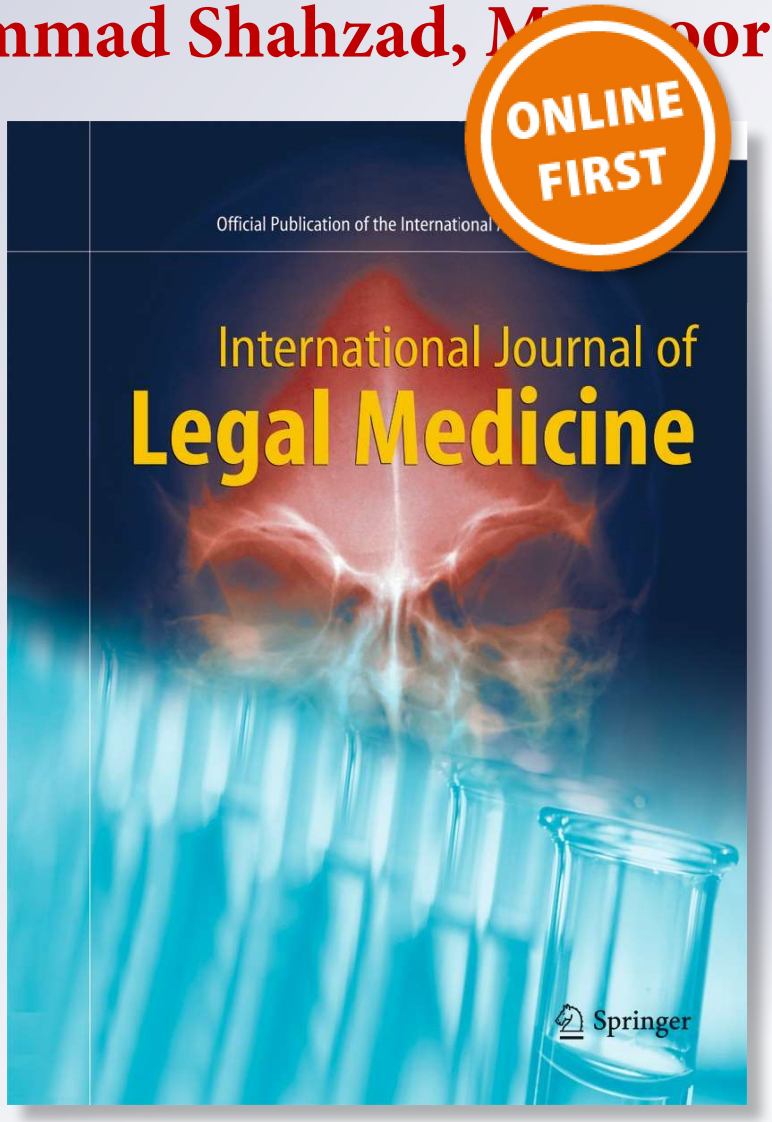
Development of new PCR multiplex system by the simultaneous detection of 10 miniSTRs, SE33, Penta E, Penta D, and four Y-STRs

Muhammad Shafique, Muhammad Saqib Shahzad, Ziaur Rahman, Muhammad Adnan Shan, Rukhsana Perveen, Muhammad Shahzad, M[illegible]oor

International Journal of Legal Medicine

ISSN 0937-9827

Int J Legal Med
DOI 10.1007/s00414-016-1372-x



Springer

Your article is protected by copyright and all rights are held exclusively by Springer-Verlag Berlin Heidelberg. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Author's personal copy

Int J Legal Med
DOI 10.1007/s00414-016-1372-x

ORIGINAL ARTICLE

Development of new PCR multiplex system by the simultaneous detection of 10 miniSTRs, SE33, Penta E, Penta D, and four Y-STRs

Muhammad Shafique[1] · **Muhammad Saqib Shahzad**[2] · **Ziaur Rahman**[1] · **Muhammad Adnan Shan**[3] · **Rukhsana Perveen**[3] · **Muhammad Shahzad**[3] · **Manzoor Hussain**[3] · **Ahmad Ali Shahid**[1] · **Tayyab Husnain**[1]

Received: 17 November 2015 / Accepted: 29 April 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract The 18 loci multiplex system has been instigated for co-amplification and fluorescent detection of Amelogenin and 17 STRs, including 10 MiniSTRs (CSF1PO, D18S51, D7S820, D2S1338, TPOX, D13S317, FGA, D5S818, D21S11, D16S539), SE33, Penta E, Penta D, and four Y-STRs (DYS385a/b, DYS438, DYS392). This multiplex system was developed for the simultaneous analysis of compromised DNA samples, Y-amelogenin marker mutation, motherless paternity issues where single allele sharing occurs at autosomal STRs in unrelated individuals, and other complex forensic cases. Selection of loci, primers, and allelic ladders were designed and created in-house with a design strategy to work in this multiplex. The multiplex system was evaluated by sensitivity, specificity, stability, precision and accuracy, case-type samples, mixture studies, PCR-based and population distribution studies to establish the robustness and reliability of the system as the current requirements of the forensic case work. Among all the markers evaluated for this study, 209 alleles including 44 variants were observed with combined power of discrimination, combined power of exclusion, and the combined probability of matching calculated as $0.99999999999999999989391633934 4$, 0.999993816173890, and 5.90019×10^{-19}, respectively. Due to highly polymorphic characteristics of these loci particularly SE33 and Penta E which are most discriminatory (PD = 0.991 and 0.983, respectively) in the Pakistani population, this multiplex would be highly valuable for individual identification in complex forensic cases and paternity issues as well as population database.

Keywords MiniSTRs · SE33 · Penta D · Penta E · Y-STRs · Forensic · PCR multiplex

Introduction

The STRs selected from human genome for forensic investigations found in the intronic regions are being used in the form of tri-, tetra- and penta-nucleotide repeats (3 to 5 repeated nucleotides), as these give a high degree of error free data while being robust enough to survive under degradation in non-ideal conditions. Moreover, STRs on the basis of high polymorphism are most rapid, precise, and well efficient in generating DNA typing results from very minute amount of samples, performing multiplex PCR amplification simultaneously. A lot of improvements have been executed in the last 20 years in the provisions of sample processing speed like extraction, amplification, genotyping, and sensitivity. Therefore, recently, a precious DNA profile can be obtained from very small quantity of biological source, as minuscule as from a cell in particular cases rather than using larger stains [1]. In most of the Europe and United Kingdom (UK), core ten loci were being used with the inclusion of two extra markers D2S1338 and D19S433 along with 8 overlapping loci to the

Electronic supplementary material The online version of this article (doi:10.1007/s00414-016-1372-x) contains supplementary material, which is available to authorized users.

