Global Scientific JOURNALS

GSJ: Volume 9, Issue 10, October 2021, Online: ISSN 2320-9186

www.globalscientificjournal.com

Systemic Review: Optimization of Short Tandem Repeat CSF1PO for Forensic Studies.

Zukhruf Asif,Husnain Karim Riaz,Mahnoor Afzal, Maryam Ali, Ammara Sameen, Alima Khan, Shehreen Sohail, Fareeha Sohail, Hamza Rana, Mukarram Farooq

Abstract:

Short tandem repeats (STRs) are small segments of 2-7 base pairs in size existing at some stage in the human genomic region. These regions exhibit excessive mutation change than any different areas of DNA that result in greater genetic diversity among humans. The number of repeats in STR markers is extraordinarily variable among the human population. STR marker evaluation is very fine at figuring out individuals at the molecular level, even from small quantities of remains in land mass disasters. STR loci in the identification selected in accordance to their dimension (CSF1PO) from NIST (National Institute of Standards and Technology). Conditions encompass for STR with the modifications in temperature conditions, magnesium ion, primers concentration, and setting-up, PCR of markers used to be carried out. PCR product investigated on the agarose gel electrophoresis. The Results confirmed that all STR loci under study are detectable as PCR. PCR effects showed that when the detection of primer and temperature measured by using the fixed concentration of magnesium, using a buffer and set Magnesium prerequisites towards adjustments in the primer awareness and temperature, when Taq polymerase enzyme is brought to test tubes, PCR bands are visible desirably.

© GSJ

Key words: Polymerase chain reaction, Primer, Tandem, STR genome, Chromosome.

Introduction:

The word mini satellite was originated in 1985.The short tandem repeats motifs that are based on the 10-60 base pairs to evaluate these [1]. They are present in numerous loci in the human genes and mainly show the less repetition at a contemptible locus. In 1982 Hamada first, narrated short tandem repeats and simple tandem repeats or microsatellite were named later for these sequences [2]. Analyzing DNA sequence is a technique and methodology used in field of forensic sciences that is mostly used in criminal investigation and paternity testing for identification. Designing and optimizing methodology is able to recognize specific regions with distinct population distribution in genomic DNA [3]. Human genome contains short repeat sequences of 2-7 base pairs are known as short tandem repeats (STR) that are reported by several studies and investigations. Genetic diversity is found in these STR regions that show a mutation rate that is higher than the mutations found in other regions of DNA in an individual [4]. In human population, the number of repeat units in the STR marker is variable. STR marker evaluation very precisely figures out individual humans at the molecular level even from very small quantities. DNA sequencing investigation is one of the strategies utilized in scientific science that normally utilized for criminal cases and paternity test concentrate to distinguish the speculated person. This requires planning and improving strategy which is equipped for perceiving the particular district with interesting populace appropriation is found in genomic DNA. In human genomic region contain short tandem repeats (STRs) which are brief portions of 2-7 base combines long detailed by a few examinations. High hereditary assorted variety is found in these regions that show transformation charge higher than some other areas of DNA apportioned among people. [4] In the human populace, the scope of repeats inside the STR marker is outstandingly factor. At the sub-atomic level, from few remainders all through a natural land mass catastrophe. They are present at many loci in the genomic DNA and there is

© GSJ

low number of repetitions at a specific locus. There is another class of repetitive DNA that consist of 1 to 10 base pairs. In 1982, Hamada first, described the existence of dinucleotide repeats poly(C-A) and poly (G-T). These repeats were named as simple sequence repeats (SSRs) by Tautz and Renz in 1984. Microsatellite and simple tandem repeats were later named these sequences [2]. Variation have been resolved disturbing a standard repeat motif pattern inside the repeat structure of numerous STR loci. By underwriting the nearness of those variations in the STRs the criminological STR loci might be expanded with the guide of separation power (DP). These methodologies have demonstrated exact and less expensive by task scale for the high caliber of DNA sample. Be that as it may, they can present an unavoidable work process bottleneck and many markers scale investigation can be anticipated [5].

