Streamlining degraded sample processing in forensic laboratories: Developing a decision tree for capillary electrophoresis and next-generation sequencing platforms

Rachel Houston, Natalia Czado, Damani Johnson, Lucio Avellaneda, Cesar Cantu, Sheree Hughes
Department of Forensic Science, Sam Houston State University, Huntsville TX, USA

Introduction

Short tandem repeat (STR) analysis is recognized worldwide as a standard approach for human identification (1). Profiles generated from crime scenes can be compared to a reference profile or searched against established STR databases and the strength of any match obtained can be calculated using available allelic frequencies. To increase the power of discrimination for individual identification and to facilitate international compatibility, the Federal Bureau of Investigation (FBI) expanded the Combined DNA Index System (CODIS) core STR loci from the original 13 loci to 20 loci (2). CODIS loci are listed in Table 1.

Table 1. The Combined DNA Index System (CODIS) core STR loci

D1S1656	D2S441	D2S1338	D3S1358	D5S818
D7S820	D8S1179	D10S1248	D12S391	D13S317
D16S539	D18S51	D19S433	D21S11	D22S1045
CSF1PO	FGA	TH01	TPOX	vWA

STR analysis was based on capillary electrophoresis (CE) technology. Recently, CE analysis of STR amplicons has been supplemented by next-generation sequencing (NGS) techniques. NGS allows for the differentiation of STR alleles by sequence rather than size, offering several crucial advantages over CE. Because alleles do not need to be distinguished by their size, smaller amplicons can be targeted, leading to improved analysis of degraded DNA. Access to sequence-level information also leads to increased allelic diversity, which in turn improves the power of discrimination of an STR profile obtained using NGS compared to CE (3).

A major focus for the Department of Forensic Science at Sam Houston State University lies in the identification of human remains. This procedure usually requires kinship testing between an unknown post-mortem (PM) sample and a family ante-mortem (AM) sample. The degree of relatedness will dictate how much information is required from the profiles, with first-degree (e.g., parent-child or full sibling) being the easiest to identify. Laboratories tasked with identifying human remains might first determine which samples should be processed with CE or with NGS. The decision affects the power of discrimination and may be based on factors such as sample type and quality, degree of relatedness and the type of investigation.

The overarching aim of this study is to address when NGS should be used, which NGS workflow should be used and where NGS offers improved results compared to CE. We wanted to determine whether CE or NGS better identified first-degree relationships based on allele recovery with degraded samples and statistical weighting of the STR profiles produced by each technology. The NGS STR chemistry tested was ForenSeq® MainstAY Kit; the CE STR chemistry tested was Investigator® 24plex QS Kit. A decision tree for the identification of unidentified human remains is shown in Figure 1.