✉ Muhammad Shafique
shafique@cemb.edu.pk

[1] Forensic DNA Typing Laboratory, National Centre of Excellence in Molecular Biology, University of the Punjab, Lahore 53700, Pakistan

[2] Faculty of Life Sciences, University of Central Punjab, Lahore, Pakistan

[3] Forensic Services Laboratory, Centre for Applied Molecular Biology, University of the Punjab, Lahore 53700, Pakistan

Published online: 11 May 2016

Author's personal copy

CODIS. In 2009, the European Network of Forensic Science Institutes (ENFSI) selected to implement five extra NCO (non CODIS) STR loci to the previously in use seven STRs of European Standard Set (ESS) to make it compatible with CODIS [2]. Likewise, the National DNA database of Germany established in 1998 comprises the reference DNA data set of 835,275 by the addition of core locus SE33 due to its very high polymorphism [3] (http://www.bka.de/nn_195364/EN/SubjectsAZ/DnaAnalysis/Statistics/dnaAnalysisStatistics__node.html?__nnn=true) [4]. DNA databases in some other countries like UK with 5.7 million DNA profiles [5], USA more than 12 million profiles (https://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-statistics), Australia >837,000 profiles (https://www.crimtrac.gov.au/dna), Israel 135,000 profiles [6], Kuwait 4.2 million profiles [7], and Interpol with 158,000 profiles were developed for the screening of potential criminals (http://www.interpol.int/INTERPOL-expertise/Forensics/DNA).

On the basis of STR characteristics, since the last one and half decade, different brands of robust commercial STR kits and non-commercial multiplexes [8–12] were developed by some modifications like increase in number of markers, addition of population specific core locus like D6S1043 for China and SE33 for German [12], inclusion of DYS391 to interpret the presence of null allele at Y-amelogenin marker in male individuals (Global Filer and PowerPlex Fusion), amplification success in the presence of inhibitors like in Identifiler Plus, reducing the STR fragment size as miniFiler, and PCR conditions to reduce time duration were conducted to overcome the experienced limitations. An additional approach to restore required information from such DNA evidence samples which become degraded due to the environmental stresses, is the reduction of fragment size of the PCR amplification products by designing primers from the flanking region as more closure towards the STR site [13, 14]. The resultant smaller fragment sized PCR amplification products obtained from degraded DNA by using redesigned primers nearer to the STR region revealed a great success which was formerly described by Whitaker et al. [15]. In this perspective, degradation studies have been conducted to evaluate the efficiency of the miniSTR primers for the reduction of product amplicons and mini-plexes were developed in concordance with other commercially available STR kits [16–19].

In this study, keeping in view of above, a new PCR multiplex was developed by co-amplification of ten miniSTRs (CSF1PO, D7S820, TPOX, D18S51, D2S1338, D13S317, FGA, D5S818, D21S11, D16S539), highly polymorphic STR markers as SE33, Penta E, Penta D, and four Y-STR loci (DYS385 a/b, DYS438 and DYS392) along with gender-identifying marker amelogenin. The position of each locus was defined appropriately according to the multiplex design configurations as their fragment size range, fluorescent dye panel distribution, and allele size overlying. Considerably, this multiplex system was improved to deal with the limitations simultaneously like degraded DNA, Y-amelogenin null allele, random allele sharing among unrelated male individuals, and fast PCR cycling procedure. This system was enabled to get successful amplification results in the presence of PCR inhibitors, highly degraded DNA, complex mixtures, and the PCR time duration of up to about 1 h.

Developmental validations were conducted as per Scientific Working Group on DNA Analysis Methods (SWGDAM) validation guidelines-2012. The validation studies like stability, sensitivity, species specificity, accuracy and precision, DNA mixtures, and studies based on PCR amplification (primer optimization, annealing temperature, PCR conditions, and PCR components) were performed using control cell line DNA 007 (Applied Biosystems, Foster City, CA) and 9948 (Promega Corporation, WI, USA). Forensic and paternity statistical parameters were evaluated by genotyping of 300 blood samples of unrelated individuals collected from the province of Punjab Pakistan with their consent. Statistical analysis revealed that the population at autosomal markers remained in Hardy-Weinberg equilibrium after Bonferroni correction except two loci (CSF1PO and TPOX) where significant deviations were observed which could be due to the trends of consanguinity in the region. Due to high discriminatory power and robustness of this system, these markers seem to be the core loci of Pakistan and provide significant information to solve complex cases in paternity, forensic, and where the DNA is highly degraded.

Materials and methods

Selection of STR markers

The STR loci were selected for the development of multiplex including ten miniSTRs, SE33, Penta E, Penta D, amelogenin for gender determination, and four Y-STRs as given in Table 1. Among these 18 markers, nine miniSTR loci (CSF1PO, D7S820, TPOX, D18S51, D13S317, FGA, D5S818, D21S11, and D16S539) were selected from CODIS core STRs [20, 21] and D2S1338 from European database (ENFSI DNA Working Group, 2014) by reducing their product sizes. The locus SE33 has been proposed to include in the CODIS core loci [22]; therefore, this locus was also integrated for the development of multiplex and to determine its polymorphism in Pakistani population.

Penta-nucleotide repeat loci Penta E and Penta D were also included in the multiplex as their highly polymorphic nature and low or no stutter artifact display [23]. Moreover, in view of Y-amelogenin marker mutation [24, 25] and the random sharing of autosomal alleles in motherless parentage analysis or between two claimers being the alleged fathers for a son [26], Y chromosome STRs DYS385a/b, DYS438, and

Author's personal copy

Table 1 Characterization of 18 STR loci with GenBank Accession, chromosome location, physical position, repeat motif, reduced size of MiniSTRs, and their final product size

Locus	GenBank Accession	Chromosome location	Physical position (Mb)	Repeat motif	Product size (STR kits)	Size reduced	Product size (bp)
Amelogenin	M55418 M55419	X: p22.1–22.3 Y: p11.2	Chr X: 11.316 Chr Y: 6.736	–	–	–	75–78
CSF1PO	X14720	5q33.1	Chr 5: 149.455	[AGAT]	304–341 (Identifiler)	220	84–120
D7S820	AC004848	7q21.11	Chr 7: 83.789	[GATA]	255–293 (ProPlus)	120	134–173
SE 33	V00481	6q14	Chr 6: 88.987	[AAAG]	–	–	197–326
Penta E	AC027004	15q26.2	Chr 15: 97.374	[AAAGA]	–	–	372–469
TPOX	M68651	2p25.3	Chr 2: 1.493	[AATG]	222–250 (Cofiler)	157	61–100
D18S51	AP001534	18q21.33	Chr 18: 60.949	[AGAA]	262–345 (ProPlus)	155	106–186
D2S1338	AC010136	2q35	Chr 2: 218.879	[TGCC]/[TTCC]	307–360 (Identifiler)	88	212–272
DYS385 a/b	AC022486	q11.222	Chr Y: 20.801 Chr Y: 20.842	[GAAA]	–	–	347–404
D13S317	AL353628	13q31.1	Chr 13: 82.722	[TATC]	216–244 (Identifiler)	121	91–131
FGA	M64982	4q28	Chr 4: 155.509	[TTTC]/[CTTT]/[TTCC]	214–355 (ProPlus)	73	139–280
DYS438	AC002531	q11.21	Chr Y: 14.938	[TTTTC]	–	–	304–336
D5S818	AC008512	5q23.2	Chr 5: 123.111	[AGAT]	134–172 (ProPlus)	28	98–143
D21S11	AP000433	21q21.1	Chr 21: 20.554	[TCTA]/[TCTG]	185–240 (ProPlus)	30	153–210
D16S539	AC024591	16q24.1	Chr 16: 86.386	[GATA]	264–304 (PP16)	29	234–276
DYS392	AC06152	q11.223	Chr Y: 22.634	[TAT]	–	–	291–328
Penta D	AP001752	21q22.3	Chr 21: 45.056	[AAAGA]	–	–	388–442