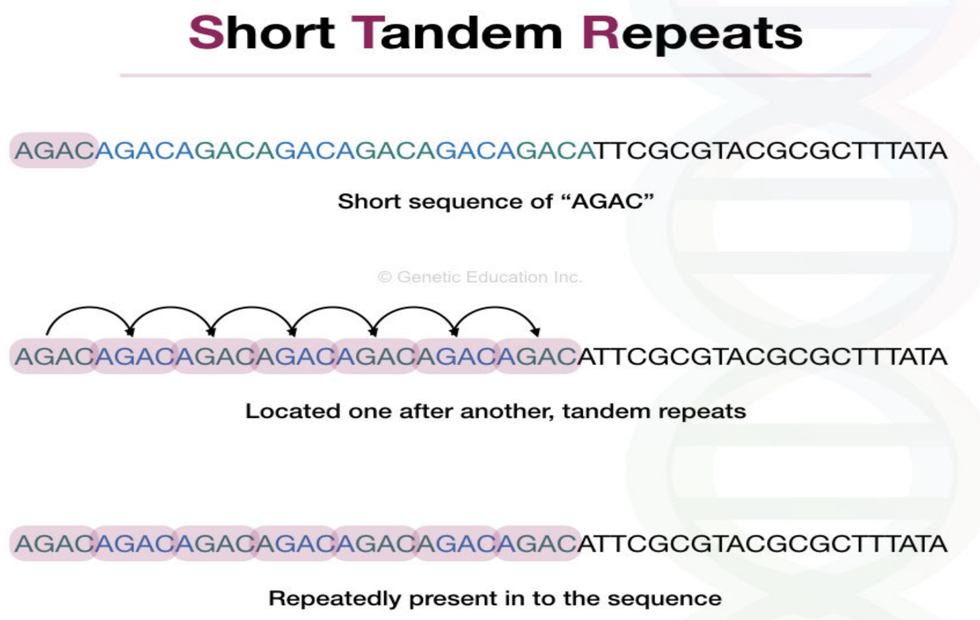

Figure 1: Short Tandem Repeats (STRs)
https://geneticeducation.co.in/wp-content/uploads/2019/07/Short-tandem-repeats-.007-e1562041458232.jpeg.

Distribution of STR in human genome:

In humans, eukaryotes and prokaryotes STRs are widely present. In human genome they are evenly distributed throughout the genome. The account for almost 3% of the genome. They don not occur frequently in sub telemetric region. They are less frequent in this region and inside the chromosome their distribution is not quite uniform [6]. Only 8 % STRs are found in the coding region. These repetitive sequences mostly found in the non-coding region. Its highest density is found in chromosome 19 in humans. In human genome STRs are found per 2,000bp. Most common STR that are found in human genomic DNA are A, AC, AAAN, AAN and AG. In order to name it take an example of D7S820 where the first letter i.e., D is for DNA the second 7 stands for the number of chromosomes on which it is located, S is for STR and 820 is the specific identifier [7].

Direct functioning of STRs in some of the host is what scientist are trying for last decade of 20th century. Most STRs are considered as junk because they have no biological use though they exist frequently in the genome. There is a huge similarity in genetic material of human beings. Difference exists only in small portions of the DNA sequence which is about 0.3%. Basically, this is the difference that can be used in identification of an individual. The portion of STR sequence is polymorphic and its amplification can be done quite easily using PCR method [8]. Changes in the nucleotide sequence in the same region in different individuals results in polymorphism that causes evolution. This has a direct effect on the allelic diversity and population distribution on a specific locus.

