Forensic Science International: Genetics 19 (2015) 172–179

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

An evaluation of the PowerSeq™ Auto System: A multiplex short tandem repeat marker kit compatible with massively parallel sequencing

Xiangpei Zeng[a,*], Jonathan King[a], Spencer Hermanson[b], Jaynish Patel[b], Douglas R. Storts[b], Bruce Budowle[a,c]

[a] Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, Texas 76107, USA
[b] Promega Corporation, 2800 Woods Hollow Road, Madison, WI 53711, USA
[c] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 3 April 2015
Received in revised form 10 June 2015
Accepted 13 July 2015

Keywords:
Short tandem repeat (STR)
Massively parallel sequencing (MPS)
PowerSeq Auto System
Single nucleotide polymorphism (SNP)
Mixture
MiSeq

ABSTRACT

Capillary electrophoresis (CE) and multiplex amplification with fluorescent tagging have been routinely used for STR typing in forensic genetics. However, CE-based methods restrict the number of markers that can be multiplexed simultaneously and cannot detect any intra-repeat variations within STRs. Several studies already have indicated that massively parallel sequencing (MPS) may be another potential technology for STR typing. In this study, the prototype PowerSeq™ Auto System (Promega) containing the 23 STR loci and amelogenin was evaluated using Illumina MiSeq. Results showed that single source complete profiles could be obtained using as little as 62 pg of input DNA. The reproducibility study showed that the profiles generated were consistent among multiple typing experiments for a given individual. The mixture study indicated that partial STR profiles of the minor contributor could be detected up to 19:1 mixture. The mock forensic casework study showed that full or partial profiles could be obtained from different types of single source and mixture samples. These studies indicate that the PowerSeq Auto System and the Illumina MiSeq can generate concordant results with current CE-based methods. In addition, MPS-based systems can facilitate mixture deconvolution with the detection of intra-repeat variations within length-based STR alleles.

1. Introduction

Short tandem repeats (STRs) are used routinely in forensic DNA human identification testing, because of their high discrimination power and relatively short amplicon size. Given the importance of STRs in forensic genetics, several commercial kits are available [1–4], each enabling the multiplex amplification of at least 15 loci. While reliable, capillary electrophoresis (CE)-based technology restricts the number of STRs that can be multiplexed. With a CE-based method, the loci, labeled with the same fluorescent dye, are separated by size (~100–500 bp), and thus only 5 or 6 loci can fit within a dye channel [5,6]. In addition, STR alleles are distinguished based on length and commonly named by the number of repeats contained within the amplicon. Any intra-allele single nucleotide polymorphisms (SNPs) or repeat motif variation (RMV) within STRs cannot be detected when using CE-based methods. Thus, the full discrimination power of some STRs cannot be realized.

Massively parallel sequencing (MPS) technology, with its high throughput, can sequence far more markers simultaneously than can be typed by CE-based methods. In addition, the capacity of MPS allows for sequencing of multiple samples in one analysis, typically 2–96 different samples, through the use of barcoding [7,8]. MPS technology identifies a locus based on sequence, and thus it can increase the discrimination power of some STRs while providing the standard nominal repeat number nomenclature. STR amplicons can be engineered to be generally shorter and more similar in size [9]. Several studies already have revealed the potential application of MPS for STR typing [9–17]. Fordyce et al. [18] described the analysis of 10 STRs on a variety of biological samples using the Ion PGM™. Gelardi et al. [19] and Rockenbauer et al. [20] characterized mutations and sequence variants in STR loci. Previously, Zeng et al. [9] described a prototype STR multiplex (17 STR loci + Amelogenin) compatible with the MiSeq system

* Corresponding author.
E-mail address: Xiangpei.Zeng@live.unthsc.edu (X. Zeng).

http://dx.doi.org/10.1016/j.fsigen.2015.07.015
1872-4973/

(Illumina, San Diego, CA, USA). The results indicated reliable STR profiles could be obtained at a sensitivity level that is comparable with current widely used CE-based methods. At the same time, several algorithms and workflows have been developed to facilitate the detection of STR alleles and intra-allele SNPs and RMV from MPS output data [14–17]. All these studies support that multiplex STR typing by MPS is a promising technology for forensic applications.

In the study herein, the performance of a new prototype kit, the PowerSeq™ Auto System (Promega Corporation, Madison, WI, USA) that is designed for MPS on the MiSeq, was evaluated. The multiplex contains the CODIS 13 core loci, the 12 core European Standard Set loci, as well as Penta D, Penta E, D2S1338, D19S433, DYS391 loci and the Amelogenin locus, i.e., the same markers in the PowerPlex Fusion system (Promega). The studies addressed are sensitivity of detection, reproducibility, mixture samples typing, and mock forensic casework-type samples.

2. Methods and materials

2.1. Sample preparation

Blood, saliva and semen samples were obtained from 18 unrelated individuals with informed consent. All samples were anonymized and collected in accordance with University of North Texas Health Science Center's Institutional Review Board. All single source samples and the blood/saliva mixture samples were extracted using the QIAamp® DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommended protocol [21]. For semen and other body fluid mixtures (blood/semen and saliva/semen mixtures), the differential extraction method according to Giusti et al. [22] was used to extract DNA. Bone samples were extracted as described by Marshall et al. [23]. Quantifiler Human DNA Quantification Kit (Thermo Fisher, Foster City, CA, USA) on an ABI Prism 7500 Sequence Detection system (Thermo Fisher) was used to estimate the quantity of extracted DNA [24].

2.2. PCR amplification

Amplification of the template DNA was performed using reagents contained within the prototype PowerSeq™ Auto System following the manufacturer's recommended protocol [25]. The loci in the multiplex are: Amelogenin, CSF1PO, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D1S1656, D21S11, D22S1045, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, DYS391, FGA, Penta D, Penta E, TH01, TPOX, and vWA. The PCR included PowerSeq™ Auto Primer Pair Mix, PowerSeq™ Auto 5X Master Mix, Amplification Grade Water, and 500 pg of total genomic DNA. Amplification was performed using the recommended thermal-cycling parameters: an initial temperature of 96 °C for 1 min; followed by 30 cycles of 94 °C for 10 s, 59 °C for 1 min, 72 °C for 30 s; and a final extension of 10 min at 60 °C. Amplified products were purified using the MinElute PCR Purification Kit (Qiagen) [26]. After PCR cleanup, the Qubit dsDNA BR kit (Thermo Fisher), which can detect 2–1000 ng, was used to determine the quantity of amplicon according the manufacturers' recommended protocols [27].

