FORENSIC SCIENCE JOURNAL SINCE 2002

Forensic Science Journal
2021;20(1):13-20
fsjournal.cpu.edu.tw
DOI:10.6593/FSJ.202112_20(1).0002

Using Next-Generation Sequencing to Analyze Mixed Forensic Specimens

Yung-Chun Lai [1], M.S. ; Man-Chen Chang [1], M.S. ; Kuan-Cheng Peng [1], M.S. ; Chun-Yen Lin [1*], Ph.D.

[1] Institute of Forensic Medicine, Ministry of Justice, No.123, Min'an St., Zhonghe Dist., New Taipei City 235016, Taiwan (R.O.C.)

Received: June 09, 2021; Accepted: September 23, 2021.

Abstract

Capillary electrophoresis (CE) has routinely been used in forensic DNA laboratories to analyze STRs and DNA sequences. However, this technique has limited ability to discriminate mixed forensic specimens. Most such specimens have very little DNA, and their mixing ratio is usually unknown. For this reason, it remains difficult to identify trace amounts of DNA in mixed samples using CE. Next-generation sequencing (NGS) can overcome the disparity problem of the DNA mixing ratio in samples by increasing the sequencing depth. In this study, 15 forensic specimens were analyzed for STR and Y-STR DNA profiles using NGS, and the results were then compared with CE data to evaluate its correctness and reliability. First, 15 samples were collected from forensic autopsy cases and were analyzed via CE. The remaining DNA samples were analyzed using NGS technology. Each sample had a different amount of DNA in the NGS analysis, depending on the condition of the forensic case. The detection rates of the STR DNA allele of NGS in 13 cases were all greater than CE, with the exception of two samples (AS1 and NA1) with low amounts of DNA (<1 ng) for NGS analysis. Furthermore, among the 15 specimens, four specimens (NA1, NA2, VS1, and VS10) were also examined for human mitochondrial HV1 and HV2 sequences. Using NGS to analyze the mitochondrial DNA of mixed forensic samples can evade the challenge of quantifying the DNA mixing ratio using CE. This study suggests that the depth of NGS sequencing must be increased, supplemented by the proportion of SNP or mitochondrial bases, to effectively solve the mixed-profile problem.

Keywords: forensic science, next-generation sequencing (NGS), capillary electrophoresis (CE), mixed forensic specimens, human identification, mitochondrial DNA

Introduction

The Illumina MiSeq instrument has been identified as one of the most popular sequencing platforms in use at present. It has the capability to provide multiple-strand DNA sequence information, which can be used to identify the DNA source [1,2]. NGS technology is suitable for DNA mixture analysis [3]. Traditional CE methods cannot analyze differences in STR DNA with the same length. If the length of the sequence is the same, it will be identified as belonging to the same pattern. In addition, the CE method cannot clearly distinguish mixed mitochondrial DNA sequences. This problem of DNA mixture is currently a challenge for forensic DNA

*Corresponding author: Chun-Yen Lin, Institute of Forensic Medicine, Ministry of Justice, No.123, Min'an St., Zhonghe Dist., New Taipei City 235016, Taiwan (R.O.C.)
E-mail: apollo591@gmail.com

identification. Because the results of CE analysis cannot be quantified, samples with lower DNA concentrations are difficult to detect. However, NGS technology can be used to analyze collections with multiple sources of DNA and identify the sources with lower DNA concentrations [4,5]. The results can be used to distinguish the mixture of DNA samples involving two or three people [6]. This study used the Illumina platform (Illumina MiSeq FGx, USA) to analyze the STR and mtDNA of 15 two-person mixture samples [7]. The results of the STR and mtDNA were compared with those of the CE and NGS.

Materials and Methods

Forensic Samples

In total, 15 samples were collected: ten vaginal swabs (VS1-VS10), two nail exogenous samples (NA1 and NA2), one anal swab (AS1), one bone sample (B1), and one tooth sample (T1). These 15 samples were collected from forensic autopsy cases. All samples were first analyzed via CE. The remaining DNA samples were analyzed using NGS. Each sample had a different amount of DNA for NGS analysis depending on its condition.

DNA Extraction and Quantification

DNA was extracted from each of forensic samples by using the PureLink™ Genomic DNA Kit (Thermo Fisher Scientific, USA) [8]. The DNA template was quantified using the Quantifiler® Human DNA Quantification Kit (Thermo Fisher Scientific, USA) [9]. The DNA products were analyzed using the 7500 Real-Time PCR System (Thermo Fisher Scientific, USA) and HID Real-Time PCR Analysis v1.3 software (Thermo Fisher Scientific, USA) for quantification analysis.