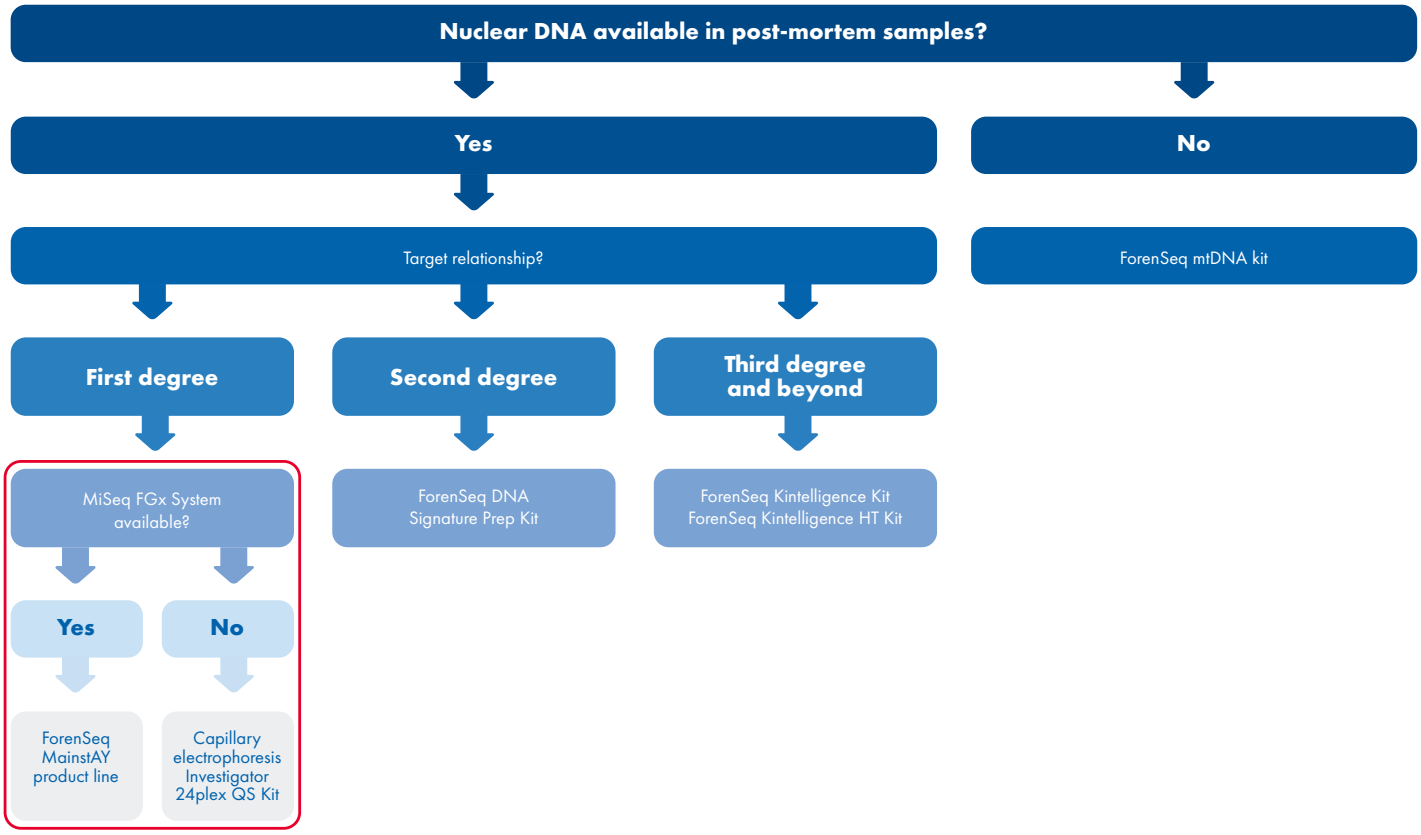

Figure 1.
Decision tree for the identification of unidentified human remains. The red box shows where labs have a choice between CE and NGS for identifying first-degree relationships. More extended relationships are not usually identified with traditional CE technology.

The ForenSeq MainstAY workflow is comprised of ForenSeq MainstAY Kit, MiSeq FGx® Sequencing System and the ForenSeq MainstAY Analysis Module in Universal Analysis Software. The workflow uses NGS technology and supports processing of up to 96 samples in one sequencing run. ForenSeq MainstAY Kit contains reagents to amplify 27 autosomal and 25 Y-STR loci as small amplicons and meets European/SWGDAM minimal Y-haplotype, Combined DNA Index System (CODIS), Interpol and ESS (European Standard Set) requirements in a single amplification. The ForenSeq MainstAY workflow has been granted National DNA Index System (NDIS) approval by the US Federal Bureau of Investigation (FBI), which allows accredited forensic DNA laboratories to process DNA casework samples and search resulting profiles against the CODIS database.

Investigator 24plex QS Kit allows amplification of the CODIS core loci, the ESS markers, plus SE33, DYS391 and Amelogenin. The kit includes a Quality Sensor to generate additional data for quality control and performance tests. Investigator 24plex QS Kit is also NDIS approved. The 22 STR markers, Amelogenin and the Quality Sensor are amplified simultaneously in a single PCR. Amplified loci then undergo capillary electrophoresis and analysis with Applied Biosystems® 3500/3500xL Genetic Analyzer and dedicated software, such as GeneMapper™ ID-X.