2.3. Library preparation

Library preparation was performed using the TruSeq DNA HT or LT sample preparation kits according to the manufacturer's protocol (Illumina) [28]. TruSeq HT kit contains 96 indices and reagents for 96 samples preparation. TruSeq LT kit (A or B set) contains 12 indices and reagents for 24 samples preparation, but otherwise, the chemistry is the same between the kits. The MinElute PCR purification kit (Qiagen) was used for size selection. The Qubit dsDNA HS kit (Thermo Fisher), which can detect 0.2–100 ng, was used to determine the quantity of each indexed DNA library, and then they were normalized to 2 nM for sequencing [29].

2.4. MPS sequencing and data analysis

Indexed DNA libraries (2 nM) were pooled and diluted to 10 pM for loading. Sequencing reagents were included in the MiSeq v2 (2×250 bp) kit (Illumina). The MiSeq re-sequencing protocol for small genome sequencing was performed [30]. Raw FASTQ files, generated by the MiSeq, were analyzed using STRait Razor v2.0 [15]. For each sample, depth of coverage (DoC) was calculated for each allele and allele coverage ratio (ACR) was determined for each heterozygous STR locus: i.e., the lower coverage allele divided by the higher coverage allele [9].

2.5. PCR sensitivity

PCR sensitivity study was conducted using DNA samples from three individuals (no. 1–3), each was amplified at different amounts of input DNA: i.e., 500, 250, 125, 62, 31, and 16 pg (i.e., 18 samples). After purification, 40 ng of amplified products were used for all library preparations; 40 ng were selected as this was the minimum amount obtained among the samples (sample no. 2, 16 pg yielded 40 ng of amplified product).

2.6. Reproducibility

DNA from six different individuals (no. 4–9) were amplified in four separate reactions to determine ACR variation, interlocus balance (the lowest coverage locus/the highest coverage locus) and coverage variation (the coverage of each locus/the total coverage of all loci). To avoid run-to-run variation, all 24 reactions were sequenced in the same run. After purification, 50 ng of amplified products were used for library preparation.

2.7. Mixture study

One male and one female DNA sample (no. 10 and 11) at a total amount of 500 pg were mixed in the following ratios: 19:1, 9:1, 6:1, 4:1, 2:1, 1:1, 1:2, 1:4, 1:6, 1:9, and 1:19 (i.e., 11 samples). Sample no. 10 was the 2800 M control DNA provided with the kit and thus also served as a positive control throughout the study. After purification, 40 ng of amplified products were used for library preparation.

2.8. Mock forensic casework study

To mimic forensic casework, three blood samples, three saliva samples, one semen sample, and two bone samples were extracted. In addition, one blood/semen mixture, one saliva/semen mixture, and two blood/saliva mixtures (in three ratios: 5:1, 1:1 and 1:5) were tested. These samples were sequenced in two runs. After purification, 50 ng and 60 ng of amplified products were used for library preparation for 15 and 12 samples, respectively. These amounts were selected as they were the minimum quantities yielded among these sets of samples.

2.9. CE STR analysis

All single source samples (no. 1–11) and mock forensic samples were typed using a CE-based method. Amplification of the template DNA was performed using reagents provided in the PowerPlex® Fusion System (Promega) according to the manufacturer's protocol [31]. Five hundred pg of total genomic DNA were

used in the PCR. Amplification was performed using the same parameters listed in Section 2.2. One microliter of amplified sample was added into 11 μl of formamide/internal lane standard mix and separated in the Applied Biosystems® 3130XL Genetic Analyzer (Thermo Fisher). GeneMapper® ID software v3.2 (Thermo Fisher) was used for data analysis. The analytical threshold used for CE analysis was 50 RFU.

3. Results and discussion

The prototype PowerSeq™ Auto System allows for multiplex amplification of 22 autosomal STRs, one Y STR and the Amelogenin locus. These loci are those of the majority of core markers required for databasing. Studies were performed to determine whether this STR multiplex kit, that generates amplicons that can be typed by MPS, can produce reliable results and whether such results are comparable to those currently obtained with a CE-based method. Studies to determine sensitivity of detection, DoC, ACR, and reproducibility were performed. In addition, mixtures and casework-type samples were sequenced. Three sequencing runs were performed in total. The first run contained 18 samples from the sensitivity study and 11 samples from the mixture study (40 ng of amplified products); the cluster density was 640 k/mm^2 and estimated yield was 6003.4 MB. The second run included 24 samples from the reproducibility study and 15 samples from the mock forensic casework study (50 ng of amplified products); the cluster density was 96 k/mm^2 and estimated yield was 935.7 MB. The third run contained 12 samples from the mock forensic casework study (60 ng of amplified products); the cluster density was 384 k/mm^2 and estimated yield was 3595 MB.

3.1. PCR sensitivity study

The sensitivity of detection of the prototype PowerSeq™ Auto System was determined by analyzing DNA samples from three individuals (no. 1–3) at six different amounts of input DNA: 500, 250, 125, 62, 31, and 16 pg. Sample no. 2 with 16 pg of input DNA yielded the lowest quantity of amplified products, i.e., 40 ng. Therefore, 40 ng of amplified products were used for library preparation for all 18 reactions. The CE data were used for concordance evaluation in the PCR sensitivity study. The sequence results at 500 pg of input DNA (i.e., the recommended input amount) were consistent with the CE data. The results showed that 500 pg of input DNA generated high DoC for all alleles (5099×–24,269×), with an overall average ACR of 0.80 ± 0.15 for these 23 STR loci and Amelogenin (Table 1, Fig. 1). A few loci showed more imbalanced ACRs with 250 and 125 pg of input DNA compared with 500 pg of input DNA. With 62 pg of input DNA, notable imbalanced ACRs were observed for the three samples: 11 loci displayed <0.4, 14 loci ranged from >0.4 to 0.6, and 34 loci displayed >0.6, with an overall average ACR of 0.63 ± 0.24. Although there was increased imbalance for heterozygous types, complete loss of an allele was not observed with 62 pg of input DNA. The greatest ACR imbalance was 0.11 in sample no. 3 at the D21S11 locus. These results indicated that 62 pg may be an initial minimum input amount for analysis. At less than 62 pg of input DNA, allele imbalance was more noted with the average ACR dropping to 0.36 ± 0.32 and examples of allele drop out were observed. These results generally are similar to those obtained with CE-based systems.