DYS392 were added in the multiplex as shown in Fig. 1. GenBank Accession, chromosome location, and repeat motif were accessed from NIST STR DNA database (http://www.cstl.nist.gov/strbase/str_fact.htm). Physical position for each locus was mapped when each primer set was blast In-Silico PCR [27] using online UCSC genome browser [28] as shown in Table 1.

Primer sequences, synthesis, and optimization

Primer sequences for miniSTR loci were selected in the size range of 61 to 276 bp. The newly designed primer sets [29] were selected for the loci D2S1338 and D16S539 using web based software Primer3version 4.0 [30] aligned with each reference sequence. While the Primer sequences for the loci CSF1PO, D7S820, TPOX, D18S51, D13S317, FGA, D5S818, D21S11, SE33, Penta E, Penta D, DYS385a/b, DYS438, DYS392, and amelogenin were preferred as previously reported [16, 31–35] as shown in supplementary table 2. The forward primer of each locus was labeled with a fluorescent dye which enabled the use of available matrix standards and reliable color separations on the ABI 3130xl Genetic Analyzer. All primer sets were synthesized from Life Technologies Inc.

Each primer set was optimized individually with different concentrations to get the amplification at same annealing temperature using control DNA 007 and 9948. Initially, a primer mix was prepared with equimolar concentration to ensure the performance of each locus. Then by changing the various primer proportions as increasing for the weaker loci and decreasing for stronger loci, a final primer mix was primed by adjusting different concentrations ranging from 0.09 to 2.25 μM in a final reaction [36] as shown in supplementary material table 2.

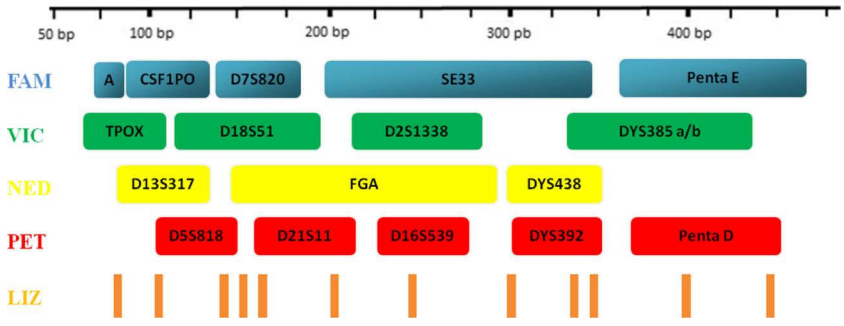


Fig. 1 Multiplex configuration schematic

Author's personal copy

DNA samples prepared for validation studies

The cell line positive control DNA 007 and 9948 (Applied Biosystems by Life Technologies) were used in the development of multiplex assay, sensitivity, stability, and PCR conditions in addition to applying as a positive control in capillary electrophoresis.

Species specificity studies were conducted on DNA samples of non-primates (dog, horse, sheep, and chick) and microbial DNA like *Escherichia coli*.

A genomic DNA of 2.7 μg was degraded using 2.5U DNase-1 (ambion® by Life Technologies) at 37 °C in a total volume of 110 μL [19] for predefined time intervals (0 and 30 s and 1, 2, 3, 4, 5, 10, 20, 40, and 80 min). Real-Time PCR was performed using Quantifiler™ Human DNA Quantification Kit (Applied Biosystems Foster City, CA, USA) to estimate the absolute quantification of all these artificially degraded DNA samples.

To study the inhibition levels of hematin during the PCR amplification, 1 ng of Control DNA 9948 was used with increasing concentrations of hematin (CHEM-IMPEX INT'L INC, USA) as 0, 50, 65, 85, 100, 125, 150, 250, 500, and 1000 μM. Similarly known quantities of humic acid (Sigma–Aldrich, MO, USA) with the escalating concentrations of 0, 25, 75, 85, 100, 125, 175, and 225 ng/μL were used in the PCR.

To perform the mixture studies, two different genotypes of DNA samples were mixed in the following different ratios as 1:15, 1:10, 1:7, 1:5, 1:3, 1:1, 3:1, 5:1, 7:1, 10:1, and 15:1. Moreover, DNA of three different samples were also mixed in the ratios of 10:1:1, 3:1:1, 1:1:1, 1:3:3, 1:10:10, 1:10:1, 1:3:1, 1:1:3, and 1:1:10 making the final concentration of 1 ng/μL.

For genetic identification of an individual, statistical parameters were evaluated to explicate DNA results and scientific certainty threshold. For this purpose blood samples of 300 unrelated individuals from the Punjab province of Pakistan were collected by informed consent.

DNA isolation and quantification

The isolation of genomic DNA from blood was carried out in the non-amplified DNA laboratory area using phenol chloroform-extraction procedures [37, 38] and purified through Microcon® centrifugal filter devices YM-100 (Millipore Corporation, USA). The extracted amount of DNA was calculated on 7500 SDS Real-Time PCR System (Applied Biosystems Foster City, CA, USA) using Quantifiler™ Human DNA Quantification Kit according to the specifications.

PCR amplification and genotyping

The DNA samples were amplified on Gold plated 96-Well GeneAmp PCR System 9700 (Applied Biosystems Foster City, CA, USA). The PCR reaction was prepared in a total volume of 12.0 μL with optimal concentrations as 1X Platinum® Multiplex PCR Master Mix, 0.09–2.25 μM Primer Mix, 0.04–0.08 ng/μL Template DNA, and T.E.$^{-4}$ buffer. The optimum PCR conditions at run mode of 9600 emulation were followed by initial denaturation for 2 min at 95 °C, 30 cycles at 95 °C for 10 s, 60 °C for 45 s, and final extension at 60 °C for 2 min completing the whole process in about 1 h.