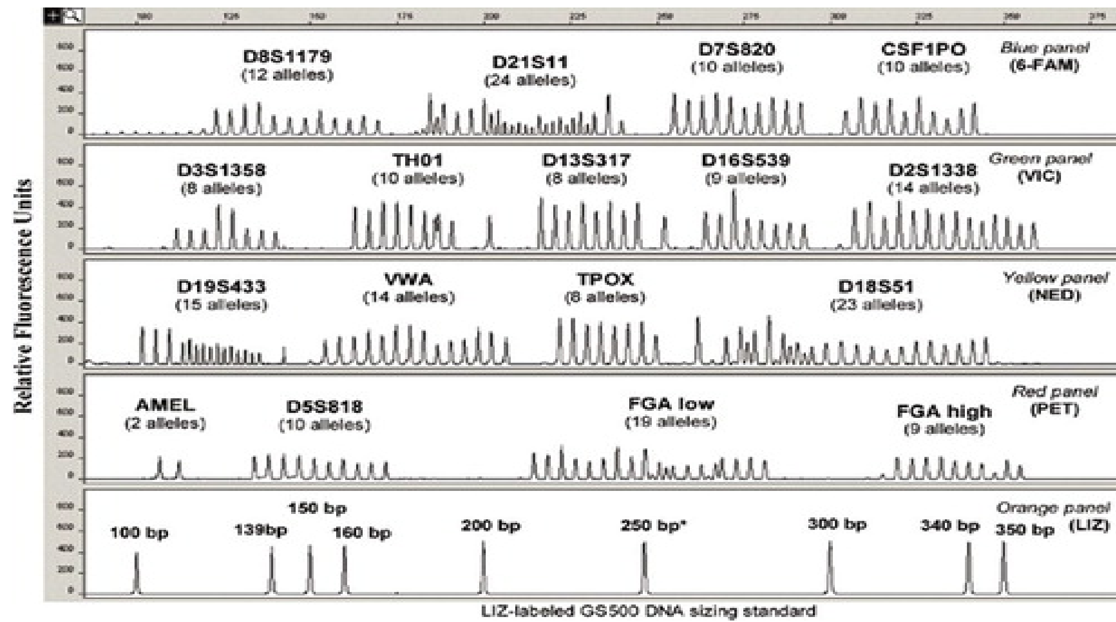

Figure 2: Short tandem repeat typing technologies used in human identity testing
https://www.future-science.com/cms/10.2144/000112582/asset/images/medium/table1.gif

Although the human genome contains thousands upon thousands of STR markers, only a small core set of loci have been selected for use in forensic DNA and human identity testing. Like using a single, common currency in a financial sense, core loci permit equivalent genetic information to be shared and compared. Commercial kits are now available to generate DNA profiles containing these core STR loci. Millions of STR profiles are generated worldwide each year by government, university, and private laboratories performing various forms of human identity testing, including DNA databasing, forensic casework, missing persons/mass disaster victim identification, or parentage testing.

Nucleotide change and forensic analysis:

Within the people, nucleotide changes occur in only the one region it can cause polymorphism that have straight reaction on allelic diversity and population giving out with in the locus, mutation can cause at a locus with the passage of time the microsatellite construction normally resolved in the amount of short tandem repeats. The very short size of the fragment which can be

obtained by the DNA prepare the chance of the test of these variations. The precise selection method can be discovered to identify the individual sensitive areas [9]. Genotype properties in STRs based on the amplification of the PCR that contain STR for genomic targets, with the help of gel electrophoresis, size followed by amplicon fragments by copying the number. In forensic sciences Microsatellite analysis is a technique used for identification. This method basically defines the allele position in repeat motif number alone on the variation which is interrupted by the aid of amplicon length, its ability to utilize the additional resolving power present in the variation of sequence in a locus. Fragments of small size like 200 base pairs makes it possible to examine the variations from small quantities of DNA and even from degraded samples. Individual identification can be done from sensitive areas and other precise identification methods can be discovered [9]. Hereditary data on center loci enables certainties to be shared and thought about. Short tandem repeats composing have interesting framework involves a gathering of the sample, DNA extraction, quantification of DNA, PCR amplification of multiplex STR loci allele separation and size dissemination, STR typing, and profiling. Casework research or circumstance together with rape and proof, a total of DNA may final product from a total of the person in question or do outline liquids and make a perplexing and extreme elucidation of the outcome. To choose the size of the STR amplicon is estimated after PCR amplification which empowers to choose the number of repeats found in each allele decided in DNA profile. These might be achieved with the guide of the estimated based absolutely partition that comprises of gel or fine electrophoresis. During PCR amplification STR are fluorescently marked, with the guide of pondering locus-specific forward and reverse primer which has a fluorescent color, recording the dye of color and time of movement of each DNA sections identified with the internal standard, each STR allele size can be determined by means of STR alleles separation [10]. The methods used have been proved to be accurate and cheap by undertaking high quality DNA sample. In case of criminal investigation large databases are used in this case 13 STR are