3.2. Reliability study

Each of six samples (no. 4–9) was amplified in four separate reactions at 500 pg of input DNA. Among the four replicates of a given sample, all STR allele calls were concordant (245/245). The ACRs ranged from a low of 0.46 ± 0.19 (for sample no. 8 at the Penta E locus) to a high of 0.95 ± 0.03, 0.95 ± 0.05 and 0.95 ± 0.04 (for sample no. 4 at the D1S1656 locus, sample no. 5 at the D7S820 locus and sample no. 9 at the Penta E locus, respectively) (Supplemental Table 1). The variation of ACRs was limited among the four reactions of six individuals with 88 of 101 heterozygous loci with a standard deviation <0.20 (arbitrarily selected value for comparison purposes only), except for the loci D12S391 (two individuals), D19S433 (two individuals), D10S1248 (one

Table 1
Comparison of ACRs that were generated with different quantities of input DNA for three samples. The blank or no entry values are homozygote types.

	No. 1–16 pg	No. 2–16 pg	No. 3–16 pg	No. 1–31 pg	No. 2–31 pg	No. 3–31 pg	No. 1–62 pg	No. 2–62 pg	No. 3–62 pg	No. 1–125 pg	No. 2–125 pg	No. 3–125 pg	No. 1–250 pg	No. 2–250 pg	No. 3–250 pg	No. 1–500 pg	No. 2–500 pg	No. 3–500 pg
AMEL	0[a]	0.2374	0[a]	0.0076	0[a]	0.7610	0.6585	0.4294	0.8415	0.7323	0.5554	0.8631	0.8256	0.6801	0.9487	0.8251	0.4260	0.6190
CSF1PO	0[a]	0.0690	0.5385	0.8348	0.3352	0.5877	0.9710	0.6055	0.6532	0.9215	0.7917	0.7649	0.6453	0.7528	0.4735	0.9508	0.8603	0.8686
D10S1248		1.0000	0.2500		0.5000	0.0024		0.7657	0.9529		0.8618	0.4052		0.3479	0.9705		0.6934	0.6627
D12S391	0[a]	0.0015	0[a]	0.0009	0.0007	0.0194	0.6373	0.9499	0.7907	0.5544	0.9971	0.5006	0.6869	0.8485	0.8085	0.9591	0.6477	0.7926
D13S317	0.2500	0.0092	0[a]	0.0700	0.5848	0.3892	0.5117	0.6292	0.4123	0.4966	0.9108	0.5173	0.9236	0.7583	0.9440	0.9038	0.7625	0.9988
D16S539	0.8319	0.1268	0.0281	0.5737	0.0144	0.2289	0.3258	0.6750	0.9687	0.7598	0.3137	0.7303	0.8655	0.6649	0.8428	0.7524	0.8859	0.8174
D18S51	0.6950	0[a]	0.0002	0.0052	0.0033	0.5456	0.6486	0.8625	0.4662	0.8416	0.4962	0.6523	0.6166	0.7096	0.8445	0.4029	0.8657	0.8251
D19S433	0.2587	0.0010		0.4707	0.9793		0.3960	0.7056		0.6873	0.8226		0.7977	0.8153		0.6402	0.8713	
D1S1656		0.5018	0[a]		0.0023	0[a]		0.4659	0.6109		0.1763	0.7401		0.8590	0.8871		0.9712	0.6321
D21S11	0.5545	0[a]	0.5000	0.0005	0.5876	0.0019	0.5853	0.4882	0.1060	0.9934	0.5736	0.4903	0.5383	0.9521	0.4247	0.9041	0.9107	0.7281
D22S1045	0.0008	0.0833		0.0002	0.3119		0.4962	0.1492		0.6537	0.4522		0.5726	0.8145		0.5657	0.9659	
D2S1338	0.0004	0.0009	0.1821	0.0031	0.8219	0.4206	0.2212	0.7358	0.9209	0.6880	0.6575	0.7723	0.7040	0.6798	0.8934	0.9482	0.9621	0.7821
D2S441	0.0005	0.6681		0.3863	0.1934		0.9862	0.9217		0.9055	0.2763		0.7335	0.8620		0.9531	0.6562	
D3S1358	0.0992	0.1538	0.6000	0.4736	0.0005	0.3674	0.6236	0.4450	0.8499	0.6398	0.9529	0.1152	0.7708	0.8850	0.4294	0.9945	0.7174	0.9987
D5S818	0.0854	0.1168	0.0053	0[a]	0.1005	0[a]	0.4465	0.4864	0.7201	0.5694	0.7272	0.9615	0.8254	0.8573	0.6361	0.8657	0.9985	0.9887
D7S820	0.8922	0.5000		0.0049	0.8808		0.6571	0.8846		0.8654	0.3910		0.8974	0.7049		0.6810	0.8532	
D8S1179	0[a]	0.0851	0.5000	1.0000	0.0594	0.2155	0.8802	0.1739	0.9306	0.6261	0.5936	0.9769	0.7799	0.8273	0.8812	0.5941	0.7424	0.6852
DYS391																		
FGA		0.0012	0.6667		0.3383	0.5245		0.4584	0.7432		0.9058	0.6834		0.8404	0.9092		0.6900	0.6373
PENTA D	0[a]	0.3238	0.0025	0.4777	0.4474	0.5135	0.6483	0.9664	0.5800	0.7721	0.5502	0.4918	0.8708	0.7307	0.9381	0.9514	0.9266	0.8338
PENTA E	0[a]	0[a]	0[a]	0.7488	0.6690	0.0026	0.3343	0.3624	0.9849	0.9345	0.3027	0.5945	0.5663	0.7402	0.6850	0.8154	0.6491	0.9383
TH01	0.0043		0.0006	0.3492		0.6675	0.4927		0.9654	0.4411		0.9839	0.9538		0.6437	0.7536		0.5675
TPOX		0.0013	0.9375		0.0020	0.2970		0.3322	0.2535		0.5942	0.6043		0.8069	0.6938		0.7848	0.8970
vWA		1.0000	0[a]		0.9060	0.8703		0.8530	0.3483		0.7091	0.2771		0.9898	0.7103		0.8044	0.9286
Average		0.2901			0.3556			0.6265			0.6580			0.7672			0.8018	
SD		0.3225			0.3158			0.2428			0.2170			0.1449			0.1463	

[a] The ACRs at some loci were 0 when input DNA less than 62 pg, because of alleles drop out.