STR Analysis with CE

For each sample, 1 ng DNA was amplified with the AmpFℓSTR Identifiler Kit (Thermo Fisher Scientific, USA) [10] and the AmpFℓSTR Yfiler Kit (Thermo Fisher Scientific, USA) [11]. The PCR products were analyzed using the Applied Biosystems 3500xL Genetic Analyzer (Thermo Fisher Scientific, USA) [12] and the GeneMapper ID-X v1.5 software (Thermo Fisher Scientific, USA) for the STR profiles.

MtDNA Sequencing with CE

For each sample, the HV1 and HV2 regions were amplified with primers L15997(5'-CACCA TTAGCACCCAAAGCT-3')/H16401(5'-TGATT TCACGGAGGATGGTG-3') and L29(5'-GGTCT ATCACCCTATTAACCAC-3')/H408(5'-CTGTT AAAAGTGCATACCGCCA-3'), respectively, and 1 ng template DNA was input for each region. The PCR products were purified with a DNA Clean/Extraction Kit (GeneMark, Taiwan) and subjected to cycle sequencing with the above-mentioned primers. The PCR products were sequenced in both directions with BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA). The results were analyzed using SeqScape v2.6 software (Thermo Fisher Scientific, USA), and these were compared with the Revised Cambridge Reference Sequence (rCRS, NC-012920.1 in GenBank).

STR Analysis with NGS

For each sample, 1 ng DNA was input for the NGS analysis. The library preparation steps for the NGS experiment include amplification, tagging targets, enriching targets, purifying libraries, and normalizing libraries. The libraries were established using ForenSeq™ DNA Signature Prep Kit (Verogen, USA) [13] and sequenced via the Illumina MiSeq Platform (Illumina, USA), following the manufacturer's suggestions. The results were analyzed with ForenSeq™ Universal Analysis Software v2.0 software (Verogen, USA) [14]. In principle, if the amount of sample DNA is large, 1ng DNA is used for NGS. But if the amount of sample DNA is less than 1ng, only the remaining can be provided for NGS.

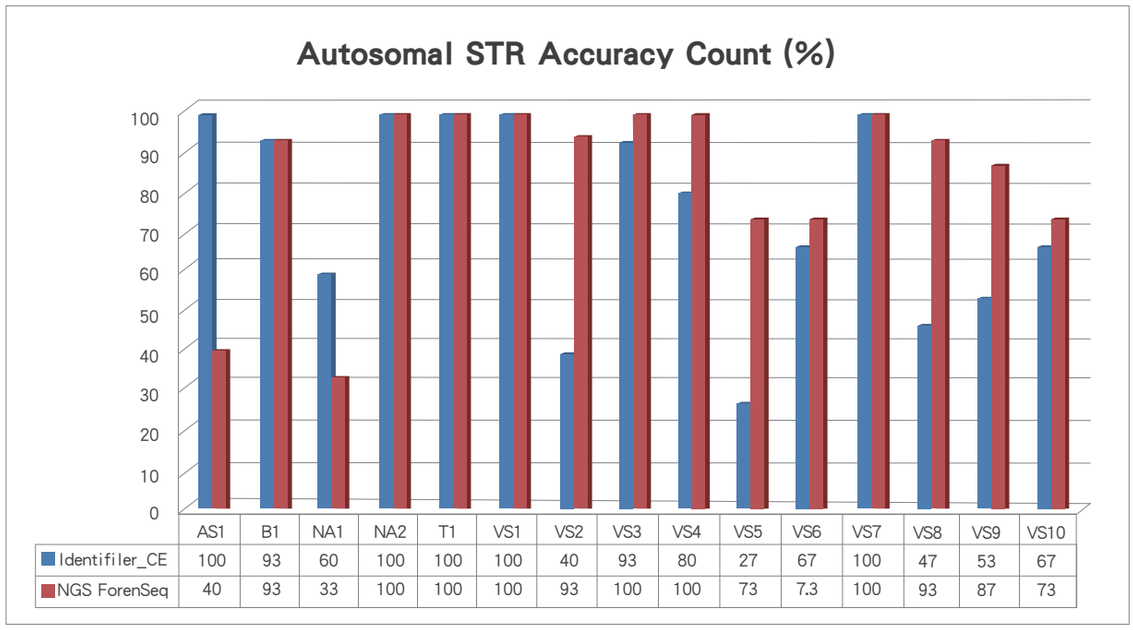
MtDNA Sequencing with NGS

The whole mtDNA genome was amplified using long-range PCR. The PCR products were produced by Nextera DNA Flex Library Prep Kit (Illumina, USA) for library preparation (for each sample, 100 ng template DNA). The steps for library preparation included tagmenting the DNA, post-tagmentation, PCR, and library cleanup. Then, the DNA quality was checked and sequenced using the Illumina MiSeq Platform [15]. The results were analyzed using ForenSeq™ Universal Analysis Software v2.0 software (Illumina, USA) and compared with the rCRS (NC-012920.1 in GenBank).