ForenSeq MainstAY Kit contains 21 autosomal STR loci smaller than 250 bp, whereas Investigator 24plex QS Kit has 10 autosomal STR loci smaller than 250 bp.

Materials and methods

Samples

Artificially degraded samples of burned human skeletal remains were obtained from the Southeast Texas Applied Forensic Science Facility. Femur fragments were thermally degraded at different levels. Samples were taken from 2 cadavers with 5 extraction replicates at each thermal degradation level: unburned; burned light brown; burned dark brown; burned black

DNA extraction and quantification

All bone samples (250 mg) were extracted with EZ1&2® DNA Investigator Kit using the extra-large volume protocol on EZ2® Connect Fx and eluted in 100 µL (8). All extracts were quantified using Investigator Quantiplex® Pro on Applied Biosystems 7500 Real-Time PCR instrument (9).

NGS STR: ForenSeq MainstAY workflow

Library preparation, sequencing on MiSeq FGx Sequencing System and analysis with Universal Analysis Software (UAS) were conducted according to manufacturer protocols (4–6).

CE STR: Investigator 24plex QS workflow

PCR and subsequent procedures followed instructions in the Investigator 24plex QS Handbook (7). Capillary electrophoresis was carried out on Applied Biosystems 3500xL Genetic Analyzer and amplified products were assigned to STR alleles using GeneMapper ID-X version 1.4 software.

Statistical weighting

Genotypes were called based on ground truth data (θ = 0.01). DNA analysis for Investigator 24plex QS Kit was performed with ArmedXpert™ software. A database of sequence-based allelic frequencies was used for Configuration C in MixtureAce™, corresponding to the sequences obtained from ForenSeq MainstAY Flanking Region Reports downloaded from UAS.

Results and discussion

Total allele recovery (%) vs. DNA concentration and vs. degradation index

As expected, more alleles were recovered with higher DNA concentrations with both Investigator 24plex QS Kit and ForenSeq MainstAY Kit. Allele recovery (%) values were also similar for both methods and reflected the degradation state of the samples with least recovery from burned black material and most from the lightly burned or unburned samples (Figure 2). Degradation values resulted in similar allele recovery for both methods (Figure 3).