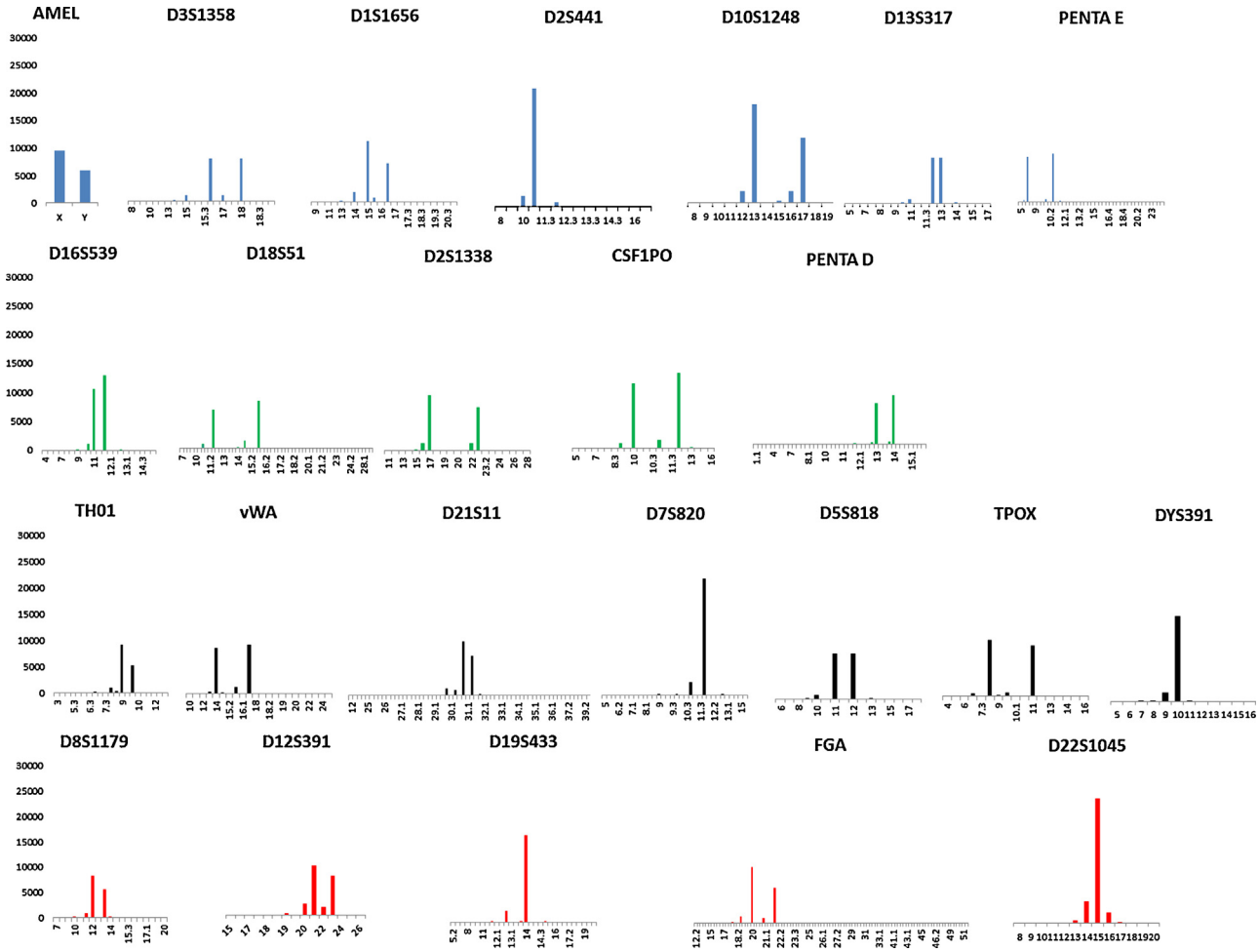

Fig. 1. A histogram portrayal of DoC by locus of one sample (No. 3) with 500 pg of input DNA. *X* axis are the alleles for a given locus, *Y* axis is the coverage for a given allele.

individual), D16S539 (one individual), D2S1338 (one individual), D2S441 (one individual), D8S1179 (one individual), FGA (one individual), Penta D (one individual), Penta E (one individual) and vWA (one individual). The interlocus balance was examined among the four reactions of each sample, and the mean and standard deviation of six samples individually were 0.28 ± 0.08, 0.30 ± 0.08, 0.34 ± 0.03, 0.28 ± 0.06, 0.40 ± 0.06, 0.35 ± 0.05. In addition, the coverage variation per locus/per sample among the four replicates was limited (standard deviation <0.02) (Supplemental Table 2). The data support that the prototype PowerSeq™ Auto System and the MiSeq system can generate consistent results in multiple reactions. While the data herein support that STR typing on MPS will become a viable methodology, additional studies still need to be performed which include increased sample testing for parameters such as stutter; allele drop out; defining proper controls for an analysis; and analyses of reproducibility between or among runs. The latter is particularly important as cluster density, which currently, varies among runs, will affect coverage and may influence stochastic effects. These tests will be performed when the kit is finalized.

3.3. Mixture study

Two individuals (one male and one female, no. 10 and 11) were selected for a mixture study and combined in different ratios. The DNA ratios were 19:1, 9:1, 6:1, 4:1, 2:1, 1:1, 1:2, 1:4, 1:6, 1:9, and 1:19. The total amount of input DNA for each reaction was 500 pg. Data from this study were utilized to determine at what ratios both contributors could be detected. Alleles lying within two repeats downstream (minus stutter) or one repeat upstream (plus stutter) were not considered for minor contributor assessment. In total, 17 alleles from sample no. 10 and 8 alleles from sample no. 11 could be considered as non-overlapping alleles. Sample no.10 is 2800 M control DNA provided with the kit, and the MPS profile was consistent with the CE results. The system was able to detect the two contributors in part at all ratios (Supplemental Table 3). All non-overlapping alleles from the minor contributor could be detected in mixtures 6:1 and 1:6 through 1:1, except two alleles dropped out at the D12S391 locus (allele 23 in the 1:4 mixture and allele 18 in the 1:6 mixture). The DoC for minor alleles from 6:1 to 2:1 mixtures was high ($\geq$305*X*), and minor alleles from the 1:2 to 1:6 mixtures had at least coverage of 241*X*. At the 9:1 and 1:9 mixtures, 5 and 15 non-overlapping alleles, respectively, from the minor contributor were detected with coverage of at least 74*X*. At the 19:1 mixture, 5 of the 8 non-overlapping alleles from no. 11 were found in the mixture ($\geq$211*X*). When the ratio was 1:19, 12 of 17 non-overlapping alleles from no. 10 were detected in the mixture ($\geq$149*X*). Fig. 2 shows that four alleles at three loci were identified from the minor contributor (no. 11) of the 19:1 mixture: i.e., 15.3, 18.3 at the D1S1656 locus; 9 at the TH01 locus; and 15 at the D12S391 locus. All alleles contributed by no. 10 and 11 were detectable in the 1:1 mixture. However, the results showed an imbalanced DoC for alleles at some loci, e.g., the coverage ratio of *X* and *Y* was almost 8:1 at Amelogenin rather than an expected 3:1 (Fig. 3). In this study, no fixed minimum coverage threshold was used for determination of a minor allele. Minor alleles were called