Results and Discussion

Analyzing the STR Results of the Mixed Specimens

In two (AS1 and NA1) out of the 15 samples, the quantity of DNA was too low (less than the threshold); thus, we could not obtain effective DNA signals. In addition, there were five specimens, that is, B1, NA2, T1, VS1, and VS7, in which the accuracy counts for CE and NGS were the same. The autosomal STR accuracy count is the number of loci in the specimen divided by the total loci. The average NGS accuracy count of the remaining eight specimens was 86.5%, and the average accuracy count for CE was 59.25%, showing a 27.25% difference. The accuracy count of NGS is indeed better than that of CE (Fig. 1). For the NA1 sample, 0.3 ng DNA was used for CE analysis, but only 0.05 ng DNA was left for NGS analysis. There was a sixfold difference between the two in terms of the quantity of DNA. For the AS1 specimens, 1 ng DNA was used for CE analysis, but only 0.1 ng DNA was left for NGS analysis. There was a tenfold difference in the quantity of DNA between the two. In the above two samples, the difference in DNA content has been deemed very large, so that the effect of CE is better than that of NGS. In addition, NGS technology can also be used to observe whether the amount of input DNA is too little to detect all profiles from the number of reads (intensity). The results of this study found that the AS1, NA1, VS2, VS3, and VS5 specimens each had less than 200,000 reads. In the analysis of a two-person mixed sample, the minor contributor usually has much lower amount of DNA within the specimen, which is sometimes even hard to detect. To effectively solve this problem, we suggest that the sequencing depth [16] should be increased. The sequencing depth, which is defined as the average number of times that each base is measured in the sequence, is a key consideration that directly affects the quality and quantity of the sequencing data.



	AS1	B1	NA1	NA2	T1	VS1	VS2	VS3	VS4	VS5	VS6	VS7	VS8	VS9	VS10
Identifiler_CE	100	93	60	100	100	100	40	93	80	27	67	100	47	53	67
NGS ForenSeq	40	93	33	100	100	100	93	100	100	73	7.3	100	93	87	73

Fig. 1 Comparison of STR results from CE and NGS.

Evaluating the Ratio of Mixed Specimens

The absence of a contributor's specimen is quite common, especially the specimen of a suspect in a criminal case. Thus, it is difficult to distinguish DNA alleles and to determine the ratio between victim and suspect DNA. In our study, both victim and suspect specimens are provided for only three cases (VS1, VS2, and AS1), from a total of 15 mixed forensic specimens. The results for the sexual sample VS1 indicate that both CE and NGS could detect the complete 16 loci (Table 1).

Table 1 The STR results for samples (VS1) using NGS.

Sample / STR Loci	Contributor 1 (Victim)	Contributor 2 (Suspect)	VS1 Vaginal swab (DNA input; reads)	
			Identifiler CE (1ng)	NGS ForenSeq (0.5 ng;1,103,826)
D8S1179	14,16	10,11	10,11,14,16	10,11,14,16
D21S11	30,32.2	30,31	30,31,32.2	30,31,32.2
D7S820	10,12	8,11	8,10,11,12	8,10,11,12
CSF1PO	10,13	11	10,11,13	10,11,13
D3S1358	15,17	15,16	15,16,17	15,16,17
TH01	7,9	9	7,9	7,9
D13S317	8,11	8,10	8,10,11	8,10,11
D16S539	9,11	12	9,11,12	9,11,12
D2S1338	19,20	18,19	18,19,20	18,19,20
D19S433	13,15	15,15.2	13,15,15.2	13,15,15.2
vWA	18,19	14,16	14,16,18,19	14,16,18,19
TPOX	8,11	8	8,11	8,11
D18S51	12,15	13,16	12,13,15,16	12,13,15,16
D5S818	10,11	9,10	9,10,11	9,10,11
FGA	22,23	21.2,24	21.2,22,23,24	21.2,22,23,24
Amelogenin	X, X	X, Y	X, Y	X, Y