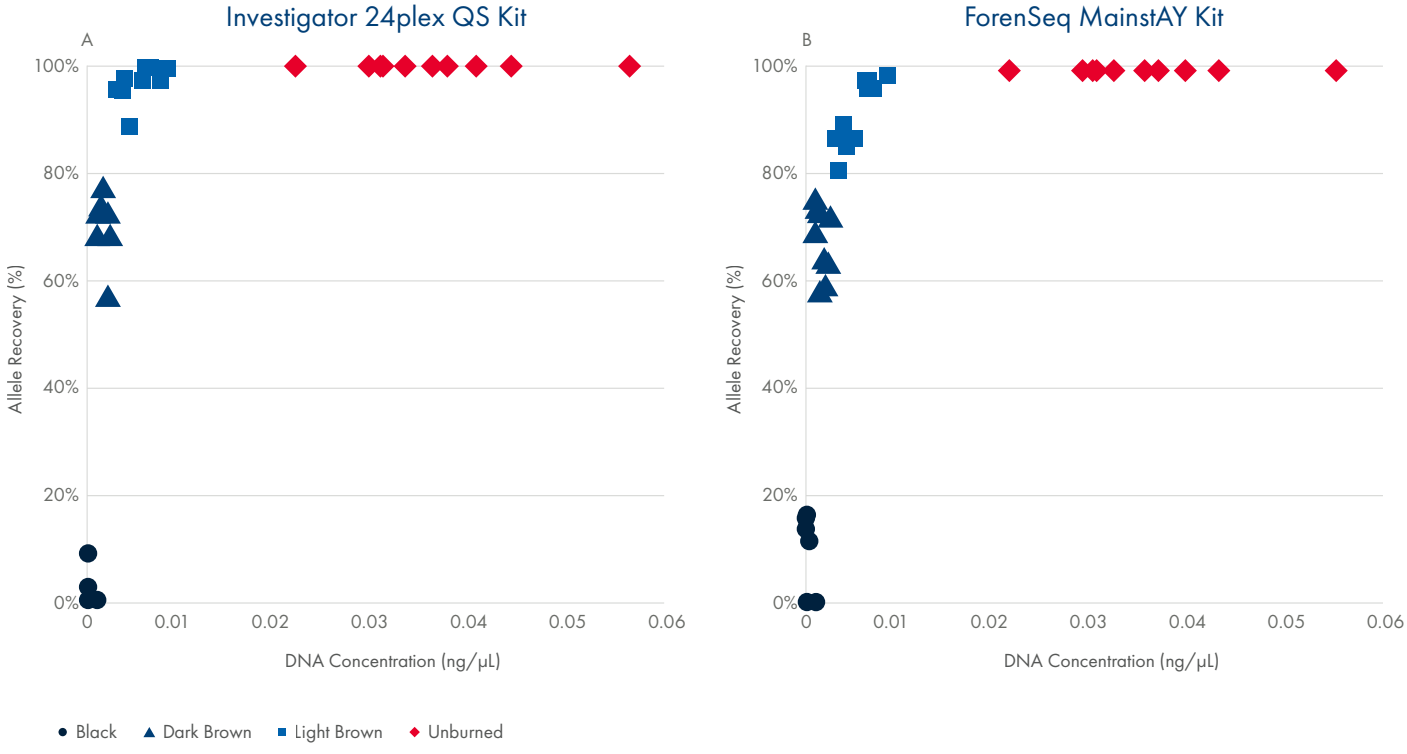

Figure 2.
Total allele recovery (%) vs. DNA concentration. (A) Allele recovery (%) vs. DNA concentration (ng/µL) with Investigator 24plex QS Kit. (B) Allele recovery (%) vs. DNA concentration (ng/µL) with ForenSeq MainstAY Kit.