Amplified product of each DNA sample was run on Genetic Analyzer ABI3130xl (Applied Biosystems) for capillary electrophoresis using G5 dye set to get the genotype. The samples were prepared by mixing 0.4 μL of GeneScan™ 500 LIZ® as a Size Standard and 13 μL of Hi-Di™ Formamide (Applied Biosystems) while 0.7 μL from each PCR product was added to the relevant well mixture of MicroAmp® Optical 96-Well Reaction Plate (Applied Biosystems). The prepared samples in the reaction plate after short spin were denatured for 5 min at the temperature of 95 °C and then snap cooled on ice for 5 min prior to run on the instrument. On completion of the electrophoretic run, the project was saved as .fsa file through the Data Collection Software V3.0. These .fsa files were imported into the GeneMapper® ID Software v3.2.1 for data analysis and interpretation of results.

Generating the allelic ladders

Allelic ladder for the loci CSF1PO, TPOX, D18S51, D2S1338, FGA, D5S818, D21S11, D16S539, DYS392, and SE33 was generated by preparing dilution (1:500) of allelic ladder from the Identifiler™ (Applied Biosystems), Y-filer™ (Applied Biosystems), and PowerPlex ES (Promega) kits [16]. From these dilutions, 0.5 μL was used to amplify for each locus and then allelic product of all loci was combined into a final allelic ladder. Allelic ladder for the locus D7S820, Penta E, DYS385a/b, D13S317, DYS438, Penta D, and amelogenin was created by selecting different alleles from the studied population. All polymorphic forms of alleles for the specific locus were amplified separately with a respective primer set. The product of each sample for different alleles were combined together to form a single allelic ladder of the locus. Then the single allelic ladders were mixed and balanced together making a final allelic ladder [39].

Population genetic distribution

Forensic and parentage statistical parameters like matching probability (MP), power of discrimination (PD), polymorphism information content (PIC), observed heterozygosity (HO), typical paternity index (TPI), and power of exclusion (PE) were evaluated using modified PowerStats v1.2 software program [40]. Minimum and null allele frequencies for valid

Author's personal copy

statistical calculations due to the presence of rare alleles were calculated by 5/2*N* where *N* is the number of samples [1]. The linkage disequilibrium, gene diversity, haplotype frequencies and diversities for YSTRs, and *p* values of Hardy-Weinberg equilibrium (HWE) test at each locus were executed using the PowerMarker v3.25 [41].

Quality control measures

The appropriate standards and controls were integrated during whole process of analysis as in extraction, amplification, and capillary electrophoresis [42]. Cell line positive control DNA 007 and 9948 (Applied Biosystems by Life Technologies) were used to assess the PCR amplification and genotyping efficiency of the assay through capillary electrophoresis.

Negative control with each set of amplification containing no DNA template was used to evaluate the PCR reagent contamination. Only 4 μL T.E.$^{-4}$ buffer in lieu of DNA template was taken in addition to all other PCR reagents. This control was also processed through post-amplification steps.

Reagent blank control containing all extraction reagents except any template DNA was also processed through each step of extraction, quantitation, amplification, and capillary electrophoresis.

Results and discussion

Multiplex assay configuration

The development of multiplex assay was followed by the optimization of major limitations like assay layout as per size overlapping, singleplex PCR reaction for each primer set, compatibility of primer pairs to the core set, monitoring the non-specific product amplification, impartial DNA profile by variable concentrations of each primer set, incomplete adenylation, the occurrence of PCR artifacts, annealing temperatures, cycle numbers, PCR time duration, DNA concentrations, and PCR reagents [11, 16]. The cell line DNA controls 007 and 9948 were used as a standard with each amplification batch of the samples. The electropherogram of genotyping profile generated for control DNA 9948 and from the population sample has been shown in Figs. 2 and 3.

Developmental validations

The new developed system is how enough proficient in generating reliable results; however, the procedure must be evaluated by conducting the validation studies. The critical features and limitations are revealed by validation procedure which is essential to interpret the perfect results for forensic analysis [43–45].

Developmental validations were conducted in replica of three reactions to achieve optimal performance of PCR primers, thermal cycler parameters like annealing temperature, PCR cycle numbers, and to assess the sensitivity, species specificity, stability like DNA degradation, and PCR inhibition, mixture studies, accuracy, and precision.

Thermal cycler parameters

PCR conditions were synchronized with respect to the denaturation, annealing/extension temperatures, and time duration. By altering the denaturation, annealing/extension temperatures, and times, each locus was confirmed to get the optimal and specific PCR results with the most favorable concentration of at least 0.5 to 1 ng of DNA. Control DNA 007 (Applied Biosystems) with concentration of 0.5 ng was amplified through the annealing/extension temperature range of 55 to 65 °C. The most non-specific products were observed at 55 °C amplifying 26 alleles. By increasing the annealing temperature, amplification efficiency of the loci was too improved with the decrease of non-specific products and artifacts. DNA profile was obtained with amplification proficiency of the loci with 28 alleles at 56, 29 alleles at 57, 58, and 59 °C, and complete DNA profile was obtained at optimal temperature of 60 °C without any non-specific product. When the annealing/extension temperature was increased beyond the optimal limits, low amplification or allele dropout was observed as illustrated in Fig. 4.

Number of PCR cycles

Number of PCR cycles also shares in counteracting the artifacts from the DNA profile; therefore, amplification of the multiplex was tested by 25, 28, 30, 35, and 40 cycle numbers. A complete DNA profile was generated without artifacts in 30 numbers of cycles at optimum peak threshold. Low DNA profile and allele dropout were observed in ≤29 cycles. High peak threshold and non-specificities in the DNA profile were perceived in >30 cycles as illustrated in Fig. 5.

Sensitivity

The optimal amount for efficient amplification of DNA was determined by amplifying different concentrations of control DNA 007 in the range of 2 to 0.031 ng. All loci were amplified successfully at threshold of 100 RFU from 0.25 to 2 ng total DNA in a reaction volume of 12 μL. At the DNA concentration of 0.125 ng, 26 alleles were obtained at threshold 100 RFU and 3 alleles were ≥70 RFU as illustrated in graphical Fig. 6.