We calculated the percentage of the samples as follows: the number of reads for individual specimens was divided by the number of the total reads. The total reads are the sum of the number of reads of the victim and the suspect. Therefore, the percentages for victim and suspect for the VS1 sample were 45.1% and 54.9%, respectively. Analysis of the VS2 results showed that only six complete loci were detected via CE, and the other nine loci were incomplete. Meanwhile, 14 complete loci were detected using NGS, and only one (vWA) was incomplete, as shown in Table 2. The ratio between victim and suspect as calculated by the reads obtained from the NGS were 38.1% and 61.9%, respectively. The DNA amount of the AS1 sample was too low to provide sufficient data in NGS for further interpretation. To avoid the problem caused by DNA quantity next time, the method to increase the sequence depth mentioned above should be applied.

Table 2 The STR results for the sample (VS2) using NGS.

Sample / STR Loci	Contributor 1 (Victim)	Contributor 2 (Suspect)	VS2 Vaginal swab (DNA input; reads)	
			Identifiler CE (1ng)	NGS ForenSeq (0.5 ng;125,849)
D8S1179	15	14,16	14,15,16	14,15,16
D21S11	29,32.2	29,32.2	29,32.2	29.32.2
D7S820	9,11	11	11	9,11
CSF1PO	10,12	11,13	11,13	10,11,12,13
D3S1358	14,16	15,17	14,15,16,17	14,15,16,17
TH01	7,9	8,9	8,9	7,8,9
D13S317	8,11	8,11	8,11	8,11
D16S539	9,13	9,12	9,12	9,12,13
D2S1338	17,19	25,26	25,26	17,19,25,26
D19S433	13	13,15	13,15	13,15
vWA	17,18	14	14	14
TPOX	8	8,9	8,9	8,9
D18S51	13,14	13,18	13,18	13,14,18
D5S818	11,13	11,12	11,12	11,12,13
FGA	21,22	18,21	18,21	18,21,22
Amelogenin	X, X	X, Y	X, Y	X, Y

Note: The gray background indicated that the complete mixed alleles cannot be obtained using CE or NGS.

Differentiating the Sequences of the Same Allele Type from Different Contributors

When using CE as an analytical method, if the same allele is observed in STR, no difference in sequence can be distinguished. However, the NGS can distinguish the sequence variation, which then helps in identifying the different contributors. For example, the allele of the D3S1358 locus was found to be the same allele type 15, but its sequence was different in the VS1 specimen with NGS. To take another example, the allele of the D21S11 locus was found to be 29, but its sequence was different in the VS2 specimen (Table 3). Thus, if similar situations are encountered in the future, to overcome the inability of CE in differentiating the same allele type with sequence variation, the NGS may help identify the mixture that comes from different contributors.

Table 3 Using NGS technology in analyzing the same allele type of mixed specimens with different base sequences.

Sample	Loci	Allele	Different base number	Sequence
VS1	D3S1358	15	1	TCTATCTGTCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA
	D3S1358	15		TCTATCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA
VS2	D21S11	29	13	TCTATCTATCTATCTATCTGTCTGTCTGTCTGTCTGTCTGTCTATCTATCTATATCTATCTATCTATCATCTATCTATCCATATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA
	D21S11	29		TCTATCTATCTATCTATCTATCTATCTGTCTGTCTGTCTGTCTGTCTATCTATCTATATCTATCTATCTATCATCTATCTATCCATATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTA

Comparing the MtDNA Sequencing Results of CE and NGS

Heteroplasmy may occur in mitochondria. Therefore, distinguishing between heterogeneous or two-person mixed mitochondria samples has been deemed of most importance. First, it is necessary to check whether the profile of the sample in the STR is mixed. Second, it should be established whether the mitochondrial sequence of the sample exhibits base mixing and what its mixing ratio is. If the above-mentioned two points are observable, the specimen is mixed. If the STR gives a single profile, and only part of the mitochondria base has two kinds of sequences, the mitochondria in the sample are very likely to be heterogeneous.

