Forensic Sci Int Genet. 2022;56:102612.

8- Boutrand L, Egyed B, Füredi S, Mommers N, Mertens G, Vandenberghe A. Variations in primer sequences are the origin of allele drop-out at loci D13S317 and CD4. Int J Legal Med. 2001;114:295–7.

9- AL-Rubai HK, AL-Zubaidi MM, Ibrahem HK, Mohammed A, Rashed S. Revealed of a novel allele in Wasit–Iraqi population. Iraqi J Sci. 2015;56(4A):2798–806.

10- Mpoloka SW, Kgotlele T, Wally A. Determination of allele frequencies in nine short tandem repeat loci of five human sub-populations in Botswana. Afr J Biotechnol. 2008;7(8).

11- Gao S, He L, Li Z, Zhang Y, Deng Z. Analysis of genetic polymorphisms and a novel tri-allelic cloning sequence for the D13S317 locus among an ethnic Han Chinese population. Zhonghua Yi Xue Yi Chuan Xue Za Zhi. 2024;41(1):42–4.

12- Gettings KB, Ballard D, Bodner M, Borsuk LA, King JL, Parson W, Phillips C. Report from the STRAND Working Group on the 2019 STR sequence nomenclature meeting. Forensic Science International: Genetics. 2019 Nov 1;43:102165.

13- Almyah MK, Albadran AI. Screening of LEP gene polymorphisms as a risk factor for obesity and type 2 diabetes in Iraqis. Mol Biol Res Commun. 2019;8(4):156–65.

14- Gettings KB, Borsuk LA, Ballard D, Bodner M, Budowle B, Devesse L, King J, Parson W, Phillips C, Vallone PM. STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. Forensic Science International: Genetics. 2017 Nov 1;31:111-7.

15- Carracedo Á, Butler JM, Gusmão L, Linacre A, Parson W, Roewer L, Schneider PM. New guidelines for the publication of genetic population data. Forensic Science International: Genetics. 2013 Feb 1;7(2):217-20.

التسلسل المباشر لمنطقة D13S317 لتحليل وتحديد بنية التكرارات القصيرة (STR) لدى العراقيين

ميسون خضير دراج المياح

قسم تقنيات المختبرات الطبية،

كلية التقنيات الصحية والطبية، الجامعة التقنية الجنوبية

الملخص

يلعب تحليل تسلسل الحمض النووي دورًا حاسمًا في الأبحاث الوراثية، إلا أن التحديات لا تزال قائمة في تحقيق دقة عالية خاصة عند وجود تسلسلات مختلطة. تهدف هذه الدراسة إلى تقييم كفاءة تطبيق تحليل التسلسلات المختلطة في تحليل تسلسلات الحمض النووي التي تحتوي على قواعد متغايرة الزيجوت، بالإضافة إلى التحليل اليدوي. تم استخلاص الحمض النووي من 31 فردًا عراقيًا لا تربطهم صلة قرابة، وتم تضخيم منطقةD13S317 ، وهي إحدى مناطق STR المتغيرة بشكل كبير بين الأفراد. تم تحليل التسلسل المختلط باستخدام أداة تحليل التسلسلات المختلطة، بالإضافة إلى التفسير اليدوي للمخطط الكروماتوغرافي. أظهرت النتائج أن أكثر الأليلات تكرارًا هو TATC 12 بنسبة (52.27%)، يليه TATC 9 بنسبة (20.45%). وكان أقل الأليلات تكرارًا هو TATC10 بنسبة (4.55%). كما تم الكشف عن أليلات مختلفة بدرجة عالية من التباين. كشف التحليل أنه على الرغم من فائدة التطبيق، إلا أن هناك تحديات في التمييز بين الاختلافات الدقيقة على مستوى القواعد نتيجة التشويش في الكروماتوغرام. وعلى الرغم من هذه التحديات، تسلط الدراسة الضوء على أهمية تطوير تقنيات تحليل تسلسلي دقيقة، خصوصًا في الحالات التي لا تتوفر فيها طرق أخرى.

الكلمات المفتاحية :التنوع الجيني، التكرارات القصيرةSTR ، منطقةD13S317 ، تسلسل الحمض النووي