Species specificity

As per SWGDAM guidelines, nonhuman DNA was also evaluated to determine the specificity of the loci used in the

Author's personal copy

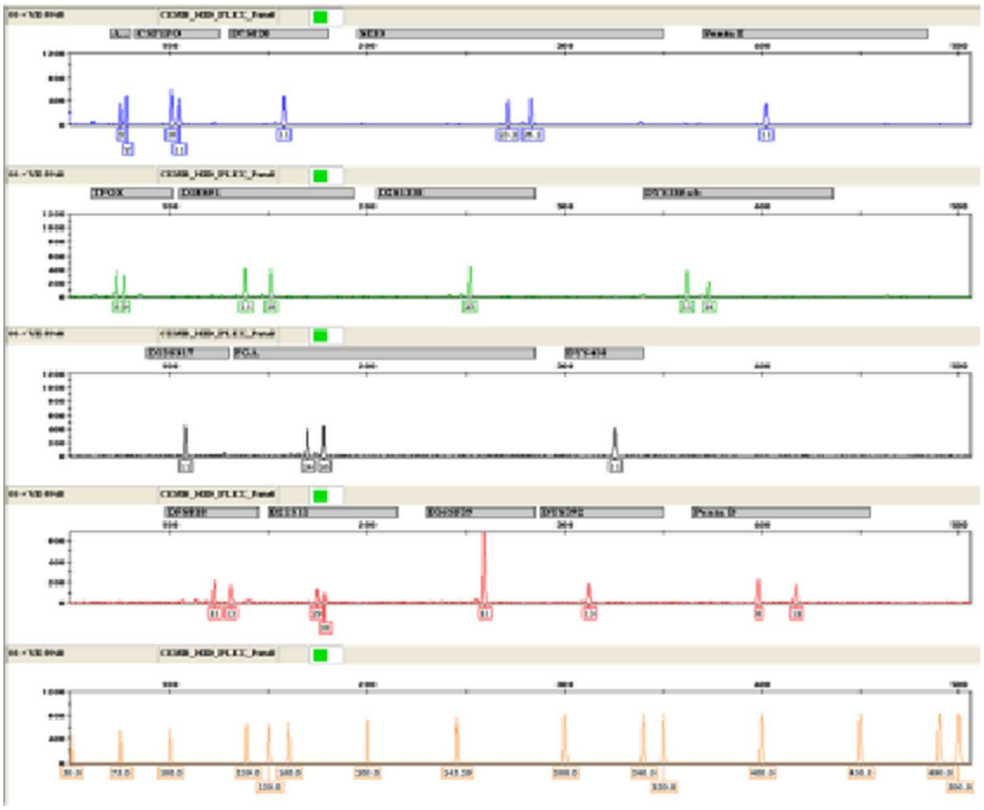


Fig. 2 Electropherogram of control DNA9948

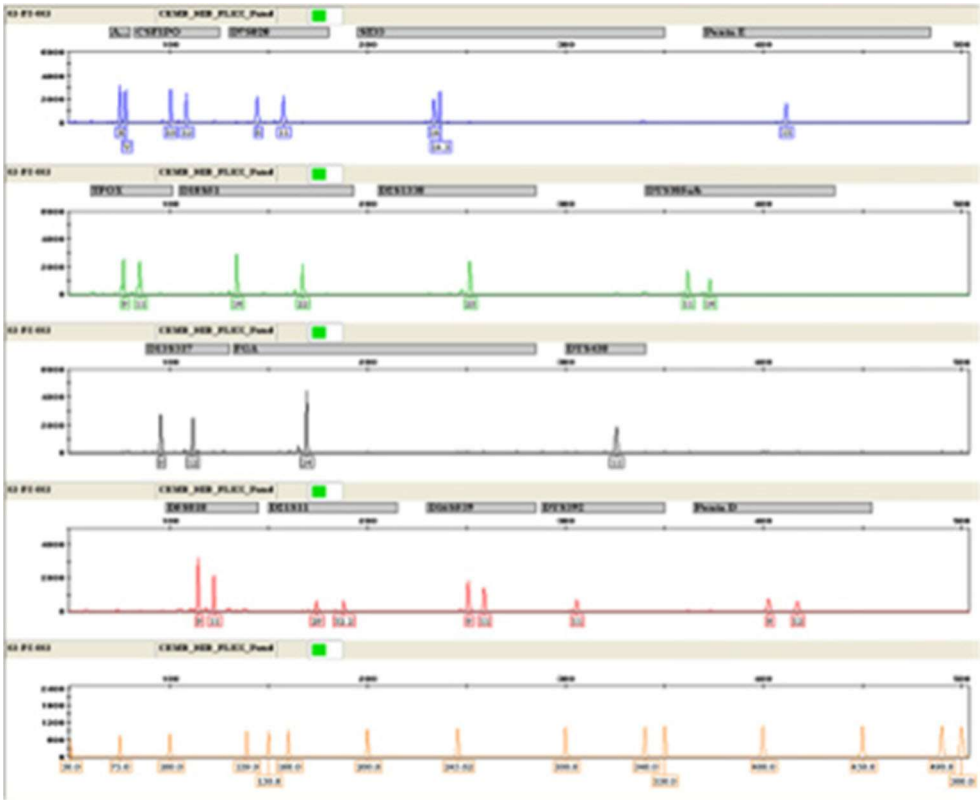


Fig. 3 DNA profile of population sample

Author's personal copy

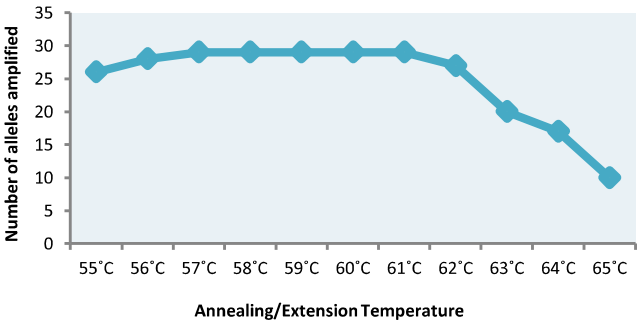


Fig. 4 Number of alleles amplified vs annealing/extension temperature

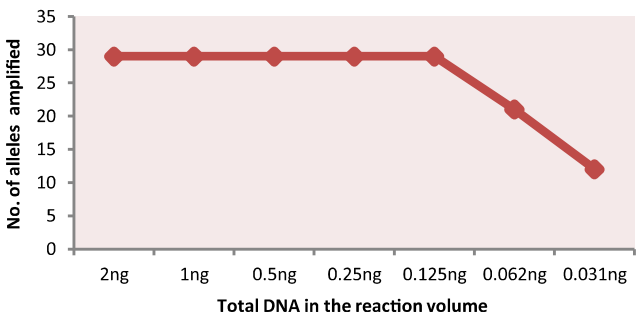


Fig. 6 PCR sensitivity with different DNA concentrations

multiplex system. In this perspective, DNA samples of animals usually found close to the humans like dog, horse, sheep, and chick were amplified with total concentration of 2.0 ng each. In addition to this, microbial DNA like *E. coli* with about 100 copies was also amplified. All these nonhuman DNA samples did not generate any significant PCR amplification as shown in Fig. 7.