used in a standard way to maintain a DNA profile that is used by US EBI agency. To keep going the DNA profile, the 13 STR core is used for the criminal inquiry. Specific assay or thoughtful for high level polymorphic markers, the small fragment STR loci is exaggerate vigorous robust by applying PCR. VWA, TPOX, TH01, FGA, D21S11, D18S51, D13S317, D16S539, D7S820, D8S1179, D3S1358, D5S818, CSF1PO, these 15 STR loci or gender resolve locus can be amplified or reorganize by the PCR [11]. To describe the database of the FBI program that hold up the criminal justice DNA database which is incorporate DNA index system (CODIS). In this study, CSF1PO, TPOX, D7S820 these three STR loci of allele frequencies can be examine in 4 unassociated individuals as well as D7S820, CSF1PO, or TPOX were achieve by the multiplex polymerase chain reaction.

Overview of STR and microsatellites:

© GSJ

Short tandem repeats have one or another intergenic or intragenic area, as well as genes, which contain 3% of human genome. Most of the microsatellites are more changeable to show sequence or length polymorphism. But the others are maintained or provide many information which is available in the marker for the population genetics, mapping or interconnection studies [12,13]. Basically, the repeated DNA sequences hold full of human genome by bear on polymerase chain reaction STRs can easily be identified which can be used as a DNA marker it is possible that the STRs and the PCR product are similar to find the result easily. A particular position of trinucleotide replicate is merge of nucleotide are organized duplicative such as (CAG, CGG, that can be replicate). The distinction in size between two diverse STR alleles may be educational the bigger the distinction, the more the quantity of transformation occasions. Goldstein in 1995 stated that STR loci will eventually allow for a high-resolution description of the human evolutionary phylogeny [14]. STR are the repeat sequences that not only necessarily vary in the length of the repeat unit and just the number of repeats but also in the rigor and

severity with which they conform to the repeat pattern. STRs can be further divided into several categories and these categories can be based on the repeat pattern. Not all of these alleles of an STR locus contain complete repeat units. Even simple repeats can also contain non consensus alleles that are in between alleles with full repeat units. Goldstein in 1995 anticipated that STR loci would at last permit a high-resolution portrayal of the human transformative history.

Microsatellite have non-coding or coding region that are present in prokaryotes as well as euchromatin of the eukaryotes. SSR divided into two categories in cereal the first one is repetitive sequence or the second one is unique sequences. In the oats the repetitive sequence is less polymorphic. SSR basically smaller than 12bp nucleotide which have mutation potential SSR is microsatellite. When the mutation rate increase in SSR it led to change in gene expression [15]. The types of microsatellites are Di, tri and tetra nucleotides, Compound. Perfect. Imperfect.

Methodology:

Our total amount of samples that were four blood samples were collected from university of central Punjab Lahore. We obtained 4ml blood sample in EDTA in sterilized collection vials. The specifics of individuals were noted in a prescribed consent from signed by the volunteer who were participating. One aliquot (300mL) of whole blood sample in micro centrifuge tube was preserved at -70°C. For each sample as backup source and remaining blood was preserved at -20°C. Sterile tubes and micro centrifuge tubes were used for preservation. The extraction of DNA from blood sample is done to extract genomic DNA.

Procedure:

Addition 200µl EDTA blood samples were thawed at 37ºC for 10 minutes and vortex it. Wash with washing buffer (20Mm TRIS HCL with 7.5) for all the removal of material of blood except white blood cells. This removes all red blood cells and serum protein following 20 minutes mix by centrifuge at 3,000 rpm for 15 minutes. Then discard the supernatant and

save the pallet. Repeat the process for 2 to 3 times until the pinkish color of blood is removed. Add 100µl lysis buffer, 15µl proteinase K and 10% SDS. Incubate at 37°C overnight in shaking incubator. Equal volume of PCI is added in Eppendorf tubes followed by centrifugation at 3000 rpm for 15 minutes and save the pallet. Equal volume of Isopropanol and centrifuge for 15 min at 3000rpm.Discard supernatant add 70% chilled ethanol and centrifuge at 5000rpm for 5 minutes. Dry the pellet and add 5micro liter RNASE incubate it and water bath at 60°C for 1:30 hours. If the pallet is visible than add 50µl TE pH 8. For complete suspension place on water bath at 70°C for 10-15 minutes.