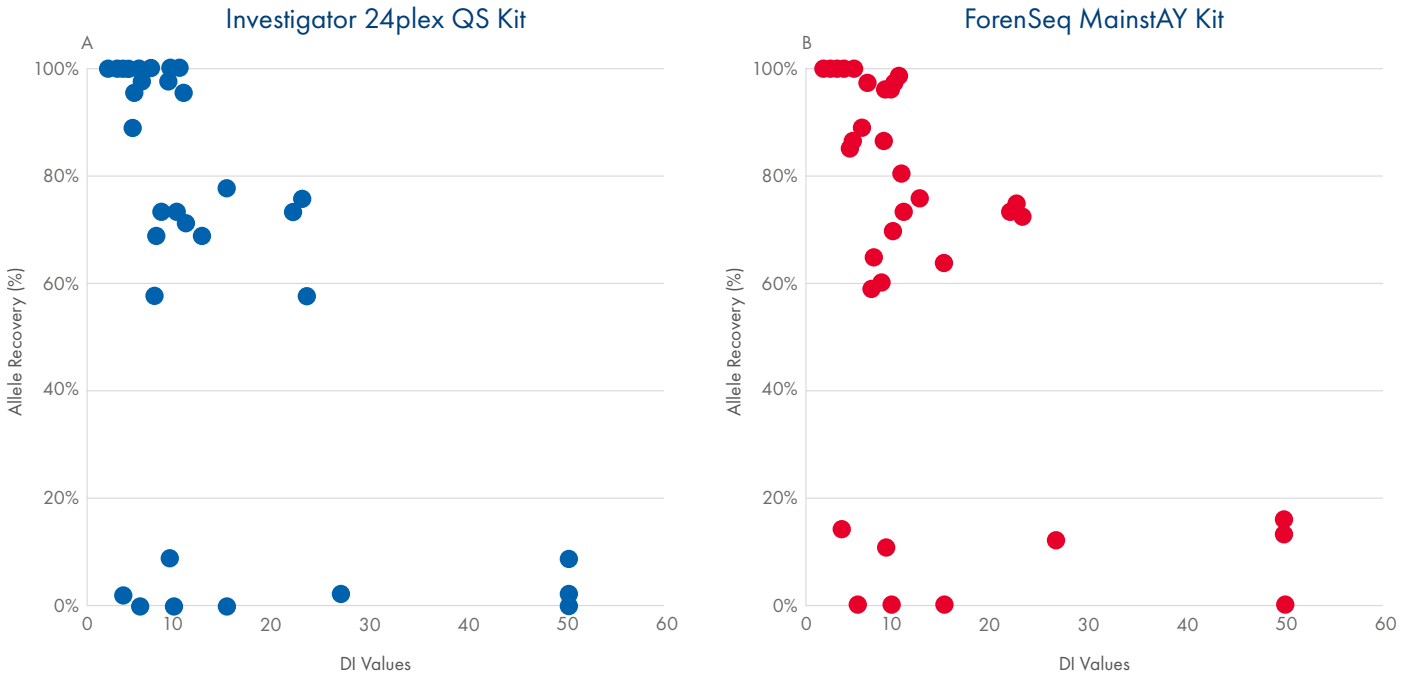

Figure 3.
Total allele recovery (%) vs. degradation index (DI). (A) Allele recovery (%) vs. degradation index values with Investigator 24plex QS Kit. (B) Allele recovery (%) vs. degradation index values with ForenSeq MainstAY Kit.

Statistical evaluation

Standard methods of forensic analysis were used to generate likelihood ratios (LRs) for statistical evaluation of results from CE STR and NGS STR methods. The software ArmedXpert was used to generate LRs for data obtained using Investigator 24plex QS Kit, whereas MixtureAce was used for data obtained using ForenSeq MainstAY Kit. MixtureAce is a plug-in for ArmedXpert and leverages the same user interface while allowing users to work with NGS methods.

Likelihood ratios from high-level profiles from unburned bone are shown in Table 2. ForenSeq MainstAY Kit returned substantially stronger likelihood ratios overall than Investigator 24plex QS Kit.

Likelihood ratios of low-level profiles from burned black bone are shown in Table 3. The likelihood ratios obtained with ForenSeq MainstAY Kit range from 1 in 111 million to 1 in 755 million. These data provide much better discrimination with more relevance to forensics than the likelihood ratios obtained with Investigator 24plex QS Kit ranging between a low of 1 in 721 to a high of 1 in 2100.

Table 2. High-level profile: Likelihood ratios obtained from profiles obtained from samples of unburned bone

	High-level profile: Likelihood ratios			
STR Method	**African American**	**Asian**	**Caucasian**	**Hispanic**
ForenSeq MainstAY Kit	1 in 1.82E + 56	1 in 1.08E+50	1 in 8.41E+51	1 in 4.20E+50
Investigator 24plex QS Kit	1 in 2.20E+32	1 in 3.78E+31	1 in 1.25E+30	1 in 7.55E+29

Table 3. Low-level profile: Likelihood ratios obtained from profiles obtained from samples of burned black bone

	Low-level profile: Likelihood ratios			
STR Method	**African American**	**Asian**	**Caucasian**	**Hispanic**
ForenSeq MainstAY Kit	1 in 7.55E+8	1 in 2.07E+7	1 in 2.65E+8	1 in 1.11E+8
Investigator 24plex QS Kit	1 in 1.34E+3	1 in 2.10E+3	1 in 721	1 in 8.95E+2

For the least degraded samples of unburned bone, likelihood ratios from NGS STR profiles were many orders of magnitude higher and more discriminating than those from CE STR. At the low level, with profiles obtained from the most degraded samples, NGS STR was more informative than CE STR.

Conclusion

Degradation values on their own do not accurately predict genotyping success of thermally degraded samples.

The ForenSeq MainstAY workflow can lead to more probative results for challenging and degraded samples with a smaller input volume and the added statistical power of sequencing.

Analysis of STRs with NGS rather than CE leads to a higher power of discrimination based on the increased information obtained from access to sequence-level allelic variation. This increased power of discrimination in turn results in a greater chance of identifying the donor of a profile with NGS rather than CE.