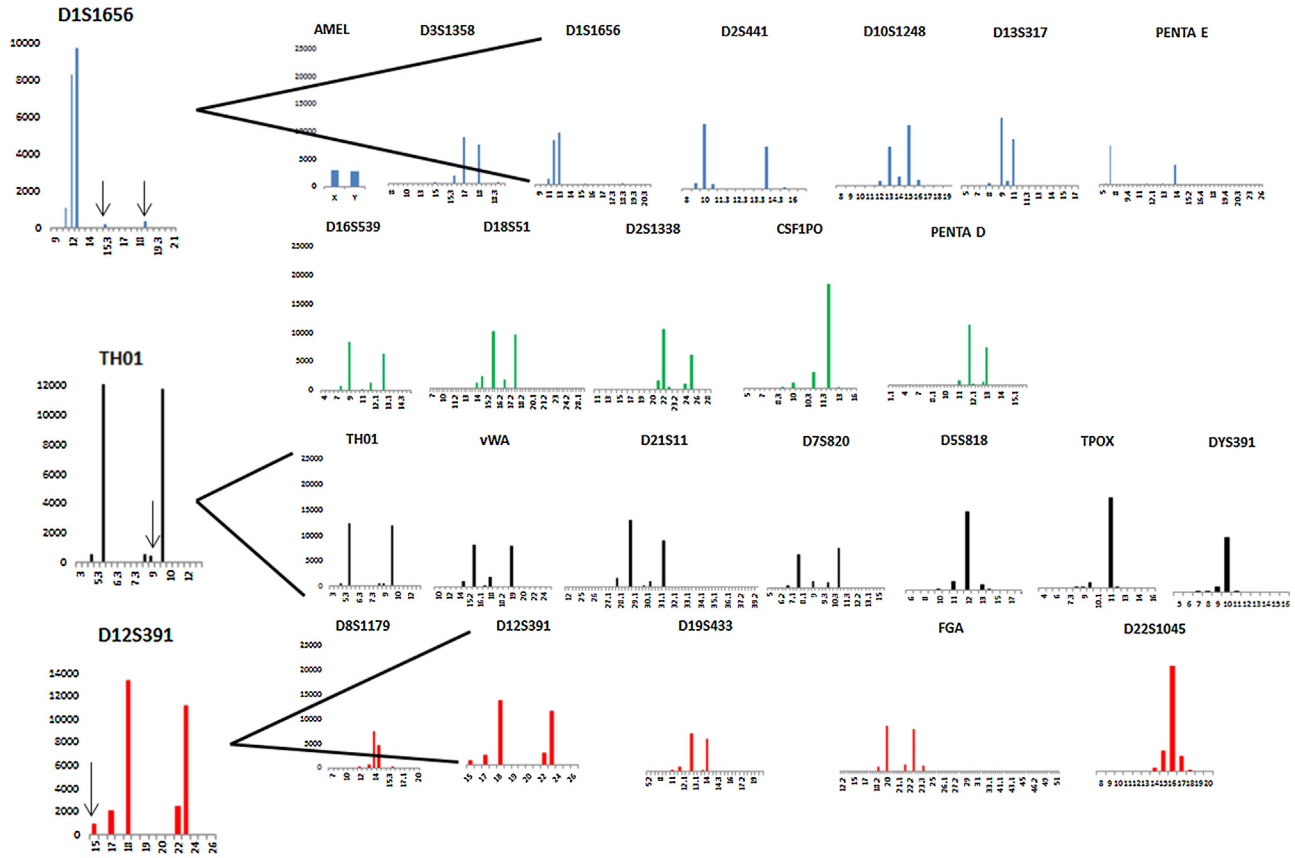

Fig. 2. A histogram portrayal of the DoC by locus of 19:1 mixture of two individuals (no. 10/11). The arrows indicate alleles from the minor contributor (no. 11): alleles 15.3 and 18.3 at D1S1656, allele 9 at TH01, allele 15 at D12S391. *X* axis are the alleles for a given locus, *Y* axis is the coverage for a given allele.

if their coverage was at least twice the signal within a locus not attributable to stutter. Noise varies by locus. More work needs to be done to determine the noise levels of each locus and, thus, better define dynamic thresholds necessary for interpretation. However, for the purposes herein the simple dynamic threshold described above was used operationally to determine if an allele was detectable. The results indicated that the prototype PowerSeq™ Auto System and the MiSeq system could detect partial minor STR profiles up to a 19:1 mixture.