Stability studies

The capability of the system to generate the DNA profile from compromised evidence samples like degraded or PCR inhibited was observed. To obtain the fragment degradation less than 300 bp, DNA was treated with DNase-1 enzyme from the intervals of 0 to 5, 10, 20, 40, and 80 min. These degraded DNA samples were subsequently amplified in triplicates to observe the amplification performance of the loci in the multiplex. However, at the interval of 0 min, all loci were amplified successfully. The amplification of two loci (Penta E and Penta D) were observed with threshold of <100 RFU at the intervals of 0.5, 1, 2, 3, and 4 min which were dropped out at the interval of 5 min. After 5 min, increase in drop out of larger amplicons like Penta E, Penta D, DYS385a/b, DYS438, and DYS392 were seen at 10, 20, 40, and 80 min time intervals as illustrated in Fig. 8 with plus-minus standard deviation error bars and average number of loci amplified in the triplicate.

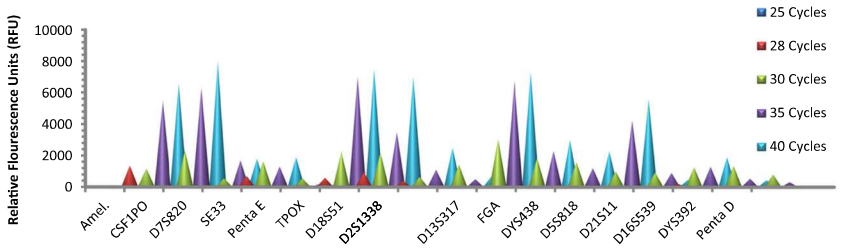
Evaluating the inhibitory effects of hematin, it was observed that each locus in the multiplex amplified successfully up to the concentration of 85 μM. The inhibition of hematin was observed partially at the concentration of 100 μM where only one locus (Penta E) was dropped out from the multiplex while four loci (Penta E, Penta D, D21S11, and DYS392) were inhibited at 125 μM. The more number of loci failed to amplify with the increase of hematin concentration by inhibiting the amplification of Penta E, Penta D, D21S11, DYS385a/b, DYS438, and DYS392 at 150 μM. All other loci except amelogenin, CSF1PO, TPOX, D18S51, D13S317, and D5S818 failed to amplify due to the inhibition effect of 250 μM hematin (Fig. 9). Likewise, amplification of all the loci was inhibited in the samples with hematin concentration of 500 and 1000 μM. Error bars in Fig. 9 show the plus/minus standard deviation and average percentage of loci amplified in the triplicate PCR.

When the inhibitory effects of humic acid were examined, we came to know that no inhibition was seen during the amplification of samples with humic acid concentration up to 100 ng. But the amplification of loci SE33, Penta E, Penta D, DYS385a/b, DYS438, and DYS392 was inhibited at the concentration of 125 ng. Similarly, all loci except CSF1PO, D7S820, TPOX, D18S51, D13S317, FGA, D5S818, amelogenin, and CSF1PO, TPOX, D18S51, D13S317, D5S818, and amelogenin were failed to amplify at humic acid concentration of 175 and 225 ng, respectively, as illustrated in Fig. 10 with average percentage of loci amplified and plus-minus standard deviation error bars.

Mixture studies

A mixed DNA sample can be identified by the presence of more than two alleles at a single locus, if stutter peak height is greater than its percentage ratio and considerably imbalance peak heights of the heterozygous alleles. Due to the presence of Y-STRs in the system, number of male individuals in the mixture may also be recognized. However, two DNA samples

Fig. 5 Amplification efficiency in different number of PCR cycles



Author's personal copy

Fig. 7 Genotype results of Nonhuman DNA (dog, horse, sheep, chick, and *E. coli*.) along with human control DNA007

(male A:female B) were mixed together in the ratios of 1:15, 1:10, 1:7, 1:5, 1:3, 1:1, 3:1, 5:1, 7:1, 10:1, and 15:1. Three DNA samples (female A: male B: male C) were also mixed in the ratios of 10:1:1, 3:1:1, 1:1:1, 1:3:3, 1:10:10, 1:10:1, 1:3:1, 1:1:3, and 1:1:10. Each sample of the ratio was combined in such a way that final DNA concentration remained 1 ng/μL in the mixed sample. Then 1 ng DNA from each prepared ratio was used to conduct mixture studies in the triplet PCR amplification. After the analysis of genotype data obtained from genetic analyzer, it was shown that all the alleles of minor DNA component generated in the mixed ratios of 1:3, 1:1, and 3:1 were at threshold of ≥100 RFU. In the mixture of 1:5 and 5:1, the Penta E and Penta D were detected at <100 RFU but >50 RFU with average percent of 94.4 ± 1.08. Alleles of the minor component generated in 10:1, 1:7, 7:1, and 1:10 ratios were at threshold level of 50 RFU particularly at Penta D locus where found lower than 50 RFU in the ratio of 1:10 by average of 90.7 % ± 1.78. In the DNA mixture ratio of 1:15, amplification of all the loci was observed below the threshold of 50 RFU.

Moreover, by the inclusion of Y-STRs (DYS385a/b, DYS438 and DYS392) in this system, it would be possible to get the information about the number of male individuals contributing DNA in the mixture by observing more than two alleles at DYS385a/b and more than one allele at DYS438 or DYS392. Therefore, at DYS385a/b in all ratios of more than one male DNA mixtures studied, 3 alleles were observed. Penta E and Penta D were also observed with very low (<1 %) or no stutter percentage in the samples that would be beneficial in mixed DNA profile interpretation.

Allelic ladders

Allelic ladders for the loci D7S820, Penta E, DYS385a/b, D13S317, DYS438, and Penta D were constructed by selecting different alleles from the population samples. Each polymorphic allele for the specific locus was separately amplified in a single PCR reaction. Then all amplified alleles of a locus were combined into single allelic ladder of the specific locus. Allele no. 7 was incorporated into the locus of DYS438 which was not included in commercial kits like Y-filer and