The NA1 and VS1 specimens were processed with NGS, and it was found that their STR profiles were mixed instead of heterogeneous. CE and NGS were then used to analyze the mtDNA sequences of the exogenous nail sample (NA1). The results of CE showed only two kinds of base found at 16183 position. However, as per the NGS results, there were 11 positions with two kinds of base, such as positions 146, 150, and 152 (Fig. 2). In addition, the vaginal swab (VS1) analyzed via CE showed only two kinds of base in six positions, as compared to the 11 positions shown via NGS (Table 4). NGS can break through the dilemma that CE can only examine the presence of two kinds of bases in mixed specimens without quantifying the ratio. Therefore, NGS technology was able to significantly improve our ability to detect mixed specimens.

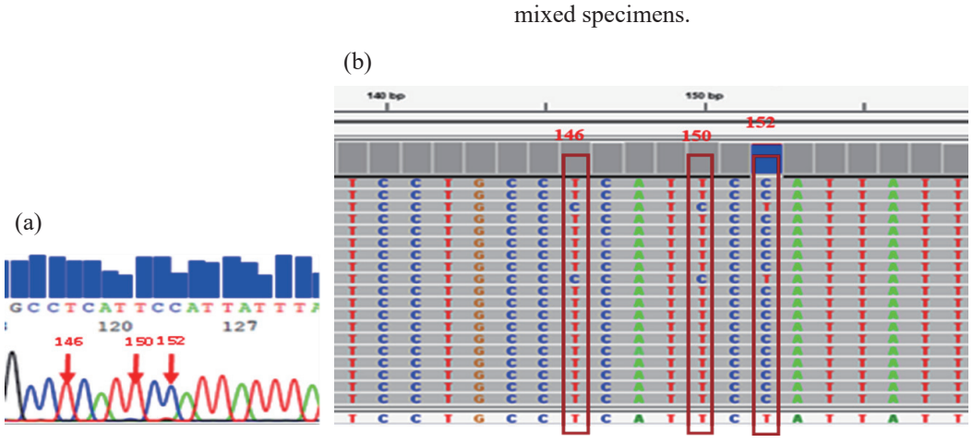


Fig. 2 MtDNA results of exogenous nail mixture sample from CE (a) and NGS (b).

Table 4 Comparison of mtDNA results of exogenous nail sample (NA1) and vaginal swab (VS1).

Position / Analysis method	146	150	152	249	16148	16183	16189	16256	16298	16327	16362
NA1-CE	T	T	C	A	T	A/C	C	T	T	C	C
NA1-NGS (ratio)	C:T (5:95)	C:T (5:95)	C:T (95:5)	Del:A (4:96)	C:T (23:77)	A:C (29:71)	C:T (65:35)	C:T (22:78)	C:T (16:84)	C:T (88:12)	C:T (89:11)
VS1-CE	C/T	C/T	C/T	Del	C/T	A/C	C/T	C	C	T	T
VS1-NGS (ratio)	C:T (74:26)	C:T (75:25)	C:T (22:76)	Del:A (74:26)	C:T (92:8)	A:C (94:6)	C:T (7:93)	C;T (92:8)	C:T (86:14)	C:T (14:84)	C:T (16:83)

Comparing the Ability of CE and NGS to Identify STR and Sequence in Mixed Samples

In the analysis of STR with CE, if the fluorescence signal is deemed too strong, the fluorescence emission spectra will partially overlap, which will then lead to misinterpretation by the instrument, resulting in false signals when it is interpreted as a fluorescence of different colors. Therefore, in a mixed sample with a large disparity, it remains impossible to increase the amount of products needed to achieve the purpose of analyzing a little DNA. However, NGS can achieve the purpose of analyzing a little DNA by increasing the depth of the sequencing. In addition to providing STR profiles of PCR products, NGS can also supply the actual STR sequences. It can use limited DNA to provide a larger amount of information for human identification. However, all of the products present their results at the same time using CE. It remains difficult to determine how many kinds of sequences there are. By contrast with CE, NGS can analyze each sequence of PCR product and establish the proportions of different sequences. This method is suitable for quantitative analysis.

Comparing the Cost of Time and Consumables

It takes several weeks to purify PCR products, add indexes, and sequence with the NGS platform. In addition, the consumables are also quite expensive. Although it requires significant time and consumables relative to the traditional approaches of Identifiler or Yfiler kits, NGS can acquire more information with the same amount of DNA.

Because the cost of NGS technology is much greater than CE, more quality control steps are necessary before performing the NGS platform. The quality control steps increase significant costs, but they can effectively advance the success rate of the experiment. This quality control step confirms whether a sufficient amount of DNA after the PCR products was purified. The PCR products can be qualified via BioDrop or Qubit and can be quantified using Bioanalyzer or Qsep.