3.4. Mock forensic casework study

Three blood stains, three saliva stains, one semen stain and two bone samples were examined. Eight samples yielded clear and full profiles in both MPS and CE. For one bone sample (2A1), MPS detected three more alleles than CE: i.e., allele 8 at the Penta D locus, allele 12 at the Penta E locus, and allele 26 at the FGA locus. The results showed that the system was promising for the detection of single source forensic samples. To simulate mixtures that may originate from a crime scene, two blood/saliva mixtures, one saliva/semen mixture, and one blood/semen mixture were prepared as stains in three different volume ratios: 5:1, 1:1 and 1:5. Saliva/semen and blood/semen mixtures were extracted using the differential extraction method. In the blood/saliva #1 mixture, MPS could detect two more alleles (allele 16 at the D12S391 locus and allele 9 at the Penta D locus) in the 5:1 mixture, one more allele (allele 16 at the D12S391 locus) at the 1:1 mixture, and the same profile at the 1:5 mixture compared with the CE-based method (Supplemental Table 4). In the blood/saliva #2 mixture, MPS detected eight more alleles but failed to detect allele 11 at the TPOX locus compared with CE at the 5:1 mixture (Supplemental Table 5). MPS detected two more alleles (18.3 at the D12S391 locus, 11 at the D22S1045 locus) at 1:1 and generated the same profile as CE at the 1:5 mixture. For the saliva/semen mixture, MPS and CE generated clear and full profiles of the male contributor in the sperm fraction in all three mixture ratios, except MPS detected one allele from the female contributor (allele 15 at the Penta E locus) at the 1:5 mixture (Supplemental Table 6). The epithelial fraction displayed mixtures of the male and female contributors. Five and two more alleles were detected by MPS at the 5:1 and 1:5 mixtures, respectively, and the two systems generated the same profile (i.e., alleles) at the 1:1 mixture (Supplemental Table 7). For the blood/semen mixture, the sperm fraction showed a full profile of the male contributor with both platforms in all three mixture ratios (Supplemental Table 8). The epithelial fraction showed different results between MPS and CE. MPS detected seven and two more alleles at the 5:1 and 1:5 mixtures, respectively (Supplemental Table 9). MPS detected one more allele but failed to detect allele 23 at the D12S391 locus and allele 15 at D1S1656, compared with CE at the 1:1 mixture. In summary, MPS could detect the same or more alleles in the mixture samples compared with the CE-based method. However, a number of alleles of the other contributor were located within stutter positions (two repeats downstream or one repeat upstream), and were not considered in this analysis. Population studies are needed to determine the stutter ratio for each locus in this multiplex STR system to facilitate interpretation of mixture samples.

One advantage of MPS is that intra-repeat variation within STRs can be detected. Among the four mixtures described above, the same nominal alleles (based on length) could be differentiated by SNPs at four loci: D8S1179, D2S441, D2S1338, and D21S11 (Supplemental Table 10). Brinkmann et al. [32], Heinrich et al.

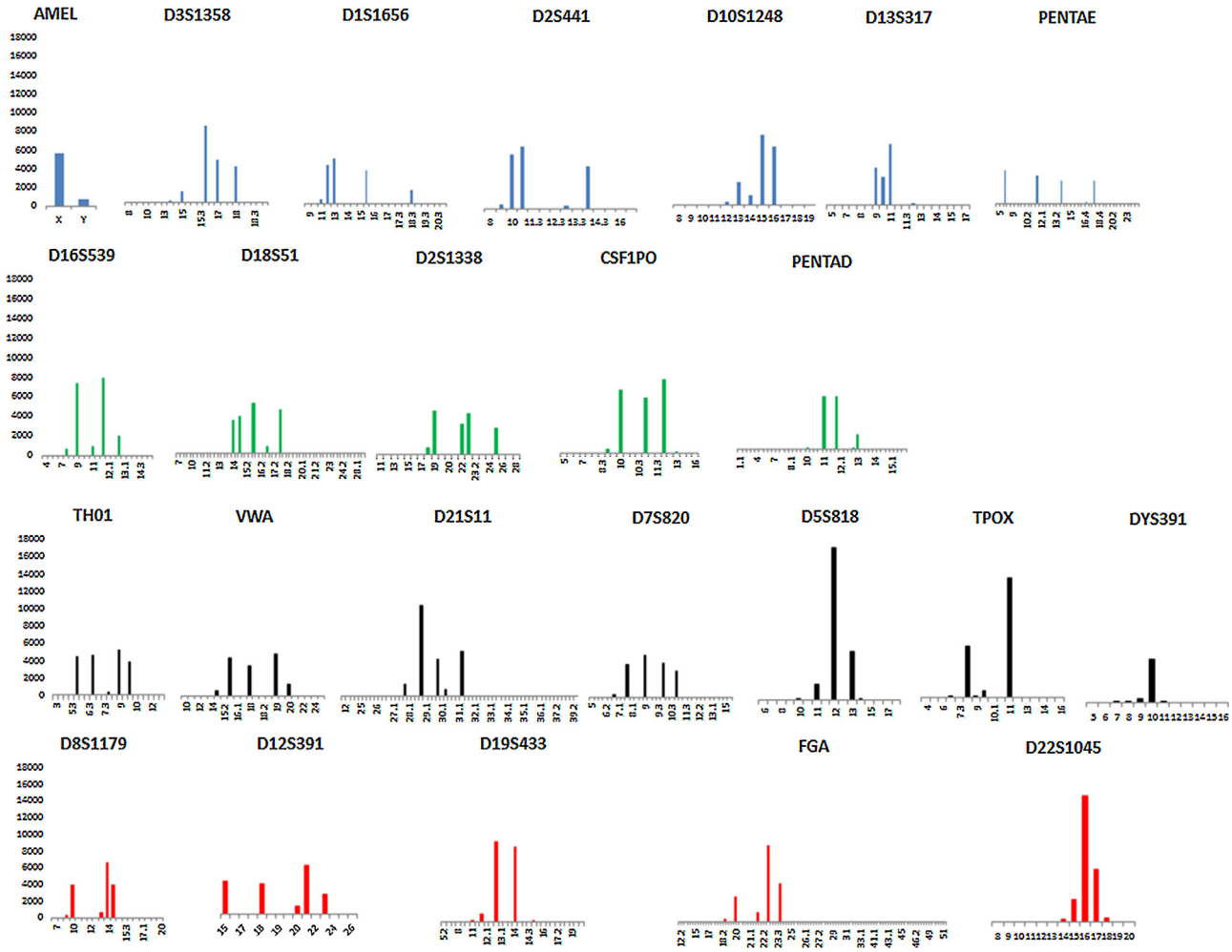

Fig. 3. A histogram portrayal of the DoC by locus of 1:1 mixture of two individuals (no. 10/11). *X* axis are the alleles for a given locus, *Y* axis is the coverage for a given allele.

[33], and Oberacher et al. [34], described these same sequences. Supplemental Table 10 shows three scenarios of intra-repeat variations at the D8S1179, D2S441, D2S1338, and D21S11 loci. The first scenario determined that a homozygous individual at the D8S1179 locus was actually a heterozygote. In the blood/saliva #2 mixture (1:1) at the D8S1179 locus, there were two types of allele 13: $(TCTA)_2TCTG(TCTA)_{10}$ and $(TCTA)_{13}$. This variation can increase discrimination power and could be useful in some kinship tests.