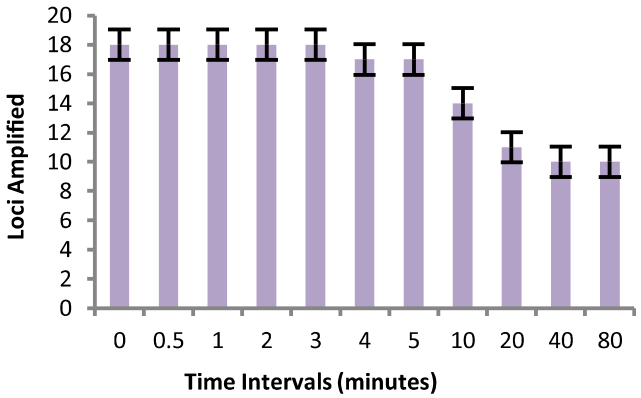


Fig. 8 DNA degradation studies

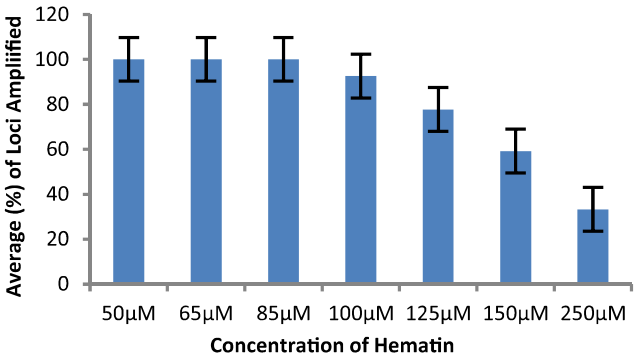


Fig. 9 Inhibitory effects of hematin

Author's personal copy

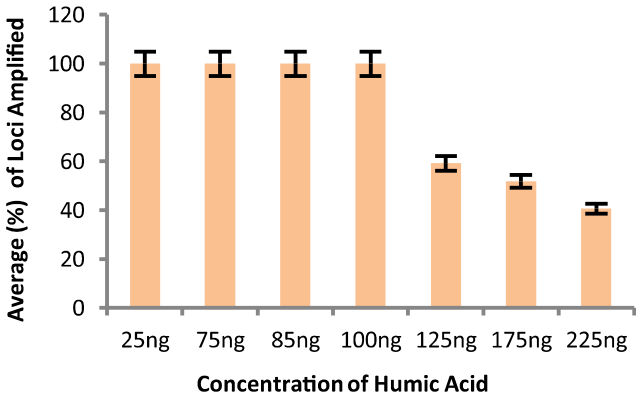


Fig. 10 Inhibitory effects of humic acid

PowerPlex Y. Variant allele 10.2 was also included into the ladder of Penta D locus. Allelic ladder of D7S820 and D13S317 were prepared from the population samples to overcome the incomplete adenylation process at each allele. The bin set for alleles of each allelic ladder was constructed according to their new fragment base pair sizes obtained from the genetic analyzer. The allelic ladders of the remaining loci were generated by amplifying the prepared dilution (1:500) of allelic ladder from the Identifiler™ (Applied Biosystems), PowerPlex ES (Promega), and Y-filer™ (Applied Biosystems) kits [16].

Size precision and accuracy

Size precision and accuracy were determined from various injections of samples run on ABI3130xl Genetic Analyzer. For accurate sizing of amplified fragments, GS500 LIZ (Applied Biosystems) was used as an internal lane size standard during capillary electrophoresis. For the accurate allele designation, the size variations were normally compared among samples' alleles and allelic ladder's alleles on the ABI3130xl Genetic Analyzer. All samples' alleles were retained by constructing the relevant bins within ±0.5 bp from a corresponding allele in the allelic ladder to designate the accurate allele number.

Forensic competency and statistical parameters

By calculating the allelic frequencies in the population, subsequently, it would be applicable to interpret the matching probability of the suspected standard samples, either included or excluded as being the source of biological evidence in the samples tested. For this purpose, relevant forensic and paternity statistical parameters were investigated by genotyping 300 DNA samples of unrelated individuals from the province of Punjab Pakistan. The allele repeat number 8 was found to be the most frequent with frequency of 0.386 at TPOX while among the Y-STRs, allele number 11 was found with maximum frequency of 0.7220 at locus DYS392 (supplementary table 3). Among all the markers evaluated for this study, 209 alleles were observed showing polymorphism information content (PIC) in the range of 0.654 (TPOX) to 0.945 (SE33) in autosomal STRs and 0.426 (DYS392) to 0.859 (DYS385a/b) in Y-STRs. All autosomal STR loci showed the power of discrimination >0.86 with maximum value of 0.9913 at SE33 and 42 alleles including 23 variants even more than the German population [46]. The second most discriminatory locus was Penta E with PD value of 0.9832 and 20 alleles (supplementary table 4). The combined power of discrimination, combined power of exclusion and the combined probability of matching were calculated as $0.999999999999999999893916339344$, 0.999993816173890, and 5.90019×10^{-19}, respectively. At the probability level of 0.05, there were significant deviations from Hardy-Weinberg equilibrium at five autosomal loci as shown in supplementary table 4, but after Bonferroni correction, no deviation was observed. The results of locus to locus pairwise linkage disequilibrium showed that at p value of 0.05, significant deviations were observed at 45 pairs of loci, but after Bonferroni correction ($p < 0.003$), there was no deviation seen. By the analysis of Y-STRs used in this multiplex, a total of 110 haplotypes were unique in 287 male DNA samples. Among these haplotypes, 62 were observed once and 48 were repeated more than one time. The unique haplotype frequencies and their diversities have been given in the supplementary table 6. The haplotype diversity for each locus was calculated as 0.875, 0.665, and 0.455 at DYS385a/b, DYS438, and DYS392, respectively.

Conclusion

The purpose of this new multiplex system is the simultaneous amplification of 10 miniSTRs, conventional STRs like SE33, Penta E, Penta D, and four Y-STRs providing DNA profiling information of an individual from those biological evidence samples where DNA is highly degraded and even in the presence of inhibitors during the 1 h of amplification procedure. In addition to this, inclusion of Y-STR markers exclusively supports the multiplex system in those samples where mutation at Y-amelogenin marker occurs in male individuals. Y-STRs will also be helpful in those samples where random sharing of allele occurs in unrelated individuals at autosomal STRs especially in parentage analysis and in forensic complex cases like mixture DNA interpretation involving more than two individuals. The locus SE33 used in this multiplex system due to its highly polymorphic nature and discriminatory power even more than the German population [46] seems to be the core locus of Pakistan. On the basis of lower probability of matching, highly polymorphic makeup, robustness of the system, and by DNA database development of these markers for the major populations of the country, this system would be greatly helpful for the inclusion or exclusion of an individual in criminal cases and to resolve paternity issues.