For the analysis of the samples, it takes about 30 hours to complete the NGS experiment. It takes more time and manpower to analyze data by its means than with CE. On the whole, the cost and expense of analysis of STR DNA by NGS are higher than that by CE. However, per base, the cost of NGS sequencing is lower. Further, NGS technology can provide allele sequences that are compatible with STR DNA databases, which could provide more accurate repetitive sequence and fragment length sequence information. This can actually allow more resources to be invested in NGS to make the technology more mature and to lower its cost. NGS can overcome certain problems that CE cannot solve.

Conclusions

NGS techniques have been determined to have the ability to analyze the mixture and overcome the problems of CE. Most evidence that is collected from criminal cases exists only in trace amounts. This evidence is very valuable, so the means of obtaining the most information with a small amount of specimens becomes very essential. In a single experiment, NGS can obtain additional information, such as sequences

of alleles, compared to CE. In addition, if CE is used for mixed forensic specimens, the DNA of the minor contributor cannot be distinguished if the amount of DNA in the minor contributor is very small. To address this problem, the amount of DNA input must be increased. However, CE may cause a pull-up situation, which then results in analytical difficulties. If the NGS technology is used to increase the depth of sequencing, none of the problems mentioned above would arise. The reads of the minor contributor in the mixed sample can be increased, and the accuracy of the minor contributor profile can also be improved. The NGS technology also can sequence the HV1 and HV2 regions of mitochondrial DNA in the mixture specimens. NGS technology can confirm the composition ratio of the mixture source better than CE. In summary, this study establishes a method for analyzing mixed DNA specimens in forensic samples in hope of effectively solving problems in case there is mixture of forensic samples.

Acknowledgments

This study was supported by the grants No.107-1301-05-04-02 from Ministry of Justice, Taiwan, ROC.

References

1. Bruijns B, Tiggelaar R, Gardeniers H. Massively parallel sequencing techniques for forensics: A review. Electrophoresis 2018; 39(21):2642-54.
2. Brandhagen MD, Just RS, Irwin JA. Validation of NGS for mitochondrial DNA casework at the FBI Laboratory. Forensic Sci Int Genet 2020; 44:102151.
3. Momota F, Tsuji A, Ishiko A, Ikeda N. Examination of the usefulness of next-generation sequencing in mixed DNA samples. Leg Med (Tokyo) 2021; 51:101874.
4. Bose N, Carlberg K, Sensabaugh G, Erlich H, Calloway C. Target capture enrichment of nuclear SNP markers for massively parallel sequencing of degraded and mixed samples. Forensic Sci Int Genet 2018; 34:186-96.
5. Ragazzo M, Carboni S, Caputo V, Buttini C, Manzo L, Errichiello V, Puleri G, Giardina E. Interpreting mixture profiles: comparison between precision ID GlobalFiler™ NGS STR panel v2 and traditional methods. Genes (Basel) 2020; 11(6):591.
6. Wei JCM, Zhao Z, Li SC, Ng YK. NGS-based likelihood ratio for identifying contributors in two-and three-person DNA mixtures. Comput Biol Chem 2018; 74: 428-33.
7. Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, Walichiewicz P, Silva D, Pham N, Caves G, Bruand J. Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. Forensic Sci Int Genet 2017; 28:52-70.
8. PureLink™ Genomic DNA Kits for purification of genomic DNA user manual. MA USA: Thermo Fisher Scientific, 2007.
9. Quantifiler™ Human DNA Quantification Kit User's Manual. Waltham, MA USA: Thermo Fisher Scientific, 2006.
10. AmpFlSTR® Identifiler™ PCR Amplification Kit User's Manual. Waltham, MA USA: Thermo Fisher Scientific, 2001.
11. AmpFlSTR® Yfiler™ PCR Amplification Kit User's Manual. Waltham, MA USA: Thermo Fisher Scientific, 2004.
12. Applied Biosystems™ 3500/3500xL Genetic Analyzer User Guide. Waltham, MA USA: Thermo Fisher Scientific, 2018.
13. ForenSeq™ DNA Signature Prep Reference Guide. San Diego, CA USA: Verogen, 2015.
14. ForenSeq™ Universal Analysis Software v2.0 Reference Guide. San Diego, CA: Verogen, 2020.
15. Huang CH, Huang TY, Chung FC, Hsu CC, Peng GC, Chang YT, Lin CY. Applications of NGS on mitochondrial DNA analysis for forensic samples. Forensic Science J 2017; 16(1):51-8.
16. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet 2014; 15(2):121-32.