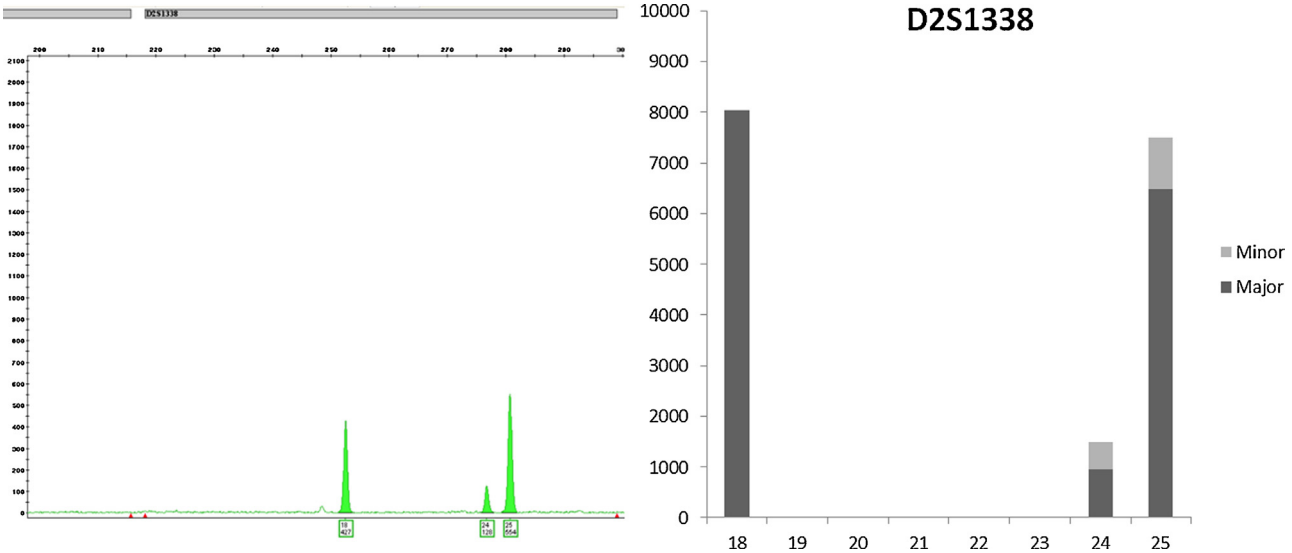

Fig. 4. The CE (left panel) and MPS (right panel) profiles of D2S1338 locus in epithelial fraction of the blood/semen mixture (5:1). MPS data revealed that alleles 24 and 25 had two different motif compositions.

The second scenario of intra-repeat variation made it possible to distinguish a minor contributor from stutter. The SNPs distinguished contribution of two types of 10 repeat length read alleles at the D2S441 locus (blood/saliva #2 mixture, 1:1). One species was stutter from allele 11 (stutter ratio = 11%) and had the repeat composition of $(TCTA)_{10}$. The other species was an allele derived from the minor contributor and had a variant at the second TCTA repeat (A/G). The third scenario is two individuals with the same nominal alleles at the D2S1338 locus were distinguished. Two different motif compositions were found for allele 25 at the D2S1338 locus in the blood/semen mixture (5:1, epithelial fraction): $(TGCC)_8(TTCC)_{14}GTCC(TTCC)_2$ with coverage of 6489*X* and $(TGCC)_7(TTCC)_{15}GTCC(TTCC)_2$ with coverage of 1010*X*. The results indicated the major contributor and minor contributor shared the same allele 25 (Supplemental Table 10, Fig. 4). Allele 24 also had two types of motif compositions. Part of the contribution to allele 24 was stutter (956*X*) from allele 25 of the major contributor and had the repeat composition of $(TGCC)_8(TTCC)_{13}GTCC(TTCC)_2$. The rest of the signal at allele 24 was from the minor contributor (536*X*) with the repeat of $(TGCC)_7(TTCC)_{14}GTCC(TTCC)_2$. Interpretation of the mixture is more difficult by CE and multiple explanations can describe the mixture when single-source data were not factored into the analysis. However, with MPS data, it was evident that the genotype of the major contributor was 18, 25 (8040*X*, 6489*X*) and the minor was 24, 25. These observations demonstrate that intra-repeat variation within STR alleles may facilitate mixture deconvolution, and future studies will focus on characterization of slippage events.

4. Conclusion

In this study, the prototype PowerSeq™ Auto System (23 STR loci + Amelogenin) was evaluated. This system is the MPS version of PowerPlex™ Fusion kit. The PCR sensitivity study demonstrated that single source complete profiles could be obtained using as little as 62 pg of input DNA, although ACRs were more imbalance than when the optimum input DNA of 500 pg was used. The profiles generated were reproducible and consistent among multiple typing replicates for a given individual. In the mixture study, this system could detect partial STR profiles of the minor contributor up to a 19:1 mixture. Different types of single source samples (blood, saliva and semen) and mixture samples were assayed using this system and obtained full or partial profiles. These results support that this STR multiplex system and the MiSeq system can provide results that are concordant and comparable in performance with current CE-based methods for PCR sensitivity, reproducibility, mixture interpretation, and typing samples similar to those of forensic samples. Furthermore, this MPS-based system was able to enhance mixture interpretation with the detection of intra-repeat variations within STR alleles. Future studies will focus on generating population data to describe the additional variation of SNPs or RMVs within STR alleles and determine stutter ratios. In addition, higher throughput studies will be performed to assess the feasibility of sequencing 96 samples simultaneously using the TruSeq HT kit.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2015.07.015.