- **Electrophoresis by agarose Gel:**

For electrophoresis the gel is made as 0.8% agarose solution. It was prepared in order to confirm the extracted DNA. Mix the extracted DNA with a dye i.e., loading dye and loaded them into the wells of the agarose gel. Now view it under the documentation system to visualize the gel.

© GSJ

- **PCR Master Mix Composition:**

Master Mix consists of Nuclease Free Water and PCR Master Mix two2X. PCR Master Mix *ie* a combination of two solution containing Taq DNA polymerase, dNTPs, MgCl2 and reaction buffers at most effective concentrations for environment friendly amplification of DNA templates through PCR.

- **Buffers and Solutions Composition:**

1. PCR Master Mix amount 50ml
2. Taq DNA polymerase of pH 8.5 400uM
3. dATP, dGTP. dCTP, dTTP 3mM

- **Polymerase Chain Reaction:**

PCR amplifies the DNA at an exponential rate. It consists of two profiles reaction profile and temperature profile. For the amplification of the CSF1PO was done by using the primers. Amplification was carried out using the master mix, primers and DNA template PCR is performed using a thermo cycler. The DNA concentration in a working solution of about 10 ng/ µL in ddH_20 was confirmed by spectrophotometric evaluation at 260 nm. For optimization, the concentration of the genomic DNA,1X buffer plus MgCl2, STR primers and Taq DNA polymerase were optimized for the STR. The primers were synthesized from genetic research. Taq polymerase, collectively, with 5X PCR buffer, MgCl2 and dNTP were synthesized locally. The amplification was performed once in the thermal cycler PTC-100 for 30 cycles.

Polymerase Chain Reaction Components	
Step 1	PCR tubes were label with permanent marker.
Step 2	PCR tube added PCR water (double distilled water) 3.5 ul
Step 3	Added chilled 12.5ul master mix consist of dnTPs, Taq polymerase, buffer
Step 4	Added 2ul forward primer and 2ul reverse primer
Step 5	We added 5ul of extracted DNA sample
Step 6	Placed PCR tubes into the PCR machine and adjusted the temperature

Table 1: PCR reaction component

Condition of PCR for amplification:

Amplification of the genomic DNA were later confirmed in agarose gel 2.5% stained with ethidium bromide. Due to the small size of the base pair, the alleles did no resolve in the gel agarose. Therefore, charge concentrations for polyacrylamide gel electrophoresis (PAGE) were performed as a function of the brightness of the band. For accurate allele determination, 5-8% PAGE was used. The gel was used using a 0.5X TBE working buffer in an apparatus followed by the silver staining method.

Constituents	Amount	Temperature	Time
Primer forward	3µl	95 °C	3 minutes
Primer backward	3µl	93 °C	45 seconds
Double distilled water	2.5µl	58 °C	45 seconds
Master mix	12.5µl	72 °C	45 seconds
DNA sample	2µl	72 °C	10 minutes
$MgCl_2$	2µl	4 °C	10 minutes
Total	25µl		

Table 2: Composition profile by changing its annealing temperature

Result:

Sample was collected from a student from university. This study was performed with three other research students' samples. In order to avoid any confusion our sample was named Z1. DNA extraction was carried out of the sample. Extraction of DNA was confirmed using gel electrophoresis method. This was performed to get the conformation. After the extraction of DNA, the samples were amplified in thermo cycler PCR machine by using PCR machine by using STR locus; CSF1PO as shown below.