Author's personal copy

References

1. Butler JM (2005) Forensic DNA typing: biology, technology, and genetics of STR markers. Academic Press
2. Butler JM, Hill CR (2012) Biology and genetics of new autosomal STR loci useful for forensic DNA analysis. Forensic Sci Rev 24: 15–26
3. Wiegand P, Budowle B, Rand S, Brinkmann B (1993) Forensic validation of the STR systems SE 33 and TC 11. Int J Leg Med 105(6):315–320
4. Jobling MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. Nat Rev Genet 5(10):739–751
5. National DNA Database statistics, Q1 2015 to 2016. National DNA Database statistics. UK Government Home Office
6. Zamir A, Dell'Ariccia-Carmon A, Zaken N, Oz C (2012) The Israel DNA database—the establishment of a rapid, semi-automated analysis system. Forensic Sci Int Genet 6(2):286–289
7. Nick V (2015) Kuwait to institute mandatory DNA testing for all residents. Huffington Post. Retrieved 10 October 2015
8. Hill CR, Kline MC, Coble MD, Butler JM (2008) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. J Forensic Sci 53(1):73–80
9. Hill CR, Butler JM, Vallone PM (2009) A 26plex autosomal STR assay to aid human identity testing. J Forensic Sci 54(5): 1008–1015
10. Jiang X, Guo F, Jia F, Jin P, Sun Z (2013) Development of a 20-locus fluorescent multiplex system as a valuable tool for national DNA database. Forensic Sci Int Genet 7(2):279–289
11. Guo F, Shen H, Tian H, Jin P, Jiang X (2014) Development of a 24-locus multiplex system to incorporate the core loci in the Combined DNA Index System (CODIS) and the European Standard Set (ESS). Forensic Sci Int Genet 8(1):44–54
12. Zhang S, Bian Y, Tian H, Wang Z, Hu Z, Li C (2015) Development and validation of a new STR 25-plex typing system. Forensic Sci Int Genet 17:61–69
13. Wiegand P, Kleiber M (2001) Less is more–length reduction of STR amplicons using redesigned primers. Int J Leg Med 114(4–5): 285–287
14. Tsukada K, Takayanagi K, Asamura H, Ota M, Fukushima H (2002) Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01, TPOX, CSF1PO, and vWA loci. Leg Med 4(4):239–245
15. Whitaker JP, Clayton TM, Urquhart AJ, Millican ES, Downes TJ, Kimpton CP, Gill P (1995) Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. Biotechniques 18(4): 670–677
16. Butler JM, Shen Y, McCord BR (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 48(5):1054–1064
17. Drábek J, Chung DT, Butler JM, McCord BR (2004) Concordance study between Miniplex assays and a commercial STR typing kit. J Forensic Sci 49(4):859–860
18. Bender K, Farfán MJ, Schneider PM (2004) Preparation of degraded human DNA under controlled conditions. Forensic Sci Int 139(2):135–140
19. Grubwieser P, Mühlmann R, Berger B, Niederstätter H, Pavlic M, Parson W (2006) A new "miniSTR-multiplex" displaying reduced amplicon lengths for the analysis of degraded DNA. Int J Leg Med 120(2):115–120
20. Budowle B, Moretti TR, Niezgoda SJ, Brown BL (1998) CODIS and PCR-based short tandem repeat loci: law enforcement tools. Proceedings of the Second European Symposium on Human Identification. Promega Corporation 73–88
21. Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. J Forensic Sci 51(2): 253–265
22. Hares DR (2012) Expanding the CODIS core loci in the United States. Forensic Sci Int Genet 6(1):e52–e54
23. Bacher J, Hennes LF, Gu T, Tereba A, Micka KA, Sprecher CJ, Lins AM, Amiott EA, Rabbach DR, Taylor JA, Helms C, Keller HD, Schumm JW (1999) Pentanucleotide repeats: highly polymorphic genetic markers displaying minimal stutter artifact. In Proceedings from the Ninth International Symposium on Human Identification
24. Santos FR, Pandya A, Tyler-Smith C (1998) Reliability of DNA-based sex tests. Nat Genet 18(2):103–103
25. Thangaraj K, Reddy AG, Singh L (2002) Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? Int J Leg Med 116(2):121–123
26. Gonzalez-Andrade F, Sanchez D, Penacino G, Jarreta BM (2009) Two fathers for the same child: a deficient paternity case of false inclusion with autosomic STRs. Forensic Sci Int Genet 3(2):138–140
27. Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12(4):656–664
28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. Genome Res 12(6):996–1006
29. Shafique M, Shahzad MS, Shan MA, Ali A, Rahman Z, Husnain T (2015) Genetic polymorphism of ten MiniSTR loci in the population of Punjab Pakistan for forensic application. Int J Leg Med 129(5):937–942
30. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Bioinform Methods Protocol 365–386
31. Haas-Rochholz H, Weiler G (1997) Additional primer sets for an amelogenin gene PCR-based DNA-sex test. Int J Leg Med 110(6): 312–315
32. Polymeropoulos MH, Rath DS, Xiao H, Merril CR (1992) Tetranucleotide repeat polymorphism at the human beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2). Nucleic Acids Res 20(6):1432–1432
33. De Knijff P, Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann F, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, Roewer L, Chromosome Y (1997) Microsatellites: population genetic and evolutionary aspects. Int J Leg Med 110(3):134–140
34. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Masibay A, Rabbach DR, Amiott EA, Sprecher CJ (2002) Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 47(4):773–785
35. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF (2002) A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci Int 129(1):10–24
36. Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH (1997) Multiplex PCR: critical parameters and step-by-step protocol. Biotechniques 23(3):504–511
37. Signer E, Kuenzle CC, Thomann PE, Hübscher U (1988) DNA fingerprinting: improved DNA extraction from small blood samples. Nucleic Acids Res 16(15):7738
38. Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning: a laboratory manual, vol 545. Cold Spring Harbor Laboratory, Cold Spring Harbor
39. Griffiths RAL, Barber MD, Johnson PE, Gillbard SM, Haywood MD, Smith CD, Arnold J, Burke T, Urquhart AJ, Gill P (1998) New reference allelic ladders to improve allelic designation in a multiplex STR system. Int J Leg Med 111(5):267–272

Author's personal copy

40. Tereba A (1999) Tools for analysis of population statistics. Profiles DNA 2(3):14–16
41. Lui K, Muse SV (2005) PowerMarker: integrated analysis environment for genetic marker data. Bioinformatics 21(9):2128–2129
42. Presley LA, Mudd JL (1996) The development of quality assurance measures in forensic DNA typing. In 16th Congress of the International Society for Forensic Haemogenetics, Santiago de Compostela, 12–16, 675–677
43. Sparkes R, Kimpton C, Gilbard S, Carne P, Andersen J, Oldroyd N, Thomas D, Urquhart A, Gill P (1995) The validation of a 7-locus multiplex STR test for use in forensic casework (II) Artefacts, casework studies and success rates. Int J Legal Med 109(4):195–204
44. Sparkes R, Kimpton C, Watson S, Oldroyd N, Barnett L, Arnold J, Thompson C, Hale R, Chapman J, Urquhart A, Gill P, Clayton T (1996) The validation of a 7-locus multiplex STIR test for use in forensic casework. Int J Leg Med 109(4):186–194
45. Wallin J, Buoncristiani M, Lazaruk K, Fildes N, Holt C, Walsh P (1998) TWGDAM validation of the AmpFlSTR Blue PCR amplification Kit for forensic casework analysis. J Forensic Sci 43(4): 854–870
46. Schmid D, Anslinger K, Rolf B (2005) Allele frequencies of the ACTBP2 (= SE33), D18S51, D8S1132, D12S391, D2S1360, D3S1744, D5S2500, D7S1517, D10S2325 and D21S2055 loci in a German population sample. Forensic Sci Int 151(2):303–305