References

[1] P.J. Collins, L.K. Hennessy, C.S. Leibelt, R.K. Roby, D.J. Reeder, P.A. Foxall, Developmental validation of a single-tube amplification of the 13CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR identifiler PCR amplification kit, J. Forensic Sci. 49 (2004) 1265–1277.
[2] B.E. Krenke, A. Tereba, S.J. Anderson, E. Buel, S. Culhane, C.J. Finis, et al., Validation of a 16-locus fluorescent multiplex system, J. Forensic Sci. 47 (2002) 773–785.
[3] P. Martin, L.F. de Simon, G. Luque, M.J. Farfan, A. Alonso, Improving DNA data exchange: validation studies on a single 6 dye STR kit with 24 loci, Forensic Sci. Int. Genet. 13 (2014) 68–78.
[4] K. Oostdik, K. Lenz, J. Nye, K. Schelling, D. Yet, S. Bruski, et al., Developmental validation of the PowerPlex((*R*)) fusion system for analysis of casework and reference samples: a 24-locus multiplex for new database standards, Forensic Sci. Int. Genet. 12 (2014) 69–76.
[5] K. Lazaruk, P.S. Walsh, F. Oaks, D. Gilbert, B.B. Rosenblum, S. Menchen, et al., Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument, Electrophoresis 19 (1998) 86–93.
[6] J.M. Butler, E. Buel, F. Crivellente, B.R. McCord, Forensic DNA typing by capillary electrophoresis using the ABI prism 310 and 3100 genetic analyzers for STR analysis, Electrophoresis 25 (2004) 1397–1412.
[7] W. Parson, C. Strobl, G. Huber, B. Zimmermann, S.M. Gomes, L. Souto, et al., Evaluation of next generation mtGenome sequencing using the ion torrent personal genome machine (PGM), Forensic Sci. Int. Genet. 7 (2013) 543–549.
[8] J.L. King, B.L. LaRue, N.M. Novroski, M. Stoljarova, S.B. Seo, X. Zeng, et al., High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq, Forensic Sci. Int. Genet. 12 (2014) 128–135.
[9] X. Zeng, J.L. King, M. Stoljarova, D.H. Warshauer, B.L. LaRue, A. Sajantila, et al., High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing, Forensic Sci. Int. Genet. 16 (2015) 38–47.
[10] D.M. Bornman, M.E. Hester, J.M. Schuetter, M.D. Kasoji, A. Minard-Smith, C.A. Barden, et al., Short-read, high-throughput sequencing technology for STR genotyping, BioTech. Rapid Dispatches (2012) 1–6.
[11] S.L. Fordyce, M.C. Ávila-Arcos, E. Rockenbauer, C. Børsting, R. Frank-Hansen, F.T. Petersen, et al., High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform, Biotechniques 51 (2011) 127–133.
[12] M. Scheible, O. Loreille, R. Just, J. Irwin, Short tandem repeat sequencing on the 454 platforms, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e357–e358.
[13] M. Scheible, O. Loreille, R. Just, J. Irwin, Short tandem repeat typing on the 454 platform: strategies and consideration for targeted sequencing of common forensic markers, Forensic Sci. Int. Genet. 12 (2014) 107–119.
[14] D.H. Warshauer, D. Lin, K. Hari, R. Jain, C. Davis, B.L. Larue, et al., STRait razor: a length-based forensic STR allele-calling tool for use with second generation sequencing data, Forensic Sci. Int. Genet. 7 (2013) 409–417.
[15] D.H. Warshauer, J.L. King, B. Budowle, STRait razor v2.0: the improved STR Allele identification tool–razor, Forensic Sci. Int. Genet. 14 (2015) 182–186.
[16] C. Van Neste, M. Vandewoestyne, W. Van Criekinge, D. Deforce, F. Van Nieuwerburgh, My-forensic-loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing, Forensic Sci. Int. Genet. 9 (2014) 1–8.
[17] C. Van Neste, Y. Gansemans, D. De Coninck, D. Van Hoofstat, W. Van Criekinge, D. Deforce, et al., Forensic massively parallel sequencing data analysis tool: implementation of MyFLq as a standalone web- and illumina BaseSpace (®)-application, Forensic Sci. Int. Genet. 15 (2015) 2–7.
[18] S.L. Fordyce, H.S. Mogensen, C. Børsting, R.E. Lagacé, C.W. Chang, N. Rajagopalan, et al., Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™, Forensic Sci. Int. Genet. 14 (2015) 132–140.
[19] C. Gelardi, E. Rockenbauer, S. Dalsgaard, C. Børsting, N. Morling, Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles, Forensic Sci. Int. Genet. 12 (2014) 38–41.
[20] E. Rockenbauer, S. Hansen, M.R. Andersen, E.F. Heerup, C. Børsting, S.L. Fordyce, et al., Characterization of sequence variations in the D21S11 locus in Danes, Somalis and Greenlanders by second generation sequencing, Forensic Sci. Int. Genet. Suppl. Ser. 4 (2013) e302–e303.
[21] QIAamp® DNA Mini and Blood Mini Handbook, 2012, http://www.qiagen.com/resources/resourcedetail?id=67893a91-946f-49b5-8033-394fa5d752ea&lang=en.
[22] A. Giusti, M. Baird, S. Pasquale, I. Balazs, J. Glassberg, Application of deoxyribonucleic acid (DNA) polymorphisms to the analysis of DNA recovered from sperm, J. Forensic Sci. 31 (1986) 409–417.
[23] P.L. Marshall, M. Stoljarova, S.E. Schmedes, J.L. King, B. Budowle, A high volume extraction and purification method for recovering DNA from human bone, Forensic Sci. Int. Genet. 12 (2014) 155–160.
[24] User guide: Quantifiler® Human and Y Human Male DNA Quantification Kits, 2014, http://tools.lifetechnologies.com/content/sfs/manuals/cms_041395 pdf.
[25] User guide: PowerSeq™ Auto system, 2014, Promega provided.
[26] MinElute® Handbook, 2008. h3d0ea6c2287d&lang=en.
[27] B.R. Qubit® dsDNA , Assay Kits,. https://tools. lifetechnologies. com/content/sfs/manuals/mp 32,850 pdf. 2011
[28] TruSeq® DNA PCR-Free Sample Preparation Guide, 2013. http://supportres.illumina.com/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqdnapcrfree/truseq-dna-pcr-free-sample-prep-guide-15036187-b.pdf.

[29] Qubit™ dsDNA HS. Assay Kits, 2010. http://tools.lifetechnologies.com/content/sfs/manuals/mp32851 pdf.
[30] MiSeq® System User Guide, 2013. http://supportres.illumina.com/documents/documentation/system_documentation/miseq/miseq_system_user_guide_15027617_l.pdf.
[31] Technical Manual PowerPlex® Fusion System, 2014. http://www.promega.com/~/media/files/resources/protocols/technical%20manuals/101/powerplex%20fusion%20system%20protocol.pdf.
[32] B. Brinkmann, A. Sajantila, H.W. Goedde, H. Matsumoto, K. Nishi, P. Wiegand, Population genetic comparisons among eight populations using allele frequency and sequence data from three microsatellite loci, Eur. J. Hum. Genet. 4 (1996) 175–182.
[33] M. Heinrich, H. Felske-Zech, B. Brinkmann, C. Hohoff, Characterisation of variant alleles in the STR systems D2S1338, D3S1358 and D19S433, Int. J. Legal Med. 119 (2005) 310–313.
[34] H. Oberacher, F. Pitterl, G. Huber, H. Niederstätter, M. Steinlechner, W. Parson, Increased forensic efficiency of DNA fingerprints through simultaneous resolution of length and nucleotide variability by high-performance mass spectrometry, Hum. Mutat. 29 (2008) 427–432.