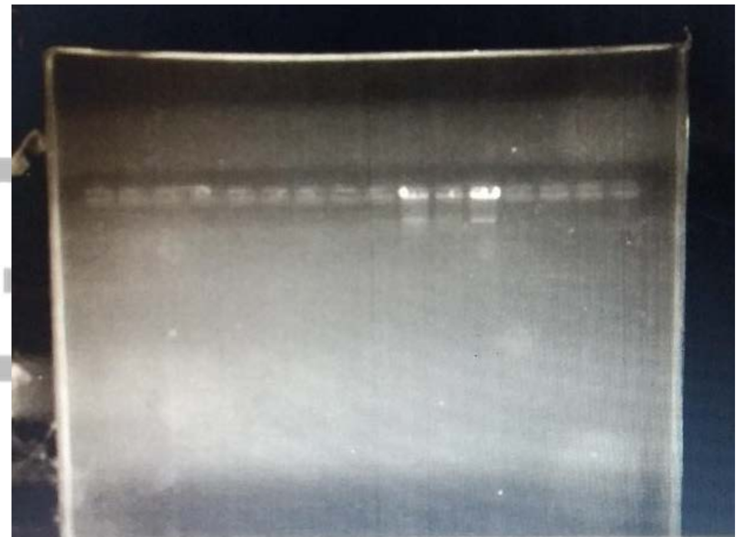

Figure 3: Collection of blood sample **Figure 4: Result of extracted DNA sample**

DNA markers and STR analysis:

For genotyping approaches markers are used. To increase their genetic essential quality in the field of botany, microsatellite is used. Microsatellite have non-coding or coding region, which are present in prokaryotes as well as euchromatin of the eukaryotes. SSR divided into two categories in cereal the first one is repetitive sequence or the second one is unique sequences. In the oats the repetitive sequence is less polymorphic. SSR basically smaller than 12bp nucleotide

which have mutation potential SSR is microsatellite. When the mutation rate increase in SSR it led to change in gene expression. It has been just about a long time since the 13 hereditary markers that shaped the center of the FBI Research center's Consolidated Ordering Framework (CODIS) were chosen in November 1997. Utilizing DNA to recognize two people is a precarious issue, since near 99.9 percent of our DNA is equivalent to every other person's DNA. DNA that codes for proteins can't change much without rendering the proteins ineffectual. The four nucleotide bases that make up the foundation of DNA give directions to collecting the amino acids in proteins by being in an exact arrangement, with every three-base gathering coding for a particular amino corrosive. In the event that that DNA base succession is modified, the grouping of amino acids in the subsequent protein can likewise be adjusted. Thus, since protein work gets from a particular amino corrosive grouping, the protein may not work. Among the 3 million or so DNA bases that don't code for proteins are areas with numerous duplicates of short rehashing groupings of these bases, which make up the DNA spine (for instance, TATT). These groupings rehash a variable number of times in various people. Such loci are classified "variable number of short tandem repeats," and they are the premise of STR analysis. An assortment of these can give almost certain proof measurably of an individual's character in light of the fact that the probability of two unrelated individuals having a similar number of rehashed arrangements in these districts turns out to be progressively little as more areas are examined. CSF1PO is a STR repeat is tetra nucleotide repeat c-fms proto-oncogene for the CSF-1 receptor on the long arm of chromosome number 5. Usually found alleles contain an AGAT repeat of 5 to 17 repeats [16]. CSF1PO is one of the thirteen core loci used in the CODIS database, and alleles found for this short tandem repeat locus contain 6 to 15 repeats of the tetra nucleotide AGAT.

CSFIPO on a locus	
Location on chromosome	5q33.3-34

© GSJ

Locus according to GenBank	HUMCSF1PO is Human c-fms proto-oncogene for CSF-1 receptor gene is the locus
Repeat sequence 5'-3'	AGAT (*) is the repeat unit

Table 3: Information about position of CSFIPO on a locus

Conclusion:

At any locus of STR, it is expected to find the presence of two alleles, one inherited maternally and other paternally. These alleles are seen as a two-banded pattern, as seen in a heterozygote, or a single banded pattern as seen in a homozygote. Rarely, a three-banded pattern can be seen at a single locus in a multiplex STR profile and that is not a result of a mixture. This can occur by an additional chromosomal occurrence or a duplication of the locus. Three-banded patterns are already reported at the TPOX, CSF1PO, FGA, D5S818, D21S11, and D18S51 loci. The presence tri-allelic patterns are mostly because of genetic replication in tandem of a short part of chromosomal DNA; or incorrect separation during meiosis or mitosis because of a chromosomal nondisjunction which leads to a trisomy. Autosomal short tandem multiplexed polymerase chain re action (PCR) system for 17 autosomal loci D1S1656, D3S3045, D3S477, D8S1435, D10S1418, D11S2368, D11S2368, D11S2368, D11S2368, D11S2368, D11S268, D11S2368, D11S2368, D19S253 and D22-GATA198B05

References:

1. Jeffery, M.A., and P. Gill. 1985. Encoded evidence: DNA in forensic analysis. *Nature review genetics.* 5:739.

2. Edawards, A., H.A. Hammond, L. Jin, C.T. Caskey, and R. Chakraborty. 1991. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics.* 12:241-253.
3. Callery, P.S., B. Boswell, P.M. Gannett, R.L. Haining, M. Sanga, P. Tirumalai, and T.S. Tracy. 2006. D HPLC method for forensic. *US Department of Justice and prepared the following final report 17.*
4. Kobilinsky, L.F., T.F. Liotti, and J. Oeser-sweat. 2005. DNA: Forensic and legal applications. Wiley Online Library.
5. Buchan, J.C., E.A. Archie, R.C. Van Horn, C.J. Moss, and S.C. Alberts. 2005. Locus effects and sources of error in noninvasive genotype. *Molecular ecology Notes*. 5:680-683.
6. Koreth, J., J.J. O`LEARY, and J. O`D. McGEE. 1996. Microsatellite and PCR genomic analysis. *The journal of pathology.* 178:239-248.
7. Nadir, E., H. Margalit, T. Gallily, and S.A. ben-sasson.1996. microsatellite spreading in the human genome: evolutionary mechanisms and structural implications. Proceeding of the National Academy of sciences. 93:6470-6475.
8. Veeramah, K.R., and M.F. Hammer. 2014. The impact of whole genome sequencing on the reconstruction of human population history. *Nature Reviews Genetics.* 15:149.
9. Hoban, S.M., O.E. Gaggiotti, and G. Bertorelle. 2013. The number of markers and samples are needed for detecting bottleneck under realistic scenarios, with and without recovery: a simulation-based study, *Molecular ecology.* 22:3444-3450.
10. Wiegand, P., and M. kleiber. 2001. Less in more length reduction of STR amplicons using redesigned primers. *International journal of legal medicine.* 114:285-287.
11. Goodwin, W., A. Linacre, and S. Hadi. 2011. An introduction to forensic genetics. John Wiley & Sons. Congram, D., and D.W. Steadman. 2008. Distinguished guests or agents of ingerence: foreign participation in Spanish Civil War grave excavation. *Complutum.* 19:161-173.
12. Romualdi, C., D. Balding, I.S. Nasidze, G. Risch, M. Robichaux, S.T. Sherry, M. Stoneking, M.A. Batzer, and G. Barbujani. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Research.* 12:602-612.
13. *Hammond, H.A., Jin Y. Zhong, C.T Caskey, and R. Chakraborty. 1994. Evolution of 13 short tandem repeat loci use* in personal identification applications.

14. Goldstein, D.B., A.R. Linares, L.L. Cavalli-Sforza, and M.W. Feldman. 1995. An evolution of genetic distances for use with microsatellite loci. Genetics. 139:463-471.
15. Vieira, Maria Lucia Carneiro, Santini, Luciane, Dinzi, Augusto Lima, & Munhoz, Carla de Frietas. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetic and Molecular Biology,* 39(3), 312-328.
16. Butler, J.M., A.E. Decker, M.C. Kline, and P.M. Vallone. 2005. Chromosomal duplication along the Y- chromosome and their potential impact on Y-STR interpretation. *Journal of Forensic Science.* 50: JFS2004481-2004487