Ordering Information

Product	Contents	Cat. no.
ForenSeq MainstAY Kit (96)	Includes all the required reagents for 96 reactions to prepare sequencing libraries generating data for mainstream casework and forensic genetic genealogy confirmatory testing	V16000142
ForenSeq MainstAY Kit (384)	Includes all the required reagents for 384 reactions to prepare sequencing libraries generating data for mainstream casework and forensic genetic genealogy confirmatory testing	V16000128
ForenSeq MainstAY SE Kit (96)	Includes all the required reagents for 96 reactions to prepare sequencing libraries generating data for mainstream casework and forensic genetic genealogy confirmatory testing including the same markers as in ForenSeq MainstAY Kit + SE33	V16000183
MiSeq FGx Sequencing System	Desktop instrument with two run modes for a range of forensic genomics applications within a validated NGS workflow	15048976
ForenSeq Analysis Software Server, Monitor	Software pre-installed as a dedicated server specific for forensic genomics for run setup, sample management, analysis and report generation. This product includes server, mouse, keyboard and monitor.	9003364
MiSeq FGx Reagent Micro Kit	Supports up to 5 million paired-end reads for small batch sizes and faster turnaround times	20021681

Ordering Information

Product	Contents	Cat. no.
EZ2 Connect Fx System	Benchtop instrument for automated isolation of nucleic acids from up to 24 samples in parallel, using sealed prefilled cartridges; includes 2x EZ2 Connect racks (EZ2 Connect Fx Tip Rack and the EZ2 Connect Fx Tip Rack – Flip Cap Tubes), EZ2 Connect Fx Cartridge Rack and 1-year warranty on parts and labor.	9003220
EZ1&2 DNA Investigator Kit (48)	For 48 preps: Reagent Cartridges (DNA Investigator), Disposable Filter-Tips, Disposable Tip-Holders, Sample Tubes (2 ml), Elution Tubes (1.5 ml), Buffer G2, Proteinase K, Carrier RNA	952034
7 mL Large-Volume Tubes (48)	Two bags of 24 large-volume tubes (7 mL)	951954
Investigator Quantiplex Pro Kit (200)	For use on Applied Biosystems Real-Time Systems: Quantiplex Pro Reaction Mix, Quantiplex Pro Primer Mix, Control DNA M1, QuantiTect® Nucleic Acid Dilution Buffer	387216
Investigator 24plex QS Kit (100)	Primer Mix, Fast Reaction Mix including Taq DNA Polymerase, Control DNA, allelic ladder 24plex, DNA size standard 24plex (BTO), and nuclease-free water	382415
Investigator 24plex QS Kit (400)	Primer mix, Fast Reaction Mix including Taq DNA Polymerase, Control DNA, allelic ladder 24plex, DNA size standard 24plex (BTO), and nuclease-free water	382417

Learn more about NGS for HID in your lab. Visit **qiagen.com/MainstAY**

References:

1. Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM, Smerick JB, Budowle B. Validation of short tandem repeats (STRs) for forensic usage: Performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. *J Forensic Sci.* 2001; 46: 647–660.
2. Hares DR. Selection and implementation of expanded CODIS core loci in the United States. *Forensic Sci Int Genet.* 2015; 17: 33–34. 10.1016/j.fsigen.2015.03.006
3. Gorden EM, Sturk-Andreaggi K, Warnke-Sommer J, Hazelwood A, Barritt-Ross S, Marshall C. Next generation sequencing of STR artifacts produced from historical bone samples. *Forensic Sci. Int. Genet.* 2020; 49: 102397. 10.1016/j.fsigen.2020.102397
4. *ForenSeq MainstAY Product Line Reference Guide.* April 2022.
5. *MiSeq FGx Sequencing System Reference Guide.* February 2021.
6. *Universal Analysis Software MainstAY Product Line Module Version 2 Reference Guide.* June 2022.
7. *Investigator 24plex QS Handbook.* February 2021.
8. *EZ1&2 DNA Investigator Kit Handbook:* For automated purification of DNA from forensic and human ID samples using the EZ2 Connect Fx instrument. July 2023.
9. *Investigator Quantiplex Pro Handbook for Applied Biosystems 7500 Real-Time PCR Systems:* For quantification of human and male DNA in forensic samples. January 2023.